A Unified Framework for Low Autocorrelation Sequence Design via Majorization–Minimization

Licheng Zhao, Junxiao Song, Prabhu Babu, and Daniel P. Palomar, *Fellow, IEEE*

***Abstract*—In this paper, we consider the low autocorrelation sequence design problem. We optimize a unified metric over a general constraint set. The unified metric includes the integrated sidelobe level (ISL) and the peak sidelobe level (PSL) as special cases, and the general constraint set contains the unimodular constraint, Peak-to-Average Ratio (PAR) constraint, and similarity constraint, to name a few. The optimization technique we employ is the majorization–minimization (MM) method, which is iterative and enjoys guaranteed convergence to a stationary solution. We carry out the MM method in two stages: in the majorization stage, we propose three majorizing functions: two for the unified metric and one for the ISL metric; in the minimization stage, we give closed-form solutions for algorithmic updates under different constraints. The update step can be implemented with a few Fast Fourier Transformations (FFTs) and/or Inverse FFTs (IFFTs). We also show the connections between the MM and gradient projection method under our algorithmic scheme. Numerical simulations have shown that the proposed MM-based algorithms can produce sequences with low autocorrelation and converge faster than the traditional gradient projection method and the state-of-the-art algorithms.**

***Index Terms*—Sequence design, low autocorrelation, unified framework, majorization minimization.**

I. Introduction

SEQUENCES with low autocorrelation sidelobes enjoy a wide range of applications in wireless communications and signal processing. Some important engineering applications include Code-Division Multiple Access (CDMA) cellular systems [1], radar systems, security systems, and even cryptography systems [2]. We define $\mathbf{x} \triangleq \{x_n\}_{n=1}^N \in \mathbb{C}^N$ as a complex-valued sequence of length N. Our objective is to design a sequence whose autocorrelation sidelobes are jointly at a low level. The aperiodic and periodic autocorrelations of the sequence $\mathbf{x}$ are given as

$$r_k = \sum_{n=1}^{N-k} x_n x_{n+k}^* = r_{-k}^* \quad \text{(aperiodic)} \tag{1}$$

and

$$r_k = \sum_{n=1}^{N} x_n x_{(n+k)(\mathrm{mod}\, N)}^* = r_{-k}^* \quad \text{(periodic)} \tag{2}$$

for $k = 0, 1, \cdots, N-1$. Usually, the sequence to be designed has a limited energy budget, so we fix the sequence energy without loss of generality.

A. Related Work

In order to measure the goodness of the autocorrelation sidelobes of a sequence, researchers have put forward various metrics. One example is the integrated sidelobe level (ISL)

$$\mathrm{ISL} = \sum_{k=1}^{N-1} |r_k|^2, \tag{3}$$

or equivalently the merit factor (MF) [3]

$$\mathrm{MF} = \frac{r_0^2}{2\sum_{k=1}^{N-1} |r_k|^2} = \frac{\|\mathbf{x}\|_2^4}{2\mathrm{ISL}}. \tag{4}$$

With the sequence energy fixed, maximizing the MF is equivalent to minimizing the ISL. As an extension to the ISL, Stoica *et al.* [4] also proposed the Weighted ISL (WISL) metric

$$\mathrm{WISL} = \sum_{k=1}^{N-1} w_k |r_k|^2, \tag{5}$$

where $\{w_k\}_{k=1}^{N-1}$ are the nonnegative weights. This metric is useful when the r_k's are not equally important and we want to suppress some r_k's in particular. Another example of the popular metrics is the peak sidelobe level (PSL)

$$\mathrm{PSL} = \max_{k=1,2,\cdots,N-1} \{|r_k|\}. \tag{6}$$

Both the ISL and PSL can be viewed as the p-norm of $|r_k|$'s with (6) as the limit of large p, and thus these two metrics can be unified.

In practice, there are many constraints to be considered apart from the aforementioned energy budget constraint. In terms of modulus, [4]–[6] have mentioned the constant modulus (unimodular, to be accurate) constraint, and [7]–[9] handled PAR constraint. In terms of phase, [10] focused on the polyphase constraint, and [11], [12] studied the similarity constraint. Phase constraints are imposed together with the constant modulus constraint.

There is a rich literature on finding low autocorrelation sequences. The existing construction methods are either analytical or computational. Using analytical methods, we are able to construct sequences with closed-form expressions, such as the Frank sequences [13], the Chu sequences [14], and the Golomb sequences [15]. Traditional computational methods are either based on search algorithms (exhaustive search or stochastic search) or evolutionary algorithms. They become computationally expensive when the sequence length goes up to a few thousand. In addition, the evolutionary algorithms are

Manuscript received May 24, 2016; revised August 10, 2016 and September 12, 2016; accepted October 4, 2016. Date of publication October 21, 2016; date of current version November 17, 2016. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Tao Jiang. This work was supported by the Hong Kong RGC 16206315 research grant and by the Hong Kong RGC Theme-based Research Scheme (TRS) Grant T21-602/15R.

The authors are with the Hong Kong University of Science and Technology, Hong Kong (e-mail: lzhaoai@ust.hk; jsong@connect.ust.hk; eeprabhubabu@ust.hk; palomar@ust.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2016.2620113

heuristic methods in nature whose convergence is not guaranteed. To summarize, the existing sequence construction methods cannot effectively design extremely long sequences (up to 10^4) with very low autocorrelation, and that is why modern optimization techniques are desperately needed.

Recently, a number of optimization-oriented approaches [4], [16]–[18] have been put forward to design long sequences with low autocorrelation sidelobes. Among them, the Cyclic Algorithm New (CAN) algorithm [4] enjoys a satisfactory performance in producing unimodular sequences of length up to several million. Rather than minimizing the ISL, the CAN algorithm minimizes a simpler but "almost equivalent" metric, changing the quartic objective function to a quadratic one. When it comes to the WISL metric, there is no simple "almost equivalent" reformulation available. Hence, as is shown in [4], the Weighted CAN (WeCAN) algorithm performs much worse than CAN.

B. Contribution

The major contributions of this paper (with respect to [5], [6]) lie in the following three aspects.

1) The first contribution is that we propose a unified framework to design low autocorrelation sequences, which provides a global perspective for the sequence design problem. In terms of the objective, we unify the existing ISL, WISL, and PSL metrics; in terms of the constraint, we consider the constant modulus constraint, PAR constraint, polyphase constraint, similarity constraint, and so on. We can solve the sequence design problem in a unified manner instead of dealing with each metric and each constraint separately in a specific way.
2) The second contribution is that we systematically apply the Majorization Minimization (MM) method [19]. We carry out the MM method in two separate stages: majorization function construction and minimization solution derivation. In the majorization stage, we construct a total of three majorizing functions. Two of them apply to the unified metric, and the remaining one, which is not mentioned in [5], [6], applies to the more specific ISL metric. Its achieved ISL level is as good as the other two majorizing functions, and the convergence speed is the fastest, 2–4 times as fast as the second fastest, according to the simulation results. In the minimization stage, we provide efficient closed-form solutions to the minimization problems depending on the constraints. The minimization stage is by nature solving a projection problem, which is not pointed out in [5], [6]. We maintain that the whole MM process is more than a simple combination of existing works. Furthermore, the MM update step can be implemented with a few FFT (IFFT) operations which are highly efficient.
3) The third contribution is that we establish the connections between the MM and gradient projection method under our algorithmic scheme, which is very insightful. The update step of the MM method has exactly the same structure as that of the gradient projection method. However, the MM method is still superior because of theoretical stationarity convergence guarantee. Apart from that, the MM method has faster convergence speed than the gradient projection method, as can be seen in the simulation results.

C. Organization and Notation

The rest of the paper is organized as follows. In Section II, we introduce the problem formulation. In Section III, a brief preliminary description of the MM method is given. In Sections IV and V, we construct majorization functions and derive minimization solutions, respectively. In Section VI, we show the connections between the MM and gradient projection method under our implementation scheme. In Section VII, we elaborate on the algorithmic implementation. Finally, Section VIII presents numerical simulations, and the conclusions are given in Section IX.

The following notation is adopted. Boldface upper-case letters represent matrices, boldface lower-case letters denote column vectors, and standard lower-case letters stand for scalars. $\mathbb{R}$ ($\mathbb{C}$) denotes the real (complex) field. $|\cdot|$ denotes the absolute value. $\|\cdot\|_p$ denotes the p-norm of a vector. $\nabla(\cdot)$ represents the gradient of a vector function (the way to derive the complex-valued gradient follows [20]). $\mathbf{I}$ stands for the identity matrix. $\mathbf{X}_{ij}$ denotes the (i,j)th element of the matrix $\mathbf{X}$. $\mathbf{X}^T$, $\mathbf{X}^*$, $\mathbf{X}^H$, $\mathrm{Tr}(\mathbf{X})$, $\lambda_{\max}(\mathbf{X})$, $\lambda_{\min}(\mathbf{X})$, and $\mathrm{vec}(\mathbf{X})$ denote the transpose, complex conjugate, conjugate transpose, trace, the largest eigenvalue, the smallest eigenvalue and stacking vectorization of $\mathbf{X}$, respectively. $\mathrm{Diag}(\mathbf{x})$ is a diagonal matrix with $\mathbf{x}$ filling its principal diagonal and $\mathrm{diag}(\mathbf{X})$ is the vector consisting of all the diagonal elements of matrix $\mathbf{X}$. Finally, $\odot$ stands for the Hadamard product.

II. Problem Statement

A. Objective Function

We propose a unified metric named Weighted Peak or Integrated Sidelobe Level (WPISL) as follows:

$$\mathrm{WPISL} = \sum_{k=1}^{N-1} w_k \left| r_k \right|^p, \tag{7}$$

where $2 \le p < +\infty$, and $\{w_k\}_{k=1}^{N-1}$ are nonnegative weights. This metric readily includes the ISL (or WISL) and the PSL as special cases: let $p = 2$, and we get the WISL metric; further let $w_k = 1$, $\forall k$, and we get the ISL metric; let $p \to +\infty$ and $w_k = 1, \forall k$, and we get $\left(\sum_{k=1}^{N-1} |r_k|^p\right)^{1/p} \to \max_k \{|r_k|\}$, i.e., the PSL metric.

B. Constraints of Interest

We have already introduced the energy budget constraint, i.e., $\|\mathbf{x}\|_2^2 = c_e^2$ (constant). Besides that, some additional constraints may also be of interest, as are described in the following.

1) Strict Constant Modulus Constraint: In [4]–[6], strict constant modulus constraint or, in particular, strict unimodular constraint is of interest due to the limitations of hardware components and/or efficiency of amplifiers [21]. This constraint is expressed as: for $n = 1, 2, \cdots, N$,

$$|x_n| = c_m = \frac{c_e}{\sqrt{N}} \ (\text{constant, specially } c_m = 1). \tag{8}$$

This constraint is denoted as C_1 for short.

2) ϵ-Uncertainty Constant Modulus Constraint: Sometimes the strict constant modulus constraint is too harsh and we may want to relax the constant modulus within a small ϵ-uncertainty

TABLE I
MISCELLANEOUS CONSTRAINTS AND THEIR NOTATIONS

Modulus Constraint	Index	Phase Constraint	Index
Strict constant modulus constraint	C_1	Discrete phase constraint	C_4
ϵ-Uncertainty constant modulus constraint	C_2	Similarity constraint	C_5
PAR constraint	C_3		

region, which is: for $n = 1, 2, \cdots, N$,

$$c_m - \epsilon_1 \leq |x_n| \leq c_m + \epsilon_2 \ (0 \leq \epsilon_1 \leq c_m, \, 0 \leq \epsilon_2) . \quad (9)$$

This constraint is a modest relaxation of the previous constraint C_1, and it is denoted as C_2 for short.

3) PAR Constraint: The signal PAR measures the ratio of the largest signal magnitude to its average power [7]–[9]:

$$\mathrm{PAR}(\mathbf{x}) = \frac{\max_n |x_n|^2}{\|\mathbf{x}\|_2^2 / N}, \quad (10)$$

and $1 \leq \mathrm{PAR}(\mathbf{x}) \leq N$. We require $\mathrm{PAR}(\mathbf{x}) \leq \rho$ (a particular threshold), so that the analog-to-digital converters and the digital-to-analog converters in the system can have lower dynamic range, and fewer linear power amplifiers are needed. Since we have already assume $\|\mathbf{x}\|_2 = c_e$, the PAR constraint is equivalent to: for $n = 1, 2, \cdots, N$,

$$|x_n| \leq c_p \ (\text{constant}) , \quad (11)$$

where $c_e/\sqrt{N} \leq c_p \leq c_e$. This constraint is more relaxed than the previous two in that the modulus is not lower bounded. When $c_p = c_e/\sqrt{N}$, the PAR constraint degenerates into strict constant modulus constraint. When $c_p > c_e/\sqrt{N} \, (= c_m)$, the PAR constraint becomes a large uncertainty set around c_m ($\epsilon_1 = c_m$ and $\epsilon_2 = c_p - c_m$). This constraint is denoted as C_3 for short.

4) Discrete Phase Constraint: Besides modulus constraints, phase constraints are also desired. One commonly used constraint is the discrete phase constraint, also known as the polyphase constraint [10]. It is expressed as: for $n = 1, 2, \cdots, N$,

$$\arg(x_n) \in \{\phi_1, \phi_2, \cdots, \phi_I\} , \quad (12)$$

where $\phi_1, \phi_2, \cdots, \phi_I$ are I fixed angles. This constraint is denoted as C_4 for short, and is often accompanied by the strict constant modulus constraint C_1.

5) Similarity Constraint: Sometimes we want the designed sequence to lie in the neighborhood of a reference one which already has good properties [11], [12]. It is written as $\|\mathbf{x} - \mathbf{x}_r\|_\infty \leq \delta$ $(0 \leq \delta \leq 2)$ with $\mathbf{x}_r$ being the reference sequence, which is equivalent to $|x_n - x_{r,n}| \leq \delta$, $\forall n$. Similarly, this constraint is often accompanied by the strict constant modulus constraint C_1, and can be further simplified as: for $n = 1, 2, \cdots, N$,

$$\arg(x_n) \in [\gamma_n, \gamma_n + \vartheta] , \quad (13)$$

where $\gamma_n = \arg(x_{r,n}) - \arccos\left(1 - \delta^2/2\right)$, and $\vartheta = 2\arccos\left(1 - \delta^2/2\right)$. Note that when $\delta = 0$, we get $\vartheta = 0$, indicating that $\mathbf{x} = \mathbf{x}_r$; when $\delta = 2$, we get $\vartheta = 2\pi$, and this constraint is always satisfied. This constraint is denoted as C_5 for short.

For clarity, we summarize all the aforementioned constraints and their notations in Table I.

C. Problem Formulation

The problem formulation consists of the minimization of the WPISL metric in (7) subject to the constraints mentioned in Table I, and it reads

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & \sum_{k=1}^{N-1} w_k \left| r_k(\mathbf{x}) \right|^p \\ \text{subject to} & \mathbf{x} \in \mathcal{X}, \end{array} \quad (14)$$

where $r_k(\mathbf{x}) = \mathbf{x}^H \mathbf{U}_k \mathbf{x}$, for the aperiodic case, i.e., (1), $\mathbf{U}_k \in \mathbb{R}^{N \times N}$ is a Toeplitz matrix with only the kth diagonal entries being 1 and 0 elsewhere; while for the periodic case, i.e., (2), $\mathbf{U}_k$ is a Toeplitz matrix with only the kth and $(k-N)$th diagonal entries being 1 and 0 elsewhere, and

$$\mathcal{X} = \left\{ \mathbf{x} \in \mathbb{C}^N \,\middle|\, \|\mathbf{x}\|_2^2 = c_e^2 \right\} \cap \left(\cap_i C_i \right) , \quad (15)$$

which means only a few C_i's are activated.

Remark 1: In the following, we focus on the aperiodic case and the results can be easily extended to the periodic case with minor modifications.

III. PRELIMINARIES: MAJORIZATION MINIMIZATION METHOD

In this section, we are going to briefly introduce the MM method, and its algorithmic framework [19].

A. The MM Method

The MM method can be applied to solve the following general optimization problem:

$$\begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \in \mathcal{X}, \end{array} \quad (16)$$

where f is differentiable on the whole $\mathbb{C}$ space and $\mathcal{X}$ is some constraint set. Rather than minimizing $f(\mathbf{x})$ directly, we consider successively solving a series of simple optimization problems. The algorithm initializes at some feasible starting point $\mathbf{x}^{(0)}$, and then iterates as $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots$ until some convergence criterion is met. For any iteration, say, the lth iteration, the update rule is

$$\mathbf{x}^{(l+1)} \in \arg\min_{\mathbf{x} \in \mathcal{X}} \bar{f}\left(\mathbf{x}; \mathbf{x}^{(l)}\right) , \quad (17)$$

where $\bar{f}\left(\mathbf{x}; \mathbf{x}^{(l)}\right)$ is the majorizing function of $f(\mathbf{x})$ at $\mathbf{x}^{(l)}$, which satisfies

1) $\bar{f}\left(\mathbf{x}; \mathbf{x}^{(l)}\right) \geq f(\mathbf{x}), \forall \mathbf{x} \in \mathcal{X}$;
2) $\bar{f}\left(\mathbf{x}^{(l)}; \mathbf{x}^{(l)}\right) = f\left(\mathbf{x}^{(l)}\right)$.

In words, $\bar{f}\left(\mathbf{x}; \mathbf{x}^{(l)}\right)$ is a tight global upper bound of $f(\mathbf{x})$ in the constraint set and also coincides with $f(\mathbf{x})$ at $\mathbf{x}^{(l)}$.

