Waveform Design for MIMO Radar with Matrix Completion

Shunqiao Sun, Student Member, IEEE, and Athina P. Petropulu, Fellow, IEEE

Abstract

It was recently shown that MIMO radars with sparse sensing and matrix completion (MC) can significantly reduce the volume of data required for accurate target detection and estimation. Based on sparsely sampled target returns, forwarded by the receive antennas to a fusion center, a matrix, referred to as the data matrix, can be partially filled, and subsequently completed via MC techniques. The completed data matrix can then be used in standard array processing methods to estimate the target parameters. This paper studies the applicability of MC theory on the data matrix arising in colocated MIMO radars using uniform linear arrays. It is shown that the data matrix coherence, and consequently the performance of MC, is directly related to the transmit waveforms. Among orthogonal waveforms, the optimum choices are those for which, any snapshot across the transmit array has a flat spectrum. The problem of waveform design is formulated as an optimization problem on the complex Stiefel manifold, and is solved via the modified steepest descent method, or the modified Newton algorithm with nonmonotone line search. It is shown that, under the optimal waveforms, the coherence of the data matrix is asymptotically optimal with respect to the number of transmit and receive antennas.

Index Terms

MIMO radar, matrix completion, waveform design, complex Stiefel manifold

S. Sun and A. P. Petropulu are with the Department of Electrical and Computer Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA (E-mail: {shunq.sun, athinap}@rutgers.edu).

This work was supported by the Office of Naval Research under grant ONR-N-00014-12-1-0036 and NSF under grant ECCS-1408437. Partial results of this paper have been published in 48th Annual Asilomar Conference on Signals, Systems, and Computers (Asilomar’14) [1] and IEEE Global Conference on Signal and Information Processing (GlobalSIP’14) [2].
I. INTRODUCTION

Unlike traditional phased-array radars which transmit fully correlated signals through their transmit antennas, multiple-input multiple-output (MIMO) radars [3] transmit mutually orthogonal signals. The orthogonality allows the receive antennas to separate the transmitted signals via matched filtering. The target parameters are obtained by processing the phase shifts of the received signals. MIMO radars offer a high degree of freedom [4] and consequently, enable improved resolution. Since the transmitted signals are orthogonal, the transmit beam is not focused on a particular direction [5], thus resulting in decrease of illumination power. Pulse compression techniques are typically used to strengthen the receive signal power. The transmit pulse can be coded or modulated, e.g., phase-coded pulse or linear frequency modulated (LFM) pulse [6].

The idea of using low-rank matrix completion (MC) techniques in MIMO radars (termed as MIMO-MC radars) was first proposed in [7], [8] as means of reducing the volume of data required for accurate target detection and estimation. In particular, [8] considers a colocated pulse MIMO radar scenario with uniform linear arrays (ULAs) at the transmitter and the receiver. Each receive antenna samples the target returns and forwards the obtained samples to a fusion center. Based on the data received, the fusion center formulates a matrix, referred to as “data matrix”, which can then be used in standard array processing methods for target detection and estimation. If the data samples are obtained in a Nyquist fashion, and the number of targets is small relative to the number of transmit and receive antennas, the data matrix is low-rank [7]. Thus, it can be recovered from a small subset of its uniformly spaced elements via matrix completion (MC) techniques. By exploiting the latter fact, the MIMO-MC radar receive antennas obtain a small number of samples at uniformly random sampling times. Based on knowledge of the sampling instances, the fusion center populates the data matrix in a uniformly sparse fashion, and subsequently recovers the full matrix via MC techniques. This is referred to as Scheme II in [8]. Alternatively, the receive antennas can perform matched filtering with a randomly selected set of transmit waveforms, and forward the results to the fusion center, partially populating the data matrix. The data matrix is subsequently completed via MC. This is referred to as Scheme I in [8]. Both Schemes I and II reduce the number of samples that need to be forwarded to the fusion center. If the transmission occurs in a wireless fashion, this translates to savings in power and bandwidth. As compared to MIMO radars based on sparse signal recovery [9], [10], [11], MIMO-MC radars achieve similar performance but without requiring a target space grid.

Details on the general topic of matrix completion and the conditions for matrix recovery can be
found in [12], [13], [14]. The conditions for the applicability of MC on Scheme I of [8] ULA can be found in [15]. In that case, and under ideal conditions, the transmit waveforms do not affect the MC performance. On the other hand, for Scheme II of [8], it was shown in [8], [16] that the transmit waveforms affect the matrix completion performance, as they directly affect the data matrix coherence [13]; a larger coherence implies that more samples need to be collected for reliable target estimation. That observation motivates the work in this paper, where we explore the relationship between matrix coherence and transmit waveforms for Scheme II of [8], and design waveforms that result in the lowest possible coherence. Waveform design in the context of MIMO radars has been extensively studied. For example, in [17], [18], the waveforms are designed to maximize the mutual information between target impulse response (which are assumed known) and reflected signals. In [19], waveforms with good correlation properties are designed under the unimodular constraint. In [20], clutter mitigation is considered in waveform design for ground moving-target indication (GMTI). Orthogonal phase-coded waveforms are designed in [21] via simulated annealing. Frequency hopping waveforms are designed in [22] to reduce the sidelobe of the radar ambiguity function [23]. For MIMO radars with sparse signal recovery, e.g., in [24], [25], the goal of waveform design is to reduce the coherence of sensing matrix.

In this paper, we aim to design waveforms so that the coherence of the receive data matrix attains its lowest possible value. Since the waveforms are constrained to be orthogonal, we formulate the design problem as an optimization problem on the complex Stiefel manifold [26], and for its solution employ the modified steepest descent algorithm [27] and the modified Newton algorithm. The local gradient and Hessian of the cost functions are derived in closed form. Under the obtained optimized waveforms, numerical results show that, as the number of transmit antennas increases, the matrix coherence approaches its smallest possible value, i.e., 1.

The rest of this paper is organized as follows. Some background on noisy matrix completion and MIMO-MC radars is provided in Section II. The matrix coherence analysis results are presented in Section III, while the waveform design is addressed in Section IV. Simulations results are given in Section V. Finally, Section VI provides some concluding remarks.

**Notation:** We use lower-case and upper-case letters in bold denote vectors and matrices, respectively. See Table I for other notations used in the paper.
II. BACKGROUND

A. Matrix Completion

In this section we provide a brief overview of the problem of recovering a rank-$r$ matrix $M \in \mathbb{C}^{n_1 \times n_2}$ based on partial knowledge of its entries [12], [13], [14].

Let us define the observation operation $Y = \mathcal{P}_\Omega (M)$ as

$$[Y]_{ij} = \begin{cases} [M]_{ij}, & (i,j) \in \Omega \\ 0, & \text{otherwise} \end{cases}$$

(1)

where $\Omega$ is the set of indices of observed entries with cardinality $m$. According to [13], when $M$ is low-rank and meets certain conditions (see (A0) and (A1), later in this section), $M$ can be estimated by solving a nuclear norm optimization problem

$$\min \|X\|_*$$
$$\text{s.t.} \quad \mathcal{P}_\Omega (X) = \mathcal{P}_\Omega (M)$$

(2)

The conditions for successful matrix completion involve the notion of coherence, which is defined as follows.
\textbf{Definition 1.} (Matrix Coherence \cite{12}): Let $U$ be a subspace of $\mathbb{C}^{n_1}$ of dimension $r$ that is spanned by the set of orthogonal vectors $\{u_i \in \mathbb{C}^{n_1}\}_{i=1,\ldots,r}$, $P_U$ be the orthogonal projection onto $U$, i.e., $P_U = \sum_{1 \leq i \leq r} u_i u_i^H$, and $e_i$ be the standard basis vector whose $i$-th element is 1. The coherence of $U$ is defined as

$$
\mu(U) = \frac{n_1}{r} \max_{1 \leq i \leq n_1} \|P_U e_i\|^2 = \frac{n_1}{r} \max_{1 \leq i \leq n_1} \left\| U^{(i)} \right\|^2 \in \left[ 1, \frac{n_1}{r} \right],
$$

where $U^{(i)}$ denotes the $i$-th row of matrix $U = [u_1, \ldots, u_r]$.

It can be found from Definition 1 that the minimum possible coherence that any given matrix can achieve is one. Let the compact singular value decomposition (SVD) of $M$ be $M = \sum_{k=1}^{r} \rho_k u_k v_k^H$, where $\rho_k, k = 1, \ldots, r$ are the singular values, and $u_k$ and $v_k$ the corresponding left and right singular vectors, respectively. Let $U, V$ be the subspaces spanned by $u_k$ and $v_k$, respectively. Matrix $M$ has coherence with parameters $\mu_0$ and $\mu_1$ if

(A0) $\max (\mu(U), \mu(V)) \leq \mu_0$ for some positive $\mu_0$.

(A1) The maximum element of the $n_1 \times n_2$ matrix $\sum_{1 \leq i \leq r} u_i v_i^H$ is bounded by $\mu_1 \sqrt{r/(n_1 n_2)}$ in absolute value, for some positive $\mu_1$.

In fact, it was shown in \cite{12} that if (A0) holds, then (A1) also holds with $\mu_1 \leq \mu_0 \sqrt{r}$.

Now, suppose that matrix $M \in \mathbb{C}^{n_1 \times n_2}$ satisfies (A0) and (A1). The following theorem gives a probabilistic bound for the number of entries, $m$, needed to estimate $M$.

\textbf{Theorem 1.} \cite{12} Suppose that we observe $m$ entries of the rank--$r$ matrix $M \in \mathbb{C}^{n_1 \times n_2}$, with matrix coordinates sampled uniformly at random. Let $n = \max\{n_1, n_2\}$. There exist constants $C$ and $c$ such that if

$$
m \geq C \max \left\{ \mu_1^2, \mu_0^{1/2}, \mu_1, \mu_0 n^{1/4} \right\} nr \beta \log n
$$

for some $\beta > 2$, the minimizer to the program of (2) is unique and equal to $M$ with probability at least $1 - cn^{-\beta}$.

For $r \leq \mu_0^{-1} n^{1/5}$ the bound can be improved to

$$
m \geq C \mu_0 n^{6/5} r \beta \log n,$$

without affecting the probability of success.
Theorem 1 implies that the lower the coherence parameter $\mu_0$, the fewer entries of $M$ are required to estimate $M$. Consequently, the intuitive behind the coherence definition is that for a given matrix, in order to use as small as possible number of its entries for fully recovery with MC, the singular vector of the matrix should be uncorrelated to the standard basis, in other words, the singular vectors need to be “sufficiently spread” [12]. To apply the matrix completion techniques to recover a given matrix, the coherence should be as close to unity as possible. In this paper, we say the coherence of a given matrix is optimal so long as it achieves its lowest bound.

