Regularized Covariance Estimation for Polarization Radar Detection in Compound Gaussian Sea Clutter

Lei Xie®, Member, IEEE, Zishu He®, Jun Tong®, Tianle Liu®, Jun Li, and Jiangtao Xi®, Senior Member, IEEE

Abstract—This article investigates regularized estimation of Kronecker-structured covariance matrices (CMs) for polarization radar in sea clutter scenarios where the data are assumed to follow the complex elliptically symmetric (CES) distributions with a Kronecker-structured CM. To obtain a well-conditioned estimate of the CM, we add penalty terms of Kullback–Leibler divergence to the negative log-likelihood function of the associated complex angular Gaussian (CAG) distribution. This is shown to be equivalent to regularizing Tyler’s fixed-point equations by shrinkage. A sufficient condition that the solution exists is discussed. An iterative algorithm is applied to solve the resulting fixed-point iterations, and its convergence is proven. In order to solve the critical problem of tuning the shrinkage factors, we then introduce two methods by exploiting oracle approximating shrinkage (OAS) and cross-validation (CV). The proposed estimator, referred to as the robust shrinkage Kronecker estimator (RSKE), is shown to achieve better performance compared with several existing methods when the training samples are limited. Simulations are conducted for validating the RSKE and demonstrating its high performance by using the IPX 1998 real sea data.

Index Terms—Covariance matrix (CM) estimation, cross-validation (CV), Kronecker product structure, polarization detection, sea clutter, shrinkage estimation.

I. INTRODUCTION

TARGET detection in the different scenarios (embracing land, space, atmosphere, and seas) is a fundamental problem in radar [1]–[7]. However, the presence of clutter poses significant challenges, especially in the sea scenario where the heterogeneity of clutter is particularly significant. Therefore, the sea clutter suppression is a recurrent topic for target detection [8]–[11].

Polarization refers to the orientation of the electric and magnetic fields in the plane perpendicular to the direction of wave propagation. Multiple polarization states of a signal can provide more information about a target. The resulting polarization diversity has proven to be a useful tool for radar detection in the presence of clutter, especially when discrimination via Doppler frequency is not possible [4], [12]–[18]. In polarization array radar, the steering vector can be expressed as the Kronecker product of a polarization component and a space–time component.

The covariance matrix (CM) estimation is at the core of the target detection [19]–[24]. The most common CM estimator is the sample covariance matrix (SCM), which is the maximum likelihood estimator (MLE) of the CM for Gaussian data. However, the Gaussian model does not fit the real sea clutter well due to its heavy tail. Instead, the compound-Gaussian (CG) distributions, which is a subclass of the complex elliptically symmetric (CES) distributions [25], have been widely used in modeling the sea clutter returns in radar applications [26]–[29]. The SCM suffers poor performance for data with outliers or heavily tailed distributions due to the lack of robustness. To tackle the heavily tailed data, one class of approaches is to censor the training samples with the aim to exclude outliers from the CM estimation [30]–[36]. Another class of methods is based on robustification. In particular, for CES distributions, various robust CM estimators based on the M-estimator have been developed and characterized [37]–[43]. With such estimators, outlying training samples are usually given small weights when an estimate of the CM is produced.

The SCM also requires an abundant number of samples to achieve satisfactory performance. Many modern applications involve high-dimensional variables whose statistical characteristics remain stationary over a short observation period, where the large sample support assumption does not hold. Regularization provides an effective strategy to improve the CM estimation for addressing the challenge of training shortage. In particular, a class of linear shrinkage algorithms has been introduced [44]–[47], and their integration into robust CM estimators for CES-distributed data has been investigated in the recent works [48]–[52]. These algorithms estimate the CM by shrinking an estimate of the CM $\hat{\Sigma}$ toward a better-conditioned target matrix $T$. There can be various choices for $\hat{\Sigma}$ and $T$. For example, one can choose $\hat{\Sigma}$ as the SCM and Tyler’s estimator [39] for Gaussian and non-Gaussian data, respectively. Moreover, different types of target matrices $T$ can be used, including the identity and diagonal targets. The linear shrinkage estimators can reduce the requirement of samples and provide positive-definite CM estimates. The choice of shrinkage factors is a fundamental problem for
shrinkage estimators. Various criteria and methods have been studied. In particular, Ledoit and Wolf (LW) [45] propose an approach that asymptotically minimizes the mean square error (mse). Then, Chen et al. [47] improve the LW approach using the Rao–Blackwell theorem and design the Rao–Blackwell LW (RBLW) estimator. The oracle approximating shrinkage (OAS) method is proposed in [47]. Both estimators have closed-form expressions and are easily computed. The problem of determining the shrinkage factors can also be cast as a model selection problem, and thus, generic model selection techniques, such as cross-validation (CV) [53], can be applied. The main challenges faced by CV include the choice of the cost function and the heavy computational cost in its direct implementation. Some efforts are made in [54] and [55] to address these challenges for linear shrinkage estimators with unstructured CM.