B. Stationarity Convergence

The MM method can guarantee convergence to a stationary solution of (16) as long as the following two additional conditions are satisfied:

3) $\nabla \bar{f}\left(\mathbf{x}^{(l)};\mathbf{x}^{(l)}\right)=\nabla f\left(\mathbf{x}^{(l)}\right)$;
4) $\bar{f}\left(\mathbf{x};\mathbf{x}^{(l)}\right)$ is continuous in both $\mathbf{x}$ and $\mathbf{x}^{(l)}$.

Interested readers may refer to [22] for detailed proof and here we only briefly mention its idea. First note that $f\left(\mathbf{x}^{(l)}\right)=\bar{f}\left(\mathbf{x}^{(l)};\mathbf{x}^{(l)}\right)\geq\min_{\mathbf{x}\in\mathcal{X}}\bar{f}\left(\mathbf{x};\mathbf{x}^{(l)}\right)=\bar{f}\left(\mathbf{x}^{(l+1)};\mathbf{x}^{(l)}\right)\geq f\left(\mathbf{x}^{(l+1)}\right)$, from which we get $f\left(\mathbf{x}^{(0)}\right)\geq f\left(\mathbf{x}^{(1)}\right)\geq f\left(\mathbf{x}^{(2)}\right)\geq\cdots$, implying the monotonicity of the sequence $\left\{f\left(\mathbf{x}^{(l)}\right)\right\}$. Then assume a subsequence of $\left\{\mathbf{x}^{(l)}\right\}$, denoted by $\left\{\mathbf{x}^{(l_j)}\right\}$, converges to a limit point $\mathbf{z}$. Thus, for all $\mathbf{x}\in\mathcal{X}$, $\bar{f}\left(\mathbf{x};\mathbf{x}^{(l_j)}\right)\geq\bar{f}\left(\mathbf{x}^{(l_j+1)};\mathbf{x}^{(l_j)}\right)\geq f\left(\mathbf{x}^{(l_j+1)}\right)\geq f\left(\mathbf{x}^{(l_{j+1})}\right)=\bar{f}\left(\mathbf{x}^{(l_{j+1})};\mathbf{x}^{(l_{j+1})}\right)$. Letting $j\to+\infty$, we obtain for all $\mathbf{x}\in\mathcal{X}$, $\bar{f}\left(\mathbf{x};\mathbf{x}^{(\infty)}\right)\geq\bar{f}\left(\mathbf{x}^{(\infty)};\mathbf{x}^{(\infty)}\right)$, which indicates

$$\bar{f}'\left(\mathbf{x}^{(\infty)};\mathbf{x}^{(\infty)},\mathbf{d}\right)\geq 0,\ \forall\mathbf{d}\in\mathcal{T}_{\mathcal{X}}\left(\mathbf{x}^{(\infty)}\right),\tag{18}$$

where $\bar{f}'$ stands for directional derivative and $\mathcal{T}_{\mathcal{X}}\left(\mathbf{x}^{(\infty)}\right)$ is the Boulingand tangent cone of $\mathcal{X}$ at $\mathbf{x}^{(\infty)}$. The definition of directional derivative of f in the direction $\mathbf{d}$ is

$$f'(\mathbf{x};\mathbf{d})=\liminf_{\lambda\downarrow 0}\frac{f(\mathbf{x}+\lambda\mathbf{d})-f(\mathbf{x})}{\lambda}.\tag{19}$$

If $\bar{f}$ is differentiable, $\bar{f}'(\mathbf{x};\mathbf{d})$ can be computed effectively as $\mathbf{d}^T\nabla\bar{f}(\mathbf{x})$. Interested readers may refer to [23], [24] for more knowledge of tangent cone. Since $\nabla\bar{f}\left(\mathbf{x}^{(\infty)};\mathbf{x}^{(\infty)}\right)=\nabla f\left(\mathbf{x}^{(\infty)}\right)$, we have $\bar{f}'\left(\mathbf{x}^{(\infty)};\mathbf{x}^{(\infty)},\mathbf{d}\right)=\mathbf{d}^T\nabla\bar{f}\left(\mathbf{x}^{(\infty)};\mathbf{x}^{(\infty)}\right)=\mathbf{d}^T\nabla f\left(\mathbf{x}^{(\infty)}\right)=f'\left(\mathbf{x}^{(\infty)};\mathbf{d}\right)$. Therefore, the limit point $\mathbf{x}^{(\infty)}$ should satisfy

$$f'\left(\mathbf{x}^{(\infty)};\mathbf{d}\right)\geq 0,\ \forall\mathbf{d}\in\mathcal{T}_{\mathcal{X}}\left(\mathbf{x}^{(\infty)}\right),\tag{20}$$

and thus $\mathbf{x}^{(\infty)}$ is a (Boulingand) stationary point.

C. Acceleration Techniques

The convergence speed of the MM method is mainly determined by the majorizing function. In some cases where the majorizing function is not well designed or cannot be better designed, we should adopt some techniques to accelerate the convergence speed.

1) Acceleration via SQUAREM: SQUAREM stands for squared iterative method, which can be viewed as an "off-the-shelf" accelerator of the MM method. It was proposed by Varadhan and Roland [25] and is claimed to have the following two advantages: 1) requiring nothing more than the MM variable updating scheme, and 2) enjoying global convergence. The detailed implementation of SQUAREM is shown in Algorithm 1. To make SQUAREM better fit in the MM method, we make the following modifications. To guarantee feasibility, we project the infeasible points back to the constraint set by the operation $\mathcal{P}_{\mathcal{X}}(\cdot)$. Moreover, to preserve the monotonicity of the MM method, we adopt the strategy of backtracking, which repeatedly halves the distance between α and -1 until the monotonicity is achieved. To see why it works, interested readers may refer to [6].

2) Acceleration via Local Majorization: This idea has previously shown up in [5]. As is mentioned above, the slow convergence speed of the MM method is due to an ill-designed majorizing function. An alternative acceleration technique is to modify the ill-designed global majorizing function to a local one. This technique can be readily applied to an objective function $f(\mathbf{x})$ that is Lipschitz continuous on the constraint set $\mathcal{X}$. In this case, the majorizing function is naturally constructed as $\bar{f}\left(\mathbf{x};\mathbf{x}^{(l)}\right)=f\left(\mathbf{x}^{(l)}\right)+\nabla^T f\left(\mathbf{x}^{(l)}\right)\left(\mathbf{x}-\mathbf{x}^{(l)}\right)+\frac{L}{2}\left\|\mathbf{x}-\mathbf{x}^{(l)}\right\|_2^2$ where L is a constant with $L\geq L_f\geq 0$ (L_f is the Lipschitz constant for the function f). In most situations, L_f is very difficult to calculate and we can only derive an upper bound, possibly quite loose. Let $\bar{f}_\theta\left(\mathbf{x};\mathbf{x}^{(l)},\theta\right)\triangleq f\left(\mathbf{x}^{(l)}\right)+\nabla^T f\left(\mathbf{x}^{(l)}\right)\left(\mathbf{x}-\mathbf{x}^{(l)}\right)+\theta\cdot\frac{L}{2}\left\|\mathbf{x}-\mathbf{x}^{(l)}\right\|_2^2$ with $0\leq\theta\leq 1$. It can be seen that $\bar{f}_\theta\left(\mathbf{x};\mathbf{x}^{(l)},1\right)=\bar{f}\left(\mathbf{x};\mathbf{x}^{(l)}\right)$, and for $0\leq\theta<1$, $\bar{f}_\theta\left(\mathbf{x};\mathbf{x}^{(l)},\theta\right)$ is a local majorizing function. Following the argument of the monotonicity of the MM method, we see that $f\left(\mathbf{x}^{(l)}\right)=\bar{f}_\theta\left(\mathbf{x}^{(l)};\mathbf{x}^{(l)},\theta\right)\geq\min_{\mathbf{x}\in\mathcal{X}}\bar{f}_\theta\left(\mathbf{x};\mathbf{x}^{(l)},\theta\right)=\bar{f}_\theta\left(\bar{\mathbf{x}}^{(l)};\mathbf{x}^{(l)},\theta\right)$ still holds, but $\bar{f}_\theta\left(\bar{\mathbf{x}}^{(l)};\mathbf{x}^{(l)},\theta\right)\geq f\left(\bar{\mathbf{x}}^{(l)}\right)$ may not. To achieve monotonicity, we start from a small-valued θ and gradually increase it (up to 1) until $\bar{f}_\theta\left(\bar{\mathbf{x}}^{(l)};\mathbf{x}^{(l)},\theta\right)\geq f\left(\bar{\mathbf{x}}^{(l)}\right)$ is satisfied. Then we set $\mathbf{x}^{(l+1)}=\bar{\mathbf{x}}^{(l)}$. The detailed implementation of this acceleration technique is elaborated in Algorithm 2.

Algorithm 1: MM Acceleration via SQUAREM.

Require: $l=0$, the initial feasible point $\mathbf{x}^{(0)}$.
repeat
  $\mathbf{x}_1^{(l)}=\arg\min_{\mathbf{x}\in\mathcal{X}}\ \bar{f}\left(\mathbf{x};\mathbf{x}^{(l)}\right)$;
  $\mathbf{x}_2^{(l)}=\arg\min_{\mathbf{x}\in\mathcal{X}}\ \bar{f}\left(\mathbf{x};\mathbf{x}_1^{(l)}\right)$;
  $\mathbf{r}=\mathbf{x}_1^{(l)}-\mathbf{x}^{(l)}$;
  $\mathbf{v}=\mathbf{x}_2^{(l)}-\mathbf{x}_1^{(l)}-\mathbf{r}$;
  $\alpha=-\left\|\mathbf{r}\right\|_2/\left\|\mathbf{v}\right\|_2$;
  $\tilde{\mathbf{x}}^{(l)}=\mathcal{P}_{\mathcal{X}}\left(\mathbf{x}^{(l)}-2\alpha\mathbf{r}+\alpha^2\mathbf{v}\right)$;
  while $f\left(\tilde{\mathbf{x}}^{(l)}\right)>f\left(\mathbf{x}^{(l)}\right)$ **do**
    $\alpha=(\alpha-1)/2$;
    $\tilde{\mathbf{x}}^{(l)}=\mathcal{P}_{\mathcal{X}}\left(\mathbf{x}^{(l)}-2\alpha\mathbf{r}+\alpha^2\mathbf{v}\right)$;
  end while
  $\mathbf{x}^{(l+1)}=\tilde{\mathbf{x}}^{(l)}$;
  $l=l+1$;
until convergence

Algorithm 2: MM Acceleration via Local Majorization.

Require: $l=0$, the initial feasible point $\mathbf{x}^{(0)}$,
  $\boldsymbol{\theta}=\left[\frac{1}{T},\frac{2}{T},\cdots,1\right]^T$ with T being the grid size.
repeat
  $i=1$;
  $\bar{\mathbf{x}}^{(l)}=\arg\min_{\mathbf{x}\in\mathcal{X}}\ \bar{f}_\theta\left(\mathbf{x};\mathbf{x}^{(l)},\theta(i)\right)$;
  while $\bar{f}_\theta\left(\bar{\mathbf{x}}^{(l)};\mathbf{x}^{(l)},\theta(i)\right)<f\left(\bar{\mathbf{x}}^{(l)}\right)$ **do**
    $i=i+1$;
    $\bar{\mathbf{x}}^{(l)}=\arg\min_{\mathbf{x}\in\mathcal{X}}\ \bar{f}_\theta\left(\mathbf{x};\mathbf{x}^{(l)},\theta(i)\right)$;
  end while
  $\mathbf{x}^{(l+1)}=\bar{\mathbf{x}}^{(l)}$;
  $l=l+1$;
until convergence

IV. Majorizing Function Construction

In this section, we are going to put forward a total of three majorizing functions. Two of them apply to the unified WPISL metric and the remaining one applies to the more specific ISL

metric. It is generally preferred that the minimization of the majorizing function has a closed-form solution. Driven by this motivation, we construct quadratic majorizing functions so that a closed-form solution is easily obtained.

A. Majorizing Functions for the WPISL Metric

We first study a scalar function x^p. It is generally known that an arbitrary power function x^p does not have a global quadratic majorizing function for $p > 2$. However, there does exist a local one, which comes from the lemma below.

Lemma 2 ([6, Lemma 10]): Let $g(x) = x^p$ with $p \geq 2$. The function $\bar{g}(x; x_0) = ax^2 + bx + ax_0^2 - (p-1)x_0^p$ is a local majorizing function of $g(x)$ at $x_0 \in [0, \bar{x})$ on the interval $[0, \bar{x}]$ where $a = [\bar{x}^p - x_0^p - px_0^{p-1}(\bar{x} - x_0)]/(\bar{x} - x_0)^2 \geq 0$ and $b = px_0^{p-1} - 2ax_0 \leq 0$. In particular, when $p = 2$, we get $a = 1$, $b = 0$, and $\bar{g}(x; x_0) = x^2 = g(x)$ (no majorization).

Remark 3: Using such a local majorizing function maintains monotonicity and will not cause infeasibility. The global minimizer of $\bar{g}(x; x_0)$ with respect to x is $-\frac{b}{2a}$. If $p > 2$, then $a > 0$ and $b \leq 0$, making $-\frac{b}{2a}$ nonnegative. In addition, we observe $-\frac{b}{2a} = -\frac{px_0^{p-1} - 2ax_0}{2a} = x_0 - \frac{p}{2a}x_0^{p-1} < x_0 < \bar{x}$. Since the global minimizer still falls within $[0, \bar{x}]$, infeasibility will not occur.

Suppose $\mathbf{x}^{(l)}$ is the designed sequence at the lth MM iteration, and we want to construct a majorizing function around $\mathbf{x}^{(l)}$ for the WPISL metric. We denote $f(\mathbf{x}) \triangleq \sum_{k=1}^{N-1} w_k |r_k(\mathbf{x})|^p$ with $r_k(\mathbf{x}) = \mathbf{x}^H \mathbf{U}_k \mathbf{x}$ and handle the summation term by term using Lemma 2. We regard $|r_k(\mathbf{x})|$ as x and $\left|r_k\left(\mathbf{x}^{(l)}\right)\right|$ as x_0, obtaining (constant terms are represented with const for simplicity)

$$f(\mathbf{x}) \leq \sum_{k=1}^{N-1} w_k \Big(a_k |r_k(\mathbf{x})|^2 + b_k |r_k(\mathbf{x})| + \text{const} \Big), \quad (21)$$

where $a_k\,(\geq 0)$ and $b_k\,(\leq 0)$ follow the structure of a and b in Lemma 2, and the corresponding interval upper limit $\bar{x}$ (just as that mentioned in Lemma 2) is different for each k: $\bar{x}_k = \begin{cases} \left(\frac{1}{w_k}\sum_{j=1}^{N-1} w_j \left|r_j\left(\mathbf{x}^{(l)}\right)\right|^p\right)^{1/p} & w_k \neq 0 \\ 0 & w_k = 0 \end{cases}$. Next, we observe $b_k \leq 0$ and $|r_k(\mathbf{x})||r_{-k}(\mathbf{x}^{(l)})| \geq \text{Re}[r_k(\mathbf{x}) \cdot r_{-k}(\mathbf{x}^{(l)})]$, so (recall $b_k \leq 0$)

$$w_k b_k |r_k(\mathbf{x})| \leq w_k b_k \text{Re}\left[\frac{r_{-k}\left(\mathbf{x}^{(l)}\right)}{\left|r_{-k}\left(\mathbf{x}^{(l)}\right)\right|} r_k(\mathbf{x})\right] =$$

$$\frac{1}{2}\mathbf{x}^H \left(w_{-k} b_{-k} \frac{r_k\left(\mathbf{x}^{(l)}\right)}{\left|r_k\left(\mathbf{x}^{(l)}\right)\right|} \mathbf{U}_{-k} + w_k b_k \frac{r_{-k}\left(\mathbf{x}^{(l)}\right)}{\left|r_{-k}\left(\mathbf{x}^{(l)}\right)\right|} \mathbf{U}_k \right) \mathbf{x}, \quad (22)$$

with $\mathbf{U}_{-k} = \mathbf{U}_k^H$, $w_{-k} = w_k$, $w_0 = 0$ and $b_{-k} = b_k$, $b_0 = 0$. Combining (21) and (22), we can further majorize $f(\mathbf{x})$ in the following manner:

$$\begin{aligned} f(\mathbf{x}) &\leq \sum_{k=1}^{N-1} w_k a_k |r_k(\mathbf{x})|^2 + \frac{1}{2}\mathbf{x}^H \mathbf{B}\mathbf{x} + \text{const} \\ &= \sum_{k=1}^{N-1} w_k a_k \left|\mathbf{x}^H \mathbf{U}_k \mathbf{x}\right|^2 + \frac{1}{2}\mathbf{x}^H \mathbf{B}\mathbf{x} + \text{const}, \end{aligned} \quad (23)$$

where

$$\mathbf{B} = \sum_{k=1-N}^{N-1} w_k b_k \frac{r_{-k}\left(\mathbf{x}^{(l)}\right)}{\left|r_{-k}\left(\mathbf{x}^{(l)}\right)\right|} \mathbf{U}_k. \quad (24)$$

Now we already have a quadratic term $\frac{1}{2}\mathbf{x}^H \mathbf{B}\mathbf{x}$, but still have an intractable quartic term $\sum_{k=1}^{N-1} w_k a_k \left|\mathbf{x}^H \mathbf{U}_k \mathbf{x}\right|^2$, which is the focus of the following majorization step. We are going to majorize the quartic term in two different ways, obtaining two majorizing functions eventually.

1) Majorizing via the Largest Eigenvalue: First we introduce a lemma addressing how to majorize a quadratic function by another quadratic function.