In practice, the observations are typically corrupted by noise, i.e., $[Y]_{ij} = [M]_{ij} + [E]_{ij}, (i, j) \in \Omega$, where, $[E]_{ij}$ represents noise. In that case, it holds that $P_\Omega (Y) = P_\Omega (M) + P_\Omega (E)$, and the completion of $M$ is done by solving the following optimization problem [14]

$$\min \|X\|_*$$
$$\text{s.t. } \|P_\Omega (X - Y)\|_F \leq \delta.$$  \hspace{1cm} (4)

Assuming that the noise is zero-mean with variance $\sigma^2$, $\delta > 0$ is a parameter defined as $\delta^2 = (m + \sqrt{8m})\sigma^2$ [12]. Let $\hat{M}$ be the solution of (4). It can be shown that the error norm $\|M - \hat{M}\|_F$ is bounded as [14]

$$\|M - \hat{M}\|_F \leq 4\sqrt{(2n_1n_2 + m)\min (n_1, n_2) \delta + 2\delta}$$  \hspace{1cm} (5)

with high probability. If $M$ has favorable coherence properties, the matrix completion for both noiseless and noisy cases will be stable. Therefore, to achieve satisfactory performance for matrix completion, it is very important to investigate the conditions under which the coherence of $M$ will be as low as possible.

B. MIMO-MC Radar

We consider the problem formulation proposed in [8] for scheme II. The scenario involves narrowband orthogonal transmit waveforms, transmitted in pulses with pulse repetition interval $T_{PRI}$ and carrier wavelength $\lambda$, $K$ far-filed targets at angles $\theta_k$, and ULAs for transmission and reception, equipped with $M_t$ transmit and $M_r$ receive antennas, respectively, and inter-element spacing $d_t$ and $r_r$, respectively (see Subsection B of Section II in [8]).

During each pulse, the $m$-th, $m \in \mathbb{N}_+^{M_t}$ antenna transmits a coded waveform containing $N$ symbols $\{s_m (n)\}, n = 1, \ldots , N$ of duration $T_b$ each, which can be written in the baseband as

$$\phi_m (t) = \sum_{n=1}^N s_m (n)\Delta \left[ \frac{t - (n - 1)T_b}{T_b} \right], t \in [0, T_\phi],$$  \hspace{1cm} (6)

January 29, 2015 DRAFT
where $s_m (n) = a_m (n) e^{j \varphi_m (n)}$, with $\{ \varphi_m (n) \}$ uniformly distributed in $[-\pi, \pi]$, and $\{ a_m (n) \}$ taking arbitrary positive values; $\Lambda (t)$ is a rectangular pulse and $T_\phi = NT_b$ is the duration of the entire pulse. We will assume that the waveforms are sufficiently narrowband, i.e., the roundtrip delay from the transmit antennas to the target and back to the receive antennas is smaller than $T_b$ and thus can be ignored.

Suppose that the receive antennas sample the target echoes with sampling interval $T_b$ and forward their samples to a fusion center. Let $X$ be the matrix formulated at the fusion center based on receive antenna samples data with each antenna contributing a row to $X$. It holds that

$$X = W + J,$$

where $J$ is an interference/noise matrix and

$$W = BD A^T S^T,$$

where $A \in \mathbb{C}^{M_t \times K}$ is the transmit steering matrix (respectively defined is $B \in \mathbb{C}^{M_r \times K}$, the receive steering matrix) defined as

$$A \overset{\Delta}{=} \begin{bmatrix} \gamma_1^1 & \gamma_2^1 & \cdots & \gamma_K^1 \\ \gamma_1^2 & \gamma_2^2 & \cdots & \gamma_K^2 \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_1^{M_t} & \gamma_2^{M_t} & \cdots & \gamma_K^{M_t} \end{bmatrix},$$

with

$$\gamma_k^m \overset{\Delta}{=} e^{j 2\pi (m-1) \alpha_k^t},$$

$$\alpha_k^t \overset{\Delta}{=} \frac{d_t \sin (\theta_k)}{\lambda}, (m, k) \in \mathbb{N}_+^{M_t} \times \mathbb{N}_+^{K},$$

where $\theta_k$ is the angle of the $k$-th target, or equivalently, $\alpha_k^t$ is the spatial frequency corresponding to the $k$-th target,

$$D \overset{\Delta}{=} \text{diag} \left( \begin{bmatrix} \beta_1 \zeta_1 & \beta_2 \zeta_2 & \cdots & \beta_K \zeta_K \end{bmatrix} \right),$$

where $\zeta_k = e^{j 2\pi \nu_k (q-1) T_{PRI}}$, with $q$ denoting the pulse index, $\nu_k = \frac{2 d_t}{\lambda}$ denoting the Doppler shift of the $k$-th target, and $\{ \beta_k \}_{k \in \mathbb{N}_+^K}, \{ \theta_k \}_{k \in \mathbb{N}_+^K}$ denoting contain target reflection coefficients and speeds, respectively. $S = [s (1), \ldots, s (N)]^T \in \mathbb{C}^{N \times M_t}$ with $s (i) = [s_1 (i), \ldots, s_{M_t} (i)]^T$, is the sampled waveform matrix, with its vertical dimension corresponding to sampling along time and its horizontal dimension corresponding to sampling across the array (sampling in space). The $i$-th row of $S$ can be thought of as the snapshot of the waveforms across the transmit antennas at sampling time $i$. Due to the assumed orthogonality of the waveforms, it holds that $S^H S = I_{M_t}$ when $N \geq M_t$ [28].
When both $M_r$ and $N$ are larger than $K$, the noise free data matrix $W$ is rank-$K$ and can be recovered from a small number of its entries via matrix completion. This fact motivated the approach of [7] [8], which calls for subsampling the target echoes at the receive antennas in a uniformly pseudo-random fashion, partially filling the data matrix at the fusion center, and then completing the data matrix via MC techniques. This approach reduces the number of samples that need to be forwarded to the fusion center. It turns out that the applicability of MC depends on the transmit waveforms. In the next section, we derive the necessary and sufficient conditions the transmit waveforms must satisfy so that the coherence of $W$ is asymptotical optimal, i.e., it approaches 1 as $M_r$ increases.

### III. Coherence of $W$ and Optimal Waveform Conditions

In this section, we analyze the coherence of $W$, for the MIMO radar system defined in the previous section. In particular,

- We provide sufficient and necessary conditions for the optimal transmit waveforms under which the coherence of $W$ attains its lowest possible value;
- Under those conditions, we show asymptotic optimality of the coherence of $W$ w.r.t. the number of transmit/receive antennas;
- We show that the coherence of $W$ does not depend on the Doppler shift.

#### A. The Coherence of $W$

Let $S_i(\alpha_k^t)$ denote the discrete-time Fourier transform (DTFT) of the $i$-th snapshot of the transmit waveforms evaluated at spatial frequency $\alpha_k^t$, i.e.,

$$S_i(\alpha_k^t) = \sum_{m=1}^{M_t} s_m(i) e^{-j2\pi m\alpha_k^t}$$  \hspace{1cm} (13)

where $\{ s_m(i) \}_{m \in \mathbb{N}_+}$ are the elements in the $i$-th row of $S$.

Before we proceed we provide a lemma that will be useful in the subsequent theorems.

**Lemma 1.** For the MIMO radar system described in Section II(B) with $d_t = \lambda/2$, and $K$ targets randomly located at angles $\{ \theta_k \in [-\pi/2, \pi/2] \}_{k \in \mathbb{N}_+ K}$, or equivalently, at spatial frequencies $\{ \alpha_k^t \in [-\frac{1}{2}, \frac{1}{2}] \}_{k \in \mathbb{N}_+ K}$, it holds that

$$\sum_{i=1}^{N} \sum_{k=1}^{K} \left| S_i(\alpha_k^t) \right|^2 = KM_t.$$  \hspace{1cm} (14)

**Proof of Lemma 1.** See the Appendix A for the proof.
The waveforms conditions under which the coherence of \( W \) attains its lowest value are summarized in the following theorem.

**Theorem 2.** *(Optimal Waveform Conditions):* Consider the MIMO radar systems as defined in Section II(B), with \( d_t = \frac{\lambda}{2} \).

The necessary condition under which the coherence of \( W \) attains its lowest possible value is
\[
\sum_{k=1}^{K} |S_i(\alpha_k^t)|^2 = \frac{KM_t}{N}, \quad \text{for all } i \in \mathbb{N}_N^+.
\]

**(15)**

A sufficient condition for the coherence of \( W \) to attain its lowest possible value, independent of the target angles is
\[
|S_i(\alpha_l^t)|^2 = \frac{M_t}{N}, \quad \text{for all } i \in \mathbb{N}_N^+ \text{ and } \alpha_l^t \in \left[-\frac{1}{2}, \frac{1}{2}\right].
\]

**(16)**

**Proof of Theorem 2.** To make the proof more tractable, we break it into two parts. In the first part, we characterize the SVD of the matrix \( W \) in order to identify the actions that are needed in order to bound its coherence. In the second part, we derive the optimal conditions of the coded orthogonal waveforms.

1) **Characterization of the SVD of \( W \):** The compact SVD of \( W \) can be expressed as
\[
W = U\Lambda V^H,
\]

**(17)**

where \( U \in \mathbb{C}^{M_r \times K} \), \( V \in \mathbb{C}^{N \times K} \) such that \( U^H U = I_K \), \( V^H V = I_K \), and \( \Lambda \in \mathbb{R}^{K \times K} \) is a diagonal matrix containing the singular values of \( W \).

Consider the QR decomposition of \( B \), i.e., \( B = Q_r R_r \), with \( Q_r \in \mathbb{C}^{M_r \times K} \), such that \( Q_r^H Q_r = I_K \) and \( R_r \in \mathbb{C}^{K \times K} \) an upper triangular matrix. Similarly, consider the QR decomposition of \( SA \), i.e., \( SA = Q_s R_s \), with \( Q_s \in \mathbb{C}^{N \times K} \), such that \( Q_s^H Q_s = I_K \) and \( R_s \in \mathbb{C}^{K \times K} \) an upper triangular matrix.

The matrix \( R_r D R_s^T \in \mathbb{C}^{K \times K} \) is rank-\( K \) and its SVD can be expressed as \( R_r D R_s^T = Q_1 \Delta Q_2^H \). Here, \( Q_1 \in \mathbb{C}^{K \times K} \) is such that \( Q_1^H = Q_1^H Q_1 = I_K \) (the same holds for \( Q_2 \)) and \( \Delta \in \mathbb{R}^{K \times K} \) is non-zero diagonal, containing the singular values of \( R_r D R_s^T \). Therefore, it holds that
\[
W = Q_r Q_1 \Delta Q_2^H Q_2^T = Q_r Q_1 \Delta (Q_1^* Q_2)^H,
\]

**(18)**

which is a valid SVD of \( W \) since \( (Q_r Q_1)^H Q_r Q_1 = I_K \) and \( (Q_1^* Q_2)^H Q_1^* Q_2 = I_K \). Via the uniqueness of the singular values of a matrix, it holds that \( \Lambda = \Delta \), thus \( U = Q_r Q_1 \) and \( V = Q_1^* Q_2 \).

