On the Design of Constant Modulus Probing Waveforms with Good Correlation Properties for MIMO Radar via Consensus-ADMM Approach

Jiangtao Wang, Yongchao Wang, Member, IEEE

Abstract—In this paper, we design constant modulus probing waveforms with good correlation properties for collocated multi-input multi-output (MIMO) radar systems. The main content is as follows: first, we formulate the design problem as a fourth order polynomial minimization problem with constant modulus constraints. Then, by exploiting introduced auxiliary variables and their inherent structures, the polynomial optimization model is equivalent to a non-convex consensus minimization problem. Second, a customized alternating direction method of multipliers (ADMM) algorithm is proposed to solve the non-convex problem approximately. In the algorithm, all the subproblems can be solved analytically. Moreover, all subproblems except one subproblem can be performed in parallel. Third, we prove that the customized ADMM algorithm is theoretically-guaranteed convergent if proper parameters are chosen. Fourth, two variant ADMM algorithms, based on stochastic block coordinate descent and accelerated gradient descent, are proposed to reduce computational complexity and speed up the convergence rate. Numerical examples show that the proposed consensus-ADMM algorithm offers better performance, especially for a large scale waveform design problem, than the state-of-the-art approaches.

Index Terms—Constant modulus probing waveform, beampattern design, MIMO radar, auto-/cross-correlation, ADMM.

I. INTRODUCTION

Multiple-input multi-output (MIMO) radar system is regarded as a promising paradigm for the next generation radar systems. Unlike the standard phased-array radar to transmit scaled versions of a single waveform, probing signals, transmitted via different antennas in the MIMO radar system, are independent. Through this additional waveform diversity, MIMO radar owns superior capabilities compared with the traditional phased-array radar, such as higher spatial resolution, more flexible beampattern, and better detection performance [1] [2]. MIMO radar system can be classified into two categories: distributed and collocated. In the former, transmitters are widely separated in space and each of them can provide an independent view of the target, which can improve detection performance [3] [4]. In the latter, antennas in the transmitter are placed in close proximity and different probing signals from various collocated antennas can generate various desired beampatterns, leading to an improved directional resolution and interference rejection capability [5]–[7].

Probing signal waveforms play a central role in the signal processing performance of a MIMO radar system. Therefore, in the past years, a lot of researchers are attracted to this research field. Authors in [8] and [9] matched the waveform covariance matrix to the desired beampattern through a semidefinite programming method, then exploited the cyclic algorithm to synthesize the constant modulus waveform and pursued good auto-/cross-correlation properties. In [10], authors formulated the waveform design problem as a fourth order polynomial minimization problem with constant modulus constraints, then proposed a quasi-Newton solving algorithm to approximate the model’s optimal solution. Moreover, the approach can be applied to the scenario of desired low correlation sidelobe levels within certain lag intervals. The authors in [11] focused on the direct or indirect control of mainlobe ripples in the beampattern design problem. They reformulated the design as a feasibility problem with the lowest system cost. To achieve a high signal to interference plus noise ratio and low sidelobe levels performance, a fixed waveform covariance matrix was proposed in [12]. However, the matrix does not exploit the full waveform diversity. In [13], the authors proposed a novel transmit beampattern matching design one-step method, which obtains the transmit signal matrix by unconstrained optimization. The drawback of the waveforms generated by this method is that their envelope is not constant modulus. To reduce the computational complexity, a closed-form covariance matrix design method was proposed in [14] based on discrete Fourier transform (DFT). The authors in [15] and [16] also applied the DFT-based technique to a planar-antenna-array, and developed a finite-alphabet constant-envelope waveforms design algorithm for the desired beampattern. However, the performance of the DFT-based method is slightly worse for a small number of antennas. The authors in [17] studied the robust transmit beampattern design problem and exploited the semidefinite relaxation technique to treat non-convex optimization problems. In [18]–[19], the authors exploited successive convex relaxation techniques to handle non-convex quadratic equality constraints in the constant modulus waveform design problem. In [21], the authors proposed a double cyclic alternating direction method of multipliers (D-ADMM) algorithm to solve the non-convex beampattern design problem and in [22], they considered the joint optimization problem of the covariance matrix and antenna position. In [23], the authors applied the majorization-minimization technique to match the desired transmit beampattern, which enjoys faster convergence than D-ADMM. Besides the above works, some researchers synthesized transmit waveforms under some practical constraints, such as mainlobe ripple constraints [24], spectral shape constraints [25], constant modulus constraints [26], similarity constraints [20] [27], and transmitted power constraints [28]. However, these works only
focus on the synthesized beampattern design problem and pay little attention to the correlation properties of the waveforms.

In this paper, we still focus on designing constant modulus probing waveforms to match the desired spatial beampatterns while suppressing the spatial auto-correlation and cross-correlation levels in the collocated MIMO radar system. The main contributions are as follows.

- **Consensus problem formulation**: the design problem is formulated as a fourth order polynomial minimization problem with constant modulus constraints. Then, by introducing auxiliary variables, it is further equivalent to a non-convex consensus minimization problem.
- **Parallel solving algorithm**: consensus-ADMM is customized to solve the non-convex consensus problem approximately. In the implementation, all the subproblems can be solved analytically. Moreover, except one subproblem, all subproblems can be performed in parallel.
- **Theoretically-guaranteed performance**: we prove that the solving algorithm is guaranteed convergent to some stationary point of the non-convex optimization problem if proper parameters are chosen.
- **Improvement strategies**: two variant ADMM algorithms, based on stochastic block coordinate descent (SBCD) and accelerated gradient descent (AGD), are proposed to reduce computational complexity and speed up the convergence rate.

The rest of the paper is organized as follows. In Section II, we present that the beampattern design problem is transformed to a non-convex consensus minimization problem. In Section III, consensus-ADMM is customized to solve the non-convex consensus minimization problem. The performance analysis, including convergence and computational complexity of the proposed consensus-ADMM algorithm, are presented in Section IV. Two variant ADMM algorithms are given to improve computational complexity and convergence performance of the solving algorithm in Section V. Finally, Section VI demonstrates the effectiveness of the proposed consensus-ADMM algorithms and the conclusions are given in Section VII.

**Notation**: bold lowercase and uppercase letters denote column vectors and matrices and italics denote scalars. \( \mathbb{R} \) and \( \mathbb{C} \) denote the real field and the complex field respectively. The superscripts \( (\cdot)^* \), \( (\cdot)^T \) and \( (\cdot)^H \) denote conjugate operator, transpose operator and conjugate transpose operator respectively. \( x_i \) denotes the \( i \)-th element of vector \( x \). \( \| \cdot \|_2 \) denotes the absolute value. The subscript \( \| \cdot \|_2 \) denotes the Euclidean vector norm and \( \| \cdot \|_F \) denotes the Frobenius matrix norm. \( \nabla(\cdot) \) represents the gradient of a function. \( \text{Re}(\cdot) \) takes the real part of the complex variable and \( \text{Tr}(\cdot) \) denotes the trace of a matrix. \( (\cdot, \cdot) \) and \( \otimes \) are the dot product operator and convolution operator respectively. \( \text{vec}(\cdot) \) vectorizes a matrix by stacking its columns on top of one another and \( \text{mat}(\cdot, N, M) \) reshapes a vector to an \( N \times M \) matrix. \( \Pi(\cdot) \) denotes the projection operator. \( \mathbb{E}[\cdot] \) performs the expectation of the variables and \( \mathbf{I} \) denotes an identity matrix.
B. Problem Formulation

We optimize MIMO radar probing waveforms based on the following considerations: first, as mentioned in (4), since the beampattern describes the spacial power distribution, we desire that it can match the directions of interest, which can decrease clutter components and extend the probing distance; second, since low auto-correlation sidelobes can increase spacial resolution and low cross-correlation levels can reduce interferences from other directions, we desire that the optimized probing waveforms have low auto-correlation sidelobes and low cross-correlation levels; third, in order to maximize the efficiency of the power amplifier in the MIMO radar transmitter, the probing waveforms should be constant modulus, i.e., $|x_{i,m}| = 1, \ i = 1, \cdots, N, \ m = 1, \cdots, M$.

Based on the above considerations, we formulate the following optimization model to design MIMO radar probing

$$
\min_{\alpha, X} \ e(\alpha, X) + P_c(X),
$$

subject to $|x_{i,m}| = 1, \ i = 1, \cdots, N, \ m = 1, \cdots, M$, (6)

where

$$
e(\alpha, X) = \sum_{\theta \in \Theta} \left| \alpha \tilde{P}_\theta - a_\theta^H X^H X a_\theta \right|^2,
$$

and $w_{ac}$ and $w_{cc}$ are preset positive weights and $T$ is the time delay parameter set of interest. In the objective function of model (6), the first term $e(\alpha, X)$ represents the mismatching square error between the designed beampattern and the desired beampattern $P_\theta$ and $\alpha$ is a scaling factor that needs to be optimized. The second term $P_c(X)$ relates to the auto-correlation sidelobes and cross-correlation levels at the considered spacial directions. Because $P_\theta^2$ and $\alpha^2$ are not included. It is difficult to solve (6) directly since its quartic objective function and constant modulus constraints are non-convex. In the following, by exploiting its inherent structure, we show how to design an efficient solving algorithm to pursue theoretically-guaranteed solutions.