Lemma 4 ([5, Lemma 1]): Given $\mathbf{M} \succeq \mathbf{M}_0$ and $\mathbf{x}_0$, the quadratic function $\mathbf{x}^H \mathbf{M}_0 \mathbf{x}$ is majorized by $\mathbf{x}^H \mathbf{M}\mathbf{x} + 2\text{Re}\left[\mathbf{x}^H (\mathbf{M}_0 - \mathbf{M})\mathbf{x}_0\right] + \mathbf{x}_0^H (\mathbf{M} - \mathbf{M}_0)\mathbf{x}_0$ at $\mathbf{x}_0$.

With some simple transformations, the quartic term $\sum_{k=1}^{N-1} w_k a_k \left|\mathbf{x}^H \mathbf{U}_k \mathbf{x}\right|^2$ can be rewritten as

$$\frac{1}{2}\text{vec}^H(\mathbf{X}) \cdot \mathbf{L} \cdot \text{vec}(\mathbf{X}), \quad (25)$$

where

$$\mathbf{L} = \sum_{k=1-N}^{N-1} w_k a_k \text{vec}(\mathbf{U}_{-k}) \text{vec}^H(\mathbf{U}_{-k}), \quad (26)$$

$\mathbf{X} = \mathbf{x}\mathbf{x}^H$, $a_{-k} = a_k$, and $a_0 = 0$. It is easy to see that (25) is quadratic in $\mathbf{X}$ and by applying Lemma 4 with $\mathbf{M}_0 = \mathbf{L}$ and $\mathbf{M} = \lambda_{\max}(\mathbf{L})\mathbf{I}$, we can get the following result.

Lemma 5: The expression $\sum_{k=1}^{N-1} w_k a_k |\mathbf{x}^H \mathbf{U}_k \mathbf{x}|^2 + \frac{1}{2}\mathbf{x}^H \mathbf{B}\mathbf{x}$ is majorized by

$$\frac{1}{2}\lambda_{\max}(\mathbf{L}) \|\mathbf{x}\|_2^4 + \mathbf{x}^H \left(\mathbf{R} - \lambda_{\max}(\mathbf{L})\mathbf{x}^{(l)}\mathbf{x}^{(l)H}\right)\mathbf{x} + \text{const}, \quad (27)$$

where

$$\mathbf{R} = \sum_{k=1-N}^{N-1} \frac{p}{2} w_k \left|r_k\left(\mathbf{x}^{(l)}\right)\right|^{p-2} r_{-k}\left(\mathbf{x}^{(l)}\right) \mathbf{U}_k. \quad (28)$$

Proof: See Appendix A for the detailed proof. ■

We apply Lemma 4 again on $\mathbf{x}^H \left(\mathbf{R} - \lambda_{\max}(\mathbf{L})\mathbf{x}^{(l)}\mathbf{x}^{(l)H}\right)\mathbf{x}$ with $\mathbf{M} = \lambda_u \mathbf{I}$ where λ_u is chosen such that $\lambda_u \geq \lambda_{\max}(\mathbf{R})$ and finally get

$$\begin{aligned} &\mathbf{x}^H \left(\mathbf{R} - \lambda_{\max}(\mathbf{L})\mathbf{x}^{(l)}\mathbf{x}^{(l)H}\right)\mathbf{x} \\ &\leq \lambda_u \mathbf{x}^H \mathbf{x} + 2\text{Re}\Big[\mathbf{x}^H \big(\mathbf{R} - \lambda_{\max}(\mathbf{L})\mathbf{x}^{(l)}\mathbf{x}^{(l)H} \\ &\quad - \lambda_u \mathbf{I}\big)\mathbf{x}^{(l)}\Big] + \text{const} \\ &= \lambda_u \|\mathbf{x}\|_2^2 - 2\lambda_u \text{Re}\left[\mathbf{y}_1^H \mathbf{x}\right] + \text{const}, \end{aligned} \quad (29)$$

where

$$\mathbf{y}_1 = \left(1 + \frac{\lambda_{\max}(\mathbf{L})}{\lambda_u}\left\|\mathbf{x}^{(l)}\right\|_2^2\right)\mathbf{x}^{(l)} - \frac{1}{\lambda_u}\mathbf{R}\mathbf{x}^{(l)}. \quad (30)$$

To conclude, the first majorizing function we propose is

$$\begin{aligned} \bar{f}_1\left(\mathbf{x}; \mathbf{x}^{(l)}\right) &= \frac{1}{2}\lambda_{\max}(\mathbf{L})\|\mathbf{x}\|_2^4 + \lambda_u \|\mathbf{x}\|_2^2 \\ &\quad - 2\lambda_u \text{Re}\left[\mathbf{y}_1^H \mathbf{x}\right] + \text{const}. \end{aligned} \quad (31)$$

Remark 6: Recall that we have an energy budget constraint $\|\mathbf{x}\|_2^2 = c_e^2$, which means $\frac{1}{2}\lambda_{\max}(\mathbf{L})\|\mathbf{x}\|_2^4 + \lambda_u\|\mathbf{x}\|_2^2 = \text{const}$. Constructing a majorizing function like this will bring convenience to the minimization step.

2) Majorizing via a Diagonal Matrix: Here we propose an alternative way of majorization, which is based on the fact that $\mathbf{L}$ is both positive semidefinite and nonnegative. We can majorize $\sum_{k=1}^{N-1} w_k a_k |r_k(\mathbf{x})|^2 + \frac{1}{2}\mathbf{x}^H\mathbf{B}\mathbf{x}$ in another way by choosing a different $\mathbf{M}$ matrix, cf. Lemma 7 below.

Lemma 7 ([6, Lemma 5], [26, Lemma 2]): Let $\mathbf{L}$ be a real symmetric nonnegative matrix. Then, $\text{Diag}(\mathbf{L}\mathbf{1}) \succeq \mathbf{L}$.

With the help of Lemma 7, we can majorize $\sum_{k=1}^{N-1} w_k a_k|\mathbf{x}^H\mathbf{U}_k\mathbf{x}|^2 + \frac{1}{2}\mathbf{x}^H\mathbf{B}\mathbf{x}$ in another way (continuing from (23)). The key idea is to apply Lemma 4 with $\mathbf{M}_0 = \mathbf{L}$ and $\mathbf{M} = \text{Diag}(\mathbf{L}\mathbf{1})$.

Lemma 8: The expression $\sum_{k=1}^{N-1} w_k a_k|\mathbf{x}^H\mathbf{U}_k\mathbf{x}|^2 + \frac{1}{2}\mathbf{x}^H\mathbf{B}\mathbf{x}$ is majorized by

$$\frac{1}{2}\mathbf{x}^H\big(\mathbf{E}\odot(\mathbf{x}\mathbf{x}^H)\big)\mathbf{x} + \mathbf{x}^H\Big(\mathbf{R} - \mathbf{E}\odot\Big(\mathbf{x}^{(l)}\mathbf{x}^{(l)H}\Big)\Big)\mathbf{x} + \text{const}, \tag{32}$$

with $\mathbf{R}$ given in (28) and

$$\mathbf{E} = \sum_{k=1-N}^{N-1} w_k a_k (N-|k|)\mathbf{U}_{-k}. \tag{33}$$

Proof: See Appendix B for the detailed proof. ■

We apply Lemma 4 again on $\mathbf{x}^H\left(\mathbf{R} - \mathbf{E}\odot\left(\mathbf{x}^{(l)}\mathbf{x}^{(l)H}\right)\right)\mathbf{x}$ with $\mathbf{M} = \lambda_u\mathbf{I} - \lambda_l\mathbf{I}$ where λ_u and λ_l are chosen such that $\lambda_u \geq \lambda_{\max}(\mathbf{R})$ and $\lambda_l \leq \lambda_{\min}\left(\mathbf{E}\odot\left(\mathbf{x}^{(l)}\mathbf{x}^{(l)H}\right)\right)$, respectively:

$$\begin{aligned}
&\mathbf{x}^H\left(\mathbf{R} - \mathbf{E}\odot\left(\mathbf{x}^{(l)}\mathbf{x}^{(l)H}\right)\right)\mathbf{x} \\
\leq\ &(\lambda_u - \lambda_l)\mathbf{x}^H\mathbf{x} \\
&+ 2\text{Re}\Big[\mathbf{x}^H\Big(\mathbf{R} - \mathbf{E}\odot\left(\mathbf{x}^{(l)}\mathbf{x}^{(l)H}\right) \\
&- \lambda_u\mathbf{I} + \lambda_l\mathbf{I}\Big)\mathbf{x}^{(l)}\Big] + \text{const} \\
=\ &(\lambda_u - \lambda_l)\left(\|\mathbf{x}\|_2^2 - 2\text{Re}\left[\mathbf{y}_2^H\mathbf{x}\right]\right) + \text{const}.
\end{aligned} \tag{34}$$

where

$$\mathbf{y}_2 = \left(\mathbf{I} + \frac{\mathbf{E}\odot\left(\mathbf{x}^{(l)}\mathbf{x}^{(l)H}\right)}{\lambda_u - \lambda_l}\right)\mathbf{x}^{(l)} - \frac{1}{\lambda_u - \lambda_l}\mathbf{R}\mathbf{x}^{(l)}. \tag{35}$$

To conclude, the second majorizing function we propose is

$$\begin{aligned}
\bar{f}_2\left(\mathbf{x};\mathbf{x}^{(l)}\right) &= \frac{1}{2}\mathbf{x}^H\left(\mathbf{E}\odot(\mathbf{x}\mathbf{x}^H)\right)\mathbf{x} \\
&+ (\lambda_u - \lambda_l)\left(\|\mathbf{x}\|_2^2 - 2\text{Re}\left[\mathbf{y}_2^H\mathbf{x}\right]\right) + \text{const}.
\end{aligned} \tag{36}$$

Remark 9: When the strict constant modulus constraint is imposed, i.e., $|x_n| = c_m$, we have i) $\frac{1}{2}\mathbf{x}^H(\mathbf{E}\odot(\mathbf{x}\mathbf{x}^H))\mathbf{x} = \frac{1}{2}\text{Tr}((\mathbf{E}\odot(\mathbf{x}\mathbf{x}^H))\cdot(\mathbf{x}\mathbf{x}^H)) \overset{(a)}{=} \frac{1}{2}\text{Tr}(((\mathbf{x}\mathbf{x}^H)\odot(\mathbf{x}^*\mathbf{x}^T)\mathbf{E}) = \frac{c_m^4}{2}\mathbf{1}^T\mathbf{E}\mathbf{1} = \text{const}$ where (a) $\text{Tr}(\mathbf{A}^T(\mathbf{B}\odot\mathbf{C})) = \text{Tr}((\mathbf{A}^T\odot\mathbf{B}^T)\mathbf{C})$, cf. [27, Ch. 3, Section VI, Theorem 7(a)], ii) $\|\mathbf{x}\|_2^2 = c_e^2 = Nc_m^2$ and iii) $\lambda_{\min}(\mathbf{E}\odot(\mathbf{x}^{(l)}\mathbf{x}^{(l)H})) = c_m^2\lambda_{\min}(\text{Diag}(\mathbf{x}^{(l)}/c_m)\cdot\mathbf{E}\cdot\text{Diag}^H(\mathbf{x}^{(l)}/c_m)) = c_m^2\lambda_{\min}(\mathbf{E})$. It will also bring convenience to the minimization step.

B. A Majorizing Function for the ISL Metric

In particular, when $p = 2$, and $w_k = 1$ for all k, $f(\mathbf{x}) = \sum_{k=1}^{N-1}|r_k(\mathbf{x})|^2$. We are going to propose a majorizing function for this specific metric. We can easily obtain that $\sum_{k=1}^{N-1}\left|\mathbf{x}^H\mathbf{U}_k\mathbf{x}\right|^2 \overset{\mathbf{X}=\mathbf{x}\mathbf{x}^H}{=} \frac{1}{2}\text{vec}^H(\mathbf{X})\Big[\sum_{k=1-N}^{N-1}\text{vec}(\mathbf{U}_k)\text{vec}^H(\mathbf{U}_k) - \text{vec}(\mathbf{I})\text{vec}^H(\mathbf{I})\Big]\text{vec}(\mathbf{X})$. For majorization, we also need the following lemma.

Lemma 10: The following equation holds:

$$\sum_{k=1-N}^{N-1}\text{vec}(\mathbf{U}_k)\text{vec}^H(\mathbf{U}_k) = \frac{1}{2N}\sum_{i=1}^{2N}\text{vec}\left(\mathbf{f}_i\mathbf{f}_i^H\right)\text{vec}^H\left(\mathbf{f}_i\mathbf{f}_i^H\right) \tag{37}$$

where $\mathbf{f}_i = [1, \exp(j\omega_i(2-1)), \cdots, \exp(j\omega_i(N-1))]^T \in \mathbb{C}^{N\times 1}$ and $\omega_i = \frac{2\pi}{2N}(i-1)$ for $i = 1, 2, \cdots, 2N$.

From Lemma 10, we can get

$$f(\mathbf{x}) = \frac{1}{4N}\sum_{i=1}^{2N}\left|\mathbf{f}_i^H\mathbf{x}\right|^4 - \frac{1}{2}\|\mathbf{x}\|_2^4. \tag{38}$$

We observe that the first term is the summation of 4th power ($p = 4$), and we can follow the trick as is mentioned in Section IV-A. First we apply Lemma 2:

$$\frac{1}{4N}\sum_{i=1}^{2N}\left|\mathbf{f}_i^H\mathbf{x}\right|^4 \leq \frac{1}{4N}\sum_{i=1}^{2N}\Big(a_i\left|\mathbf{f}_i^H\mathbf{x}\right|^2 + b_i\left|\mathbf{f}_i^H\mathbf{x}\right| + \text{const}\Big), \tag{39}$$

where $a_i\ (\geq 0)$ and $b_i\ (\leq 0)$ follow the structure of a and b in Lemma 2, and the corresponding $\bar{x}$ is $\left(\sum_{i=1}^{2N}\left|\mathbf{f}_i^H\mathbf{x}^{(l)}\right|^4\right)^{1/4}$. With $b_i\left|\mathbf{f}_i^H\mathbf{x}\right| \leq b_i\text{Re}\left[\frac{\mathbf{x}^{(l)H}\mathbf{f}_i}{\left|\mathbf{f}_i^H\mathbf{x}^{(l)}\right|}\mathbf{f}_i^H\mathbf{x}\right]$, we further have (leaving out the constant and scaling factor in (39))

$$\begin{aligned}
&\sum_{i=1}^{2N}\Big(a_i\left|\mathbf{f}_i^H\mathbf{x}\right|^2 + b_i\left|\mathbf{f}_i^H\mathbf{x}\right|\Big) \\
\leq\ &\sum_{i=1}^{2N}\left(a_i\mathbf{x}^H\mathbf{f}_i\mathbf{f}_i^H\mathbf{x} + b_i\text{Re}\left[\frac{\mathbf{x}^{(l)H}\mathbf{f}_i}{\left|\mathbf{f}_i^H\mathbf{x}^{(l)}\right|}\mathbf{f}_i^H\mathbf{x}\right]\right) \\
=\ &\mathbf{x}^H\mathbf{G}\mathbf{x} + \text{Re}\left[\mathbf{x}^{(l)H}\mathbf{H}\mathbf{x}\right]
\end{aligned} \tag{40}$$

where $\mathbf{G} = \sum_{i=1}^{2N}a_i\mathbf{f}_i\mathbf{f}_i^H$ and $\mathbf{H} = \sum_{i=1}^{2N}\frac{b_i}{\left|\mathbf{f}_i^H\mathbf{x}^{(l)}\right|}\mathbf{f}_i\mathbf{f}_i^H$. A second majorization is still needed (using Lemma 4):

$$\begin{aligned}
&\mathbf{x}^H\mathbf{G}\mathbf{x} + \text{Re}\left[\mathbf{x}^{(l)H}\mathbf{H}\mathbf{x}\right] \overset{(a)}{\leq} \lambda_v\mathbf{x}^H\mathbf{x} \\
&- 2\text{Re}\left[\mathbf{x}^H\left(\lambda_v\mathbf{I} - \mathbf{G} - \frac{1}{2}\mathbf{H}\right)\mathbf{x}^{(l)}\right] + \text{const} \\
&\overset{(b)}{=} \lambda_v\mathbf{x}^H\mathbf{x} - 2\text{Re}\left[\mathbf{x}^H(\lambda_v\mathbf{I} - \mathbf{S})\mathbf{x}^{(l)}\right] + \text{const} \\
&\overset{(c)}{=} \lambda_v\|\mathbf{x}\|_2^2 - 2\lambda_v\text{Re}\left[\mathbf{y}_3^H\mathbf{x}\right] + \text{const}
\end{aligned} \tag{41}$$

TABLE II
PROPOSED MAJORIZING FUNCTIONS

Majorizing Function	Expression	Intended Metric
$\bar{f}_1\left(\mathbf{x};\mathbf{x}^{(l)}\right)$	$\frac{1}{2}\lambda_{\max}\left(\mathbf{L}\right)\left\|\mathbf{x}\right\|_2^4+\lambda_u\left\|\mathbf{x}\right\|_2^2-2\lambda_u\mathrm{Re}\left[\mathbf{y}_1^H\mathbf{x}\right]+\mathrm{const.}$	WPISL
$\bar{f}_2\left(\mathbf{x};\mathbf{x}^{(l)}\right)$	$\frac{1}{2}\mathbf{x}^H\left(\mathbf{E}\odot\left(\mathbf{x}\mathbf{x}^H\right)\right)\mathbf{x}+\left(\lambda_u-\lambda_l\right)\left(\left\|\mathbf{x}\right\|_2^2-2\mathrm{Re}\left[\mathbf{y}_2^H\mathbf{x}\right]\right)+\mathrm{const.}$	WPISL
$\bar{f}_3\left(\mathbf{x};\mathbf{x}^{(l)}\right)$	$-\frac{1}{2}\left\|\mathbf{x}\right\|_2^4+\frac{\lambda_u}{4N}\left\|\mathbf{x}\right\|_2^2-\frac{\lambda_u}{2N}\mathrm{Re}\left[\mathbf{y}_3^H\mathbf{x}\right]+\mathrm{const.}$	ISL

where (a) $\lambda_v \geq \lambda_{\max}\left(\mathbf{G}\right)$, (b)

$$\mathbf{G}+\frac{1}{2}\mathbf{H}=\sum_{i=1}^{2N}\left(a_i+\frac{b_i}{2\left|\mathbf{f}_i^H\mathbf{x}^{(l)}\right|}\right)\mathbf{f}_i\mathbf{f}_i^H$$
$$=\sum_{i=1}^{2N}2\left|\mathbf{f}_i^H\mathbf{x}^{(l)}\right|^2\mathbf{f}_i\mathbf{f}_i^H=\mathbf{S}, \tag{42}$$

and (c)

$$\mathbf{y}_3=\mathbf{x}^{(l)}-\tfrac{1}{\lambda_v}\mathbf{S}\mathbf{x}^{(l)}. \tag{43}$$

To conclude, the third majorizing function we propose is

$$\bar{f}_3\left(\mathbf{x};\mathbf{x}^{(l)}\right)=\frac{1}{4N}\left(\lambda_v\left\|\mathbf{x}\right\|_2^2-2\lambda_v\mathrm{Re}\left[\mathbf{y}_3^H\mathbf{x}\right]\right)$$
$$-\frac{1}{2}\left\|\mathbf{x}\right\|_2^4+\mathrm{const.} \tag{44}$$

Remark 11: The energy budget constraint $\left\|\mathbf{x}\right\|_2^2=c_e^2$ makes two terms of $\bar{f}_3\left(\mathbf{x};\mathbf{x}^{(l)}\right)$ constant.