Let \( Q_r^i \) denote the \( i \)-th row of \( Q_r \). The coherence of the row space of \( W \) is
\[
\mu(U) = \frac{M_r}{K} \sup_{i \in \mathbb{N}_M} \left\| Q_r^i Q_1 \right\|_2^2.
\]
\[ M_r \sup_{i \in \mathbb{N}_N} \left\| Q_r^i \right\|_2^2 \]

(19)

It can be seen from (19) that \( \mu(U) \) is determined by \( Q_r \), which is only related to the receive steering matrix \( B \) and is independent of the transmit waveform \( S \). In the MIMO radar systems under ULA configuration with \( d_r = \frac{\lambda}{2} \), under the assumption that the target angles set \( \{ \theta_k \}_{k \in \mathbb{N}_N^+} \) are distinct with minimal spatial frequency separation \( \xi \), it was shown in [15] that

\[ \mu(U) \leq \frac{\sqrt{M_r}}{\sqrt{M_r} - (K - 1) \sqrt{\beta_{M_r}(\xi_r)}}. \]

(20)

where \( \beta_{M_r}(\xi_r) \) is the Fejér kernel (see (33)).

Let \( Q_s^{*}(i) \) and \( S_s^{*}(i) \), \( i \in \mathbb{N}_N^+ \) denote the \( i \)-th row of \( Q_s^* \) and \( S^* \), respectively. For the coherence of the row space of \( W \) we have

\[ \mu(V) = \frac{N}{K} \sup_{i \in \mathbb{N}_N^+} \left\| Q_s^{*}(i) Q_s \right\|_2^2 \]

\[ = \frac{N}{K} \sup_{i \in \mathbb{N}_N^+} \left\| Q_s^{*}(i) \right\|_2^2 \]

\[ = \frac{N}{K} \sup_{i \in \mathbb{N}_N^+} \left\| S_s^{*}(i) A^* (R_s^*)^{-1} \right\|_2^2 \]

\[ \leq \frac{N}{K} \sup_{i \in \mathbb{N}_N^+} \left\| S_s^{*}(i) A^* \right\|_2^2 \]

(21)

where

\[ \sigma_{\min}^2(R_s^*) = \lambda_{\min}\left((R_s^*)^H R_s^*\right) \]

\[ = \lambda_{\min}(R_s^H R_s) \]

\[ = \lambda_{\min}(R_s^H Q_s^H Q_s R_s) \]

\[ = \lambda_{\min}((SA)^H SA) \]

\[ = \lambda_{\min}(A^H S^H SA) \]

\[ = \lambda_{\min}(A^H A). \]

(22)

Here, we use the symbol \( \lambda_{\min}(\cdot) \) to denote the minimal eigenvalue of a matrix. In addition, we apply the fact that the eigenvalues of a Hermitian matrix are real, and the eigenvalues of \( X^* \) are the complex conjugate of the eigenvalues of \( X \). Thus, if \( X \) is Hermitian, its eigenvalues are equal to eigenvalues of \( X^* \).
In the MIMO radar systems under ULA configuration with $d_t = \frac{\lambda}{2}$, under the assumption that the target angles are distinct with minimal spatial frequency separation $\xi_t$, it was shown in [15] that

$$\lambda_{\text{min}} \left( A^H A \right) \geq M_t - (K - 1) \sqrt{M_t \beta_{M_t}}(\xi_t) \quad (23)$$

where $\beta_{M_t}(\xi_t)$ is the kernel defined in (33). Therefore, regarding the coherence of the column space of $W$, we have

$$\mu(V) \leq N \sup_{i \in N_K^+} \frac{\| S^{(i)} A^* \|^2}{M_t - (K - 1) \sqrt{M_t \beta_{M_t}}(\xi_t)} \quad (24)$$

Next, we focus on finding the minimum of the supremum of $\| S^{(i)} A^* \|^2$ over $i \in N_K^+$, which results in the optimal waveform conditions.

2) Optimal Waveform Conditions: The transmit steering matrix has the Vandermonde form given by equation (9). Then we have

$$A A^H = \begin{bmatrix} K & \delta_{1,2} & \cdots & \delta_{1,M_t} \\ \delta_{2,1} & K & \cdots & \delta_{2,M_t} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{M_t,1} & \delta_{M_t,2} & \cdots & K \end{bmatrix}, \quad (25)$$

where $\delta_{m',m} = \sum_{k=1}^K e^{j2\pi[(m'-1)-(m-1)]\alpha_k^t} = \sum_{k=1}^K e^{j2\pi(m'-m)\alpha_k^t}$. Consequently, it holds that

$$\| S^{(i)} A^* \|^2 = \left( S^{(i)} A^* \right)^H S^{(i)} A^*$$

$$= S^{(i)} A^* A^T \left( S^{(i)} \right)^H$$

$$= S^{(i)} A A^H \left( S^{(i)} \right)^H$$

$$= \sum_{k=1}^K \sum_{m=1}^{M_t} \sum_{m'=1}^{M_t} s_m^i(i)e^{-j2\pi m\alpha_k^t} s_{m'}^*\left(i\right) e^{j2\pi m'\alpha_k^t}$$

$$= \sum_{k=1}^K \sum_{m=1}^{M_t} s_m^i(i)e^{-j2\pi m\alpha_k^t}^2$$

$$= \sum_{k=1}^K |S_i^t(\alpha_k^t)|^2, \quad (26)$$

where $S_i^t(\alpha_k^t)$ was defined in (13).
Therefore, the coherence bound of the row space of $W$ is

$$
\mu(V) \leq \frac{N}{K M_t - (K - 1) \sqrt{M_t/\beta M_t(\xi_t)}}.
$$

(27)

The lowest possible coherence bound of $\mu(V)$ can be achieved by finding waveforms that minimize $\sup_{i \in \mathbb{N}_N^+, k = 1} |S_i(\alpha_k^t)|^2$. This can be formulated as a min-max optimization problem subject to the constraint given in Lemma 1, i.e.,

$$
\min_S \left( \max_{i \in \mathbb{N}_N^+} \sum_{k=1}^{K} |S_i(\alpha_k^t)|^2 \right)
\text{s.t. } \sum_{i=1}^{N} \sum_{k=1}^{K} |S_i(\alpha_k^t)|^2 = KM_t
$$

(28)

Since $\sum_{k=1}^{K} |S_i(\alpha_k^t)|^2 \geq 0$, for $i \in \mathbb{N}_N^+$, the optimal solution of the min-max optimization problem is

$$
\sum_{k=1}^{K} |S_i(\alpha_k^t)|^2 = \frac{KM_t}{N}, \text{ for all } i \in \mathbb{N}_N^+.
$$

(29)

The solution, as shown in (29) depends on the specific target spatial angles $\{\alpha_k^t\}_{k \in \mathbb{N}_K^+}$. Since these angles are not known, we need to consider every possible angle $\theta_l$ in the angle space $[-\pi/2, \pi/2]$, or every $\alpha_l^t \in [-\frac{1}{2}, \frac{1}{2}]$. Thus, the optimal waveforms should sufficiently satisfy:

$$
|S_i(\alpha_l^t)|^2 = \frac{M_t}{N}, \text{ for all } i \in \mathbb{N}_N^+ \text{ and } \forall \alpha_l^t \in \left[-\frac{1}{2}, \frac{1}{2}\right].
$$

(30)

The condition of (30) indicates that the power spectrum of each snapshot should be flat in the spatial frequency range $\alpha_l^t \in [-\frac{1}{2}, \frac{1}{2}]$, and thus each waveform snapshot must be white noise type sequence with variance $M_t/N$. This completes the proof of Theorem 2.

Theorem 3. (Coherence of $W$): Consider the MIMO radar system as presented in Section II-B and $K$ distinct targets. Let the minimum spatial frequency separation of the targets be

$$
\lambda = \min_{(i,j) \in \mathbb{N}_N^+ \times \mathbb{N}_N^+} \frac{d_h}{\lambda} \left( \sin \theta_i - \sin \theta_j \right), h \in \{t, r\}.
$$

(31)

and assume that

$$
|x| \geq \xi_h \neq 0, h \in \{t, r\}.
$$

(32)
Let us also define

$$\beta_{M_h}(x) = \frac{\sin^2(\pi M_h x)}{M_h \sin^2(\pi x)}, \quad (33)$$

For $d_t = d_r = \frac{1}{2}$ and under the optimal waveform conditions stated in Theorem 2, as long as

$$K \leq \min_{h \in \{t, r\}} \left\{ \sqrt{\frac{M_t}{\beta_{M_h}(\xi_h)}} \right\}, \quad (34)$$

the matrix $W$ obeys the conditions (A0) and (A1) with

$$\mu_0 \Delta = \max_{h \in \{t, r\}} \left\{ \frac{\sqrt{M_h}}{\sqrt{M_h} - (K - 1) \sqrt{\beta_{M_h}(\xi_h)}} \right\}, \quad (35)$$

$$\mu_1 \Delta = \max_{h \in \{t, r\}} \left\{ \frac{\sqrt{M_h K}}{\sqrt{M_h} - (K - 1) \sqrt{\beta_{M_h}(\xi_h)}} \right\} \quad (36)$$

with probability 1.

**Proof of Theorem 3.** Following Theorem 2, for waveforms that satisfy the necessary condition (29), it holds that

$$\mu(V) \leq \inf_{S} \left( \frac{N K}{K} \sup_{i \in \mathbb{N}_K} \frac{\sum_{k=1}^{K} \left| S_i \left( \alpha_k \right) \right|^2}{M_t - (K - 1) \sqrt{M_t \beta_{M_t}(\xi_t)}} \right)$$

$$= \frac{N}{K} \frac{KM_t}{M_t - (K - 1) \sqrt{M_t \beta_{M_t}(\xi_t)}}$$

$$= \frac{\sqrt{M_t}}{\sqrt{M_t} - (K - 1) \sqrt{\beta_{M_t}(\xi_t)}}. \quad (37)$$

Consequently, under the optimal waveforms conditions, and via inequality (20), we have

$$\mu_0 \Delta = \max (\mu(U), \mu(V))$$

$$= \max_{h \in \{t, r\}} \left\{ \frac{\sqrt{M_h}}{\sqrt{M_h} - (K - 1) \sqrt{\beta_{M_h}(\xi_h)}} \right\}. \quad (38)$$

It was shown in [12] that in the general case, $\mu_1 = \mu_0 \sqrt{K}$ always holds true. Then, one can choose

$$\mu_1 \Delta = \mu_0 \sqrt{K} = \max_{h \in \{t, r\}} \left\{ \frac{\sqrt{M_h K}}{\sqrt{M_h} - (K - 1) \sqrt{\beta_{M_h}(\xi_h)}} \right\}. \quad (39)$$

Consequently, the conditions (A0) and (A1) hold.
B. Remarks

1) Asymptotic Optimal Coherence of $\mathbf{W}$

It should be noted that kernel $\beta_{M_h}(x), h \in \{t, r\}$ is a periodic function of $x$. In Fig. 1, we plot the value of $\beta_{M_h}(x)$ for $x \geq 0$ and $M_h = 10$.