First, let $X$’s phase $\Phi$ be the new variable, i.e., $x_{i,m} = e^{j\phi_{i,m}}$. Then, we can drop constant modulus constraints and rewrite (6) as the following minimization problem

$$
\min_{\alpha, \Phi} \ e(\alpha, X(\Phi)) + P_c(X(\Phi)),
$$

subject to $\alpha \in (0, \alpha_{\text{max}}]$. (8)

Second, we define the following quantities

$$
a_{\theta, \rho} = \text{vec}(a_\theta a_\rho^H), \ p = \sum_{\theta \in \Theta} \tilde{P}_\theta^2,
$$

$$
q = -\sum_{\theta \in \Theta} \tilde{P}_\theta a_{\theta, \rho}, \ A = \sum_{\theta \in \Theta} a_{\theta, \rho} a_{\theta, \rho}^H.
$$

Then, the first term $e(\alpha, X(\Phi))$ in (8) can be equivalent to

$$
e(\alpha, X(\Phi)) = \mathbf{v}^H (\alpha, \Phi) \mathbf{Q} \mathbf{v}(\alpha, \Phi),
$$

where

$$
\mathbf{v}(\alpha, \Phi) = \begin{bmatrix} \text{vec}(X^H (\Phi) X(\Phi)) \end{bmatrix},
$$

$$
\mathbf{Q} = \begin{bmatrix} \mathbf{p} & \mathbf{q}^H \end{bmatrix}.
$$

In addition, define the matrices $B_n(\Phi), n \in T$, and for $n = 0$

$$
B_0(\Phi) = \begin{bmatrix} 0 & w_{cc} P_{\theta_1, \theta_2, n} & \cdots & w_{cc} P_{\theta_1, K, n} \\
\vdots & \ddots & \ddots & \vdots \\
w_{cc} P_{K, \theta_1, n} & w_{cc} P_{K, \theta_2, n} & \cdots & 0 \end{bmatrix},
$$

and for $n \in T \setminus\{0\}$

$$
B_n(\Phi) = \begin{bmatrix} w_{ac} P_{\theta_1, \theta_1, n} & w_{cc} P_{\theta_1, \theta_2, n} & \cdots & w_{cc} P_{\theta_1, K, n} \\
\vdots & \ddots & \ddots & \vdots \\
w_{cc} P_{K, \theta_1, n} & w_{cc} P_{K, \theta_2, n} & \cdots & w_{ac} P_{K, K, n} \end{bmatrix},
$$

where $K$ is the set $\hat{\Theta}$’s size. Then, $P_c(X(\Phi))$ in (8) can be rewritten as

$$
P_c(X(\Phi)) = \sum_{n \in T} \left\| B_n(\Phi) \right\|_F^2.\ (12)
$$

To facilitate the subsequent derivations, we further define

$$
h(\alpha, \Phi) = \mathbf{v}^H (\alpha, \Phi) \mathbf{Q} \mathbf{v}(\alpha, \Phi),
$$

$$
f_n(\Phi) = \left\| B_n(\Phi) \right\|_F^2,\ (13b)
$$

and introduce a set of auxiliary variables $\{\Phi_n | n \in T\}$. Then, problem (8) can be formulated as the following consensus-like problem

$$
\min_{\alpha \in \mathbb{R}^N, \{\Phi_n | n \in T\}} h(\alpha, \Phi) + \sum_{n \in T} f_n(\Phi_n),
$$

subject to $\Phi = \Phi_n, \ n \in T$, $\alpha \in (0, \alpha_{\text{max}}]$. (14)

In comparison with (8), model (14) allows subfunction $h(\alpha, \Phi)$ or $f_n(\Phi_n)$ to handle its local variable independently when $\Phi$ or $\Phi_n$ is fixed. In the next section, by exploiting this key observation, we design an efficient algorithm with theoretically-guaranteed convergence, named consensus-ADMM, to solve (14) approximately. To the best of our knowledge, it is the first time that a parallel algorithm structure is introduced to match the desired beampattern for the MIMO radar system, which means that the proposed consensus-ADMM algorithm is more suitable for the large scale MIMO radar waveforms design problem. Moreover, convergence analysis and improved variants of the proposed consensus-ADMM algorithm are also considered.

III. CONSENSUS-ADMM SOLVING ALGORITHM

The augmented Lagrangian function of problem (14) can be written as

$$\mathcal{L}(\alpha, \Phi, \{\Phi_n, \Lambda_n, n \in T\}) = h(\alpha, \Phi) + \sum_{n \in T} \left( f_n(\Phi_n) + \langle \Lambda_n, \Phi_n - \Phi \rangle + \rho_n \left\| \Phi_n - \Phi \right\|_F^2 \right)\ (15)$$. 

where $\Lambda_n$ and $\rho_n$ are the Lagrangian multiplier and penalty parameters respectively. To facilitate discussions later, we define the following functions

$$L_n(\Phi, \Phi_n, \Lambda_n) = f_n(\Phi_n) + \langle \Lambda_n, \Phi_n - \Phi \rangle + \frac{\rho_n}{2} \| \Phi - \Phi_n \|^2, \quad n \in \mathcal{T}.$$  

(16)

Based on (15) and (16), the classical consensus-ADMM algorithm framework can be described as

$$\{ \alpha^{k+1}, \Phi^{k+1} \} = \arg \min_{\alpha \in [0, \alpha_{\max}]} \mathcal{L}(\alpha, \Phi, \{ \Phi_n^k, \Lambda_n^k, n \in \mathcal{T} \}), \quad (17a)$$

$$\Phi_n^{k+1} = \arg \min_{\Phi_n} L_n(\Phi^{k+1}, \Phi_n, \Lambda_n^k), \quad n \in \mathcal{T}, \quad (17b)$$

$$\Lambda_n^{k+1} = \Lambda_n^k + \rho_n(\Phi_n^{k+1} - \Phi^{k+1}), \quad n \in \mathcal{T}, \quad (17c)$$

where $k$ is the iteration number.

**Remarks on (17):** Since $\mathcal{L}(\alpha, \Phi, \{ \Phi_n^k, \Lambda_n^k, n \in \mathcal{T} \})$ and $L_n(\Phi^{k+1}, \Phi_n, \Lambda_n^k)$ are non-convex, it is difficult to implement (17a) and (17b). However, we have the following lemma to characterize Lipschitz properties of $\nabla h(\alpha, \Phi)$ and $\nabla f_n(\Phi)$ (see proof in Appendix A).

**Lemma 1:** gradients $\nabla h(\alpha, \Phi)$ and $\nabla f_n(\Phi_n)$ are Lipschitz continuous, i.e.,

$$\| \nabla h(\alpha, \Phi) - \nabla h(\alpha, \Phi') \|_F \leq L_\alpha |\alpha - \alpha'|, \quad (18)$$

$$\| \nabla f_n(\Phi) - \nabla f_n(\Phi') \|_F \leq L_n |\Phi - \Phi'|, \quad (19)$$

where constants $L_\alpha \geq 2p, L > 2(2M-1)(2M-1+\alpha_{max}\bar{P}_{\max}+M^2N)|\Theta|$ and $L_n > 4\rho_n^2\bar{M}^2(2+MN+N)K^2$.