For clarity, we summarize all the aforementioned majorizing functions and their intended metrics in Table II.

Remark 12: We would like to provide more insight into the quartic term of $\bar{f}_1$ and $\bar{f}_2$. Let's start from that of $\bar{f}_1$:

$$\begin{aligned}&\frac{1}{2}\lambda_{\max}\left(\mathbf{L}\right)\left\|\mathbf{x}\right\|_2^4\\&=\frac{1}{2}\mathbf{x}^H\left(\lambda_{\max}\left(\mathbf{L}\right)\mathbf{x}^H\mathbf{x}\right)\mathbf{x}\\&=\frac{1}{2}\mathbf{x}^H\left(\lambda_{\max}\left(\mathbf{L}\right)\cdot\mathbf{1}\mathbf{1}^T\odot\mathbf{x}^H\mathbf{x}\right)\mathbf{x}\\&=\frac{1}{2}\mathbf{x}^H\left(\mathrm{mat}\left(\lambda_{\max}\left(\mathbf{L}\right)\mathbf{1}\right)\odot\mathbf{x}^H\mathbf{x}\right)\mathbf{x},\end{aligned} \tag{45}$$

where $\mathrm{mat}\left(\mathbf{L}\mathbf{1}\right)$ reshapes a length-N^2 vector into an $N\times N$ matrix (reverse operation of $\mathrm{vec}\left(\cdot\right)$); now $\bar{f}_2$: from (73), we know $\frac{1}{2}\mathbf{x}^H\left(\mathbf{E}\odot\left(\mathbf{x}\mathbf{x}^H\right)\right)\mathbf{x}=\frac{1}{2}\mathbf{x}^H\left(\mathrm{mat}\left(\mathbf{L}\mathbf{1}\right)\odot\left(\mathbf{x}\mathbf{x}^H\right)\right)\mathbf{x}$. Recall that when constructing $\bar{f}_1$, we choose $\mathbf{M}=\lambda_{\max}(\mathbf{L})\mathbf{I}=\mathrm{Diag}(\lambda_{\max}(\mathbf{L})\mathbf{1})$ and $\bar{f}_2$, $\mathbf{M}=\mathrm{Diag}\left(\mathbf{L}\mathbf{1}\right)$. Thus, the quartic term of $\bar{f}_1$ and $\bar{f}_2$ can be unified as $\frac{1}{2}\mathbf{x}^H\left(\mathbf{Z}\odot\left(\mathbf{x}\mathbf{x}^H\right)\right)\mathbf{x}$ where $\mathbf{Z}=\mathrm{mat}\left(\mathrm{diag}\left(\mathbf{M}\right)\right)$.

V. MINIMIZATION SOLUTION DERIVATION

Now we move on to the minimization step. We should carefully choose majorizing functions so that the minimization solution can be easily obtained. In the following, we would like to specify the general constraint set $\mathcal{X}$ in (15) and discuss four sets of combinations of the constraints mentioned in Table I. They are: $\mathcal{X}_1=\left\{\mathbf{x}\in\mathbb{C}^N\middle|\left\|\mathbf{x}\right\|_2^2=c_e^2\right\}$, $\mathcal{X}_2=\left\{\mathbf{x}\in\mathbb{C}^N\middle|\left\|\mathbf{x}\right\|_2^2=c_e^2\right\}\cap C_3$, $\mathcal{X}_3=\left\{\mathbf{x}\in\mathbb{C}^N\middle|\left\|\mathbf{x}\right\|_2^2=c_e^2\right\}\cap C_2$, and $\mathcal{X}_4=\left\{\mathbf{x}\in\mathbb{C}^N\middle|\left\|\mathbf{x}\right\|_2^2=c_e^2\right\}\cap C_1\cap(C_4,\ C_5\ \text{optional})$. Let's look into the four cases one by one.

A. Fixed Energy Constraint Only ($\mathcal{X}=\mathcal{X}_1$)

In this case, we adopt majorizing function $\bar{f}_1\left(\mathbf{x};\mathbf{x}^{(l)}\right)$ for the WPISL metric, and $\bar{f}_3\left(\mathbf{x};\mathbf{x}^{(l)}\right)$ for the ISL metric. Since $\lambda_u,\lambda_v>0$ and $\left\|\mathbf{x}\right\|_2^2=c_e^2$, the minimizing problems in both scenarios reduce to:

$$\begin{aligned}\underset{\mathbf{x}}{\text{minimize}}\quad&-\mathrm{Re}\left[\mathbf{y}_j^H\mathbf{x}\right]\\\text{subject to}\quad&\left\|\mathbf{x}\right\|_2^2=c_e^2\end{aligned} \tag{46}$$

where $j=1,3$. The (update) solution is

$$\mathbf{x}^{(l+1)}=\frac{c_e}{\left\|\mathbf{y}_j\right\|_2}\mathbf{y}_j, \tag{47}$$

following the Cauchy-Schwartz inequality.

B. PAR Constraint With Fixed Energy ($\mathcal{X}=\mathcal{X}_2$)

In this case, we adopt majorizing function $\bar{f}_1\left(\mathbf{x};\mathbf{x}^{(l)}\right)$ for the WPISL metric, and $\bar{f}_3\left(\mathbf{x};\mathbf{x}^{(l)}\right)$ for the ISL metric. The minimizing problem is:

$$\begin{aligned}\underset{\mathbf{x}}{\text{minimize}}\quad&-\mathrm{Re}\left[\mathbf{y}_j^H\mathbf{x}\right]\\\text{subject to}\quad&\left\|\mathbf{x}\right\|_2^2=c_e^2\\&\left|x_n\right|\leq c_p,\ \forall n\end{aligned} \tag{48}$$

where $j=1,3$. The (update) solution is already given in [7, Algorithm 2] by using the Karush–Kuhn–Tucker (KKT) condition. The phases of $\mathbf{x}^{(l+1)}$ follow those of $\mathbf{y}_j$. Denote the number of nonzero elements of $\mathbf{y}_j$ as $M\left(\leq N\right)$, and the set containing all the corresponding indexes as $\mathcal{M}$. With $c_e/\sqrt{N}\leq c_p$, this problem always has a feasible solution.

- If $Mc_p^2<c_e^2\leq Nc_p^2$, the solution is:

$$\left|x_n^{(l+1)}\right|=\begin{cases}c_p & \forall n\in\mathcal{M},\\ \sqrt{\frac{c_e^2-Mc_p^2}{N-M}} & \forall n\notin\mathcal{M};\end{cases} \tag{49}$$

- If $c_e^2\leq Mc_p^2$, the solution is:

$$\left|\mathbf{x}^{(l+1)}\right|=\left[\beta\left|\mathbf{y}_j\right|\right]_0^{c_p} \tag{50}$$

where β satisfies $\left\|\left[\beta\left|\mathbf{y}_j\right|\right]_0^{c_p}\right\|_2=c_e$ ($|\cdot|$ denotes elementwise absolute value, $[\mathbf{x}]_{\underline{v}}^{\bar{v}}$ means projecting $\mathbf{x}$ elementwisely onto the interval $[\underline{v},\bar{v}]$) . Observing that $h_1\left(\beta\right)=\left\|\left[\beta\left|\mathbf{y}_j\right|\right]_0^{c_p}\right\|_2$ is a strictly increasing function on $\left[0,\frac{c_p}{\min_{n\in\mathcal{M}}\{|y_{j,n}|\}}\right]$, there is a unique β satisfying $h_1(\beta)=c_e$.

C. ϵ-Uncertainty Constant Modulus Constraint With Fixed Energy ($\mathcal{X} = \mathcal{X}_3$)

In this case, we adopt majorizing function $\bar{f}_1\left(\mathbf{x};\mathbf{x}^{(l)}\right)$ for the WPISL metric, and $\bar{f}_3\left(\mathbf{x};\mathbf{x}^{(l)}\right)$ for the ISL metric. The minimizing problem is:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & -\operatorname{Re}\left[\mathbf{y}_j^H\mathbf{x}\right] \\ \text{subject to} \quad & \|\mathbf{x}\|_2^2 = c_e^2 \\ & c_m - \epsilon_1 \le |x_n| \le c_m + \epsilon_2,\ \forall n \end{aligned} \tag{51}$$

where $j = 1, 3$. The (update) solution is analogous with the previous case. The phases of $\mathbf{x}^{(l+1)}$ follow those of $\mathbf{y}_j$. M and $\mathcal{M}$ follow the same definitions as above. The problem has a feasible solution if and only if $N\left(c_m-\epsilon_1\right)^2 \le c_e^2 \le N\left(c_m+\epsilon_2\right)^2$. The solution of $\left|\mathbf{x}^{(l+1)}\right|$ is:

- If $M(c_m+\epsilon_2)^2 + (N-M)(c_m-\epsilon_1)^2 < c_e^2 \le N(c_m+\epsilon_2)^2$, The solution is:

$$\left|x_n^{(l+1)}\right| = \begin{cases} c_m+\epsilon_2 & \forall n \in \mathcal{M}, \\ \sqrt{\frac{c_e^2 - M\left(c_m+\epsilon_2\right)^2}{N-M}} & \forall n \notin \mathcal{M}; \end{cases} \tag{52}$$

- If $N(c_m-\epsilon_1)^2 \le c_e^2 \le M(c_m+\epsilon_2)^2 + (N-M)(c_m-\epsilon_1)^2$, the solution is:

$$\left|\mathbf{x}^{(l+1)}\right| = \left[\beta\left|\mathbf{y}_j\right|\right]_{c_m-\epsilon_1}^{c_m+\epsilon_2} \tag{53}$$

where β satisfies $\left\|\left[\beta|\mathbf{y}_j|\right]_{c_m-\epsilon_1}^{c_m+\epsilon_2}\right\|_2 = c_e$. Observing that $h_2(\beta) = \left\|\left[\beta|\mathbf{y}_j|\right]_{c_m-\epsilon_1}^{c_m+\epsilon_2}\right\|_2$ is a strictly increasing function on $\left[\frac{c_m-\epsilon_1}{\max_{n\in\mathcal{M}}\{|y_{j,n}|\}}, \frac{c_m+\epsilon_2}{\min_{n\in\mathcal{M}}\{|y_{j,n}|\}}\right]$, there is a unique β satisfying $h_2(\beta) = c_e$.

D. Strict Constant Modulus Constraint With Miscellaneous Phase Constraints ($\mathcal{X} = \mathcal{X}_4$)

In this case, we adopt majorizing function $\bar{f}_1\left(\mathbf{x};\mathbf{x}^{(l)}\right)$ and $\bar{f}_2\left(\mathbf{x};\mathbf{x}^{(l)}\right)$ for the WPISL metric, and $\bar{f}_3\left(\mathbf{x};\mathbf{x}^{(l)}\right)$ for the ISL metric. Some of the nontrivial issues in minimizing $\bar{f}_2\left(\mathbf{x};\mathbf{x}^{(l)}\right)$ are 1) $\lambda_u - \lambda_l \ge \lambda_{\max}(\mathbf{R}) - \lambda_{\min}\left(\mathbf{E}\odot\left(\mathbf{x}^{(l)}\mathbf{x}^{(l)H}\right)\right) \overset{(a)}{=} \lambda_{\max}(\mathbf{R}) - c_m^2\lambda_{\min}(\mathbf{E}) \overset{(b)}{>} 0$ where (a) is due to Remark 9, and (b) $\mathbf{R}$ [cf. (28)] and $\mathbf{E}$ [cf. (33)] are nonzero Hermitian matrices with their diagonal entries being all zero (because $w_0 = 0$); 2) $\frac{1}{2}\mathbf{x}^H\left(\mathbf{E}\odot\left(\mathbf{x}\mathbf{x}^H\right)\right)\mathbf{x} = \text{const}$, cf. Remark 9. The minimizing problem is:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & -\operatorname{Re}\left[\mathbf{y}_j^H\mathbf{x}\right] \\ \text{subject to} \quad & |x_n| = c_m,\ \forall n \\ & \arg(x_n) \in \Phi_n,\ \forall n \end{aligned} \tag{54}$$

where $j = 1, 2, 3$, and Φ_n is a phase constraint set for x_n which will be specified later. The (update) solution is expressed elementwisely: $\forall n$,

$$x_n^{(l+1)} = c_m \exp\left(j\varphi_n\right) \tag{55}$$

with

- C_4, C_5 neither included, $\Phi_n = [0, 2\pi)$, $\varphi_n = \arg(y_{j,n})$;
- C_4 included, $\Phi_n = \{\phi_1, \phi_2, \cdots, \phi_I\}$, $\varphi_n = \arg\min_{\{\phi_i\}}\left(|\phi_i - \arg(y_{j,n})|\right)$;
- C_5 included, $\Phi_n = [\gamma_n, \gamma_n + \vartheta]$, $\varphi_n = \left[\arg(y_{j,n})\right]_{\gamma_n}^{\gamma_n+\vartheta}$.

VI. Connections With Gradient Projection Method

In this section, we are going to show the connections with the gradient projection method. A gradient projection step takes the following format:

$$\mathbf{x}^{(l+1)} = \mathcal{P}_{\mathcal{X}}\left(\mathbf{x}^{(l)} - [\text{step size}]\cdot\nabla f(\mathbf{x})\right), \tag{56}$$

where the notation $\mathcal{P}_{\mathcal{X}}$ denotes projection onto $\mathcal{X}$. The gradient of the WPISL metric $f(\mathbf{x}) = \sum_{k=1}^{N-1} w_k\left|\mathbf{x}^H\mathbf{U}_k\mathbf{x}\right|^p$ is given by

$$\nabla f(\mathbf{x}) = \mathbf{R}\mathbf{x}, \tag{57}$$

with $\mathbf{R}$ given in (28), and that of the ISL metric $f(\mathbf{x}) = \sum_{k=1}^{N-1}\left|\mathbf{x}^H\mathbf{U}_k\mathbf{x}\right|^2$ is

$$\nabla f(\mathbf{x}) = \frac{1}{4N}\mathbf{S}\mathbf{x}^{(l)} - c_e^2\mathbf{x}^{(l)}, \tag{58}$$

with $\mathbf{S}$ given in (42). We will relate the gradient projection step to our MM methods. Let's start from the following theorem.

Theorem 13: The solution of the following problem can be expressed as $\mathcal{P}_{\mathcal{X}}(\mathbf{y})$:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & -\operatorname{Re}\left[\mathbf{y}^H\mathbf{x}\right] \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{59}$$

where $\mathcal{X}$ is given in (15). For all $c > 0$, $\mathcal{P}_{\mathcal{X}}(c\mathbf{y}) = \mathcal{P}_{\mathcal{X}}(\mathbf{y})$ holds. In particular, when $\mathcal{X} = \mathcal{X}_4 = \left\{\mathbf{x}\in\mathbb{C}^N \,\middle|\, \|\mathbf{x}\|_2^2 = c_e^2\right\} \cap C_1 \cap (C_4,\ C_5 \text{ optional})$, $\mathcal{P}_{\mathcal{X}}(\mathbf{c}\odot\mathbf{y}) = \mathcal{P}_{\mathcal{X}}(\mathbf{y})$ holds with $\mathbf{c} = [c_1, c_2, \cdots, c_N]^T$ elementwise positive.