Due to the independence between space–time domain and polarization domain, the polarization-space–time CM also has the Kronecker structure [18], [56]–[58]. Exploiting this structural knowledge about the CM can also significantly reduce the number of unknown parameters and improve its high estimation accuracy under limited training data [59]–[67]. Particularly, Wiesel et al. [59] propose a robust estimator for Kronecker-structured CM and prove that a globally optimal solution can be found, Sun et al. [60] propose a majorization minimization (MM) solution to the Kronecker MLE (KMLE), and Lu and Zimmerman [68] introduce the maximum likelihood (ML) estimation of Kronecker-structured CM with the presence of Gaussian clutter. An extension of KMLE is also studied for compound Gaussian clutter with inverse Gamma-distributed texture, and the Kronecker normalized SCM (KNSCM) is proposed in [69] to estimate the CM. Although both KMLE and KNSCM provide considerable performance with abundant samples, they still noticeably suffer from performance degradation when the samples are limited.

A. Contributions

In this article, we consider the estimation of Kronecker-structured CM for polarized sea clutter data under low sample supports. In order to improve the performance, in this case, we introduce the Kullback–Leibler divergence penalty to the negative log-likelihood function for the CM estimation. We then derive a robust shrinkage Kronecker estimator (RSKE) that aims to achieve well-conditioned and highly accurate CM estimates. With RSKE, the structural knowledge is exploited together with robustification and regularization techniques. Based on the findings of the previous studies in [25], [48], [50], [54], [71], [72], and others, we investigate the existence of RSKE, its iterative solver and convergence, and also the choice of the shrinkage factors. We then study the performance of the RSKE for the polarization-space–time adaptive processing (PSTAP) in radar applications. The contributions of this article can be summarized as follows.

1) We propose to apply RSKE to polarization radar detection in compound Gaussian sea clutter. We show that the RSKE can be interpreted as the minimizer of a negative log-likelihood function penalized by the Kullback–Leibler divergence. Based on this, the condition for the existence of RSKE is established under some mild assumptions, which provides insights into the relationship between the dimensionality, sample size, and shrinkage factors.

2) We study an iterative solver involving two fixed-point equations to find RSKE and prove its convergence. Following the MM framework, we prove the monotonic decrease in the penalized log-likelihood function over iterations. We show that, with fixed shrinkage factors and arbitrary positive-definite initial estimates, the iterative solver converges.

3) We address the critical challenge of shrinkage factor choice in order to exploit the potential of RSKE. We introduce data-driven methods that automatically tune the linear shrinkage factors, based on OAS and CV. The OAS method adopts a minimum mse (MMSE) criterion and plug-in estimates of the oracle shrinkage factors. For the CV methods, we start with a quadratic loss for leave-one-out CV (LOOCV) and derive analytical solutions of the shrinkage factors, which can approach the performance of the Oracle solutions that minimize the mse of CM estimation. The complexities of these different methods are analyzed. It is found that the analytical CV solutions successfully address the key challenge of the high computational complexity of general applications of CV, and the resulting RSKE has a complexity similar to that of the KMLE.

B. Organization

The remainder of this article is organized as follows. Section II introduces the signal model, the RSKE, and its existence and iterative solution. Section III gives the choices of the shrinkage factors. Section IV presents simulation results to show the performance of CM estimation. Finally, Section V gives the conclusions.

II. ROBUST SHRINKAGE KRONECKER ESTIMATOR

In this section, we introduce the robust shrinkage estimator for Kronecker-structured CMs. We first discuss the motivation, then give the condition for its existence, and, finally, introduce the iterative solver and its convergence property.

A. Signal Model

Consider a pulsed Doppler radar deploying a uniform linear array (ULA) of \( N_t \) antennas, each of which can measure electromagnetic wave in \( N_p \) polarization channels [56], [69], [73]. A burst of \( N_t \) identical pulses at a constant pulse repetition frequency (PRF) of \( f_r \) is transmitted during the coherent processing interval (CPI). The received signals of all the polarization channels at each sensor in the cell under test (CUT) are downconverted to the baseband or to an intermediate frequency in all the pulses at each sensor. They are then processed by the corresponding matched
filters, and sampled and stacked into an $N$-dimensional vector $y \in \mathbb{C}^{N \times 1}$, where $N = N_p N_c N_t$. Let $y_i, i = 1, 2, \ldots, L$, be $L$ independent identically distributed (i.i.d.) signal-free secondary data, arising from adjacent range cells.