For $d_t = d_r = \frac{\lambda}{2}$, the spatial frequency separation corresponding to both transmit and receive arrays, defined in (31), satisfy $|x| \in \left(0, \frac{1}{2}\right]$. If $\xi \Delta \equiv \max \{\xi_t, \xi_r\} \neq 0$, we can find a small constant $\xi$ and $0 < \xi < \frac{1}{\min\{M_t, M_r\}}$ such that the Dirichlet kernel $\frac{\sin(\pi M_h \xi)}{\sin(\pi \xi)} = O(1)$ and the kernel $\beta_{M_h}(\xi)$ satisfies $\sqrt{\beta_{M_h}(\xi)} = \frac{\sin(\pi M_h \xi)}{\sqrt{2\pi M_h} \sin(\pi \xi)} = O\left(\frac{1}{\sqrt{M_h}}\right)$. Consequently, the values of $\beta_{M_h}(\xi)$ decrease as $M_h, h \in \{t, r\}$ increase. Then for any fixed $K$, if $\sqrt{M_h} \geq K \sqrt{\beta_{M_h}(\xi)}, h \in \{t, r\}$, or equivalently

$$M_h \geq K \frac{\sin(\pi M_h \xi)}{\sin(\pi \xi)} = O(K),$$

both (20) and (37) hold. Consequently, under the optimal waveform conditions, it holds that

$$\lim_{M_t \to \infty} \mu(V) \leq \lim_{M_t \to \infty} \frac{\sqrt{M_t}}{\sqrt{M_t} - (K - 1) \sqrt{\beta_{M_h}(\xi)}} = 1.$$  \hspace{1cm} (41)

Since $\mu(V) \geq 1$, via definition 1, it must hold that under the optimal waveform conditions $\mu(V) = 1$ in the limit w.r.t. $M_t$. Similarly, it must hold that $\mu(U) = 1$ in the limit w.r.t. $M_r$. As a result, the coherence of $\mathbf{W}$ is asymptotically optimal.

It should be noted that the spatial frequency separation requirements defined in (32) is not restrictive. For example, in a ULA with $M$ antennas, the spatial frequency separation of targets should be larger than the resolution of the array, i.e., $\frac{1}{M}$. As it can be seen in the proof of Theorem 3, Theorem 3 holds even when the spatial frequency separation of the targets is less than the resolution of the array.

The asymptotic optimality of the coherence of $\mathbf{W}$ holds for a ULA configuration with $d_t = \frac{\lambda}{2}, d_r = \frac{M \lambda}{2}$ too. Such configuration is of interest because it achieves $M_t, M_r$ degrees of freedom [4]. In the following we show the asymptotic optimality of the coherence for this case. Under the optimal waveforms, if the minimum spatial frequency separation corresponding to transmit array satisfies $|x| \geq \xi_t$ for $0 < \xi_t < \frac{1}{M_t}$ such that $\sqrt{\beta_{M_t}(\xi_t)} = O\left(\frac{1}{\sqrt{M_t}}\right)$, it still holds that

$$\lim_{M_t \to \infty} \mu(V) \leq \lim_{M_t \to \infty} \frac{\sqrt{M_t}}{\sqrt{M_t} - (K - 1) \sqrt{\beta_{M_t}(\xi_t)}} = 1.$$  \hspace{1cm} (42)

For $d_r = \frac{M \lambda}{2}$, the spatial frequency separation corresponding to the receive array, defined in (31), satisfies $|x| \in \left(0, \frac{M}{2}\right]$. The period of kernel $\beta_{M_h}(x), h \in \{t, r\}$ is 1 and among one period, e.g., $x \in [k, k + 1]$, its value is symmetric about $x = k + \frac{1}{2}$ for $k = 0, \cdots, \frac{M}{2} - 1$ (see Fig. 1). In addition, the value of kernel $\beta_{M_h}(x), h \in \{t, r\}$ decreases nonmonotonely when $x \in [k, k + \frac{1}{2}]$ and increases nonmonotonely.
when \( x \in \left[ k + \frac{1}{2}, k + 1 \right] \). We can find a minimal separation \( \xi_r \in \left( 0, \frac{1}{M_r} \right) \) such that \( \frac{\sin(\pi M_r \xi_r)}{\sin(\pi \xi_r)} = \mathcal{O}(1) \) and thus \( \sqrt{\beta_{M_r}}(\xi_r) = \mathcal{O}\left( \frac{1}{\sqrt{M_r}} \right) \). If the minimum spatial frequency separation satisfies

\[
|x| \in \bigcup_k \left[ k + \xi_r, k + 1 - \xi_r \right], \text{ for } k = 0, 1, \ldots, \frac{M_t}{2} - 1,
\]

it holds that

\[
\lim_{M_r \to \infty} \mu(U) \leq \lim_{M_r \to \infty} \frac{\sqrt{M_r}}{\sqrt{M_r} - (K - 1) \sqrt{\beta_{M_r}}(\xi_r)} = 1.
\]

Therefore, the asymptotic optimality of coherence of \( \mathbf{W} \) holds.

The minimal spatial frequency separation condition defined in (43) can be viewed in a probabilistic fashion. If \( x \) is uniformly distributed in \( \left[ 0, \frac{M_t}{2} \right] \), the probability that the condition of (43) holds is \( 1 - 2\xi_r > 1 - \frac{2}{M_r} \); the probability tends to 1 as \( M_r \) increases.

2) Coherence and Doppler Shift

It can be easily seen from Theorem 3 that the coherence of \( \mathbf{W} \) does not depend on the Doppler shift \( \{ \nu_k \}_{k \in \mathbb{N}_K} \).

C. Comparative Study of Hadamard and Gaussian Orthogonal Waveforms

This section provides a comparative study of Hadamard and randomly generated Gaussian Orthogonal (G-Orth) waveforms in terms of the resulting coherence of \( \mathbf{W} \) and the corresponding matrix completion performance. Let us consider a ULA transmit array with carrier frequency \( f_c = 1 \times 10^9 \text{Hz} \), \( d_t = \frac{\lambda}{2} \), \( M_t = 40 \) and \( N = 64 \). The corresponding waveform matrix \( \mathbf{S} \in \mathbb{C}^{N \times M_t} \) contains Hadamard or G-Orth waveforms in its columns. Two targets are considered at \( \theta_1 = 0^\circ, \theta_2 = 20^\circ \), or equivalently, \( \alpha_1^t = 0, \alpha_2^t = \frac{1}{2} \sin \left( \frac{\pi}{9} \right) \). The target angle search space is \( [-90^\circ, 90^\circ] \) and the corresponding spatial frequency range is \( \alpha_k^t \in \left[ -\frac{1}{2}, \frac{1}{2} \right] \). The power spectra of the \( i \)-th waveform snapshot, \( i \in \mathbb{N}_N^+ \), corresponding to Hadamard and G-Orth waveforms are shown in Fig. 2.

It can be seen in Fig. 2 (a) that some power spectra of the snapshots corresponding to Hadamard waveforms attain high values at certain angles, while for the G-Orth case all values are small and spread evenly across all snapshots and angles (see Fig. 2 (c)). Thus, according to (27), the coherence bound \( \mu(V) \) corresponding to Hadamard waveforms should be larger than that corresponding to G-Orth waveforms. The lower coherence bound should result in better MC recovery performance for the G-orth waveforms. Indeed, this can be seen in Fig. 3, where the corresponding matrix recovery error is plotted versus the portion of entries of \( \mathbf{W} \) that were used, i.e., \( \rho \). It can be seen in Fig. 3, that for \( \rho > 0.25 \) the recovery error corresponding to G-orth waveforms drops below the inverse SNR level, denoting very good matrix
recovery performance and even some noise smoothing. In the simulations, the data matrix corresponding to the two targets is recovered via the SVT algorithm [29] for \( M_r = 128, M_t = 40, N = 64 \) and \( \text{SNR} = 25\text{dB} \). The simulation results are averaged over 50 independent runs. In each run, the G-Orth waveforms are randomly generated.

**IV. WAVEFORM DESIGN UNDER SPATIAL POWER SPECTRA CONSTRAINTS**

Theorem 2 states that, among the class of orthogonal waveforms, and for MIMO radars using ULAs, the optimal waveform matrix should have rows that are white-noise type functions, i.e., the waveform snapshots across the transmit antennas should be white. In this section, we propose a scheme to optimally design the transmit waveform matrix for MIMO-MC radars. In particular, the design problem is formulated as optimization on matrix manifolds [30]. Due to the orthogonality constraint on the transmit waveforms, i.e., the columns of matrix \( S \), the matrix manifold is the complex Stiefel manifold, which is non-convex. The solution can be obtained via the modified steepest descent algorithm [27], or the modified Newton algorithm with a nonmonotone line search method [31].

**A. Problem Formulation**

Let us discretize the angle space \([−\pi/2, \pi/2]\) into \( L \) phases \( \{\theta_l\}_{l \in \mathbb{N}_L^+} \), corresponding to the spatial frequencies \( \{\alpha^i_t\}_{i \in \mathbb{N}_N^+} \). Let \( c_{il} = S^* A^* (\theta_i) \) for \( i \in \mathbb{N}_N^+ \). According to the optimal condition (30), it holds that

\[
|c_{il}|^2 = \left| S^i_t \left( \alpha^i_t \right) \right|^2 = \frac{M_t}{N}, \quad i \in \mathbb{N}_N^+, \quad l \in \mathbb{N}_L^+.
\]

Define \( A^* = \left[ A^*(\theta_1), \ldots, A^*(\theta_L) \right] \) and \( F = S^* A^* \). It holds that \( [F \odot F^*]_{il} = |c_{il}|^2 \). Based on (30), let us define the objective function

\[
f(S) = \left\| F \odot F^* - \frac{M_t}{N} \mathbf{1}_N \mathbf{1}_L^T \right\|_F^2.
\]

The waveform design problem is formulated as

\[
\min f(S) \quad \text{s.t.} \quad S^H S = I_{M_t}.
\]

Due to the orthogonal constraint, \( S \) belongs to the complex Stiefel manifold \( S(N, M_t) \), defined as

\[
S(N, M_t) = \left\{ S \in \mathbb{C}^{N \times M_t} : S^H S = I_{M_t} \right\}.
\]
The nonconvexity of the orthogonal constraint on the complex Stiefel manifold makes the waveform design problem challenging. In the following we adopt the modified steepest descent algorithm [27], or the modified Newton algorithm on the Stiefel manifold to solve the problem of (47).

B. Derivative and Hessian of Cost Function $f(S)$

In this subsection, we will address the derivative and Hessian of the cost function $f(S)$ defined in (46) w.r.t. the variables $S$. First, based on the second order Taylor series approximation (see [27]), the cost function $f(S)$:

$$f(S + \delta Z) = f(S) + \delta \Re \{ \text{tr} \left( Z^H D_S \right) \} + \frac{\delta^2}{2} \text{vec}(Z)^T H_S \text{vec}(Z) + \frac{\delta^2}{2} \Re \{ \text{vec}(Z)^T C_S \text{vec}(Z) \} + O(\delta^3),$$

(49)

where $D_S \in \mathbb{C}^{N \times M_t}$ is the derivative of $f$ evaluated at $S$, and the matrix pair $H_S, C_S \in \mathbb{C}^{NM_t \times NM_t}$ are the Hessian of $f$ evaluated at $S$. To ensure uniqueness, we require $H_S = H_S^H, C_S = C_S^T$.