Proof: See Appendix C for the detailed proof. ■

When we minimize the majorizing function $\bar{f}_1\left(\mathbf{x};\mathbf{x}^{(l)}\right)$, the minimization solution is given as

$$\begin{aligned} \mathbf{x}^{(l+1)} &= \mathcal{P}_{\mathcal{X}}(\mathbf{y}_1) \\ &= \mathcal{P}_{\mathcal{X}}\left(\left(1 + \frac{\lambda_{\max}(\mathbf{L})}{\lambda_u}\left\|\mathbf{x}^{(l)}\right\|_2^2\right)\mathbf{x}^{(l)} - \frac{1}{\lambda_u}\mathbf{R}\mathbf{x}^{(l)}\right) \\ &= \mathcal{P}_{\mathcal{X}}\left(\mathbf{x}^{(l)} - \frac{1}{\lambda_u + \lambda_{\max}(\mathbf{L})c_e^2}\nabla f\left(\mathbf{x}^{(l)}\right)\right). \end{aligned} \tag{60}$$

When we minimize the majorizing function $\bar{f}_2\left(\mathbf{x};\mathbf{x}^{(l)}\right)$ over $\mathcal{X} = \mathcal{X}_4$, the minimization solution is

$$\begin{aligned} \mathbf{x}^{(l+1)} &= \mathcal{P}_{\mathcal{X}}(\mathbf{y}_2) \\ &= \mathcal{P}_{\mathcal{X}}\left(\left(\mathbf{I} - \frac{\mathbf{R}-\mathbf{E}\odot\left(\mathbf{x}^{(l)}\mathbf{x}^{(l)H}\right)}{\lambda_u - \lambda_l}\right)\mathbf{x}^{(l)}\right) \\ &= \mathcal{P}_{\mathcal{X}}\left(\mathbf{x}^{(l)} - \frac{1}{\lambda_u - \lambda_l}\nabla f\left(\mathbf{x}^{(l)}\right) + \frac{\mathbf{E}\odot\left(\mathbf{x}^{(l)}\mathbf{x}^{(l)H}\right)}{\lambda_u - \lambda_l}\mathbf{x}^{(l)}\right) \\ &= \mathcal{P}_{\mathcal{X}}\left(\left(\mathbf{1} + \frac{c_m^2}{\lambda_u - \lambda_l}\mathbf{E}\mathbf{1}\right)\odot\mathbf{x}^{(l)} - \frac{1}{\lambda_u - \lambda_l}\nabla f\left(\mathbf{x}^{(l)}\right)\right) \\ &\overset{(a)}{=} \mathcal{P}_{\mathcal{X}}\left(\mathbf{x}^{(l)} - \frac{1}{\lambda_u - \lambda_l}\left(\mathbf{1} + \frac{c_m^2}{\lambda_u - \lambda_l}\mathbf{E}\mathbf{1}\right)^{-1}\odot\nabla f\left(\mathbf{x}^{(l)}\right)\right), \end{aligned} \tag{61}$$

where (a) $\lambda_u - \lambda_l > 0$, $\mathbf{E}$ is a nonnegative matrix, and $(\cdot)^{-1}$ is an elementwise inverse operator. When we minimize majorizing

function $\bar{f}_3\left(\mathbf{x};\mathbf{x}^{(l)}\right)$, the minimization solution is

$$\begin{aligned}\mathbf{x}^{(l+1)} &= \mathcal{P}_{\mathcal{X}}\left(\mathbf{y}_3\right)\\ &= \mathcal{P}_{\mathcal{X}}\left(\left(\mathbf{I}-\tfrac{1}{\lambda_v}\mathbf{S}\right)\mathbf{x}^{(l)}\right)\\ &= \mathcal{P}_{\mathcal{X}}\left(\mathbf{x}^{(l)}-\tfrac{4N}{\lambda_v}\left(\nabla f\left(\mathbf{x}^{(l)}\right)+c_e^2\mathbf{x}^{(l)}\right)\right) \qquad (62)\\ &\overset{(a)}{=} \mathcal{P}_{\mathcal{X}}\left(\mathbf{x}^{(l)}-\tfrac{4N}{\lambda_v-4Nc_e^2}\nabla f\left(\mathbf{x}^{(l)}\right)\right),\end{aligned}$$

where (a) $\lambda_v > 4Nc_e^2$ and the proof is given in Appendix D.

It is very explicit that $\mathcal{P}_{\mathcal{X}}\left(\mathbf{y}_1\right)$ and $\mathcal{P}_{\mathcal{X}}\left(\mathbf{y}_3\right)$ have the same structure as the update step of the gradient projection method, but $\mathcal{P}_{\mathcal{X}}\left(\mathbf{y}_2\right)$ is different because a diagonal matrix is used during the construction of the majorizing function $\bar{f}_2$. Basically, it is a tradeoff between two concerns. The first is the simplicity of the minimization problem after majorization (we want closed-form solutions) and the second is the goodness of the majorizing function when it is used to approximate the original objective function. Depending on the constraint set, sometimes we can only use majorizing functions of the forms of $\bar{f}_1$ and $\bar{f}_3$ to get a closed-form solution in each iteration, and for some other constraint sets we may choose the majorizing function $\bar{f}_2$ which admits a closed-form solution in each iteration and at same time approximates the original objective better.

In addition, when we adopt the MM method, the update solution of projection on a nonconvex set has convergence guarantee, which results from the nature of the MM method. Moreover, the MM method proves to be superior in terms of convergence speed judging from the simulation results in the simulation section. This is because the MM step size is automatically computed from the majorizing function and it is adaptively changing, while for the traditional gradient projection method, some backtracking line search strategy needs to be used.

VII. Algorithmic Implementation

In this section, we will look into the algorithmic implementation. We observe that all the minimization solutions are closely related to $\mathbf{y}_j$'s $(j=1,2,3)$, and thus we need to compute them efficiently to reduce complexity and save CPU time.

A. Computation of $\mathbf{y}_1$

The expression of $\mathbf{y}_1$ is $\left(1+\frac{\lambda_{\max}(\mathbf{L})}{\lambda_u}\left\|\mathbf{x}^{(l)}\right\|_2^2\right)\mathbf{x}^{(l)}-\frac{1}{\lambda_u}\mathbf{R}\mathbf{x}^{(l)}$, and we need to compute $\lambda_{\max}\left(\mathbf{L}\right)$, λ_u, and $\mathbf{R}\mathbf{x}^{(l)}$. Firstly, according to [6, Lemma 2], $\lambda_{\max}\left(\mathbf{L}\right)=\max_k\left\{w_k a_k\left(N-k\right)|k=1,2,\cdots,N-1\right\}$, which is of complexity $\mathcal{O}\left(N\right)$.[1] Then, we compute $\lambda_u \geq \lambda_{\max}\left(\mathbf{R}\right)$ where $\mathbf{R}$ [cf. (28)] is a Toeplitz matrix. The value λ_u follows that in [6, Lemma 3]. The computation process needs 3 FFT(IFFT) operations. In particular, for the ISL metric, only 1 FFT operation is needed. Lastly, we compute $\mathbf{R}\mathbf{x}^{(l)}$. This process additionally requires 1 IFFT. All the relevant details are elaborated in Appendix E.

[1]The computation of a_k's is not counted for the moment. The a_k's are readily computed with (78) and (79) since they are functions of autocorrelations $r_k\left(\mathbf{x}^{(l)}\right)$.

Claim 14: The overall process of computing $\mathbf{y}_1$ takes 4 FFT(IFFT) for WPISL metric and 2 FFT(IFFT) for the ISL metric, which is of complexity $\mathcal{O}\left(N\log N\right)$.

B. Computation of $\mathbf{y}_2$

The expression of $\mathbf{y}_2$ is $\left(\mathbf{I}+\frac{\mathbf{E}\odot\left(\mathbf{x}^{(l)}\mathbf{x}^{(l)H}\right)}{\lambda_u-\lambda_l}\right)\mathbf{x}^{(l)}-\frac{1}{\lambda_u-\lambda_l}\mathbf{R}\mathbf{x}^{(l)}$, and we need to compute λ_u, λ_l, $\mathbf{R}\mathbf{x}^{(l)}$, and $\left(\mathbf{E}\odot\left(\mathbf{x}^{(l)}\mathbf{x}^{(l)H}\right)\right)\mathbf{x}^{(l)}$. With the previous subsection, we still have to additionally compute λ_l and $\left(\mathbf{E}\odot\left(\mathbf{x}^{(l)}\mathbf{x}^{(l)H}\right)\right)\mathbf{x}^{(l)}$. Because $\bar{f}_2\left(\mathbf{x};\mathbf{x}^{(l)}\right)$ is only used when each element of $\mathbf{x}^{(l)}$ has strict constant modulus c_m, we can take advantage of this extra property. Firstly, we compute $\lambda_l \leq \lambda_{\min}\left(\mathbf{E}\odot\left(\mathbf{x}^{(l)}\mathbf{x}^{(l)H}\right)\right)$ where $\mathbf{E}$ [cf. (33)] is a Toeplitz matrix. The value λ_l follows that in [6, Lemma 3]. The computation process needs 1 FFT(IFFT) operations. Then, we compute $\left(\mathbf{E}\odot\left(\mathbf{x}^{(l)}\mathbf{x}^{(l)H}\right)\right)\mathbf{x}^{(l)}$. Only 1 additional IFFT is needed per iteration. In particular, for the ISL metric, no additional FFT(IFFT) operation is needed. The details are elaborated in Appendix F.

Claim 15: The overall process of computing $\mathbf{y}_2$ takes 6 FFT(IFFT) for WPISL metric and 2 FFT(IFFT) for the ISL metric, which is of complexity $\mathcal{O}\left(N\log N\right)$.

C. Computation of $\mathbf{y}_3$

The expression of $\mathbf{y}_3$ is $\mathbf{x}^{(l)}-\frac{1}{\lambda_v}\mathbf{S}\mathbf{x}^{(l)}$, and we need to compute λ_v and $\mathbf{S}\mathbf{x}^{(l)}$. Firstly, we compute $\lambda_v \geq \lambda_{\max}\left(\mathbf{G}\right)$ where $\mathbf{G}=\sum_{i=1}^{2N}a_i\mathbf{f}_i\mathbf{f}_i^H$ is also a Toeplitz matrix. The value λ_v can also be derived from [6, Lemma 3]. The computation complexity of λ_v is $\mathcal{O}\left(N\right)$. Then, we compute $\mathbf{S}\mathbf{x}^{(l)}$ where $\mathbf{S}$ is in (42). This process requires 2 FFT(IFFT). The details are elaborated in Appendix G.

Claim 16: The overall process of computing $\mathbf{y}_3$ takes 2 FFT(IFFT) for the ISL metric, which is of complexity $\mathcal{O}\left(N\log N\right)$.

VIII. Numerical Simulations

We give some numerical results in this section. All experiments were performed on a PC with a 3.20 GHz i5-4570 CPU and 8GB RAM. We would like to specify the unified WPISL metric to be the ISL, WISL, and PSL to compare the performance with various benchmarks and the gradient projection method (if applicable). The Armijo step size rule is used to implement the gradient projection method. We initialize the algorithms with either a random sequence in the constraint set or some known sequence (Frank, Golomb, etc.). Since we are working on a nonconvex problem, initialization does affect the performance. Empirically, initiating from an existing sequence endowed with low autocorrelation sidelobes could reach good performance. However, random initialization may not do as well. The stopping criterion is: $\left|\mathrm{WPISL}\left(\mathbf{x}^{(l+1)}\right)-\mathrm{WPISL}\left(\mathbf{x}^{(l)}\right)\right|/\max\left(1,\mathrm{WPISL}\left(\mathbf{x}^{(l)}\right)\right) \leq \mathrm{Tol}$, where Tol is the tolerant precision. In case the stopping criterion is too harsh, we set the maximum number of iterations to be MaxIter.

A. ISL Minimization

Set $p=2$, $w_k=1$, $\forall k$, and we get the ISL metric. The first experiment is to present the ISL values or the merit factors $(\mathrm{MF}=\frac{N^2}{2\mathrm{ISL}})$ of different algorithms under different sequence

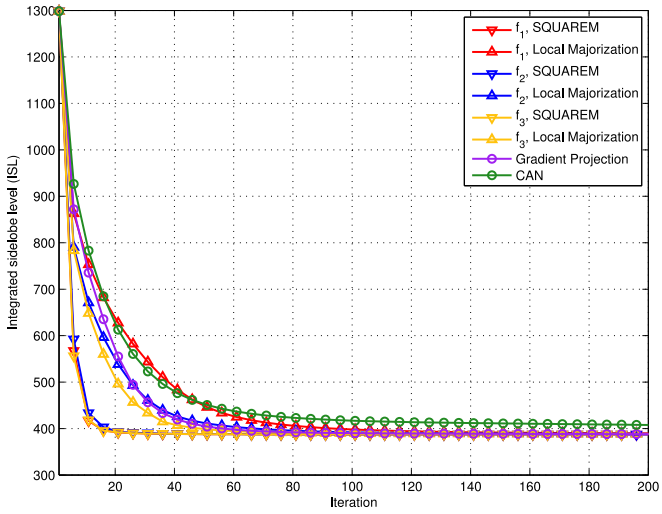


Fig. 1. ISL value versus iterations.

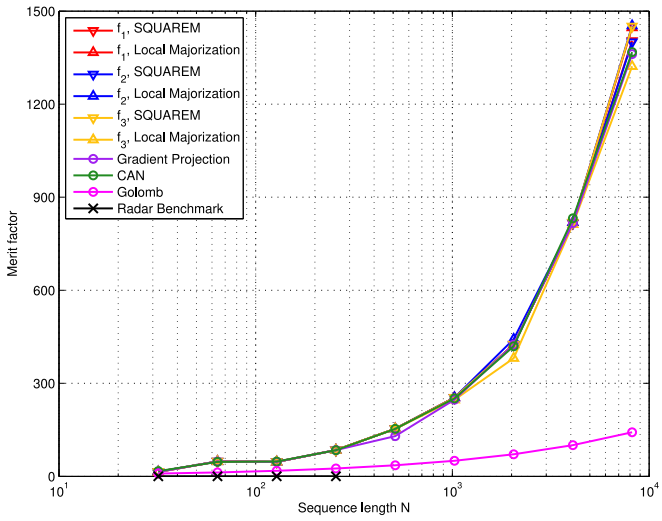


Fig. 2. Merit factor versus sequence length.

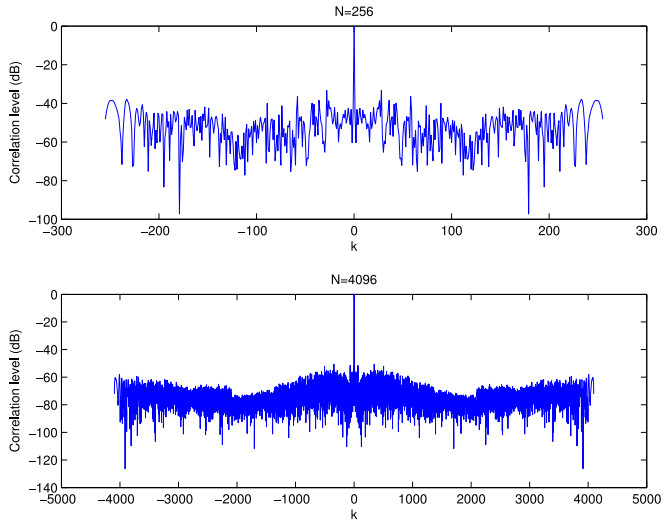


Fig. 3. Correlation level of two sequences: $N = 256$ and $N = 4096$. Majorizing function: $\bar{f}_1$, accelerated by SQUAREM.

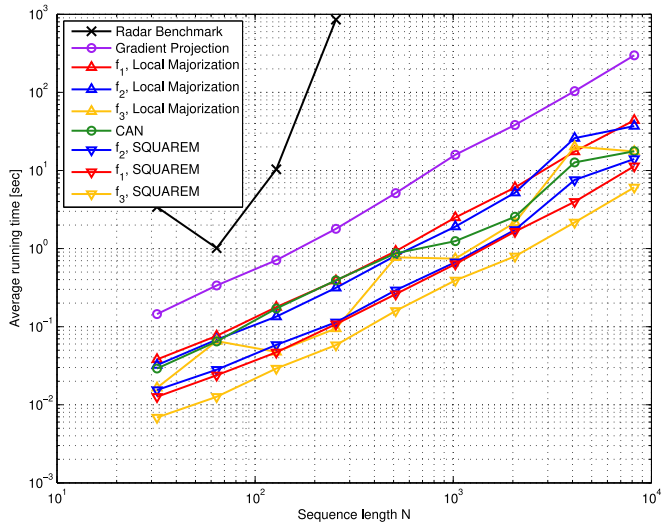


Fig. 4. Average running time versus sequence length.

lengths: $N = 2^5, 2^6, \cdots, 2^{13}$. The constraint set is (strict) unimodular constraint: $\mathcal{X} = \left\{ \mathbf{x} \in \mathbb{C}^N \,\middle|\, |x_n| = 1, \forall n \right\}$. For the proposed MM-based algorithm, we have 6 choices: majorizing functions $\bar{f}_1, \bar{f}_2, \bar{f}_3$ with 2 acceleration techniques. The benchmark is the CAN algorithm, and the gradient projection method is applicable here. The initial sequence is the Golomb sequence, $\mathrm{Tol} = 10^{-8}$, and $\mathrm{MaxIter} = 5 \times 10^4$. In Fig. 1, we show the convergence property of different algorithms. All the algorithms display a monotonic decreasing property, and the SQUAREM-accelerated MM algorithms converge the fastest. Next, we would present the results of merit factor and we additionally add a benchmark from a recent radar paper [28]. The authors provide an algorithm for designing space-time transmit code and receive filter. The algorithm alternately optimizes transmit code and receive filter, but it lacks theoretical guarantee for monotonic property and stationarity convergence though empirical convergence is observed. In Fig. 2, we see a significant increase at all sequence lengths in merit factor except the radar benchmark if the algorithms are initialized by some known sequence. Both the MM-based algorithms and CAN can achieve high merit factor, and thus they are equally good. To see the sequence correlation level, we plot two examples from the first experiment in Fig. 3 ($\text{correlation level} = 20\log_{10}|r_k/r_0|$, $k = 1-N, \cdots, N-1$.). The radar benchmark only has four data points due to its prohibitive computational cost, as will be shown later.