Based on the above Lemma, $\mathcal{L}(\alpha, \Phi, \{ \Phi_n^k, \Lambda_n^k, n \in \mathcal{T} \})$ and $L_n(\Phi^{k+1}, \Phi_n, \Lambda_n^k)$ can be upper-bounded by the following strongly convex functions

$$\mathcal{L}(\alpha, \Phi, \{ \Phi_n^k, \Lambda_n^k, n \in \mathcal{T} \}) \leq \mathcal{U}(\alpha, \Phi, \{ \Phi_n^k, \Lambda_n^k, n \in \mathcal{T} \}), \quad (19a)$$

$$L_n(\Phi^{k+1}, \Phi_n, \Lambda_n^k) \leq \mathcal{U}_n(\Phi^{k+1}, \Phi_n, \Lambda_n^k), \quad n \in \mathcal{T}, \quad (19b)$$

where

$$\mathcal{U}(\alpha, \Phi, \{ \Phi_n^k, \Lambda_n^k, n \in \mathcal{T} \}) = h(\alpha, \Phi) + \langle \nabla h(\alpha, \Phi), \Phi - \Phi_k \rangle + \langle \nabla h(\alpha, \Phi), \alpha - \alpha_k \rangle - \frac{1}{2} \alpha \| \Phi - \Phi_k \|^2 + \frac{L_\alpha}{2} |\alpha - \alpha_k|^2 + \sum_{n \in \mathcal{T}} L_n(\Phi, \Phi_n, \Lambda_n^k),$$

and

$$\mathcal{U}_n(\Phi^{k+1}, \Phi_n, \Lambda_n^k) = f_n(\Phi_n^{k+1}) + \langle \nabla f_n(\Phi_n^{k+1}), \Phi_n - \Phi_n^{k+1} \rangle + \rho_n + \frac{L_n}{2} \| \Phi_n - \Phi_n^{k+1} \|^2. \quad (20)$$

Then, instead of solving (17a) and (17b) directly, we propose the following customized consensus-ADMM algorithm (22).

$$\{ \alpha^{k+1}, \Phi^{k+1} \} = \arg \min_{\alpha \in [0, \alpha_{\max}]} \mathcal{U}(\alpha, \Phi, \{ \Phi_n^k, \Lambda_n^k, n \in \mathcal{T} \}), \quad (22a)$$

$$\Phi_n^{k+1} = \arg \min_{\Phi_n} \mathcal{U}_n(\Phi^{k+1}, \Phi_n, \Lambda_n^k), \quad (22b)$$

$$\Lambda_n^{k+1} = \Lambda_n^k + \rho_n(\Phi_n^{k+1} - \Phi^{k+1}). \quad (22c)$$

Since $\mathcal{U}(\alpha, \Phi, \{ \Phi_n^k, \Lambda_n^k, n \in \mathcal{T} \})$ and $\mathcal{U}_n(\Phi^{k+1}, \Phi_n, \Lambda_n^k)$ are strongly convex quadratic functions, optimal solutions of the problems (22a) and (22b) can be obtained by solving the following linear equations

$$\nabla \alpha \mathcal{U}(\alpha, \Phi, \{ \Phi_n^k, \Lambda_n^k, n \in \mathcal{T} \}) = 0, \quad (23a)$$

$$\nabla \mathcal{U}_n(\Phi^{k+1}, \Phi_n, \Lambda_n^k) = 0, \quad (23b)$$

$$\nabla \Phi^{k+1} \mathcal{U}(\Phi^{k+1}, \Phi_n, \Lambda_n^k) = 0. \quad (23c)$$

which lead to

$$\alpha^{k+1} = \prod_{\alpha \in [0, \alpha_{\max}]} \left( \frac{\alpha - L \alpha}{\alpha + \rho_n} \right), \quad (24a)$$

$$\Phi^{k+1} = \Phi^{k+1} - \frac{L \Phi_k - \nabla h(\alpha, \Phi_k)}{\rho_n + L}. \quad (24b)$$

Combining (22c) and (24), we summarize the customized consensus-ADMM algorithm in Table I.

**IV. ANALYSIS**

**A. Convergence Issue**

We have the following theorem to show convergence properties of the proposed consensus-ADMM algorithm in Table I.

**Theorem 1:** (Guaranteed Convergence) $\forall n \in \mathcal{T}$, if the penalty parameters $\rho_n$ and Lipschitz constants $L_n$ satisfy $\rho_n \geq 9L_n$, the proposed consensus-ADMM algorithm is convergent, i.e.,

$$\lim_{k \to +\infty} \alpha^k = \alpha^*, \quad \lim_{k \to +\infty} \Phi^k = \Phi^*.$$

Moreover, $(\alpha^*, \Phi^*)$ is some stationary point of problem (8), i.e., it satisfies the following inequalities

$$\langle \nabla h(\alpha^*, \Phi^*), \alpha - \alpha^* \rangle \geq 0, \quad \alpha \in [0, \alpha_{\max}], \quad \langle \nabla \mathcal{U}_n(\Phi^*, \Phi^*), \Phi^* - \Phi^* \rangle \geq 0. \quad (25)$$

**Remarks on Theorem 1:** it shows that the proposed consensus-ADMM algorithm is theoretically-guaranteed to be convergent to some stationary point of model (8) under the
conditions $\rho_n \geq 9L_n$, $n \in T$. Here we should note that these kinds of conditions are easily satisfied since $L_n$’s upper bound can be obtained according to Lemma 1 and penalty parameters $\rho_n$ can be set accordingly. To prove theorem 1 in Appendix B we give Lemmas 4 and 5 and their proofs in advance. In it, we first show that the difference value between the original functions and their upper bound functions are lower-bounded in Lemma 2. Next, we show that $\|A_k^{k+1} - A_k^k\|_F$ is upper-bounded in Lemma 3. Based on this result, we show that if proper parameters are chosen, the augmented Lagrangian function $\mathcal{L}(\alpha, \Phi^k, \{\Phi^k, \Lambda^k_n, n \in T\})$ is sufficient descent in every iteration and has a lower bound in Lemma 4 and 5 respectively, which lead $\mathcal{L}(\alpha, \Phi^k, \{\Phi^k, \Lambda^k_n, n \in T\})$ to be convergent as $k \to +\infty$. Based on these lemmas, we prove theorem 1 in Appendix C.

### B. Implementation Analysis

In the following, we show $\nabla_a h(\alpha, \Phi)$, $\nabla_\Phi h(\alpha, \Phi)$ and $\nabla f_n(\Phi)$ can be computed efficiently by exploiting their inside structures such as sparsity and convolution.

First, we focus on $\nabla_a h(\alpha, \Phi)$ and $\nabla_\Phi h(\alpha, \Phi)$, which can be expressed as (27a) and (27b) respectively.

$$\nabla_a h(\alpha, \Phi) = 2\text{Re}\left(\frac{\partial h}{\partial \alpha}(\alpha, \Phi) \cdot \Phi v(\alpha, \Phi)\right), \quad (27a)$$

$$\nabla_\Phi h(\alpha, \Phi) = \text{mat}(2\text{Re}\left(\frac{\partial h}{\partial \Phi}(\alpha, \Phi) \cdot \Phi v(\alpha, \Phi)\right), N, M) \quad (27b)$$

where $\frac{\partial h}{\partial \alpha}(\alpha, \Phi) = [1; 0]$ and

$$\frac{\partial h}{\partial \Phi}(\alpha, \Phi) = \begin{bmatrix} \frac{\partial h}{\partial \phi_{1,1}}(\alpha, \Phi) & \ldots & \frac{\partial h}{\partial \phi_{1,M}}(\alpha, \Phi) \\ \vdots & \ddots & \vdots \\ \frac{\partial h}{\partial \phi_{N,1}}(\alpha, \Phi) & \ldots & \frac{\partial h}{\partial \phi_{N,M}}(\alpha, \Phi) \end{bmatrix}^H. \quad (28)$$

In (28), $\frac{\partial h}{\partial \phi_{i,m}}(\alpha, \Phi)$ can be calculated through

$$\frac{\partial h}{\partial \phi_{i,m}}(\alpha, \Phi) = \begin{bmatrix} \partial (\Phi^H \Phi) \Phi(\alpha, \Phi) \\ \partial (\Phi^H \Phi)(\alpha, \Phi) \end{bmatrix} \quad (29)$$

where $i = 1, \ldots, N$, $m = 1, \ldots, M$, and $\partial (\Phi^H \Phi)(\alpha, \Phi)$ is computed through (30).

Since $Q \in \mathbb{C}^{(M^2+1) \times (M^2+1)}$ and $v \in \mathbb{C}^{M^2+1}$, it takes around $(M^2 + 1)^2$ complex multiplications to compute $Qv(\alpha, \Phi)$. Since $\frac{\partial h}{\partial \alpha}(\alpha, \Phi) = [1; 0]$, we don’t need to do extra multiplications to obtain $\nabla_a h(\alpha, \Phi)$. Moreover, since $\partial (\Phi^H \Phi)(\alpha, \Phi)$ involves $2M - 1$ nonzero elements (see (30)), then totally there are $(2M - 1)MN$ nonzero elements in $\frac{\partial h}{\partial \Phi}(\alpha, \Phi)$, which means that it takes $(M^2 + 1)^2 + (2M - 1)MN$ complex multiplications to compute $\nabla_\Phi h(\alpha, \Phi)$. Then, we can see that the total computation cost on $\nabla_a h(\alpha, \Phi)$ and $\nabla_\Phi h(\alpha, \Phi)$ is roughly $O(M^4 + 2M^2N)$.