The complex-valued derivative $D_S$ is used in the modified steepest descent method. To calculate the Newton direction with the standard Newton method [32], we will use the second order Taylor series expansion of the function $f$:

$$f(s + \delta z) = f(s) + \delta z^T d + \frac{\delta^2}{2} z^T H z + O(\delta^3),$$

(50)

where the vector $s \in \mathbb{R}^{2NM_t}$ is defined as

$$s \triangleq \begin{pmatrix} s_{re} \\ s_{im} \end{pmatrix} \triangleq \begin{pmatrix} \Re \{ \text{vec}(S) \} \\ \Im \{ \text{vec}(S) \} \end{pmatrix}.$$  

(51)

In the above, $d \in \mathbb{R}^{2NM_t}$ is the derivative of $f(s)$ evaluated at $s$, and $H \in \mathbb{R}^{2NM_t \times 2NM_t}$ is the Hessian of $f(s)$ evaluated at $s$ (for definitions, see Section IV-D). The derivatives and Hessians developed in the above two Taylor series expansion forms of (49) and (50) can be transformed into each other [33].

In the following, we list the derivative and Hessian of the cost function $f(S)$. The derivation details are given in Appendix B.

$$D_S = 2 \left\{ \left[ (S^* A^*) \odot (SA) - N \right] \odot Y^T \right\} A^H,$$

(52)

$$H_S = 4P_{M_t \times N} \left( I_N \otimes A^* \right) \text{diag} \left( \text{vec} \left( 2H^T - N^T \right) \right) \left( I_N \otimes A^T \right) P_{N \times M_t},$$

(53)

$$C_S = 4P_{M_t \times N} \left( I_N \otimes A \right) \text{diag} \left( \text{vec} \left( Y^* \odot Y^* \right) \right) \left( I_N \otimes A^T \right) P_{N \times M_t},$$

(54)
where \( N = \frac{M_t}{N} 1_{N} 1_{N}^{T}, \ \tilde{H} = (S^{*} A^{*}) \odot (SA) \in \mathbb{R}^{N \times L} \) and \( Y = A^{T} S^{T} \in \mathbb{C}^{L \times N} \). In the above, \( P_{N \times M_t} \) is a commutation matrix, such that
\[
\text{vec} \left( Z^{T} \right) = P_{N \times M_t} \text{vec} (Z),
\]
which can be expressed as [34]
\[
P_{N \times M_t} = \sum_{m=1}^{N} \sum_{n=1}^{M_t} \left( E_{mn} \otimes E_{mn}^{T} \right),
\]
where \( E_{mn} \) is a matrix of dimension \( N \times M_t \) with 1 at its \( mn \)-th position and zeros elsewhere. It holds that \( P_{N \times M_t} = P_{M_t \times N}^{T} \). It is easy to verify that \( H_{S} = H_{S}^{H}, C_{S} = C_{S}^{T} \).

**C. Modified Steepest Descent on the Complex Stiefel Manifold**

Here, we apply the modified steepest descent method of [27] to solve the optimization problem of (47). Let \( T_{S} (N, M_t) \) denote the tangent space, i.e., the plane that is tangent to the complex Stiefel manifold at point \( S \in S (N, M_t) \) [26]. The inner product in the tangent space is defined using the canonical metric [26] in the complex-value case, i.e.,
\[
\langle Z_{1}, Z_{2} \rangle = \Re \left\{ \text{tr} \left[ Z_{2}^{H} \left( I - \frac{1}{2} SS^{H} \right) Z_{1} \right] \right\},
\]
for \( Z_{1}, Z_{2} \in T_{S} (N, M_t) \).

Let \( Z^{k} \in T_{S} (N, M_t) \) be the steepest descent at point \( S^{k} \in S (N, M_t) \) in the \( k \)-th iteration. The steepest descent algorithm starts from point \( S^{k} \) and moves along \( Z^{k} \) with a step size \( \delta \), i.e.,
\[
S^{k+1} = S^{k} + \delta Z^{k}.
\]

To preserve the orthogonality during the update steps, the new point \( S^{k+1} \) is projected back to the complex Stiefel manifold, i.e., \( S^{k+1} = \Pi \left( S^{k} + \delta Z^{k} \right) \), where \( \Pi \) is the projection operator. For a matrix \( S \in \mathbb{C}^{N \times M_t} \) with \( N \geq M_t \) and with SVD \( S = \tilde{U} \Sigma \tilde{V}^{H} \), the point in the Stiefel manifold that is nearest to \( S \) in the Frobenius norm sense is given by \( \Pi (S) = \tilde{U} \tilde{I}_{N,M_t} \tilde{V}^{H} \) [27].

The modified steepest descent is defined as follows [27]. Let \( g \left( Z^{k} \right) = f \left( \Pi \left( S^{k} + Z^{k} \right) \right) \) be the local cost function for \( S^{k} \in S (N, M_t) \). The gradient of \( g \left( Z^{k} \right) \) at \( Z^{k} = 0 \) under the canonical inner product (57) is
\[
\tilde{\nabla}_{S} f \left( S \right) = \nabla_{S} f \left( S \right) - S^{k} \left( \nabla_{S} f \left( S \right) \right)^{H} S^{k},
\]
where \( \nabla_{S} f \left( S \right) = D_{S} \) denotes the derivative of \( f \left( S \right) \) (see (52)). Then, the modified steepest descent is \( Z^{k} = -\tilde{\nabla}_{S} f \left( S \right) \).
The step size $\delta$ is chosen using a nonmonotone line search method based on [31], i.e., so that
\[
 f \left( \Pi \left( S^k + \delta Z^k \right) \right) \leq C_k + \beta \delta \left( \nabla_{S^k} f \left( S^k \right) , Z^k \right), \tag{60}
\]
\[
 \left( \nabla_{S^k} f \left( \Pi \left( S^k + \delta Z^k \right) \right) , Z^k \right) \geq \sigma \left( \nabla_{S^k} f \left( S^k \right) , Z^k \right). \tag{61}
\]

Here, $C_k$ is taken to be a convex combination of the function values $f \left( S^0 \right), f \left( S^1 \right), \ldots, f \left( S^k \right)$, i.e.,
\[
 C_{k+1} = \frac{\eta Q_k C_k + f \left( S^{k+1} \right)}{Q_{k+1}}, \tag{62}
\]
where $Q_{k+1} = \eta Q_k + 1$, $C_0 = f \left( S^0 \right)$ and $Q_0 = 1$. In the above, the parameter $\eta$ controls the degree of nonmonotonicity. When $\eta = 0$, the line search is the usual monotone Wolfe or Armijo line search [35]. When $\eta = 1$, then
\[
 C_k = \frac{1}{k+1} \sum_{i=0}^{k} f \left( S^{i+1} \right). \tag{63}
\]

The modified steepest descent algorithm is summarized in Algorithm 1.

**Algorithm 1 Modified steepest descent algorithm**

1. **Initialize**: Choose $S^0 \in S\left( N, M_t \right)$ and parameters $\alpha, \eta, \epsilon \in (0, 1)$, $0 < \beta < \sigma < 1$. Set $\delta = 1$, $C_0 = f \left( S^0 \right)$, $Q_0 = 1$, $k = 0$.

2. **Descent direction update**: Compute the descent direction as $Z^k = -\nabla_{S^k} f \left( S^k \right)$ via equation (59).

3. **Convergence test**: If $\left\langle Z^k, Z^k \right\rangle \leq \epsilon$, then stop.

4. **Line search update**: Compute $S^{k+1} = \Pi \left( S^k + \delta Z^k \right)$ and $\nabla_{S^k} f \left( S^{k+1} \right)$. If $f \left( S^{k+1} \right) \geq \beta \delta \left( \nabla_{S^k} f \left( S^k \right) , Z^k \right) + C_k$ and $\left( \nabla_{S^k} f \left( S^{k+1} \right) , Z^k \right) \leq \sigma \left( \nabla_{S^k} f \left( S^k \right) , Z^k \right)$, then set $\delta = \alpha \delta$ and repeat Step 4.

5. **Cost update**: $Q_{k+1} = \eta Q_k + 1$, $C_{k+1} = \frac{\eta Q_k C_k + f \left( S^{k+1} \right)}{Q_{k+1}}$.

6. Perform update $S^{k+1} = \Pi \left( S^k + \delta Z^k \right)$, $k = k + 1$. Go to Step 2.

**D. Modified Newton Algorithm on the Complex Stiefel Manifold**

With expressions for the derivative and Hessian of the cost function given in (52), (53) and (54), we can now formulate the Newton method [32] to solve the waveform design problem of (47).

First, the Newton search direction is calculated as follows. Let $Z^k \in \mathbb{C}^{N \times M_t}$ denote the Newton search direction in the $k$-th iteration. In a similar way as in [36], we arrange the complex-valued elements of

---

January 29, 2015 DRAFT
\( Z^k \) into a real-valued vector \( z^k \) of length \( 2NM \), defined as
\[
z^k \triangleq \begin{pmatrix} z^k_{re} \\ z^k_{im} \end{pmatrix} \triangleq \begin{pmatrix} \Re \{ \text{vec} (Z^k) \} \\ \Im \{ \text{vec} (Z^k) \} \end{pmatrix}.
\] (64)

Let us also define the real-valued vector
\[
d^k \triangleq \begin{pmatrix} \Re \{ \text{vec} (D^k) \} \\ \Im \{ \text{vec} (D^k) \} \end{pmatrix},
\] (65)

and the real-valued matrix
\[
H^k \triangleq \begin{bmatrix} \Re \{ H^k + C^k \} & -\Im \{ H^k + C^k \} \\ \Im \{ H^k - C^k \} & \Re \{ H^k - C^k \} \end{bmatrix}.
\] (66)

Following the standard Newton method, the vector \( z^k \) is computed as
\[
z^k = -\left[ H^k + \sigma_k I \right]^{-1} d^k,
\] (67)

where \( \sigma_k \geq 0 \) is chosen to make the matrix \( H^k + \sigma_k I \) positive definite. Consequently, the complex-valued Newton search direction can be found as \( Z^k = \text{mat}_{N \times M} \left( z^k_{re} + jz^k_{im} \right) \), corresponding to the inverse vector operation defined in (64).