Radar detection is a binary hypothesis testing problem, where hypotheses $H_0$ and $H_1$ correspond to target absence and presence, respectively. We first ignore noise in the received signal, which approximates the case of high clutter-to-noise ratio (CNR). The received signal can then be approximately modeled as [69], [74]

$$
H_0: \quad y = c_0
$$

$$
H_1: \quad y = as + c_0
$$

(1)

where $a$ denotes the complex amplitude of the target signal, $s$ denotes the steering vector of target, and $\{c_i\}$ denote the clutter returns in the CUT and adjacent cells. In sea clutter scenarios, experimental trials have shown a good fitting of the compound Gaussian model to the heterogeneous clutter measurements [27], [29]. The received clutter can then be modeled using a positive texture and a Gaussian vector referred to as the speckle, i.e.,

$$
c_0 = \sqrt{\tau_1} u_0 \in \mathbb{C}^{N \times 1}
$$

$$
c_i = \sqrt{\tau_1} u_i \in \mathbb{C}^{N \times 1}
$$

(2)

where $\tau_1$ is the texture and $u_i$ is the speckle component. We assume that $\mathbb{E} \left( \tau_1 \right) = \infty$, $\forall l$, so that the CM of $c_i$ exists, where $\mathbb{E} (\cdot)$ denotes the mathematical expectation. We assume that all the clutter patches are associated with the same terrain, and thus, $u_i$ are zero-mean and i.i.d. with a shared CM $\mathbf{R}$, i.e., $u_i \sim \mathcal{CN}(0, \mathbf{R})$. For conciseness, we here drop the subscript $l$ of $u_i$ while discussing its CM $\mathbf{R}$ in the following.

The clutter signal for a polarimetric radar can be expressed as the sum of $N_c$ clutter patches in the same range cell, i.e.,

$$
\mathbf{c} = \sum_{i=1}^{N_c} \mathbf{c}_i = \sum_{i=1}^{N_c} \mathbf{a}_i \otimes \mathbf{p}_i \in \mathbb{C}^{N \times 1}
$$

(3)

where $\mathbf{a}_i \in \mathbb{C}^{N_t \times 1}$ and $\mathbf{p}_i \in \mathbb{C}^{N_s \times 1}$ denote the space–time steering vector and the polarization scattering vector of the $i$th clutter patch, respectively. Similarly, the polarization–space–time steering vector of the target can be written as $s = \mathbf{a}_t \otimes \mathbf{p}_t$, where $\mathbf{a}_t$ denotes the target space–time steering vector that depends on the direction and velocity of the target.

Following [75], we assume that $N_p = 3$ and $\mathbf{p}_t$ consist of three complex elements: HH, VV, and HV, i.e.,

$$
\mathbf{p}_t = \begin{bmatrix} p_{c, hh}^{(i)} & p_{c, vv}^{(i)} & p_{c, hv}^{(i)} \end{bmatrix}^T
$$

(4)

with $(\cdot)^T$ denoting the transpose. Furthermore, we assume that $\mathbf{p}_t$ follows a complex Gaussian distribution with zero mean and CM $\mathbf{R}_p$ [69], [75]

$$
\mathbb{E} \left( \mathbf{p}_t (\mathbf{p}_t)^H \right) = \mathbf{R}_p = \begin{bmatrix} 1 & \rho_c \sqrt{\gamma_c} & 0 \\ \rho_c \sqrt{\gamma_c} & \gamma_c & 0 \\ 0 & 0 & \delta_c \end{bmatrix}
$$

(5)

with $(\cdot)^H$ denoting the conjugate transpose, $\epsilon_i = \mathbb{E}(\Re (p_{c, hh}^{(i)}^2))$, $\delta_c = \mathbb{E}(\Re (p_{c, hh}^{(i)}))^2)/\mathbb{E}((p_{c, hh}^{(i)})^2)$, $\gamma_c = \mathbb{E}(\Re (p_{c, hv}^{(i)})^2)/\mathbb{E}((p_{c, hv}^{(i)})^2)$, and $\rho_c = \mathbb{E}(\Re (p_{c, hh}^{(i)} p_{c, hv}^{(i)})))/\mathbb{E}((p_{c, hh}^{(i)})^2)$.