Second, we notice that the elements in $\nabla f_n(\Phi)$ can be obtained through

$$\frac{\partial f_n(\Phi)}{\partial \phi_{i,m}} = \begin{cases} \sum_{\theta_i, \theta_j \in \Theta} 2w_c \text{Re}(P^*_{\theta_i, \theta_j, n} \frac{\partial P_{\theta_i, \theta_j, n}}{\partial \phi_{i,m}}), & n = 0, \\
\sum_{\theta_i, \theta_j \in \Theta} 2w_c \text{Re}(P^*_{\theta_i, \theta_j, n} \frac{\partial P_{\theta_i, \theta_j, n}}{\partial \phi_{i,m}}), & n \in T \setminus 0. \end{cases} \quad (31)$$

To compute $P_{\theta_i, \theta_j, n}$ and $\frac{\partial P_{\theta_i, \theta_j, n}}{\partial \phi_{i,m}}$ for every $\theta_i, \theta_j$, we define $s_{\theta_i} = X_{\theta_i}$, and $s_{\theta_j}$, denoting the reversing of the order of the entries of $s_{\theta_i}$. Then, $\forall n \in T$, $P_{\theta_i, \theta_j, n}$ can be obtained through $s_{\theta_i} \otimes s_{\theta_j}$. It means that the cost to obtain all the $P_{\theta_i, \theta_j, n}$ is roughly $|T|K^2N$ complex multiplications, where $K$ is the number of considered spatial directions. Moreover, the corresponding $\frac{\partial P_{\theta_i, \theta_j, n}}{\partial \phi_{i,m}}$ can be obtained through $s_{\theta_i} \otimes s_{\theta_j}$ or $s_{\theta_j} \otimes s_{\theta_i}$. Since there is only one nonzero element in either $\frac{\partial P_{\theta_i, \theta_j, n}}{\partial \phi_{i,m}}$ or $\frac{\partial P_{\theta_i, \theta_j, n}}{\partial \phi_{i,m}}$, it takes only two complex multiplications to compute $\frac{\partial P_{\theta_i, \theta_j, n}}{\partial \phi_{i,m}}$. So, we can obtain all of them through $2|T|K^2MN$ complex multiplications. Taking complex multiplication operations to obtain $\frac{\partial f_n(\Phi)}{\partial \phi_{i,m}}$ into account, we can see that the total cost of computing $\frac{\partial f_n(\Phi)}{\partial \phi_{i,m}}$, $\forall n \in T$ is roughly $O(|T|K^2MN)$.

Observing the proposed consensus-ADMM algorithm in Table 1 and the corresponding (24) and (22c), we can see that the main computation cost lies in computing $\nabla_a h(\alpha, \Phi)$, $\nabla_\Phi h(\alpha, \Phi)$ and $\nabla f_n(\Phi)$, which are much larger than the other terms. Thus, we can conclude that the total cost in each ADMM iteration is roughly $O(M^4 + 2M^2N + 2|T|K^2MN)$.

### V. Improvements

#### A. Reduce Complexity

In the proposed consensus-ADMM algorithm, $\Phi_n, \forall n \in T$, are updated independently (or in parallel) as are the Lagrangian multipliers $\Lambda_n$. This fact admits us to update only a part of the variables $\{\Phi_n, \Lambda_n, n \in T\}$ in each ADMM iteration to reduce computational complexity.

Specifically, consider a randomized updating strategy called stochastic block coordinate descent (SBCD) $\mathbf{31}$. In the $k$-th iteration, let $N^k$ denote some $T$’s subset. We choose elements from $T$ to construct $N^k$ with the probability

$$\text{Pr}(n \in N^k) = \frac{p_n \geq p_{\text{min}} > 0}{(2)}$$

If some $n \in N^k$, the corresponding variables $\Phi^k_n$ and $\Lambda^k_n$ are updated using (24) and (22c) respectively. Otherwise, we just set $\Phi^k_n = \Phi^k_n$, $\Lambda^k_n = \Lambda^k_n$. In this way, it is obvious that computational complexity in each ADMM iteration can be reduced significantly. Moreover, this kind of implementation strategy can still guarantee that the iteration algorithm converges with high probability to some stationary point of problem (8) under proper conditions (we provide a sketch of the proof in Appendix D).

#### B. Speed Up Convergence

Besides the computational complexity in each iteration, convergence speed is another concern from a practical viewpoint. In this paper, inspired by Nesterov’s accelerated gradient

1 Usually, criteria of selecting $N^k$ is to guarantee every element in $T$ is implemented equally in probability.
To initialize all approaches, we develop its variant in the following
\[ \hat{\Phi}_n^{k+1} = \arg\min_{\Phi_n} U_n (\Phi_n^{k+1}, \Phi_n, \Lambda_n^k), \quad (33a) \]
\[ \Phi_n^{k+1} = \hat{\Phi}_n^{k+1} + \gamma^k \left( \hat{\Phi}_n^{k+1} - \Phi_n^k \right), \quad (33b) \]
where \( \gamma^k = \frac{k-1}{k+1} \) and \( r \geq 3 \) is some preset constant. The algorithm starts from \( \hat{\Phi}_1^1 = \Phi_1^1 \). Here, we should note that the proposed method is heuristic. It is difficult to prove that it can improve convergence speed theoretically. However, its practical performance is much better than the original one in Table I. We present the corresponding simulation results in the following section.

VI. SIMULATION RESULTS

In this section, numerical results are presented to illustrate the performance of the proposed MIMO radar beampattern design algorithm. We consider the number of antennas as \( M = 8, 16, 128 \) with the length of each sequence \( N = 64, 128, 1024 \) respectively. The set of the spatial angles covers \((-90^\circ, 90^\circ)\) with spacing \( 0.1^\circ \). The primal/dual residuals of the proposed consensus-ADMM algorithm at \( k \)-th iteration are defined as
\[ R^{k+1} = \sum_{n \in T} \| \Phi_n^{k+1} - \Phi_n^k \|_F, \quad D^{k+1} = \sum_{n \in T} \| \Phi_n^{k+1} - \Phi_n^k \|_F. \]
The termination criteria are set as both of them are less than \( 10^{-4} \) and the maximum iteration number 1000 is reached. The weights \( (w_{ae}, w_{ex}) \) are \((10, 10)\). The desired beampattern is
\[ \hat{P}(\theta) = \begin{cases} 1, & \theta \in [\theta_1 - 10^\circ, \theta_1 + 10^\circ], \ i = 1, 2, \\ 0, & \text{otherwise}, \end{cases} \quad (34) \]
where \( \theta_1 = -40^\circ \) and \( \theta_2 = 30^\circ \). The parameter \( r \) in the AGD method is 3. The random phase sequence is applied to initialize all approaches. All experiments are performed in MATLAB 2016b/Windows 7 environment on a computer with 2.1GHz Intel 4110×2 CPU and 64GB RAM.

Figure 2(a)-(b) plot the performance curves of \( e(\alpha, X) \) and \( P_i(X) \) versus the iteration number respectively for our proposed consensus-ADMM algorithms and other three state-of-the-art approaches: CA approach [9], L-BFGS method [10], and D-ADMM [21]. Here, we show \( e(\alpha, X) \) and \( P_i(X) \) separately. It is because both of them represent important features of spatial beampattern matching and correlation characteristics between spatial directions respectively. In figure 2(a), we see that \( e(\alpha, X) \) have similar convergence performance (converged within 50 iterations). However, in figure 2(b), convergence results of different approaches are quite different. Specifically, the proposed consensus-ADMM-AGD algorithm enjoys the best convergence performance. It indicates that the AGD strategy [33] can speed up convergence very well. In comparison, consensus-ADMM-SBCD-25%’s convergence is a little bit slow. However, we should note that it has lower computational complexity. In practice, the parameter 25% can be adjusted to make a tradeoff between convergence rate and computational
are symmetric to that of (30\degree) normalized spacial auto-correlation functions are symmetric forms respectively. From the figures, we can see that the characteristics of the designed MIMO radar probing waveform and the normalized cross-correlation functions of \(i, j, n\) and \(\theta, \theta_j, n\) indicates spacial auto-/cross-correlation characteristics of the designed MIMO radar probing waveforms respectively.

The figures also indicate that either increasing \(N\) and \(M\) or decreasing \(T\) can lower the correlation levels. These facts are reasonable since larger \(N\) and \(M\) or smaller \(T\) indicate more degrees of freedom in designing probing waveforms. Moreover, we can also see that the consensus-ADMM-AGD approach enjoys the best auto/cross-correlation characteristics. This fact is in accordance with the simulation result in Figure 2(b).