In the \( k \)-th iteration, the standard Newton method performs the update
\[
S^{k+1} = S^k + \delta Z^k,
\] (68)

where \( \delta \) is the step size. Since the waveforms are on the complex Stiefel manifold, to preserve the orthogonality in the modified Newton method, the new point \( S^{k+1} \) is projected back to the complex Stiefel manifold, i.e.,
\[
S^{k+1} = \Pi \left( S^k + \delta Z^k \right).
\] (69)

It should be pointed out that the Hessian matrix defined in (66) is not always positive definite. In each step we choose \( \sigma_k \) such that \( H^k + \sigma_k I \) is positive definite. If matrix \( H_k \) has nonpositive eigenvalues, \( \sigma_k \) should be larger than \( -\lambda_{\text{min}} (H_k) \), where \( \lambda_{\text{min}} \) is the minimal eigenvalue of \( H_k \). In the local area of the minimum, the modified Newton update will approach the pure Newton step. However, if \( \sigma_k \) is chosen very large, the modified Newton search direction will be close to the negative steepest descent. Last, the step size \( \delta \) could be obtained using a similar nonmonotone line search method [31]. Throughout the modified Newton method, the inner product of two matrices \( Z_1, Z_2 \in \mathbb{C}^{N \times M} \) is defined as \( \langle Z_1, Z_2 \rangle = \text{tr} \left( Z_1^H Z_2 \right) \).

The modified Newton algorithm is summarized in Algorithm 2.
Algorithm 2 Modified Newton algorithm

1: **Initialize:** Choose \(S^0 \in \mathcal{S}(N, M_t)\) and parameters \(\alpha, \eta, \epsilon \in (0, 1), 0 < \beta < \sigma < 1\). Set \(\delta = 1, C_0 = f\left(S^0\right), Q_0 = 1, k = 0.\)

2: Compute the derivative \(D_{S^k}\) with equation (52) as well as Hessian \(H_{S^k}\) with equations (53) and (54).

3: **Convergence test:** If \(\langle D_{S^k}, D_{S^k} \rangle \leq \epsilon\), then stop.

4: **Newton search direction computation:** Compute the real-valued vector \(\mathbf{d}^k\) with (65), as well as the real-valued matrix \(H^k\) with (66). Compute the vector \(\mathbf{z}^k\) with (67). The Newton search direction is arranged as \(Z^k = \text{mat}_{N \times M_t}(\mathbf{z}_{\text{re}}^k + j\mathbf{z}_{\text{im}}^k)\).

5: **Line search update:** Compute \(S^{k+1} = \Pi(S^k + \delta Z^k)\) and \(D_{S^{k+1}}\). If \(f\left(S^{k+1}\right) \geq \beta \delta \Re\left\{\langle D_{S^{k+1}}, Z^k \rangle\right\} + C_k\) and \(\Re\left\{\langle D_{S^{k+1}}, Z^k \rangle\right\} \leq \sigma \Re\left\{\langle D_{S^k}, Z^k \rangle\right\}\), then set \(\delta = \alpha \delta\) and repeat Step 5.

6: **Cost update:** \(Q_{k+1} = \eta Q_k + 1, C_{k+1} = \left[\eta Q_k C_k + f\left(S^{k+1}\right)\right] / Q_{k+1}.\)

7: Perform update \(S^{k+1} = \Pi\left(S^k + \delta Z^k\right), k = k + 1.\) Go to Step 2.

V. Numerical Results

In this section, we provide some numerical results to demonstrate the proposed waveform matrix design schemes and evaluate their performance in MIMO-MC radars. Both transmit and receive antennas are configured as ULAs with \(d_t = d_r = \lambda / 2\) and carrier frequency \(f_c = 1 \times 10^9\) Hz.

A. Performance Comparison of Waveform Design Methods

We first design the waveform matrix by applying the modified steepest descent method. We take \(N = 64, M_t = 40\) for DOA space \([-10^\circ : 1^\circ : 10^\circ]\); correspondingly, \(\alpha_t \in [-0.0868, 0.0868]\), the number of discretized angles is \(L = 21\) and the steering matrix \(\mathbf{A}\) has a dimension \(40 \times 21\). In the nonmonotone line search, we chose \(\beta = 0.01\) and \(\sigma = 0.99\). These values are selected by trial and error to satisfy the Wolfe conditions (60) and (61). In addition, we set \(\alpha = 0.5\) to adjust the step size, and \(\epsilon = 10^{-5}\) as the stopping check value. The initial step size is set to \(\delta = 0.1\). The iteration is initialized with a column-wise Hadamard matrix \(S^0 \in \mathcal{S}(64, 40)\), i.e., a matrix that has Hadamard sequences in its columns. The convergence of the proposed modified steepest descent algorithm for \(\eta = 1, 0.5, 0\) is shown in Fig. 4 (a). As it can be seen from Fig. 4 (a), the objective value \(f\) under \(\eta = 0.5\) decreases the fastest, while under \(\eta = 1\) decreases very slowly for number of iterations less than 2000. The simulation results
indicate that the performance of the line search method could be improved if the historical objective values in each iteration are partially utilized, as indicated in (62). The simulation results also show that the value of the objective function, \( f \), approaches its global minimal, i.e., 0. The corresponding optimal solution, \( S \), is not unique, and depends on the initial point and the step size. However, as it will be seen later (see Fig. 7 (a) for an example), all solutions result in very similar MC recovery performance.

Since the complex Stiefel manifold is not a convex set, there is no guarantee that the algorithms will converge to the global minimum. In the problem of (47), the number of equations is \( \frac{M_t(M_t+1)}{2} + NL \), which should be less than the total available combinations \( 2M_tN \). Consequently, to make the objective function zero, it must hold that \( L < 2M_t - \frac{M_t^2+M_t}{2N} \). Our simulations show that for the entire DOA space \([-90^\circ : 1^\circ : 90^\circ]\), corresponding to \( L = 181 \), when the dimension of \( S \) is relatively small, e.g., \( N = 64, M_t = 40 \) and therefore \( L > 2M_t - \frac{M_t^2+M_t}{2N} \approx 67 \), the objective value gets stuck to local minima; however, if the spacing increases, for example to \( 5^\circ \), corresponding to \( L = 37 \), the iteration converges to the global minimum. If the dimension is relatively large, e.g., \( N = 512, M_t = 500 \), even for small spacing, i.e., \( 1^\circ \), the objective value converges to its global minimum (see Fig. 4 (b) for \( \eta = 0 \)).

Next we provide a numerical example to compare the performance of the modified steepest descent algorithm and the modified Newton algorithm. Since the complexity of the Newton method increases with the size of the matrix, we do the comparison for \( N = 32, M_t = 20 \) for DOA space \([-5^\circ : 1^\circ : 5^\circ]\). A column-wise Hadamard waveform matrix \( S^0 \in \mathbb{C}^{32 \times 20} \) is used as initial search point for both algorithms. The performance comparison is illustrated in Fig. 5, where it can be seen that the value of the objective function \( f(S) \) under the modified Newton algorithm decreases much faster than that under the modified steepest descent algorithm.

B. Spatial Power Spectra of Optimized Waveform Snapshots

The power spectra of the optimized waveform snapshots, i.e., the rows of the optimized waveform matrix \( S \), are plotted in Fig. 6 (b) and (d) for the same parameters defined in Section V-A, i.e., \( M_t = 40, N = 64 \) for DOA space \([-10^\circ : 1^\circ : 10^\circ]\), corresponding to \( \alpha_k \in [-0.0868, 0.0868] \). In the simulation, a waveform matrix that is either column-wise G-Orth or Hadamard, is used as the initial search point, respectively. The power spectra of the rows of the initial waveform matrix fluctuate over different DOAs (see Fig. 6 (a) and (c), respectively). It can be seen in Fig. 6 (b) and (d) that the optimized waveform matrix yields almost flat row power spectra with value \( \frac{M_t}{N} = 0.625 \). According to Theorem 2, with the optimized waveforms, the coherence of matrix \( W \) nearly achieves its minimum value.
C. Matrix Recovery Error Performance

Continuing on the scenario of Section V-B, we look at the MC performance corresponding to the optimized waveform matrix as function of the portion of observed entries, $p$. We consider $K = 2$ targets located at $\theta_1 = -2^\circ, \theta_2 = 2^\circ$. The number of receive antennas is $M_r = 64$ and the SNR is set to 25dB. The simulation results are averaged over 50 independent runs, where in each run, the noise is randomly generated.

In the simulations, the data matrix is recovered via the SVT algorithm of [29]. Figure 7 (a) shows the recovery error, suggesting that the optimized waveform matrix results in significantly better performance as compared to the column-wise Hadamard matrix, especially for small values of $p$. One can see that in order to achieve an error around 5%, MC with the optimized waveforms requires about 20% of the data matrix entries, while MC with a column-wise Hadamard matrix requires more than 70% of the data matrix entries. Also, although the optimized waveforms are not unique, all solutions result in almost identical MC performance. In the same figure, we also compare the optimized waveforms against column-wise Gaussian Orthogonal (G-Orth) waveforms. One can see that the former result in lower MC recover error for smaller $p$’s, while their advantage diminishes for higher $p$’s. Further simulations suggest that range of $p$ over which the optimized waveforms have an advantage over the G-Orth waveforms shrinks as $M_t$ increases. As the number of transmit antennas increases, this observation suggests that the G-Orth waveforms behave like optimal in the sense that they achieve a comparable MC performance as the optimized waveforms without involving high computational complexity.

Although the waveform design requires angle space discretization, the sensitivity due to targets falling off grids is rather low. The recovery error comparison between $K = 2$ on-grid targets located at $\theta_1 = -2^\circ, \theta_2 = 2^\circ$ and off-grid at $\theta_1 = -2.5^\circ, \theta_2 = 2.5^\circ$ is shown in Fig. 7 (b). The other simulation settings are the same as in Fig. 7 (a). It can be found that the performance of off-grid targets is almost identical to that of on-grid targets.

D. Coherence Properties Under Optimized Waveforms

In Fig. 8, we plot the coherence $\mu(V)$ of matrix $W$ and its bound, defined in (37), versus the number of transmit antennas, for $K = 4$ targets located at $[-10^\circ, -5^\circ, 5^\circ, 10^\circ]$. The optimized waveforms for different values of $M_t$ are obtained by solving the problem of (47) via Algorithm 1 focusing on DOA space $[-10^\circ : 1^\circ : 10^\circ]$, i.e., $L = 21$. For comparison, the coherence $\mu(V)$ under the G-Orth waveform matrix is also plotted, where the results are averaged over 100 independent implementations, and in each implementation the waveforms are generated randomly. It can be found that the averaged coherence under
G-Orth waveforms is higher than the coherence under optimized waveforms over the entire $M_t$ range. On the other hand, under the optimized waveforms, our simulations show that for different number of targets, the coherence is always bounded by the bound of (37) and approaches its smallest value (not necessarily in a monotone way) when $M_t$ increases. The simulation results in Fig. 8 confirm the conclusions in Theorem 3, i.e., when the waveforms satisfy the optimal waveform conditions stated in Theorem 2, the matrix coherence $\mu(V)$ is asymptotically optimal w.r.t. $M_t$. We should note, however, that the rather big coherence difference between the optimized and the G-Orth waveforms, does not translate in to substantial difference in terms of matrix recovery error. Indeed, the G-Orth waveforms perform very closely with the optimized ones when $M_t$ becomes larger.