The second experiment is to present the average running time of different algorithms under different sequence lengths: $N = 2^5, 2^6, \cdots, 2^{13}$. The initial sequence is a random unimodular sequence and we repeat 100 times to compute the average running time. $\mathrm{Tol} = 10^{-8}$, and $\mathrm{MaxIter} = 3 \times 10^4$. The gradient projection method is implemented with the Armijo step size rule. The radar benchmark is required to initialize with some specific sequence, so only one realization is carried out. In Fig. 4, we see that the radar benchmark always takes the most CPU time and the consumption goes up to 10^3 when the sequence length is merely 256, which reflects a prohibitive computational cost in this algorithm. Also, the time consumption may not increase with the sequence length, and this is because it takes

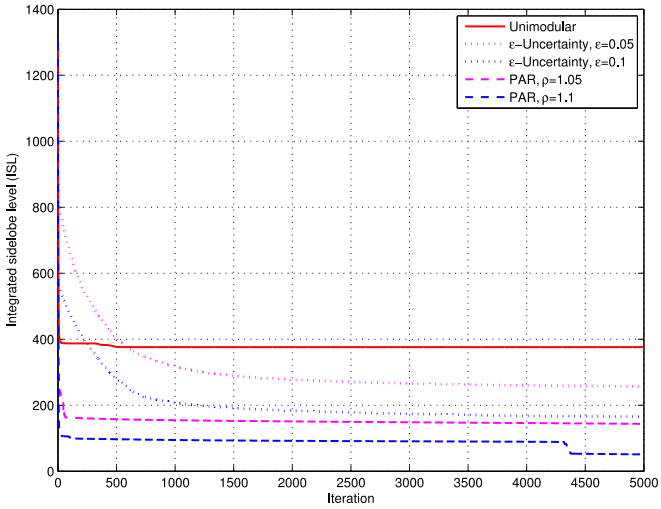


Fig. 5. ISL under different modulus constraints versus iterations.

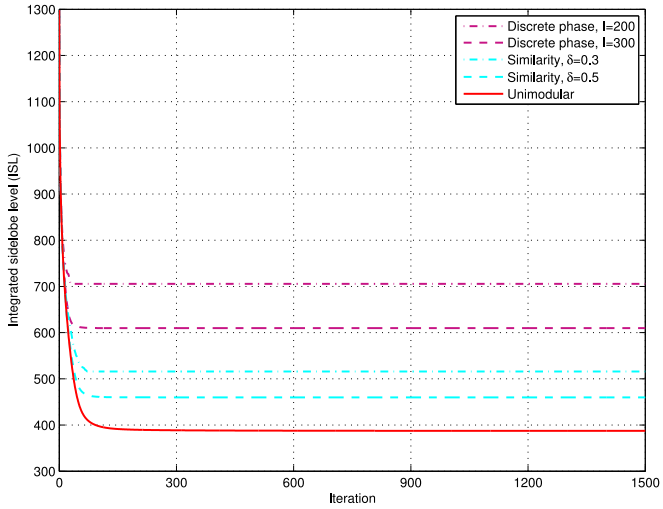


Fig. 6. ISL under different phase constraints versus iterations.

many more iterations for the $N = 2^5$ case to converge than the $N = 2^6$ case. The gradient projection method takes the second most CPU time, which is 1–2 orders of magnitude slower than the CAN and MM methods. Some of the MM-based algorithms, i.e., those accelerated by SQUAREM, are faster than the existing CAN algorithm.

Remark 17: Up to this point, we have done the comparison of our proposed algorithms and the methods in [5], [6]. It can be checked that the first two proposed majorizing functions ($\bar{f}_1$ and $\bar{f}_2$) are variants of [5], [6], with [5] proposing $\bar{f}_1$ only and [6] proposing both. However, the third majorizing function ($\bar{f}_3$) is newly proposed and not mentioned in [5], [6]. The comparison is elaborated in Figs. 1, 2, and 4. As can be seen in Figs. 1 and 2, the ISL level achieved by $\bar{f}_3$ is as good as that from $\bar{f}_1$ and $\bar{f}_2$, which indicates that we can get as good results as [5], [6]. Moreover, in Fig. 4, we observe that the convergence speed of $\bar{f}_3$ is the fastest, 2–4 times as fast as [5] (proposing only $\bar{f}_1$) and 2–6 times as fast as [6] (proposing $\bar{f}_1$ and $\bar{f}_2$). In terms of time consumption, we can achieve even better results than [5], [6].

The third experiment is to show the ISL under miscellaneous constraints, i.e., modulus constraints and phase constraints. In terms of modulus constraints, we have three different types: strict constant modulus constraint ($|x_n| = c_m$), ϵ-uncertainty constant modulus constraint ($c_m - \epsilon_1 \leq |x_n| \leq c_m + \epsilon_2$), and PAR constraint ($|x_n| \leq c_p$). We set $N = 256$, $c_m = 1$, $\epsilon_1 = \epsilon_2 = \epsilon$, and $c_p = c_m + \epsilon_2 = 1 + \epsilon$. We use MM-based algorithm to compute the ISL, using majorizing function $\bar{f}_1$ and SQUAREM acceleration. The initial sequence is the Golomb sequence, $\mathrm{Tol} = -10^{-8}$ (stopping criterion is deactivated), and $\mathrm{MaxIter} = 5 \times 10^3$. In Fig. 5, we relax the unimodular constraint gradually as shown, and there is remarkable ISL decrease even if the constraint set is relaxed a little bit. In terms of phase constraints, we also have three different types: no phase constraint, discrete phase constraint ($\arg(x_n) \in \{\phi_1, \phi_2, \cdots, \phi_I\}$), similarity constraint ($\arg(x_n) \in [\gamma_n, \gamma_n + \vartheta]$, $\gamma_n = \arg(x_{r,n}) - \arccos\left(1 - \delta^2/2\right)$, and $\vartheta = 2\arccos\left(1 - \delta^2/2\right)$). Note that all the phase constraints are accompanied by the strict constant modulus constraint. We set $N = 256$, $c_m = 1$, $\phi_i = \frac{2\pi}{I}(i-1)$, $i = 1, \cdots, I$, and the reference sequence $\mathbf{x}_r$ is the Golomb sequence. The MM-based algorithm is based on majorizing function $\bar{f}_1$ and local majorization acceleration. The initial sequence is the Golomb sequence, $\mathrm{Tol} = -10^{-8}$, and $\mathrm{MaxIter} = 1.5 \times 10^3$. In Fig. 6, we tightened the unimodular constraint gradually as shown, and we observe an increase in ISL. Analogous result can be achieved when we optimize under other metrics (WISL and PSL), and thus we do not replicate.

B. WISL Minimization

Set $p = 2$ only, and we get the WISL metric. We do one experiment is to present the convergence speed of different algorithms under $N = 100$. The weight is

$$w_k = \begin{cases} 1 & k \in \{1, \cdots, 20\} \cup \{51, \cdots, 70\} \\ 0 & \text{otherwise,} \end{cases} \tag{63}$$

such that only $r_1 \sim r_{20}$ and $r_{51} \sim r_{70}$ have small correlations. The constraint set is unimodular constraint. For the proposed MM-based algorithm, we have 4 choices: majorizing functions $\bar{f}_1, \bar{f}_2$ with 2 acceleration techniques. The benchmark is the WeCAN algorithm, and the gradient projection method is applicable here. The initial sequence is a random unimodular sequence, $\mathrm{Tol} = 10^{-10}$, and $\mathrm{MaxIter} = 10^6$. In Fig. 7, we show the monotonic decreasing property plotting WISL versus iterations. In Fig. 8, we see that all the algorithms can achieve WISL down to 10^{-10}, but all the MM-based algorithms are much faster than the gradient projection method and the existing WeCAN algorithm. To see the sequence correlation level, we refer to Fig. 9 to see one example of the converged sequence. We see that the designed sequences have low autocorrelation sidelobes at the required lags.

C. PSL Minimization

Set p to be large, $w_k = 1$, $\forall k$, and we approximately get the PSL metric. The first experiment is to present the PSL of different algorithms under different sequence lengths: $N = 10^2, 20^2, 40^2, 60^2, 80^2, 100^2$. We set $p = 100$. The constraint set is unimodular constraint. For the proposed MM-based algorithm, we have 4 choices: majorizing functions $\bar{f}_1, \bar{f}_2$ with 2 acceleration techniques. We do not have benchmarks here and

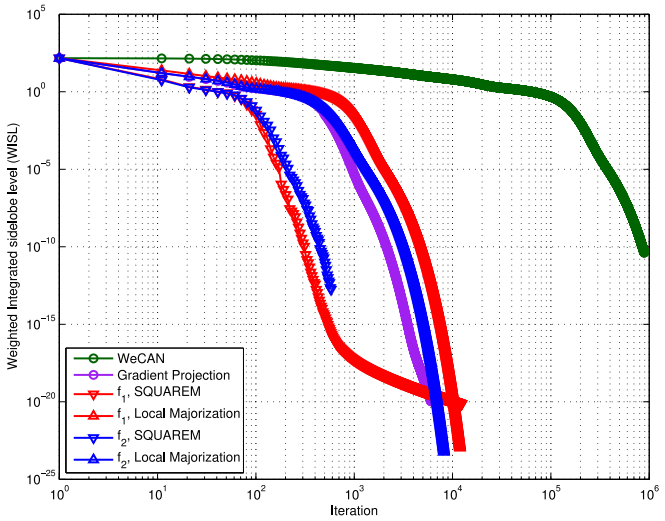


Fig. 7. Weighted ISL versus iterations.

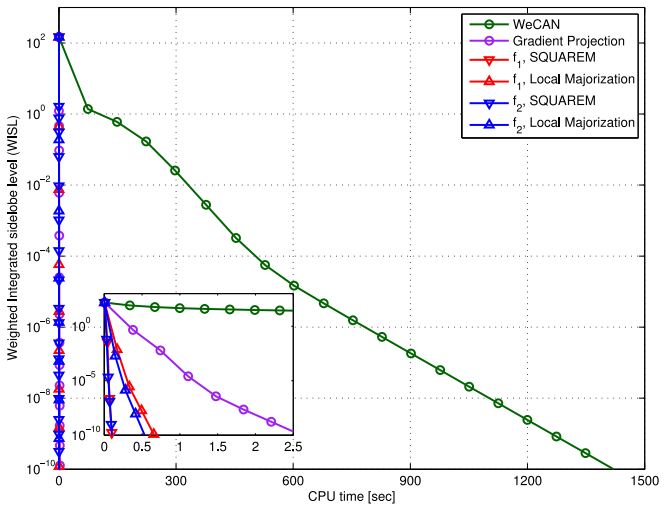


Fig. 8. Weighted ISL versus CPU time.

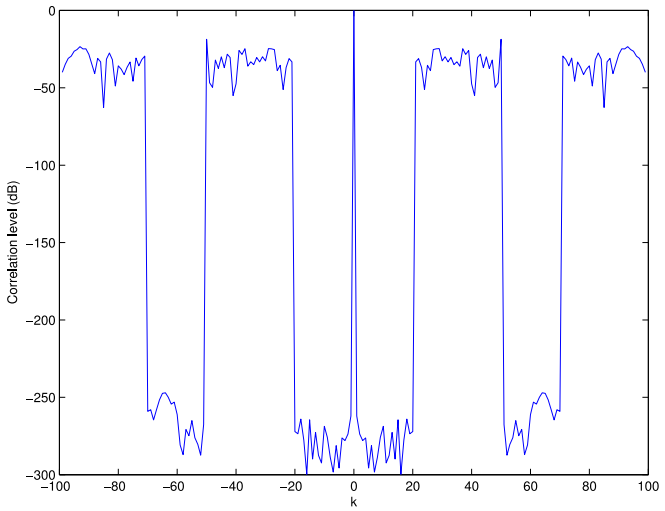


Fig. 9. Correlation level of the designed sequence. Majorizing function: $\bar{f}_1$, accelerated by SQUAREM.

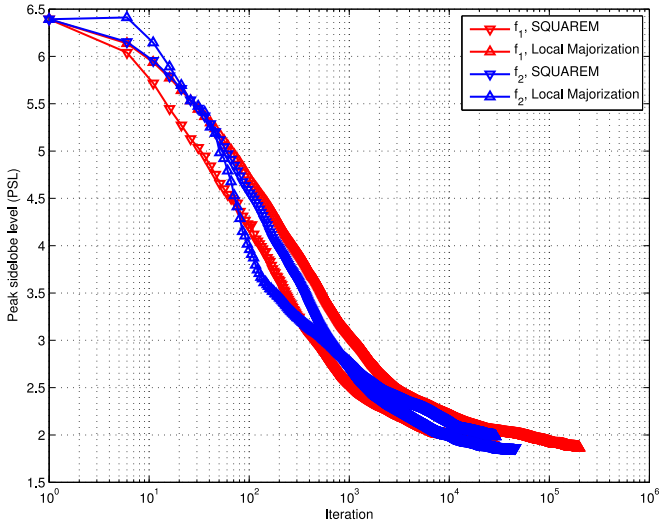


Fig. 10. PSL value versus iterations.

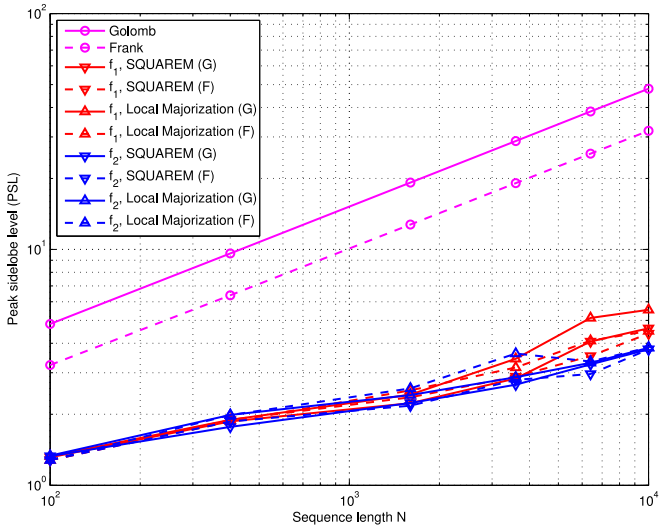


Fig. 11. PSL versus sequence length. (G): initialized by the Golomb sequence, and (F): initialized by the Frank sequence.

the gradient projection method is not applicable because of the numerical issue caused by large value of p. The initial sequences are the Golomb sequence and the Frank sequence, $\mathrm{Tol} = 10^{-10}$, and $\mathrm{MaxIter} = 2 \times 10^5$. In Fig. 10, we show the monotonic decreasing property plotting PSL versus iterations when $N = 400$ and the initial sequence is the Frank sequence. In Fig. 11, we see a significant decrease in PSL if the algorithm is initialized by some known sequence.

The second experiment is to present the convergence speed of different values of p: $p = 10, 10^2, 10^3, 10^4$. The initial sequence is the Frank sequence, $N = 400$, $\mathrm{Tol} = -10^{-10}$, and $\mathrm{MaxIter} = 5 \times 10^4$. In Fig. 12, we see that when p is small ($p = 10$), the convergence speed is fast, but the converged PSL is high; when p is large ($p = 10000$), the situation is right the opposite. Therefore, we can adopt an increasing scheme of p to get low PSL and fast convergence speed. We increase p as $2, 2^2, \cdots, 2^{13}$. The initial sequence is the Frank sequence with $N = 100^2$, $\mathrm{Tol} = 10^{-5}/p$ for each p, and $\mathrm{MaxIter} = 5 \times 10^3$.

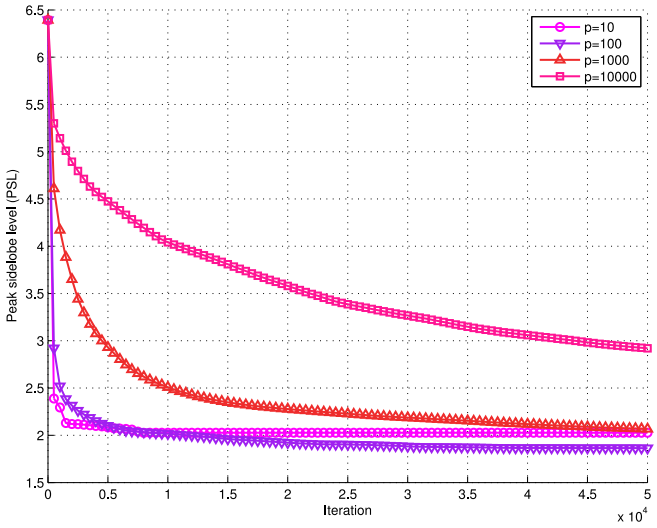


Fig. 12. PSL versus iterations (majorizing function: $\bar{f}_1$, accelerated by SQUAREM).

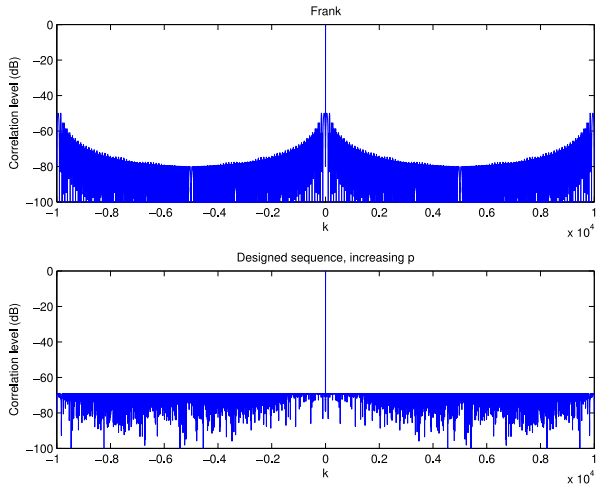


Fig. 13. Correlation level of the Frank and designed sequence (majorizing function: $\bar{f}_1$, accelerated by SQUAREM).