Figure 3 shows synthesized spacial beampatterns by our proposed consensus-ADMM approaches and three other approaches. From the figure, it can be observed that all of the approaches can match the desired spacial beampattern very well at different antenna numbers and waveform lengths. Figure 4 - Figure 6 show the normalized spacial correlation level \(C_{\theta, \theta_j, n}\) with different simulation parameters. Here, the normalized spacial correlation function \(C_{\theta, \theta_j, n}\) for a certain interval in dB is defined as

\[
C_{\theta, \theta_j, n} = 10 \log_{10} \frac{|P_{\theta, \theta_j, n}|}{\max\{|P_{\theta, \theta_j, 0}|, |P_{\theta_j, \theta, 0}|\}},
\]

where \(n \in T\) and \(\theta, \theta_j \in \hat{\Theta}\). It is obvious that for \(i = j\) and \(i \neq j\), \(C_{\theta, \theta_j, n}\) indicates spacial auto-/cross-correlation characteristics of the designed MIMO radar probing waveforms respectively. From the figures, we can see that the normalized spacial auto-correlation functions are symmetric and the normalized cross-correlation functions of (-40\degree, 30\degree) are symmetric to that of (30\degree, -40\degree). The figures also indicate

**Table II** shows averaged implementation time (per iteration) of the proposed consensus-ADMM approach and three other state-of-the-art approaches [9] [10] [21]. Here, N/A means that the iteration operations cannot be implemented in reasonable time. From the table, we can see that when \(M\) and \(N\) are small, the proposed consensus-ADMM approach enjoys the least execution time. When \(N\) and \(M\) are increased, for example,
from (64,8) to (1024,128), execution time of the consensus-ADMM approach becomes comparable to that of the L-BFGS approach. However, we should note that in the proposed consensus-ADMM algorithm, most parts can be implemented in parallel, which means that it is more suitable for large-scale applications from a practical viewpoint of implementation.3

VII. Conclusion

In this paper, we focus on designing constant modulus probing waveforms with good correlation properties for the collocated MIMO radar system. By introducing auxiliary variables and exploiting the designing problem’s inherent structure, we formulate a consensus-like optimization model. Then, the ADMM technique is customized to solve the corresponding non-convex problem approximately. We prove that the proposed ADMM approach is theoretically-guaranteed convergent if proper parameters are chosen. Simulation results demonstrate that the proposed consensus-ADMM approaches outperform the state-of-the-art approaches, especially more suitable for large-scale MIMO radar systems.

3In the real radar system, the algorithm is usually implemented using Field Programmable Gate Array (FPGA). This kind of integrated chip is very suitable for implementing an algorithm with a parallel structure.

### APPENDIX A

#### PROOF OF LEMMA 1

In the following, we prove that both $\nabla h(\alpha, \Phi)$ and $\nabla f_n(\Phi)$ are Lipschitz continuous via the definition of Lipschitz continuity. To state the proof clearly, we rewrite (9) and (11) in the following

$$a_{l,\theta} = \text{vec}(a_{l,\theta} H), \quad p = \sum_{\theta \in \Theta} \hat{P}_{\theta}^2.$$  

$$q = -\sum_{\theta \in \Theta} \hat{P}_{\theta} a_{l,\theta}, \quad A = \sum_{\theta \in \Theta} a_{l,\theta} a_{l,\theta} H.$$  

$$v(\alpha, \Phi) = \left[ \begin{array}{c} \text{vec} \left( X H(\Phi) X(\Phi) \right) \end{array} \right], \quad Q = \left[ \begin{array}{c} p \quad q^H \end{array} \right].$$

To facilitate the subsequent derivations, we denote

$$z = \text{vec} \left( X H(\Phi) X(\Phi) \right).$$

Since $A$ is Hermitian, $Qv$ can be denoted by

$$Qv = \left[ \begin{array}{c} p\alpha + q^H z \end{array} \right].$$  

First, we can obtain $\frac{\partial v(\alpha, \Phi)}{\partial \alpha} = [1, 0].$ Plugging it and (36) into (27a), we can have

$$|\nabla_\alpha h(\alpha, \Phi) - \nabla_\alpha h(\hat{\alpha}, \Phi)| = 2\text{Re}(p\alpha + q^H z - p\hat{\alpha} - q^H z)$$

$$\leq 2p, \quad \alpha, \hat{\alpha} \in [0, \alpha_{\max}].$$  

From (37), we can see that $\nabla_\alpha h(\alpha, \Phi)$ is Lipschitz continuous with the constant $L_\alpha \geq 2p$.

Second, for $\nabla_\Phi h(\alpha, \Phi)$, we have the following derivations

$$\left\| \nabla_\Phi h(\alpha, \Phi) - \nabla_\Phi h(\hat{\alpha}, \hat{\Phi}) \right\|^2_F$$

$$= \sum_{i=1}^N \sum_{m=1}^M \left| \frac{\partial h(\alpha, \Phi)}{\partial \phi_{i,m}} - \frac{\partial h(\hat{\alpha}, \hat{\Phi})}{\partial \hat{\phi}_{i,m}} \right|^2$$

$$\leq \max_{i,m} \left\{ \left| \frac{\partial h(\alpha, \Phi)}{\partial \phi_{i,m}} - \frac{\partial h(\hat{\alpha}, \hat{\Phi})}{\partial \hat{\phi}_{i,m}} \right|^2 \right\}.$$  

According to Lagrange’s mean value theorem, since $h(\alpha, \Phi)$ is continuous and differentiable, there exists some point $\hat{\phi}_{i,m}$.

#### TABLE II

| Execution time | (N, M, T) | CA approach | D-ADMM | L-BFGS | consensus-ADMM |
|---------------|-----------|-------------|--------|--------|----------------|
| (64,8,[0,16]) | 0.34s     | 316s        | 0.12s  | 0.08s  |
| (128,8,[0,16]) | 0.67s     | 950s        | 0.14s  | 0.09s  |
| (128,16,[0,32]) | 3.4s      | N/A         | 0.24s  | 0.26s  |
| (1024,128,[0,256]) | N/A       | N/A         | 5.1s   | 18.1s  |

Fig. 6. Comparison of correlation characteristics for interval [0, 256] with $M = 128, N = 1024.$
between \( \phi_{i,m} \) and \( \hat{\phi}_{i,m} \), which satisfies
\[
\frac{\partial h(\alpha, \Phi, \phi_{i,m}) - \partial h(\alpha, \hat{\Phi}, \phi_{i,m})}{\phi_{i,m} - \hat{\phi}_{i,m}} = \frac{\partial^2 h(\alpha, \Phi)}{\partial \phi_{i,m}^2}.
\]
(39)

Combining (38) and (39), we obtain
\[
\| \nabla \Phi \Phi(h(\alpha, \Phi), \Phi') - \nabla \Phi (h(\alpha, \Phi), \Phi') \|_F \leq \max_{i,m} \left\{ \left| \frac{\partial^2 h(\alpha, \Phi)}{\partial \phi_{i,m}^2} \right| \right\}.
\]
(40)

Based on (27b), we have
\[
\left| \frac{\partial^2 h(\alpha, \Phi)}{\partial \phi_{i,m}^2} \right| = \left| 2 \text{Re} \left( \frac{H \phi_{i,m}}{\partial \phi_{i,m}} \frac{\partial v}{\partial \phi_{i,m}} + \frac{\partial^2 v^H}{\partial \phi_{i,m}^2} \right) \right|
\leq 2 \left| \frac{\partial v^H}{\partial \phi_{i,m}} \frac{\partial v}{\partial \phi_{i,m}} \right| + 2 \left| \frac{\partial^2 v^H}{\partial \phi_{i,m}^2} \right|.
\]
(41)

Since \( \frac{\partial v}{\partial \phi_{i,m}} = 0 \), the first term in (41) can be derived as
\[
\left| \frac{\partial v^H}{\partial \phi_{i,m}} \frac{\partial v}{\partial \phi_{i,m}} \right| = \left| \frac{\partial z^H}{\partial \phi_{i,m}} A \frac{\partial z}{\partial \phi_{i,m}} \right|.
\]
(42)

Moreover, the vectors \( \frac{\partial \Phi}{\partial \phi_{i,m}} \) and \( \frac{\partial \Phi}{\partial R_{i,m}} \) have \( 2M - 1 \) nonzero constant modulus elements respectively (see (30)) and the maximum modulus of the elements in \( A \) is \( |\Theta| \). (42) can be further derived as
\[
\left| \frac{\partial v^H}{\partial \phi_{i,m}} \frac{\partial v}{\partial \phi_{i,m}} \right| \leq (2M - 1)^2 |\Theta|.
\]
(43)

Similarly, since \( \frac{\partial^2 v}{\partial \phi_{i,m}} = 0 \), we have
\[
\left| \frac{\partial^2 v^H}{\partial \phi_{i,m}^2} \right| \leq \left| \frac{\partial z^H}{\partial \phi_{i,m}^2} A \frac{\partial z}{\partial \phi_{i,m}} \right|.
\]
(44)

Since the maximum moduli of elements in \( q, z \) and \( A \) are \( P_\theta |\Theta|, N, \) and \( |\Theta| \) respectively. (44) can be further derived as
\[
\left| \frac{\partial^2 v^H}{\partial \phi_{i,m}^2} \right| \leq (2M - 1)(\alpha P_\theta |\Theta| + M^2 N |\Theta|).
\]
(45)

Combining (41), (43) and (45), we have
\[
\left| \frac{\partial^2 h(\alpha, \Phi)}{\partial \phi_{i,m}^2} \right| \leq 2 \left( (2M - 1)^2 + (2M - 1)(\alpha P_\theta + M^2 N) \right) |\Theta|
\leq 2(2M - 1)(2M - 1 + \alpha_{\text{max}} P_{\text{max}} + M^2 N) |\Theta|.
\]
(46)

where \( P_{\text{max}} = \max \{ P_\theta, \theta \in \Theta \} \). Combining the above results with (40), we obtain that \( \nabla \Phi h(\alpha, \Phi) \) is Lipschitz continuous with the constant \( L > 2(2M - 1)(2M - 1 + \alpha_{\text{max}} P_{\text{max}} + M^2 N) |\Theta| \).