VI. CONCLUSIONS

In this paper, we have presented an analysis of the coherence of the data matrix arising in MIMO-MC radar with ULA configurations and transmitting orthogonal waveforms. We have shown that, the data matrix attains its lowest possible coherence if the waveform snapshots across the transmit array have flat power spectra for all time instances. The waveform design problem has been approached as an optimization problem on the complex Stiefel manifold and has been solved via the modified steepest descent algorithm and the modified Newton algorithm. The numerical results have shown that as the number of antennas increases, the optimized waveforms result in optimal data matrix coherence, i.e., 1, and thus, only a small portion of samples are needed for the data matrix recovery. For a particular array, the optimal waveforms depend on the target space to be investigated; for different regions of the target space, the corresponding optimal waveforms can be constructed a priori. Since their construction involves high computational complexity, the optimal waveforms can be used as benchmark against easily constructed waveforms. For example, our simulations revealed that as the number of transmit antennas increases, simply transmitting G-Orth waveforms results in comparable matrix recovery performance as transmitting optimized waveforms. Thus, given the cost of computing the optimized waveforms, certain applications and under certain conditions may treat G-Orth waveforms as optimal.

APPENDIX A

PROOF OF LEMMA 1

Proof. Assume the MIMO radar systems are configured with ULA transmit array of size $M_t$ and inter-element spacing as $d_t$. There are $K$ targets in the far-field at DOAs $\{\theta_k \in [-\frac{\pi}{2}, \frac{\pi}{2}]\}_{k \in \mathbb{N}_K}$, corresponding
to spatial frequencies $\{\alpha_k^t \in [-\frac{1}{2}, \frac{1}{2}]\}_{k \in \mathbb{N}_K}$. Then the transmit steering matrix $A$ has the Vandermonde form defined in (9). As a result, it holds that $\text{tr}(AA^H) = KM_t$.

Suppose that orthogonal waveforms are transmitted so that $S \in \mathbb{C}^{N \times M_t}$. Since $S^H S = \mathbf{I}_{M_t}$, it holds that

$$\sum_{i=1}^{N} s^*_m(i) s_m(i) = \begin{cases} 1, & m = m' \\ 0, & m \neq m' \end{cases}, \quad m, m' \in \mathbb{N}_M^+.$$  \hfill (70)

Consequently,

$$\sum_{i=1}^{N} (S^{(i)})^H S^{(i)} = \mathbf{I}_{M_t},$$ \hfill (71)

where $S^{(i)}$ denotes the $i$-th row of $S$. Following the equation (26), it holds that

$$\sum_{i=1}^{N} \sum_{k=1}^{K} \left| S_i(\alpha_k^t) \right|^2 = \sum_{i=1}^{N} \left\| S^{(i)} A^* \right\|_2^2$$

$$= \sum_{i=1}^{N} \text{tr} \left( A^T (S^{(i)})^H S^{(i)} A^* \right)$$

$$= \sum_{i=1}^{N} \text{tr} \left( (S^{(i)})^H S^{(i)} A^* A^T \right)$$

$$= \sum_{i=1}^{N} \text{tr} \left( (S^{(i)})^H A (S^{(i)}) A^H \right)$$

$$= \text{tr} \left( \left( \sum_{i=1}^{N} (S^{(i)})^H S^{(i)} \right) A A^H \right)$$

$$= \text{tr} (A A^H)$$

$$= KM_t.$$ \hfill (72)

Thus the statements in Lemma 1 follow.

\[ \square \]

**APPENDIX B**

**DERIVATIVE AND HESSIAN OF $f(S)$**

First, we give the following two Lemmas, which will be applied in finding the the derivation.

**Lemma 2.** Let $A \in \mathbb{C}^{M_t \times L}, Z \in \mathbb{C}^{N \times M_t}, Y \in \mathbb{C}^{L \times N}, H \in \mathbb{R}^{N \times L}$ be arbitrary matrices. It can be shown that

$$\text{tr} \left\{ H \left[(A^H Z^H) \odot Y \right] \right\} = \text{tr} \left\{ Z^H \left[ (H \odot Y^T) A^H \right] \right\}.$$ \hfill (73)
Proof. See the Appendix C. □

Lemma 3. Let $\tilde{H} \in \mathbb{C}^{N \times L}$ and $G, M \in \mathbb{C}^{L \times N}$ be general matrices with arbitrary elements. It holds that

$$\text{tr} \left( (G \odot M) \tilde{H} \right) = [\text{vec}(G)]^T \text{vec} \left( M \odot \tilde{H}^T \right).$$

(74)

Proof. See the Appendix D. □

Since $F \odot F^*$ and $N$ are real-valued matrices, the objective function $f(S)$ defined in (46) can be written as

$$f(S) = \text{tr} \left\{ (F \odot F^* - N) (F \odot F^* - N)^T \right\}.$$  

(75)

In order to find the derivative and Hessian of $f(S)$, we do the following expansion:

$$f(S + \delta Z) = \text{tr} \left\{ \left[ ((S + \delta Z)^* A^*) \odot ((S + \delta Z) A) - N \right] \left[ ((S + \delta Z)^* A^*) \odot ((S + \delta Z) A) - N \right]^T \right\}$$

$$= f(S) + \delta \text{tr} \{ T \} + \delta^2 \text{tr} \{ T' \} + O(\delta^3).$$

(76)

Here,

$$T = \left\{ \left[ (S^* A^*) \odot (SA) \right] - N \right\} \left[ \left[ (S^* A^*) \odot (ZA) \right]^T + \left[ (S^* A^*) \odot (ZA) \right]^H \right]$$

$$+ \left\{ \left[ (S^* A^*) \odot (ZA) \right] + \left[ (S^* A^*) \odot (ZA) \right]^* \right\} \left[ \left[ (S^* A^*) \odot (SA) \right]^T - N^T \right] \right\},$$

(77)

$$T' = \left[ (S^* A^*) \odot (SA) \right] \left[ (Z^* A^*) \odot (ZA) \right]^T + \left\{ \left[ (S^* A^*) \odot (SA) \right] \left[ (Z^* A^*) \odot (ZA) \right]^T \right\}$$

$$+ \left\{ \left[ (S^* A^*) \odot (ZA) \right] \left[ (Z^* A^*) \odot (SA) \right]^T \right\}$$

$$+ \left\{ \left[ (S^* A^*) \odot (ZA) \right] \left[ (Z^* A^*) \odot (ZA) \right]^T \right\}$$

$$- \left[ (Z^* A^*) \odot (ZA) \right] N^T - \left\{ \left[ (Z^* A^*) \odot (ZA) \right] N^T \right\}^T.$$  

(78)

Thus, it holds that

$$\text{tr} \{ T \} = \text{tr} \left( 2 \left\{ \left[ (S^* A^*) \odot (SA) \right] - N \right\} \left\{ \left[ (S^* A^*) \odot (ZA) \right]^T + \left[ (S^* A^*) \odot (ZA) \right]^H \right\} \right)$$

$$= \Re \left\{ \text{tr} \left( H \left[ \left( A^H Z^H \right) \odot \left( A^T S^T \right) \right] \right) \right\},$$

(79)

where $H = 2 \left[ (S^* A^*) \odot (SA) - N \right] \in \mathbb{R}^{N \times L}$. Let $Y = A^T s^T$. Following Lemma 2, it holds that

$$\text{tr} \{ T \} = \Re \left\{ \text{tr} \left( Z^H \left[ \left( H \odot Y^T \right) A^H \right] \right) \right\}.$$  

(80)
In addition, it holds that
\[
\text{tr} \left( T' \right) = 2 \text{tr} \left( \left[ (S^* A^*) \odot (SA) \right] \left[ (Z^* A^*) \odot (ZA) \right] \right) + \text{tr} \left( \left[ (SA) \odot (Z^* A^*) \right] [ (S^* A^*) \odot (ZA) ]^T \right) \\
+ 2 \Re \left\{ \text{tr} \left( \left[ (S^* A^*) \odot (ZA) \right] \left[ (S^* A^*) \odot (ZA) \right]^T \right) \right\} - 2 \text{tr} \left( N \left[ (A^H Z^H) \odot (A^T Z^T) \right] \right).
\]
(81)

Now, we focus on the first term on the right side of equation (81). Following Lemma 3, it holds that
\[
\text{tr} \left( \left[ (S^* A^*) \odot (SA) \right] \left[ (Z^* A^*) \odot (ZA) \right] \right) = \text{tr} \left( \left[ (A^H Z^H) \odot (A^T Z^T) \right] \tilde{H} \right) = \left[ \text{vec} \left( A^H Z^H \right) \right]^T \text{vec} \left( A^T Z^T \odot \tilde{H}^T \right),
\]
where \( \tilde{H} = (S^* A^*) \odot (SA) \in \mathbb{R}^{N \times L} \). Further, via equations (55) and (109), it holds that
\[
\left[ \text{vec} \left( A^H Z^H \right) \right]^T = \left[ \left( I_N \otimes A^H \right) \text{vec} \left( Z^H \right) \right]^T = \left\{ \left( I_N \otimes A^H \right) \left[ \text{vec} \left( Z^T \right) \right]^* \right\}^T \\
= \left\{ \left( I_N \otimes A^H \right) \left[ P_{N \times M} \text{vec} \left( Z \right) \right]^* \right\}^T = \left[ \text{vec} \left( Z \right) \right]^H P_{M_t \times N} \left( I_N \otimes A^* \right),
\]
as well as
\[
\text{vec} \left( A^T Z^T \odot \tilde{H}^T \right) = \text{diag} \left( \text{vec} \left( \tilde{H}^T \right) \right) \text{vec} \left( A^T Z^T \right) = \text{diag} \left( \text{vec} \left( \tilde{H}^T \right) \right) \left( I_N \otimes A^T \right) \text{vec} \left( Z^T \right) = \text{diag} \left( P_{N \times L} \text{vec} \left( \tilde{H} \right) \right) \left( I_N \otimes A^T \right) P_{N \times M} \text{vec} \left( Z \right).
\]
Consequently, it holds that
\[
\text{tr} \left( \left[ (A^H Z^H) \odot (A^T Z^T) \right] \tilde{H} \right) = \left[ \text{vec} \left( Z \right) \right]^H P_{M_t \times N} \left( I_N \otimes A^* \right) \text{diag} \left( P_{N \times L} \text{vec} \left( \tilde{H} \right) \right) \left( I_N \otimes A^T \right) P_{N \times M} \text{vec} \left( Z \right).
\]
(85)

Let us focus on the second term on the right side of equation (81). It holds that
\[
\text{tr} \left( \left[ (SA) \odot (Z^* A^*) \right] \left[ (S^* A^*) \odot (ZA) \right]^T \right) = \text{tr} \left( \left[ (A^H S^H) \odot \left( A^T Z^T \right) \right]^H \left[ (A^H S^H) \odot \left( A^T Z^T \right) \right] \right) = \text{tr} \left( \left[ Y^* \odot \left( A^T Z^T \right) \right]^H \left[ Y^* \odot \left( A^T Z^T \right) \right] \right)
\]
\[
= \left[ \text{vec} \left( Y^* \odot \left( A^T Z^T \right) \right) \right]^H \text{vec} \left( Y^* \odot \left( A^T Z^T \right) \right),
\]

as well as
\[
\text{vec} \left( Y^* \odot \left( A^T Z^T \right) \right) = \text{diag} \left( \text{vec} \left( Y^* \right) \right) \text{vec} \left( A^T Z^T \right).
\]