We plot the sequence correlation level in Fig. 13 compared with the initial Frank sequence. The autocorrelation sidelobes of the Frank sequence become larger for k close to 0 and $N-1$, while those of the designed sequence are much more uniform across all lags.

D. Metric Comparison: ISL and PSL

In this subsection, we compare the results of different metrics, and we focus on ISL and PSL. The designed sequence is unimodular. Note that it is not a good idea to optimize the weighted sum of ISL and PSL because in this case the pth order term in PSL will dominate the objective, implicitly suppressing the effect of the ISL term. If we want to achieve a tradeoff between ISL and PSL, we may as well tune the order parameter p. In Fig. 14, we plot the optimized ISL and PSL value with respect to different p, ranging from 2 to 128. The initial sequence is chosen as the Frank sequence, $N = 256$, $\mathrm{Tol} = 10^{-10}$, and $\mathrm{MaxIter} = 3 \times 10^4$. The majorizing function is chosen as $\bar{f}_1$, accelerated by SQUAREM. The best tradeoff is achieved when

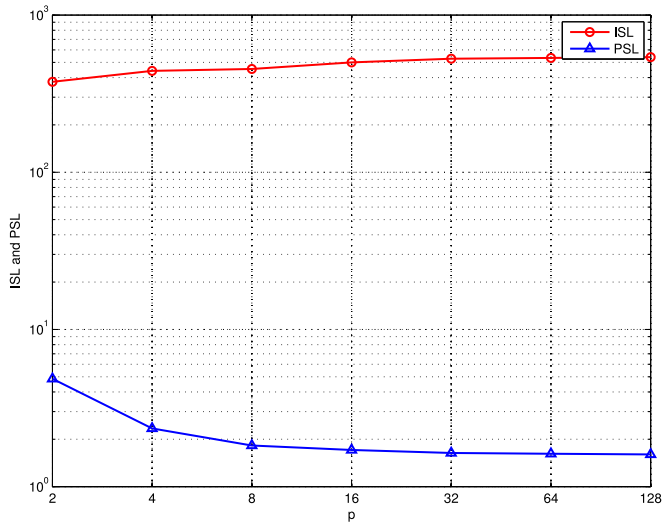


Fig. 14. (Optimized) ISL and PSL levels versus order parameter p.

p lies between 8 and 16, where the level of PSL is already low enough and that of ISL is still not too high.

IX. Conclusion

We have proposed a unified framework to design low autocorrelation sequences. We have optimized a unified metric over a general constraint set. We have carried out the MM method in two stages. In the majorizing function construction stage, we have constructed three majorizing functions. Two of them apply to the unified WPISL metric, and the remaining one applies to the specific ISL metric. In the minimization solution derivation stage, we have provided closed-form solutions to different minimization problems. Additionally, we have shown the connections between the MM and gradient projection method under our algorithmic scheme. Thereafter, we efficiently implement the MM step with FFT (IFFT) operations. Numerical simulations have shown that the proposed MM-based algorithms converge faster than the traditional gradient projection method and the state-of-the-art algorithms.

Appendix A
Proof of Lemma 5

Proof: We apply Lemma 4 with $\mathbf{M}_0 = \mathbf{L}$ and $\mathbf{M} = \lambda_{\max}(\mathbf{L})\mathbf{I}$, obtaining (constant terms are represented with const for simplicity)

$$\begin{aligned}&\frac{1}{2}\mathrm{vec}^H(\mathbf{X})\cdot\mathbf{L}\cdot\mathrm{vec}(\mathbf{X})\\ &\le \frac{1}{2}\lambda_{\max}(\mathbf{L})\,\mathrm{vec}^H(\mathbf{X})\,\mathrm{vec}(\mathbf{X}) + \mathrm{Re}\Big[\mathrm{vec}^H\left(\mathbf{X}^{(l)}\right)\cdot\\ &\quad(\mathbf{L}-\lambda_{\max}(\mathbf{L})\mathbf{I})\,\mathrm{vec}(\mathbf{X})\Big] + \mathrm{const}.\end{aligned} \tag{64}$$

Now we recover $\mathbf{X}$ and $\mathbf{X}^{(l)}$ as $\mathbf{x}\mathbf{x}^H$ and $\mathbf{x}^{(l)}\mathbf{x}^{(l)H}$, respectively:

$$\begin{aligned}&\sum_{k=1}^{N-1} w_k a_k \left|\mathbf{x}^H\mathbf{U}_k\mathbf{x}\right|^2 \le \frac{1}{2}\lambda_{\max}(\mathbf{L})\,\|\mathbf{x}\|_2^4\\ &+\mathbf{x}^H\left(\mathbf{A}-\lambda_{\max}(\mathbf{L})\,\mathbf{x}^{(l)}\mathbf{x}^{(l)H}\right)\mathbf{x}+\mathrm{const},\end{aligned} \tag{65}$$

where $\mathbf{A}$ is defined as

$$\mathbf{A}=\sum_{k=1-N}^{N-1} w_k a_k r_{-k}\left(\mathbf{x}^{(l)}\right)\mathbf{U}_k=\mathbf{A}^H. \tag{66}$$

We add $\frac{1}{2}\mathbf{x}^H\mathbf{B}\mathbf{x}$ to (65):

$$\sum_{k=1}^{N-1} w_k a_k \left|r_k\left(\mathbf{x}\right)\right|^2+\frac{1}{2}\mathbf{x}^H\mathbf{B}\mathbf{x}\le\frac{1}{2}\lambda_{\max}\left(\mathbf{L}\right)\left\|\mathbf{x}\right\|_2^4 + \mathbf{x}^H\left(\mathbf{A}+\frac{1}{2}\mathbf{B}-\lambda_{\max}\left(\mathbf{L}\right)\mathbf{x}^{(l)}\mathbf{x}^{(l)H}\right)\mathbf{x}+\text{const}, \tag{67}$$

where $\mathbf{A}+\frac{1}{2}\mathbf{B}$ can be further simplified [cf. (66) and (24)]:

$$\begin{aligned}\mathbf{A}+\frac{1}{2}\mathbf{B}&=\sum_{k=1-N}^{N-1} w_k\left(a_k+\frac{b_k}{2\left|r_{-k}\left(\mathbf{x}^{(l)}\right)\right|}\right)r_{-k}\left(\mathbf{x}^{(l)}\right)\mathbf{U}_k\\&=\sum_{k=1-N}^{N-1} w_k\left(a_k+\frac{p\left|r_k\left(\mathbf{x}^{(l)}\right)\right|^{p-1}-2a_k\left|r_k\left(\mathbf{x}^{(l)}\right)\right|}{2\left|r_{-k}\left(\mathbf{x}^{(l)}\right)\right|}\right)r_{-k}\left(\mathbf{x}^{(l)}\right)\mathbf{U}_k\\&=\sum_{k=1-N}^{N-1}\frac{p}{2}w_k\left|r_k\left(\mathbf{x}^{(l)}\right)\right|^{p-2} r_{-k}\left(\mathbf{x}^{(l)}\right)\cdot\mathbf{U}_k=\mathbf{R}.\end{aligned} \tag{68}$$

■

APPENDIX B
PROOF OF LEMMA 8

Proof: We start from the following:

$$\begin{aligned}&\frac{1}{2}\text{vec}^H\left(\mathbf{X}\right)\cdot\mathbf{L}\cdot\text{vec}\left(\mathbf{X}\right)\\&\le\frac{1}{2}\text{vec}^H\left(\mathbf{X}\right)\text{Diag}\left(\mathbf{L}\mathbf{1}\right)\text{vec}\left(\mathbf{X}\right)+\text{Re}\Big[\text{vec}^H\left(\mathbf{X}^{(l)}\right)\cdot\\&\left(\mathbf{L}-\text{Diag}\left(\mathbf{L}\mathbf{1}\right)\right)\text{vec}\left(\mathbf{X}\right)\Big]+\text{const}.\end{aligned} \tag{69}$$

Then we recover $\mathbf{X}$ and $\mathbf{X}^{(l)}$ as $\mathbf{x}\mathbf{x}^H$ and $\mathbf{x}^{(l)}\mathbf{x}^{(l)H}$, respectively. We do this term by term:

$$\begin{aligned}&\frac{1}{2}\text{vec}^H\left(\mathbf{X}\right)\text{Diag}\left(\mathbf{L}\mathbf{1}\right)\text{vec}\left(\mathbf{X}\right)\\&=\frac{1}{2}\text{vec}^H\left(\mathbf{X}\right)\left(\left(\mathbf{L}\mathbf{1}\right)\odot\text{vec}\left(\mathbf{X}\right)\right)\\&\overset{(a)}{=}\frac{1}{2}\text{Tr}\left(\mathbf{x}\mathbf{x}^H\text{mat}\left(\left(\mathbf{L}\mathbf{1}\right)\odot\text{vec}\left(\mathbf{X}\right)\right)\right)\\&=\frac{1}{2}\mathbf{x}^H\left(\text{mat}\left(\mathbf{L}\mathbf{1}\right)\odot\left(\mathbf{x}\mathbf{x}^H\right)\right)\mathbf{x}\\&\overset{(b)}{=}\frac{1}{2}\mathbf{x}^H\left(\mathbf{E}\odot\left(\mathbf{x}\mathbf{x}^H\right)\right)\mathbf{x},\end{aligned} \tag{70}$$

$$\text{Re}\left[\text{vec}^H\left(\mathbf{X}^{(l)}\right)\mathbf{L}\text{vec}\left(\mathbf{X}\right)\right]=\mathbf{x}^H\mathbf{A}\mathbf{x}, \tag{71}$$

$$\begin{aligned}&\text{Re}\left[\text{vec}^H\left(\mathbf{X}^{(l)}\right)\text{Diag}\left(\mathbf{L}\mathbf{1}\right)\text{vec}\left(\mathbf{X}\right)\right]\\&\overset{(c)}{=}\mathbf{x}^H\left(\mathbf{E}\odot\left(\mathbf{x}^{(l)}\mathbf{x}^{(l)H}\right)\right)\mathbf{x}\end{aligned} \tag{72}$$

where (a) $\text{mat}(\cdot)$ is the inverse operation of $\text{vec}(\cdot)$, (b)

$$\begin{aligned}\mathbf{E}&=\text{mat}\left(\mathbf{L}\mathbf{1}\right)=\text{mat}\left(\sum_{k=1-N}^{N-1} w_k a_k\text{vec}\left(\mathbf{U}_{-k}\right)\text{vec}^H\left(\mathbf{U}_{-k}\right)\mathbf{1}\right)\\&=\text{mat}\left(\sum_{k=1-N}^{N-1} w_k a_k\left(N-|k|\right)\text{vec}\left(\mathbf{U}_{-k}\right)\right)\\&=\sum_{k=1-N}^{N-1} w_k a_k\left(N-|k|\right)\mathbf{U}_{-k}=\mathbf{E}^H,\end{aligned} \tag{73}$$

and (c) follow (70). Therefore,

$$\sum_{k=1}^{N-1} w_k a_k\left|\mathbf{x}^H\mathbf{U}_k\mathbf{x}\right|^2\le\frac{1}{2}\mathbf{x}^H\left(\mathbf{E}\odot\left(\mathbf{x}\mathbf{x}^H\right)\right)\mathbf{x} + \mathbf{x}^H\left(\mathbf{A}-\mathbf{E}\odot\left(\mathbf{x}^{(l)}\mathbf{x}^{(l)H}\right)\right)\mathbf{x}+\text{const}. \tag{74}$$

We add $\frac{1}{2}\mathbf{x}^H\mathbf{B}\mathbf{x}$ to (74):

$$\sum_{k=1}^{N-1} w_k a_k\left|r_k\left(\mathbf{x}\right)\right|^2+\frac{1}{2}\mathbf{x}^H\mathbf{B}\mathbf{x}\le\frac{1}{2}\mathbf{x}^H\left(\mathbf{E}\odot\left(\mathbf{x}\mathbf{x}^H\right)\right)\mathbf{x} + \mathbf{x}^H\left(\mathbf{R}-\mathbf{E}\odot\left(\mathbf{x}^{(l)}\mathbf{x}^{(l)H}\right)\right)\mathbf{x}+\text{const}. \tag{75}$$

■

APPENDIX C
PROOF OF THEOREM 13

Proof: The problem (59) is equivalent to the following one due to the blanket constraint $\|\mathbf{x}\|_2^2=c_e^2$:

$$\begin{aligned}\underset{\mathbf{x}}{\text{minimize}}\quad&\frac{1}{2}\left\|\mathbf{x}-\mathbf{y}\right\|_2^2\\\text{subject to}\quad&\mathbf{x}\in\mathcal{X},\end{aligned} \tag{76}$$

which can be interpreted as a projection problem.

Judging from (59), it is obvious that $\mathcal{P}_{\mathcal{X}}\left(c\mathbf{y}\right)=\mathcal{P}_{\mathcal{X}}\left(\mathbf{y}\right)$ for $c>0$. When $\mathcal{X}=\mathcal{X}_4$, we simplify it as $\mathcal{X}_4=\left\{\mathbf{x}\in\mathbb{C}^N\middle|\,|x_n|=c_m,\,\arg\left(x_n\right)\in\Phi_n,\forall n\right\}$ where Φ_n is a phase constraint set for x_n. We express $\mathcal{P}_{\mathcal{X}}\left(\mathbf{y}\right)$ elementwisely as $\left[\mathcal{P}_{\mathcal{X}}\left(\mathbf{y}\right)\right]_n=c_m\exp\left(j\varphi_n\left(y_n\right)\right),\forall n$.

- C_4, C_5 neither included, $\Phi_n=[0,2\pi)$, $\varphi_n(y_n)=\arg(y_n)=\arg(c_n y_n)=\varphi_n(c_n y_n)$, thus $[\mathcal{P}_{\mathcal{X}}(\mathbf{y}]_n=[\mathcal{P}_{\mathcal{X}}(\text{Diag}(\mathbf{c})\mathbf{y})]_n$;
- C_4 included, $\Phi_n=\{\phi_1,\phi_2,\cdots,\phi_I\}$, $\varphi_n(y_n)=\text{argmin}_{\{\phi_i\}}(|\phi_i-\arg(y_n)|)=\arg\min_{\{\phi_i\}}(|\phi_i-\arg(c_n y_n)|)=\varphi_n(c_n y_n)$, thus $[\mathcal{P}_{\mathcal{X}}(\mathbf{y})]_n=[\mathcal{P}_{\mathcal{X}}(\text{Diag}(\mathbf{c})\mathbf{y})]_n$;
- C_5 included, $\Phi_n=[\gamma_n,\gamma_n+\vartheta]$, $\varphi_n(y_n)=[\arg(y_n)]_{\gamma_n}^{\gamma_n+\vartheta}=[\arg(c_n y_n)]_{\gamma_n}^{\gamma_n+\vartheta}=\varphi_n(c_n y_n)$, thus $[\mathcal{P}_{\mathcal{X}}(\mathbf{y})]_n=[\mathcal{P}_{\mathcal{X}}(\text{Diag}(\mathbf{c})\mathbf{y})]_n$.