Third, for \( \nabla f_n(\Phi) \), there exists
\[
\frac{\| \nabla f_n(\Phi) - \nabla f_n(\Phi') \|_F}{\| \Phi - \Phi' \|_F} \leq \max_{i,m} \left\{ \left| \frac{\partial^2 f_n(\Phi)}{\partial \phi_{i,m}^2} \right| \right\}.\]
(47)

For \( \frac{\partial^2 f_n(\Phi)}{\partial \phi_{i,m}^2} \), we have
\[
\left| \frac{\partial^2 f_n(\Phi)}{\partial \phi_{i,m}^2} \right| = \left| 2 \text{Re} \left( \frac{\partial B_n}{\partial \phi_{i,m}} \frac{\partial B_n}{\partial \phi_{i,m}} + B_n^H \frac{\partial^2 B_n}{\partial \phi_{i,m}^2} \right) \right|
\left\{ \begin{array}{c}
\leq 2 \left| \frac{\partial B_n}{\partial \phi_{i,m}} \frac{\partial B_n}{\partial \phi_{i,m}} \right| + 2 \left| B_n^H \frac{\partial^2 B_n}{\partial \phi_{i,m}^2} \right|.
\end{array} \right.
\]
(48)

Observing \( B_n, \frac{\partial B_n}{\partial \phi_{i,m}}, \) and \( \frac{\partial^2 B_n}{\partial \phi_{i,m}^2} \), we can find that their maximum moduli of the corresponding elements in them are \( w_cM^2 N, 2w_cM \) and \( 2w_c(M + 1) \) respectively, where \( w_c = \max \{ w_{ac}, w_{cc} \} \). Then, we have
\[
\left| \frac{\partial^2 f_n(\Phi)}{\partial \phi_{i,m}^2} \right| \leq 2(4w_c^2 M^2 K^2 + 2M^2 w_c^2 (M + 1) K^2)
= 4w_c^2 M^2 (2 + MN + N) K^2.
\]
(49)

Combining (47) and (49), we can see that functions \( \nabla f_n(\Phi), n \in \mathcal{T} \) are Lipschitz continuous with the constant \( L_n > 4w_c^2 M^2 (2 + MN + N) K^2 \). This concludes the proof. ■

**APPENDIX B**

**PROOF OF LEMMAS 2, 5**

**Lemma 2:** The following two inequalities exist
\[
U(\alpha, \Phi, \{ \hat{\Phi}_n^k, \Lambda_n^k \}, n \in \mathcal{T}) - \mathcal{L}(\alpha, \Phi, \{ \hat{\Phi}_n^k, \Lambda_n^k \}, n \in \mathcal{T}) \leq 2L_\alpha |\alpha - \alpha_k|^2 + 2L \| \Phi - \Phi^k_l \|^2.
\]
(50)

**Proof** For (50), its left side can be derived as
\[
U(\alpha, \Phi, \{ \hat{\Phi}_n^k, \Lambda_n^k \}, n \in \mathcal{T}) - \mathcal{L}(\alpha, \Phi, \{ \hat{\Phi}_n^k, \Lambda_n^k \}, n \in \mathcal{T})
= h(\alpha, \Phi, \Phi^k_l) - h(\alpha, \Phi) + \langle \nabla h(\alpha, \Phi), \Phi - \Phi^k_l \rangle
= h(\alpha, \Phi, \Phi^k_l) - h(\alpha, \Phi) - \langle \nabla h(\alpha, \Phi), \Phi - \Phi^k_l \rangle + \frac{L}{2} \| \Phi - \Phi^k_l \|^2 + \frac{L_\alpha}{2} |\alpha - \alpha_k|^2.
\]
(52)

Moreover, since \( \nabla h(\alpha, \Phi) \) is Lipschitz continuous, we can get
\[
\langle \nabla h(\alpha, \Phi), \Phi^k_l \rangle - h(\alpha, \Phi) \leq \frac{L}{2} \| \Phi - \Phi^k_l \|^2 + \frac{L_\alpha}{2} |\alpha - \alpha_k|^2.
\]
(53)

Plugging (53) into (52), we have the following derivations
\[
U(\alpha, \Phi, \{ \hat{\Phi}_n^k, \Lambda_n^k \}, n \in \mathcal{T}) - \mathcal{L}(\alpha, \Phi, \{ \hat{\Phi}_n^k, \Lambda_n^k \}, n \in \mathcal{T})
\leq \langle \nabla h(\alpha, \Phi), \Phi^k_l \rangle - \nabla h(\alpha, \Phi, \Phi^k_l) + L \| \Phi - \Phi^k_l \|^2
\]
(54)

Furthermore, since
\[
\langle \nabla h(\alpha, \Phi, \Phi^k_l) - \nabla h(\alpha, \Phi), \Phi^k_l \rangle - h(\alpha, \Phi, \Phi^k_l) \leq L \| \Phi - \Phi^k_l \|^2,
\]
(55)

we can get (50) from (54).

Exploiting the Lipschitz continuous property of \( \nabla f_n(\Phi) \) and similar derivations, (51) can also be obtained. ■
Lemma 3: in each iteration, \( \forall n \in \mathcal{T} \), \( \| \Lambda_{n+1} - \Lambda_n \|_F^2 \) can be bounded, i.e.,
\[
\| \Lambda_{n+1} - \Lambda_n \|_F^2 \leq 2L_n^2 \left( 2\| \Phi_{n+1} - \Phi_n \|_F^2 + 3\| \Phi_{n+1} - \Phi_n \|_F^2 \right). \tag{55}
\]

Proof The optimal solutions of the problems \((22b)\) can be obtained by solving \( \nabla f_n \Phi_n \), \( \Phi_n, \Lambda_n \) = 0, \( \forall n \in \mathcal{T} \), i.e.,
\[
\nabla f_n (\Phi_{n+1}) + \Lambda_n^k + (\rho_n + L_n)(\Phi_{n+1} - \Phi_k) = 0. \tag{56}
\]
Combining \((56)\) and \((22c)\), we can obtain
\[
\Lambda_{n+1}^k = -\nabla f_n (\Phi_{n+1}) - L_n(\Phi_{n+1} - \Phi_k). \tag{57}
\]
Plugging \((57)\) into \( \| \Lambda_{n+1} - \Lambda_n \|_F^2 \), we have the following derivations
\[
\| \Lambda_{n+1} - \Lambda_n \|_F^2 = \| \nabla f_n (\Phi_{n+1}) - \nabla f_n (\Phi_k) + L_n(\Phi_{n+1} - \Phi_k - \Phi_k + \Phi_k) \|_F^2 \\
\leq \| \nabla f_n (\Phi_{n+1}) - \nabla f_n (\Phi_k) \|_F^2 + 2L_n^2 \| \Phi_{n+1} - \Phi_k + \Phi_k \|_F^2 \\
\leq 2L_n^2 \| \Phi_{n+1} - \Phi_k \|_F^2 + 3\| \Phi_{n+1} - \Phi_k \|_F^2,
\]
where the second inequality comes from Lemma 1. This completes the proof.