Consequently, it holds that
\[
\text{tr} \left( \left[ Y^* \odot \left( A^T Z^T \right) \right]^H \left[ Y^* \odot \left( A^T Z^T \right) \right] \right)
= [\text{vec}(Z)]^H P_{M \times N} (I_N \otimes A^*) [\text{diag} \left( \text{vec} \left( Y^* \right) \right)]^H \text{diag} \left( \text{vec} \left( Y^* \right) \right) \left( I_N \otimes A^T \right) P_{N \times M_t} \text{vec}(Z).
\]

Finally, let us focus on the forth term on the right side of equation (81). Via equations (102) (83) (84) and Lemma 3, it holds that
\[
\text{tr} \left( N \left[ \left( A^H Z^H \right) \odot \left( A^T Z^T \right) \right] \right) = \text{tr} \left( \left[ \left( A^H Z^H \right) \odot \left( A^T Z^T \right) \right] N \right)
= [\text{vec} \left( A^H Z^H \right)]^T \text{vec} \left( \left( A^T Z^T \right) \odot N^T \right)
= [\text{vec}(Z)]^H P_{M \times N} (I_N \otimes A^*) \text{diag} \left( P_{N \times L} \text{vec}(N) \right) \left( I_N \otimes A^T \right) P_{N \times M_t} \text{vec}(Z).
\]

Therefore, it holds that
\[
f(S + \delta Z) = f(S) + \delta \mathbb{R} \left\{ \text{tr} \left( Z^H \left[ \left( H \odot Y^T \right) A^H \right] \right) \right\}
+ 2\delta^2 [\text{vec}(Z)]^H P_{M \times N} (I_N \otimes A^*) \text{diag} \left( \text{vec} \left( \tilde{H}^T + Y \otimes Y^* - N^T \right) \right) \left( I_N \otimes A^T \right) P_{N \times M_t} \text{vec}(Z).
\]
\[ + 2 \delta^2 \mathbb{R} \left\{ \vec{\text{vec}}(Z)^T P_{M_t \times N} (I_N \otimes A) \text{ diag} \left( \vec{\text{vec}}(Y \circ Y^*) \right) \left( I_N \otimes A^T \right) P_{N \times M_t} \text{vec}(Z) \right\} + \mathcal{O}(\delta^3). \]

By coefficient comparison between (91) and the matrix form of the second-order Taylor series (49), we finally obtain

\[ D_S = \left( H \circ Y^T \right) A^H, \]

\[ H_S = 4 P_{M_t \times N} (I_N \otimes A^*) \text{ diag} \left( \vec{\text{vec}} \left( 2 \tilde{H}^T - N^T \right) \right) \left( I_N \otimes A^T \right) P_{N \times M_t}, \]

\[ C_S = 4 P_{M_t \times N} (I_N \otimes A) \text{ diag} \left( \vec{\text{vec}} \left( Y^* \circ Y^* \right) \right) \left( I_N \otimes A^T \right) P_{N \times M_t}, \]

where \( H = 2 \left[ ( S^* A^* \circ ( S A ) - N \right], Y = A^T S^T \) and the fact \( \tilde{H}^T = Y \circ Y^* = \left( A^T S^T \right) \circ \left( A^H S^H \right) \) is applied. It is easy to verify that \( H_S = H_S^H, C_S = C_S^T \).

**APPENDIX C**

**PROOF OF LEMMA 2**

**Proof.** We use \( a_{ij}, z_{ij}, y_{ij}, h_{ij} \) to denote the \( ij \)-th element of the corresponding matrices \( A \in \mathbb{C}^{M_t \times L}, Z \in \mathbb{C}^{N \times M_t}, Y \in \mathbb{C}^{L \times N}, H \in \mathbb{R}^{N \times L} \), respectively. It holds that

\[ \text{tr} \left\{ H \left[ \left( A^H Z^H \right) \circ Y \right] \right\} = \sum_{m=1}^{M_t} \sum_{n=1}^{L} h_{mn} \left( \sum_{i=1}^{M_t} a_{in}^* z_{mi}^* \right) y_{nm} \]

\[ = \sum_{i=1}^{M_t} \sum_{m=1}^{M_t} z_{mi}^* \left( \sum_{n=1}^{L} h_{mn} y_{nm} a_{in}^* \right). \]  

(92)

Let \( D_S \in \mathbb{C}^{N \times M_t} \) such that \( \text{tr} \left\{ H \left[ \left( A^H Z^H \right) \circ Y \right] \right\} = \text{tr} \left\{ Z^H D_S \right\} \). We use \( d_{ij} \) to denote the \( ij \)-th element of \( D_S \) and it holds that

\[ \text{tr} \left\{ Z^H D_S \right\} = \sum_{i=1}^{M_t} \sum_{m=1}^{N} z_{mi}^* d_{mi}. \]  

(93)

By comparing equations (92) and (93), we have

\[ d_{mi} = \sum_{n=1}^{L} h_{mn} y_{nm} a_{in}^*. \]  

(94)

As a result, the matrix \( D_S \) has the form as

\[ D_S = \left( H \circ Y^T \right) A^H. \]  

(95)

Consequently, it holds that \( \text{tr} \left\{ H \left[ \left( A^H Z^H \right) \circ Y \right] \right\} = \text{tr} \left\{ Z^H \left[ \left( H \circ Y^T \right) A^H \right] \right\} \), which completes the proof.
APPENDIX D

PROOF OF LEMMA 3

Proof. We use $g_{ij}, m_{ij}, \tilde{h}_{ij}$ to denote the $ij$-th element of the corresponding matrices $G, M \in \mathbb{C}^{L \times N}$ and $\tilde{H} \in \mathbb{C}^{N \times L}$. Then, it holds that

$$
\text{tr}\left\{ (G \odot M) \tilde{H} \right\} = \sum_{j=1}^{L} \sum_{i=1}^{N} g_{ji} m_{ji} \tilde{h}_{ij}. \quad (96)
$$

On the other hand, it holds that

$$
\text{vec}\left( M \odot \tilde{H}^T \right) = \begin{bmatrix} m_{11} \tilde{h}_{11}, \cdots, m_{L1} \tilde{h}_{1L}, \cdots, m_{1N} \tilde{h}_{N1}, \cdots, m_{LN} \tilde{h}_{NL} \end{bmatrix}^T, \quad (97)
$$

$$
\text{vec}(G) = \begin{bmatrix} g_{11}, \cdots, g_{L1}, \cdots, g_{1N}, \cdots, g_{LN} \end{bmatrix}^T. \quad (98)
$$

Consequently, it holds that

$$
\text{vec}(G)^T \text{vec}\left( M \odot \tilde{H}^T \right) = \sum_{i=1}^{N} \sum_{j=1}^{L} g_{ji} m_{ji} \tilde{h}_{ij}
= \sum_{j=1}^{L} \sum_{i=1}^{N} g_{ji} m_{ji} \tilde{h}_{ij}. \quad (99)
$$

By comparing equations (96) and (99), we have

$$
\text{tr}\left\{ (G \odot M) \tilde{H} \right\} = \text{vec}(G)^T \text{vec}\left( M \odot \tilde{H}^T \right), \quad (100)
$$

which completes the proof. \qed

APPENDIX E

USEFUL EQUATIONS

Here, we list some useful equations for deriving the derivative and Hessian of a matrix-valued cost function with $Z \in \mathbb{C}^{N \times M_t}$. They are

$$
\|A\|_F^2 = \text{tr}\left( AA^H \right), \quad (101)
$$

$$
\text{tr}(ZH) = \text{tr}(HZ), \quad (102)
$$

$$
\text{tr}\left( A^T \right) = \text{tr}(A), \quad (103)
$$

$$
\text{tr}\left( A^H \right) = \text{tr}(A^*) = (\text{tr}(A))^*, \quad (104)
$$

$$
(H \odot Z)^T = H^T \odot Z^T, \quad (105)
$$

$$
(H \odot Z)^T = H^T \odot Z^T, \quad (106)
$$
\[ \text{tr} (\mathbf{H} \mathbf{Z}) = \left[ \text{vec} \left( \mathbf{H}^T \right) \right]^T \text{vec} (\mathbf{Z}) , \]  
(107)

\[ \text{tr} \left( \mathbf{H}^H \mathbf{Z} \right) = [\text{vec} (\mathbf{H})]^H \text{vec} (\mathbf{Z}) , \]  
(108)

\[ \text{vec} (\mathbf{H} \mathbf{Z} \mathbf{G}) = \left( \mathbf{G}^T \otimes \mathbf{H} \right) \text{vec} (\mathbf{Z}) . \]  
(109)

REFERENCES


Figure 1. The value of kernel $\beta_{M_h}(x)$ as function of $x$ for $x \geq 0$ with $M_h = 10$. 
Figure 2. (a) Snapshot power spectra, $|S_i\left(\alpha'_k\right)|^2$, for $\alpha'_k \in \left[-\frac{1}{2}, \frac{1}{2}\right]$ with $M_t = 40$ and $N = 64$, corresponding to Hadamard waveforms; (b) Magnitude of $|S_i\left(\alpha'_1\right)|^2 + |S_i\left(\alpha'_2\right)|^2$ for Hadamard waveforms; (c) Snapshot power spectra corresponding to G-Orth waveforms; (d) Magnitude of $|S_i\left(\alpha'_1\right)|^2 + |S_i\left(\alpha'_2\right)|^2$ for G-Orth waveforms.
Figure 3. Comparison of MC error corresponding to Hadamard and G-Orth waveforms, as function of the number of data matrix entries used.

Figure 4. $f$ versus iterations: (a) DOA space $[−10^\circ : 1^\circ : 10^\circ]$; (b) DOA space $[−90^\circ : \Delta \theta : 90^\circ]$, $\Delta \theta = 1^\circ, 5^\circ$. 
Figure 5. $f$ versus iteration number under the modified steepest descent algorithm and the modified Newton algorithm. The waveform design parameters are $M_t = 20$, $N = 32$ and DOA space $[-5^\circ : 1^\circ : 5^\circ]$. 
Figure 6. The power spectra $|S_i(\alpha_t^k)|^2$ for $\alpha_t^k \in [-0.0868, 0.0868]$ with $M_t = 40$ and $N = 64$ and DOA space $[-10^\circ : 1^\circ : 10^\circ]$. (a) Hadamard waveforms; (b) Optimized waveforms using Hadamard waveforms as initialization; (c) G-Orth waveforms; (d) Optimized waveforms using G-Orth waveforms as initialization.
Figure 7. MC error as function of $p$ for $M_t = 40$, $N = 64$. (a) Comparison with Hadamard and G-Orth waveforms; (b) Off-grid cases.

Figure 8. Coherence $\mu(V)$ and its bound under optimized as well as G-Orth waveforms for $K = 4$ targets located at $[-10^\circ, -5^\circ, 5^\circ, 10^\circ]$.