Therefore, $\mathcal{P}_{\mathcal{X}}\left(\mathbf{c}\odot\mathbf{y}\right)=\mathcal{P}_{\mathcal{X}}\left(\mathbf{y}\right)$ for any $\mathbf{c}$ elementwise positive. ■

APPENDIX D
PROOF OF THE CLAIM $\lambda_v>4Nc_e^2$

Proof: The proof is given as follows: $\lambda_v=\frac{1}{2}\left(\max_{1\le i\le N}2Na_{2i}+\max_{1\le i\le N}2Na_{2i-1}\right)\ge\sum_{i=1}^{2N}a_i=\sum_{i=1}^{2N}$

$\frac{\bar{x}^4-|\mathbf{f}_i^H\mathbf{x}^{(l)}|^4-4|\mathbf{f}_i^H\mathbf{x}^{(l)}|^3(\bar{x}-|\mathbf{f}_i^H\mathbf{x}^{(l)}|)}{(\bar{x}-|\mathbf{f}_i^H\mathbf{x}^{(l)}|)^2}$ with $\bar{x}=(\sum_{i=1}^{2N}|\mathbf{f}_i^H\mathbf{x}^{(l)}|^4)^{1/4}$. We know that $\frac{\bar{x}^4-|\mathbf{f}_i^H\mathbf{x}^{(l)}|^4-4|\mathbf{f}_i^H\mathbf{x}^{(l)}|^3(\bar{x}-|\mathbf{f}_i^H\mathbf{x}^{(l)}|)}{(\bar{x}-|\mathbf{f}_i^H\mathbf{x}^{(l)}|)^2}=3|\mathbf{f}_i^H\mathbf{x}^{(l)}|^2+2|\mathbf{f}_i^H\mathbf{x}^{(l)}|\bar{x}+\bar{x}^2\geq 3|\mathbf{f}_i^H\mathbf{x}^{(l)}|^2+\bar{x}^2$. Also, $\bar{x}=(\sum_{i=1}^{2N}|\mathbf{f}_i^H\mathbf{x}^{(l)}|^4)^{1/4}\geq(2N)^{1/4}(\frac{1}{2N}\sum_{i=1}^{2N}|\mathbf{f}_i^H\mathbf{x}^{(l)}|^2)^{1/2}=(2N)^{1/4}c_e$. Then, $\lambda_v\geq 3\sum_{i=1}^{2N}|\mathbf{f}_i^H\mathbf{x}^{(l)}|^2+2N\sqrt{2N}c_e^2=(6N+2N\sqrt{2N})c_e^2>4Nc_e^2$. ■

APPENDIX E
COMPUTATION PROCESS OF λ_u AND $\mathbf{R}\mathbf{x}^{(l)}$

From [29] and [6, Lemma 3], we see that in order to satisfy $\lambda_u\geq\lambda_{\max}(\mathbf{R})$, λ_u can be chosen as $\frac{1}{2}(\max_{1\leq i\leq N}\mu_{2i}+\max_{1\leq i\leq N}\mu_{2i-1})$ with $\boldsymbol{\mu}=\mathbf{F}\mathbf{c}$ where $\mathbf{F}\in\mathbb{C}^{2N\times 2N}$ is a $2N-$DFT matrix with $\mathbf{F}_{mn}=\exp(-j\frac{2\pi(m-1)(n-1)}{2N})$, $1\leq m,n\leq 2N$, and $\mathbf{c}=[c_0,c_1,\cdots,c_{2N-1}]^T$ with

$$c_k=\begin{cases}0 & k=0,N\\ \frac{p}{2}w_k\left|\mathbf{x}^{(l)H}\mathbf{U}_k\mathbf{x}^{(l)}\right|^{p-2}\cdot & \\ \left(\mathbf{x}^{(l)H}\mathbf{U}_k\mathbf{x}^{(l)}\right) & k=1,\cdots,N-1\\ c_{2N-k}^* & k=N+1,\cdots,2N-1.\end{cases}\tag{77}$$

In short, $\boldsymbol{\mu}$ can be computed as $\boldsymbol{\mu}=\mathrm{FFT}(\mathbf{c})$. Moreover, the vector $\mathbf{c}$ can also be computed with FFT. We define $\mathbf{r}=\left[r_0\left(\mathbf{x}^{(l)}\right),r_1\left(\mathbf{x}^{(l)}\right),\cdots,r_{N-1}\left(\mathbf{x}^{(l)}\right),0,r_{N-1}^*\left(\mathbf{x}^{(l)}\right),\cdots,r_1^*\left(\mathbf{x}^{(l)}\right)\right]^T$. Since $\mathbf{r}$ is comprised of the autocorrelations of $\mathbf{x}^{(l)}$, it can be efficiently computed via FFT:

$$\mathbf{r}=\mathrm{IFFT}\left(|\mathbf{t}|^2\right),\tag{78}$$

where

$$\mathbf{t}=\mathrm{FFT}\left(\left[\mathbf{x}^{(l)T},\mathbf{0}_{N\times 1}^T\right]^T\right)\tag{79}$$

and $|\cdot|^p$ denotes the elementwise absolute value to the pth power. Here we need 1 FFT and 1 IFFT. Knowing $\mathbf{r}$, we can compute $\mathbf{c}$ with simple Hadamard product:

$$\mathbf{c}=\frac{p}{2}\mathbf{w}\odot|\mathbf{r}|^{p-2}\odot\mathbf{r},\tag{80}$$

where $\mathbf{w}=[0,w_1,\cdots,w_{N-1},0,w_{N-1},\cdots,w_1]^T$, of complexity $\mathcal{O}(N)$. To sum up, the computation process is:

$$\begin{aligned}&\mathbf{t}=\mathrm{FFT}\left(\left[\mathbf{x}^{(l)T},\mathbf{0}_{N\times 1}^T\right]^T\right)\\ &\mathbf{r}=\mathrm{IFFT}\left(|\mathbf{t}|^2\right)\\ &\boldsymbol{\mu}=\mathrm{FFT}\left(\frac{p}{2}\mathbf{w}\odot|\mathbf{r}|^{p-2}\odot\mathbf{r}\right)\\ &\lambda_u=\frac{1}{2}\left(\max_{1\leq i\leq N}\mu_{2i}+\max_{1\leq i\leq N}\mu_{2i-1}\right),\end{aligned}\tag{81}$$

3 FFT(IFFT) operations. When $p=2$ and $w_k=1,\forall k$, i.e., for the ISL metric, we have $\boldsymbol{\mu}=|\mathbf{t}|^2$ and only 1 FFT operation is needed.

The Toeplitz matrix can be expressed as $\mathbf{R}=\frac{1}{2N}\mathbf{F}_{:,1:N}^H\mathrm{Diag}(\boldsymbol{\mu})\mathbf{F}_{:,1:N}$ ($\mathbf{F}_{:,1:N}$ stands for the first N columns of $\mathbf{F}$, the $2N-$DFT matrix). Then,

$$\begin{aligned}\mathbf{R}\mathbf{x}^{(l)}&=\left[\frac{1}{2N}\mathbf{F}^H\mathrm{Diag}(\boldsymbol{\mu})\mathbf{F}\begin{bmatrix}\mathbf{x}^{(l)}\\ \mathbf{0}_{N\times 1}\end{bmatrix}\right]_{1:N}\\ &=\left[\mathrm{IFFT}(\boldsymbol{\mu}\odot\mathbf{t})\right]_{1:N},\end{aligned}\tag{82}$$

where $[\cdot]_{1:N}$ means taking the first N elements of a vector, and $\mathbf{t}$ follows (79). Here we additionally need 1 IFFT.

APPENDIX F
COMPUTATION PROCESS OF λ_l AND $(\mathbf{E}\odot(\mathbf{x}^{(l)}\mathbf{x}^{(l)H}))\mathbf{x}^{(l)}$

We have $\lambda_{\min}(\mathbf{E}\odot(\mathbf{x}^{(l)}\mathbf{x}^{(l)H}))=c_m^2\lambda_{\min}(\mathbf{E})$, cf. Remark 9. From [29] and [6, Lemma 3], we see that in order to satisfy $\lambda_l\leq c_m^2\lambda_{\min}(\mathbf{E})$, λ_l is set to be $c_m^2\cdot\frac{1}{2}(\min_{1\leq i\leq N}\nu_{2i}+\min_{1\leq i\leq N}\nu_{2i-1})$ where $\boldsymbol{\nu}=\mathbf{F}\mathbf{u}$, $\mathbf{F}$ is the $2N-$DFT matrix, and $\mathbf{u}=[u_0,u_1,\cdots,u_{2N-1}]^T$ with

$$u_k=\begin{cases}0 & k=0,\ N\\ w_ka_k(N-|k|) & k=1,\cdots,N-1\\ u_{2N-k} & k=N+1,\cdots,2N-1.\end{cases}\tag{83}$$

In short, $\boldsymbol{\nu}$ can be computed as $\boldsymbol{\nu}=\mathrm{FFT}(\mathbf{u})$.

Next, we observe $\left(\mathbf{E}\odot\left(\mathbf{x}^{(l)}\mathbf{x}^{(l)H}\right)\right)\mathbf{x}^{(l)}=\mathrm{Diag}\left(\mathbf{x}^{(l)}\right)\cdot\mathbf{E}\cdot\mathrm{Diag}^H\left(\mathbf{x}^{(l)}\right)\mathbf{x}^{(l)}=c_m^2(\mathbf{E}\mathbf{1})\odot\mathbf{x}^{(l)}$, where $\mathbf{E}\mathbf{1}$ can also be implemented with FFT:

$$\begin{aligned}\mathbf{E}\mathbf{1}&=\left[\frac{1}{2N}\mathbf{F}^H\mathrm{Diag}(\boldsymbol{\nu})\mathbf{F}\begin{bmatrix}\mathbf{1}_{N\times 1}\\ \mathbf{0}_{N\times 1}\end{bmatrix}\right]_{1:N}\\ &=\left[\mathrm{IFFT}\left(\boldsymbol{\nu}\odot\mathrm{FFT}\left(\left[\mathbf{1}_{N\times 1}^T,\mathbf{0}_{N\times 1}^T\right]^T\right)\right)\right]_{1:N}.\end{aligned}\tag{84}$$

Since $\mathrm{FFT}\left(\left[\mathbf{1}_{N\times 1}^T,\mathbf{0}_{N\times 1}^T\right]^T\right)$ only needs to be computed once, so only 1 additional IFFT is needed per iteration. When $p=2$ and $w_k=1,\forall k$, i.e., the ISL metric, $\mathbf{u}$ is a constant vector ($a_k=1,\forall k$), and thus $\boldsymbol{\nu}$ and $\mathbf{E}\mathbf{1}$ are both constant vectors, which means no additional FFT(IFFT) operation is needed.

APPENDIX G
COMPUTATION PROCESS OF λ_v AND $\mathbf{S}\mathbf{x}^{(l)}$

We understand that $\mathbf{G}=\sum_{i=1}^{2N}a_i\mathbf{f}_i\mathbf{f}_i^H=\frac{1}{2N}\mathbf{F}_{:,1:N}^H\mathrm{Diag}(2N\mathbf{a})\mathbf{F}_{:,1:N}$. Following [29] and [6, Lemma 3], λ_v is set to be $N(\max_{1\leq i\leq N}a_{2i}+\max_{1\leq i\leq N}a_{2i-1})$, which is of complexity $\mathcal{O}(N)$.

Then, we compute $\mathbf{S}\mathbf{x}^{(l)}$, where $\mathbf{S}=\sum_{i=1}^{2N}2|\mathbf{f}_i^H\mathbf{x}^{(l)}|^2\mathbf{f}_i\mathbf{f}_i^H=\frac{1}{2N}\mathbf{F}_{:,1:N}^H\mathrm{Diag}(2N\mathbf{v})\mathbf{F}_{:,1:N}$ and $\mathbf{v}=2|\mathrm{FFT}([\mathbf{x}^{(l)T},\mathbf{0}_{N\times 1}^T]^T)|^2=2|\mathbf{t}|^2$ ($\mathbf{t}$ [cf. (79)]). So 1 FFT for computing $\mathbf{t}$ is needed. Thus,

$$\begin{aligned}\mathbf{S}\mathbf{x}^{(l)}&=\left[\frac{1}{2N}\mathbf{F}^H\mathrm{Diag}(2N\mathbf{v})\mathbf{F}\begin{bmatrix}\mathbf{x}^{(l)}\\ \mathbf{0}_{N\times 1}\end{bmatrix}\right]_{1:N}\\ &=\left[\mathrm{IFFT}(2N\mathbf{v}\odot\mathbf{t})\right]_{1:N}\\ &\overset{\mathbf{v}=2|\mathbf{t}|^2}{=}4N\left[\mathrm{IFFT}\left(|\mathbf{t}|^2\odot\mathbf{t}\right)\right]_{1:N}.\end{aligned}\tag{85}$$

Here comes 1 IFFT.

REFERENCES

[1] R. Turyn, "Sequences with small correlation," *Error Correcting Codes*, in Herry B. Mann, Ed. New York, NY, USA: Wiley, pp. 195–228, 1968.

[2] S. W. Golomb and G. Gong, *Signal Design for Good Correlation: For Wireless Communication, Cryptography, and Radar*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[3] M. Golay, "A class of finite binary sequences with alternate autocorrelation values equal to zero (corresp.)," *IEEE Trans. Inf. Theory*, vol. 18, no. 3, pp. 449–450, May 1972.

[4] P. Stoica, H. He, and J. Li, "New algorithms for designing unimodular sequences with good correlation properties," *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1415–1425, Apr. 2009.

[5] J. Song, P. Babu, and D. P. Palomar, "Optimization methods for designing sequences with low autocorrelation sidelobes," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 3998–4009, Aug. 2015.

[6] J. Song, P. Babu, and D. Palomar, "Sequence design to minimize the weighted integrated and peak sidelobe levels," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 2051–2064, Apr. 2016.

[7] J. Tropp *et al.*, "Designing structured tight frames via an alternating projection method," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 188–209, Jan. 2005.

[8] A. De Maio, Y. Huang, M. Piezzo, S. Zhang, and A. Farina, "Design of optimized radar codes with a peak to average power ratio constraint," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2683–2697, Jun. 2011.

[9] M. Soltanalian, M. M. Naghsh, and P. Stoica, "A fast algorithm for designing complementary sets of sequences," *Signal Process.*, vol. 93, no. 7, pp. 2096–2102, 2013.

[10] I. Mercer, "Merit factor of Chu sequences and best merit factor of polyphase sequences," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 6083–6086, Sep. 2013.

[11] A. De Maio, S. De Nicola, Y. Huang, Z.-Q. Luo, and S. Zhang, "Design of phase codes for radar performance optimization with a similarity constraint," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 610–621, Feb. 2009.

[12] G. Cui, H. Li, and M. Rangaswamy, "MIMO radar waveform design with constant modulus and similarity constraints," *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 343–353, Jan. 2014.

[13] R. L. Frank, "Polyphase codes with good nonperiodic correlation properties," *IEEE Trans. Inf. Theory*, vol. 9, no. 1, pp. 43–45, Jan. 1963.

[14] D. Chu, "Polyphase codes with good periodic correlation properties (corresp.)," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 531–532, Jul. 1972.

[15] N. Zhang and S. W. Golomb, "Polyphase sequence with low autocorrelations," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 1085–1089, May 1993.

[16] M. Soltanalian and P. Stoica, "Computational design of sequences with good correlation properties," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2180–2193, May 2012.

[17] M. M. Naghsh, M. Modarres-Hashemi, S. ShahbazPanahi, M. Soltanalian, and P. Stoica, "Unified optimization framework for multi-static radar code design using information-theoretic criteria," *IEEE Trans. Signal Process.*, vol. 61, no. 21, pp. 5401–5416, Nov. 2013.

[18] M. Soltanalian and P. Stoica, "Designing unimodular codes via quadratic optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1221–1234, Mar. 2014.

[19] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer. Statist.*, vol. 58, no. 1, pp. 30–37, 2004.

[20] A. Hjørungnes, *Complex-Valued Matrix Derivatives: With Applications in Signal Processing and Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[21] H. He, J. Li, and P. Stoica, *Waveform Design for Active Sensing Systems: A Computational Approach*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[22] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, 2013.

[23] J.-S. Pang, M. Razaviyayn, and A. Alvarado, "Computing B-stationary points of nonsmooth DC programs," arXiv:1511.01796, 2015.

[24] J. Pang, "Partially B-regular optimization and equilibrium problems," *Math. Oper. Res.*, vol. 32, no. 3, pp. 687–699, 2007.

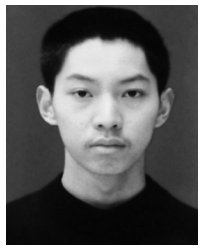
[25] R. Varadhan and C. Roland, "Simple and globally convergent methods for accelerating the convergence of any EM algorithm," *Scand. J. Statist.*, vol. 35, no. 2, pp. 335–353, 2008.

[26] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2001, pp. 556–562.

[27] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus With Applications in Statistics and Econometrics*. Hoboken, NJ, USA: Wiley, 1999.

[28] X. Yu, G. Cui, L. Kong, and V. Carotenuto, "Space-time transmit code and receive filter design for colocated mimo radar," in *Proc. IEEE Radar Conf.*, 2016, pp. 1–6.

[29] P. Jorge and S. Ferreira, "Localization of the eigenvalues of toeplitz matrices using additive decomposition, embedding in circulants, and the Fourier transform," in *Proc. Symp. Syst. Identif.*, 1994, vol. 3, pp. 271–275.



Licheng Zhao received the B.S. degree in information engineering from Southeast University, Nanjing, China, in 2014. He is currently working toward the Ph.D. degree in the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong. His research interests include optimization theory and fast algorithms, with applications in signal processing, machine learning, and financial engineering.

Junxiao Song received the B.Sc. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2011, and the Ph.D. degree in electronic and computer engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2015.

His research interests include convex optimization and efficient algorithms, with applications in big data, signal processing, and financial engineering.

Prabhu Babu received the Ph.D. degree in electrical engineering from the Uppsala University, Sweden, in 2012. From 2013–2016, he was a Post-Doctorial Fellow with the Hong Kong University of Science and Technology. He is currently with the Center for Applied Research in Electronics (CARE), Indian Institute of Technology Delhi, New Delhi, India.

Daniel P. Palomar (S'99-M'03-SM'08-F'12) received the Electrical Engineering and Ph.D. degrees from the Technical University of Catalonia, Barcelona, Spain, in 1998 and 2003, respectively.

He is a Professor in the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST), Hong Kong, which he joined in 2006. Since 2013, he has been a Fellow of the Institute for Advance Study, HKUST. He had previously held several research appointments, namely, at King's College London, London, U.K.; Stanford University, Stanford, CA, USA; Telecommunications Technological Center of Catalonia, Barcelona, Spain; Royal Institute of Technology, Stockholm, Sweden; University of Rome "La Sapienza," Rome, Italy; and Princeton University, Princeton, NJ, USA. His current research interests include applications of convex optimization theory, game theory, and variational inequality theory to financial systems, big data systems, and communication systems.

Dr. Palomar received the 2004/06 Fulbright Research Fellowship, the 2004 and 2015 (coauthor) Young Author Best Paper Awards by the IEEE Signal Processing Society, the 2015–16 HKUST Excellence Research Award, the 2002/03 best Ph.D. prize in Information Technologies and Communications by the Technical University of Catalonia (UPC), the 2002/03 Rosina Ribalta first prize for the Best Doctoral Thesis in Information Technologies and Communications by the Epson Foundation, and the 2004 prize for the best Doctoral Thesis in Advanced Mobile Communications by the Vodafone Foundation and COIT. He is a Guest Editor of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING 2016 Special Issue on Financial Signal Processing and Machine Learning for Electronic Trading and has been Associate Editor of IEEE TRANSACTIONS ON INFORMATION THEORY and of IEEE TRANSACTIONS ON SIGNAL PROCESSING, a Guest Editor of the *IEEE Signal Processing Magazine* 2010 Special Issue on Convex Optimization for Signal Processing, the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2008 Special Issue on Game Theory in Communication Systems, and theIEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2007 Special Issue on Optimization of MIMO Transceivers for Realistic Communication Networks.