Lemma 4: let \( c_n = \rho_n^3 - 7\rho_n^2 L_n - 8\rho_n L_n^2 - 32 L_n^3 \) and \( \bar{c}_n = \rho_n - 12\rho_n L_n^2 - 48 L_n^3 \). If \( c_n > 0 \) and \( \bar{c}_n > 0 \), in each consensus-ADMM iteration, the augmented Lagrangian function \( L(\cdot) \) decreases sufficiently, i.e.,
\[
L(\alpha^k, \Phi_k, \{ \Phi_n^k, \Lambda_n^k \}, n \in \mathcal{T}) \\
- L(\alpha^{k+1}, \Phi_{k+1}, \{ \Phi_n^{k+1}, \Lambda_n^{k+1} \}, n \in \mathcal{T}) \geq \sum_{n \in \mathcal{T}} \frac{1}{2\rho_n} (\bar{c}_n \| \Phi_{n+1} - \Phi_n \|_F^2 + c_n \| \Phi_{n+1} - \Phi_k \|_F^2) + \frac{L}{2} \| \Phi_{k+1} - \Phi_k \|_F^2 + \frac{L}{2} \alpha^{k+1} - \alpha^k. \tag{58}
\]

Proof Define the following quantities
\[
\Delta_{\alpha^k, \Phi} = L(\alpha^k, \Phi^k, \{ \Phi_n^k, \Lambda_n^k \}, n \in \mathcal{T}) \\
- L(\alpha^{k+1}, \Phi_{k+1}, \{ \Phi_n^{k+1}, \Lambda_n^{k+1} \}, n \in \mathcal{T}),
\]
\[
\Delta_{\Phi} = L_n(\Phi_{k+1}, \Lambda_n^{k+1}) - L_n(\Phi_{k+1}, \Lambda_n^{k+1}),
\]
\[
\Delta_{\Lambda} = L_n(\Phi_{k+1}, \Phi_{k+1}, \Lambda_n^{k+1}) - L_n(\Phi_{k+1}, \Phi_{k+1}, \Lambda_n^{k+1}).
\]
Then, we can get
\[
L(\alpha^k, \Phi, \{ \Phi_n^k, \Lambda_n^k \}, n \in \mathcal{T}) \\
- L(\alpha^{k+1}, \Phi_{k+1}, \{ \Phi_n^{k+1}, \Lambda_n^{k+1} \}, n \in \mathcal{T}) \\
= \Delta_{\alpha^k, \Phi} + \sum_{n \in \mathcal{T}} (\Delta_{\alpha^k, \Lambda} + \Delta_{\Phi}^k). \tag{59}
\]

For \( \Delta_{\alpha^k, \Phi} \), we have the following inequality
\[
\Delta_{\alpha^k, \Phi} \geq L(\alpha^k, \Phi^k, \{ \Phi_n^k, \Lambda_n^k \}, n \in \mathcal{T}) \\
- U(\alpha^{k+1}, \Phi_{k+1}, \{ \Phi_n^{k+1}, \Lambda_n^{k+1} \}, n \in \mathcal{T}). \tag{60}
\]
Moreover, according to Lemma 2, there exists
\[
U(\alpha^k, \Phi^k, \{ \Phi_n^k, \Lambda_n^k \}, n \in \mathcal{T}) = L(\alpha^k, \Phi^k, \{ \Phi_n^k, \Lambda_n^k \}, n \in \mathcal{T}).
\]
Then, the inequality \((60)\) can be further derived as follows
\[
\Delta_{\alpha^k, \Phi} \geq U(\alpha^k, \Phi^k, \{ \Phi_n^k, \Lambda_n^k \}, n \in \mathcal{T}) \\
- U(\alpha^{k+1}, \Phi_{k+1}, \{ \Phi_n^{k+1}, \Lambda_n^{k+1} \}, n \in \mathcal{T}), \tag{61}
\]
\[
L + \sum_{n \in \mathcal{T}} \rho_n \geq \frac{2}{\| \Phi_{k+1} - \Phi_k \|_F^2 + \frac{L_n}{2} \| \Phi_{k+1} - \Phi_k \|_F^2 \geq \| \alpha^{k+1} - \alpha^k \|^2 \}
\]
where the second inequality is true because the function \( U(\alpha, \Phi, \{ \Phi_n^k, \Lambda_n^k \}, n \in \mathcal{T}) \) is strongly convex with respect to \( \alpha \) and \( \Phi \).

For \( \Delta_{\Phi} \), according to Lemma 2, it can be rewritten as
\[
\Delta_{\Phi} \geq L_n(\Phi_{k+1}, \Phi_{k+1}, \Lambda_n^k) - U_n(\Phi_{k+1}, \Phi_{k+1}, \Lambda_n^k) \tag{62}
\]
Moreover, we have the following two inequalities
\[
L_n(\Phi_{k+1}, \Phi_{k+1}, \Lambda_n^k) - U_n(\Phi_{k+1}, \Phi_{k+1}, \Lambda_n^k) \geq -2L_n(\Phi_{k+1} - \Phi_k)^2, \tag{63}
\]
\[
U_n(\Phi_{k+1}, \Phi_{k+1}, \Lambda_n^k) - U_n(\Phi_{k+1}, \Phi_{k+1}, \Lambda_n^k) \geq \rho_n(\Phi_{k+1} - \Phi_k)^2, \tag{64}
\]
where the first inequality comes from Lemma 2 and the second inequality is true because the functions \( U_n(\Phi_{k+1}, \Phi_{k+1}, \Lambda_n^k) \), \( n \in \mathcal{T} \), are strongly convex with respect to \( \Phi_n \). Then, we have
\[
\Delta_{\Phi} \geq -2L_n(\Phi_{k+1} - \Phi_k)^2 + \rho_n + L_n(\Phi_{k+1} - \Phi_k)^2 \geq 0. \tag{65}
\]

For \( \Delta_{\Lambda} \), through similar derivations, we have
\[
\Delta_{\Lambda} \geq -\frac{2L_n^2}{\rho_n} (2(\Phi_{k+1} - \Phi_k)^2 + 3(\Phi_{k+1} - \Phi_k)^2). \tag{66}
\]

Plugging \((61), (65)\), and \((66)\) into \((59)\), we can obtain
\[
L(\alpha^k, \Phi, \{ \Phi_n^k, \Lambda_n^k \}, n \in \mathcal{T}) \\
- L(\alpha^{k+1}, \Phi_{k+1}, \{ \Phi_n^{k+1}, \Lambda_n^{k+1} \}, n \in \mathcal{T}) \geq \sum_{n \in \mathcal{T}} \frac{1}{2\rho_n} (\bar{c}_n \| \Phi_{n+1} - \Phi_n \|_F^2 + c_n \| \Phi_{n+1} - \Phi_k \|_F^2) + \frac{L}{2} \| \Phi_{k+1} - \Phi_k \|_F^2 + \frac{L}{2} \alpha^{k+1} - \alpha^k,
\]
where \( \bar{c}_n = \rho_n - 7\rho_n^2 L_n - 8\rho_n L_n^2 - 32 L_n^3 \) and \( c_n = \rho_n - 12\rho_n L_n^2 - 48 L_n^3 \). So, \( \forall n \in \mathcal{T} \), if \( \bar{c}_n > 0 \) and \( c_n > 0 \), then the augmented Lagrangian function \( L(\cdot) \) decreases sufficiently. This completes the proof.

Lemma 5: if \( \rho_n > 5L_n \), augmented Lagrangian function \( L(\alpha^k, \Phi_{k+1}, \{ \Phi_n^{k+1}, \Lambda_n^{k+1} \}, \forall k \), is larger than zero.

Proof First, we consider \( L_n(\Phi_{k+1}, \Phi_{k+1}, \Lambda_n^{k+1}) \). Plugging
in Lemma 5, we desire
\[ L_n(\Phi^{k+1}, \Phi^{k+1}, \Lambda^{k+1}_n) = f_n(\Phi^{k+1}) + \left(\frac{\rho_n}{2} - L_n\right)\|\Phi^{k+1} - \Phi^k\|^2_F + \langle \nabla f_n(\Phi^{k+1}), \Phi^{k+1} - \Phi^k \rangle \] (67)

Since \( \nabla f_n(\Phi) \) is Lipschitz continuous, we have the following equality for the last term in (67).
\[ \langle \nabla f_n(\Phi^{k+1}), \Phi^{k+1} - \Phi^k \rangle \geq \langle \nabla f_n(\Phi^{k+1}), \Phi^{k+1} - \Phi^k \rangle - L_n\|\Phi^{k+1} - \Phi^k\|^2_F \]

Plugging it into (67), we can get
\[ L_n(\Phi^{k+1}, \Phi^{k+1}, \Lambda^{k+1}_n) \geq f_n(\Phi^{k+1}) + \langle \nabla f_n(\Phi^{k+1}), \Phi^{k+1} - \Phi^k \rangle + \left(\frac{\rho_n}{2} - 2L_n\right)\|\Phi^{k+1} - \Phi^k\|^2_F \] (68)

Moreover, we can also exploit \( \nabla f_n(\Phi^k) \)'s Lipschitz continuous property and get the following inequality
\[ f_n(\Phi^{k+1}) \leq f_n(\Phi^{k+1}) + \langle \nabla f_n(\Phi^{k+1}), \Phi^{k+1} - \Phi^k \rangle + \frac{L_n}{2}\|\Phi^{k+1} - \Phi^k\|^2_F \]

Plugging it into (68), we can obtain
\[ L_n(\Phi^{k+1}, \Phi^{k+1}, \Lambda^{k+1}_n) \geq f_n(\Phi^{k+1}) + \langle \nabla f_n(\Phi^{k+1}), \Phi^{k+1} - \Phi^k \rangle + \frac{\rho_n - 5L_n}{2}\|\Phi^{k+1} - \Phi^k\|^2_F \] (69)

Second, plugging (69) into \( L(\alpha^{k+1}, \Phi^{k+1}, \Lambda^{k+1}_n, n \in \mathcal{T}) \), we can get
\[ L(\alpha^{k+1}, \Phi^{k+1}, \Lambda^{k+1}_n, n \in \mathcal{T}) \geq h(\alpha^{k+1}, \Phi^{k+1}) + \sum_{n \in \mathcal{T}} \left( f_n(\Phi^{k+1}) + \frac{\rho_n - 5L_n}{2}\|\Phi^{k+1} - \Phi^k\|^2_F \right) \] (70)

Since \( h(\alpha^{k+1}, \Phi^{k+1}) \geq 0 \) and \( f_n(\Phi^{k+1}) \geq 0 \), we can conclude that, if \( \rho_n > 5L_n \), \( L(\alpha^{k+1}, \Phi^{k+1}, \Lambda^{k+1}_n, n \in \mathcal{T}) \) must be larger than zero. This completes the proof. ■

**APPENDIX C**

**PROOF OF THEOREM 1**

First, we prove (55) in Theorem 1. It can be seen that in Lemma 3 we desire \( \bar{c}_n = \rho_n^2 - 7\rho_nL_n - 8\rho_nL_n^2 - 32L_n^3 \geq 0 \) and \( \bar{c}_n = \rho_n^2 - 12\rho_nL_n^2 - 48L_n^3 \geq 0 \), and in Lemma 5 we desire \( \rho_n > 5L_n \). It is easy to see that if \( \forall n \in \mathcal{T}, \rho_n \geq 9L_n \), both Lemma 3 and Lemma 5 are tenable. Then, summing both sides of the inequality (58) when \( k = 1, 2, \ldots, +\infty \), we can obtain
\[ L(\alpha^1, \Phi^1, \Lambda^1_n, n \in \mathcal{T}) \geq \sum_{k=1}^{+\infty} \sum_{n \in \mathcal{T}} \frac{1}{2\rho_n} (\bar{c}_n\|\Phi^{k+1} - \Phi^k\|^2_F + \bar{c}_n\|\Phi^{k+1} - \Phi^k\|^2_F) \]

Since \( \bar{c}_n > 0, \bar{c}_n > 0, \) and \( \lim_{k \to +\infty} L(\alpha^{k+1}, \Phi^{k+1}, \{\Phi^{k+1}_n, \Lambda^{k+1}_n, n \in \mathcal{T}\}) \geq 0 \), we can obtain (71). (72), and (73).
\[ \lim_{k \to +\infty} |\alpha^{k+1} - \alpha^k| = 0. \] (71)
\[ \lim_{k \to +\infty} \|\Phi^{k+1} - \Phi^k\|_F = 0. \] (72)
\[ \lim_{k \to +\infty} \|\Phi^{k+1} - \Phi^k\|_F = 0, \forall \ n \in \mathcal{T}. \] (73)

Plugging (72), (73) into (55)’s right side, we can get
\[ \lim_{k \to +\infty} \|\Lambda^{k+1}_n - \Lambda^*_n\|_F = 0. \] (74)

Combining (74) and (22a), we further have
\[ \lim_{k \to +\infty} \|\Phi^{k+1}_n - \Phi^*_n\|_F = 0. \] (75)

Since \( \Phi^k, \Phi^*_n \in [0, 2\pi] \) and \( \alpha^*_n \in (0, \alpha_{\max}] \), we can see that sequence, either \( \{\Phi^k\} \), \( \{\Phi^k_n, n \in \mathcal{T}\} \), or \( \{\alpha^*_n\} \), is bounded. This result infers that the sequence \( \{\Lambda^*_n\} \) is also bounded. Then, we can get the following convergence result
\[ \lim_{k \to +\infty} \alpha^k = \alpha^*, \lim_{k \to +\infty} \Phi^k = \Phi^*, \lim_{k \to +\infty} \Phi^k_n = \Phi^*_n, \lim_{k \to +\infty} \Lambda^k_n = \Lambda^*_n, \forall n \in \mathcal{T}. \] (76)

Second, we prove \((\alpha^*, \Phi^*)\) is some stationary point of problem 8.

Since the function \( \mathcal{U}(\alpha, \Phi, \{\Phi^k_n, \Lambda^k_n, n \in \mathcal{T}\}) \) is strongly convex, we have
\[ \langle \nabla \mathcal{U}(\alpha^*, \Phi^*), \{\Phi^k_n, \Lambda^k_n, n \in \mathcal{T}\} \rangle, \alpha - \alpha^* \rangle \geq 0, \] (77a)
\[ \langle \nabla \mathcal{U}(\alpha^*, \Phi^*), \{\Phi^k_n, \Lambda^k_n, n \in \mathcal{T}\} \rangle, \Phi - \Phi^* \rangle \geq 0, \] (77b)

where \( \alpha \in (0, \alpha_{\max}] \). (77) can be further equivalent to
\[ \langle \nabla \mathcal{U}(\alpha^*, \Phi^*), L(\alpha^*, \Phi^*) - L(\alpha^*, \Phi^*) - \sum_{n \in \mathcal{T}} (\rho_n(\Phi^k_n - \Phi^*) + \Lambda^*_n), \Phi - \Phi^* \rangle \geq 0 \]

When \( k \to +\infty \), plugging the convergence results (75) and (76) into (78), this can be written as
\[ \langle \nabla \mathcal{U}(\alpha^*, \Phi^*), \{\Phi^k_n, \Phi^k_n, \Lambda^k_n, n \in \mathcal{T}\} \rangle, \alpha - \alpha^* \rangle \geq 0, \alpha \in (0, \alpha_{\max}] \], (79a)
\[ \langle \nabla \mathcal{U}(\alpha^*, \Phi^*), - \sum_{n \in \mathcal{T}} \Lambda^*_n, \Phi - \Phi^* \rangle \geq 0. \] (79b)

Moreover, since \( e(\alpha, X(\Phi)) = h(\alpha, \Phi) \) and \( P_e(X(\Phi)) = \sum_{n \in \mathcal{T}} \nabla f_n(\Phi) \), (79a) and (80) can be rewritten as
\[ \langle \nabla_e e(\alpha^*, X(\Phi^*)), \alpha - \alpha^* \rangle \geq 0, \alpha \in (0, \alpha_{\max}], \]
\[ \langle \nabla e(\alpha^*, X(\Phi^*)), + \sum_{n \in \mathcal{T}} P_e(X(\Phi^*)), \Phi - \Phi^* \rangle \geq 0, \]

which completes the proof. ■
APPENDIX D

PROOF OF THE CONVERGENCE OF THE PROPOSED ALGORITHM WITH SBCD METHOD

Performing expectation on both sides of (81) and (82) respectively, we can get inequalities (81) and (82).

\[
E \left[ \mathcal{L}(\alpha^k, \Phi^k, \{\Phi^k, \Lambda^k, n \in T\}) \right] 
\geq \sum_{n \in T} p_{\text{min}} \left( c_n \|\Phi^k + L_1 n \|_F + c_n \|\Phi^k + L_1 n \|_F \right) 
+ \frac{L}{2} \|\Phi^k - \Phi^{k+1}\|_F^2 
\geq h(\alpha^{k+1}, \Phi^{k+1}) 
\]

\[
\sum_{n \in T} p_n \left( f_n \|\Phi^k + L_1 n \|_F^2 \right) \geq \frac{L_1}{2} \|\Phi^k - \Phi^{k+1}\|_F^2 \geq 0.
\]

where the probabilities \( p_{\text{min}} \) and \( p_n \) are defined in (81) and (82). (81) and (82) indicate that the augmented Lagrangian function decreases sufficiently in each consensus-ADMM iteration and it is also lower-bounded as \( k \to +\infty \) respectively. Through similar derivations in Appendix C, we can conclude that the consensus-ADMM algorithm with SBCD strategy is convergent with high probability to some stationary point of problem (8).
[37] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” Found. Trend. Mach. Learn., vol. 3, no. 1, pp. 1-122, Jan. 2011.