The space–time steering vector is expressed as

$$
\mathbf{a}^{(i)} = \mathbf{a}_t \otimes \mathbf{a}_s \in \mathbb{C}^{N_t \times N_c \times N_t}
$$

(6)

where $f_{d,i} = (2v_0/\lambda) \cos(\phi_i)$ denotes the normalized Doppler frequency, $f_{l,i} = (d/\lambda) \cos(\phi_i)$ denotes the normalized spatial frequency, $d$ the interelement spacing, $v_0$ the velocity of the platform, $\lambda$ denotes the radar wavelength, $\phi_i$ denotes the direction of the $i$th clutter patch with respect to the array, $\otimes$ denotes the Kronecker product, and

$$
\mathbf{a}_t (f_{d,i}) = [1, e^{j2\pi f_{d,i}}, \ldots, e^{j2\pi f_{d,i}(N_c-1)^{N_c}}] \in \mathbb{C}^{N_c \times 1}
$$

(7)

$\mathbf{a}_s (f_{l,i}) = [1, e^{j2\pi f_{l,i}}, \ldots, e^{j2\pi f_{l,i}(N_s-1)^{N_s}}] \in \mathbb{C}^{N_s \times 1}$ are the temporal and spatial steering vectors, respectively.

The CM of $\mathbf{u}$ can be given as [69]

$$
\mathbf{R} = \mathbb{E}(\mathbf{uu}^H) = \mathbf{R}_u \otimes \mathbf{R}_p \in \mathbb{C}^{N \times N}
$$

(8)

where the space–time and polarization CMs are, respectively, defined as

$$
\mathbf{R}_u = \sum_{i=1}^{N_c} \mathbf{a}_t (f_{d,i}) \mathbf{a}_t (f_{d,i})^H \in \mathbb{C}^{N_c \times N_c \times N_c}
$$

(9)

and

$$
\mathbf{R}_p = \begin{bmatrix} \rho_c \gamma_c & 0 & 0 \\ 0 & \gamma_c & 0 \\ 0 & 0 & \delta_c \end{bmatrix} \in \mathbb{C}^{3 \times 3}
$$

(10)

### B. Kronecker Maximum Likelihood Estimator

The CES distributions have been widely employed for modeling radar clutter, and many previous experiments have shown that they fit the measured clutter well [25]–[29], [43]. Therefore, following these studies and as will also be demonstrated in Section IV, we assume that the sea clutter $y_i$ follows the CES distribution. The probability density function (pdf) of $y_i$ is of the form

$$
p(y_i) = C_{N_S} \det(\mathbf{R})^{-1} g(y_i^H \mathbf{R}^{-1} y_i)
$$

(11)

where $g(\cdot)$ denotes the density generator and $C_{N_S}$ denotes a normalizing constant. Note that $\mathbf{R}$ is also known as the scatter matrix [25], [43].

The normalized samples $\{x_i = y_i/\|y_i\|_2\}_{i=1}^L$, which belong to a complex unit $N$-dimensional sphere, follow the complex angular Gaussian (CAG) distribution [25], [43]. The joint distribution function of $\{x_i\}_{i=1}^L$ is expressed as [25]

$$
p((x_i)) = \prod_{i=1}^L p(x_i) \propto \det(\mathbf{R})^{-L} \prod_{i=1}^L (x_i^H \mathbf{R}^{-1} x_i)^{-N}
$$

(12)

where $\det(\cdot)$ denotes the determinant. After omitting some additive constants and scaling, the negative log-likelihood
function of such a joint distribution is given by
\[ L_0(\hat{R}_s, \hat{R}_p) = \log \det(\hat{R}_s \otimes \hat{R}_p) + \frac{N}{L} \sum_{l=1}^{L} \log y_l^H (\hat{R}_s \otimes \hat{R}_p)^{-1} y_l \]  
(13)
where \( \hat{R}_s \in S_{N_s}^+ \), \( \hat{R}_p \in S_{N_p}^+ \), and we have used the fact that \( \log(y_l^H \hat{R}_p^{-1} y_l) - \log(x_l^H \hat{R}_s^{-1} x_l) = \log(||y_l||^2) \) is irrelevant to \( R = \hat{R}_s \otimes \hat{R}_p \) in the likelihood function. The above cost function \( L_0(\hat{R}_s, \hat{R}_p) \) is nonconvex in the classical definitions but is jointly geodesic-convex (g-convex) [59] with respect to \( \hat{R}_s \) and \( \hat{R}_p \). Minimizing this cost function produces the KMLE [60], [72]. In the low-sample-support cases, the solution of KMLE can suffer from significant errors and ill-conditioning. For many applications, such as beamforming and spectral estimation [76]–[82], the inverse of the CM estimate is required. Inverting an erroneous, ill-conditioned CM estimate can bring enormous errors. This motivates the design of accurate, well-conditioned CM estimators.

C. Regularization via KL Divergence Penalty

In this section, we introduce a penalized estimator that promotes well-conditioned estimates of the sub-CMs \( R_s \) and \( R_p \). We adopt penalty terms of the Kullback–Leibler divergence for Gaussian distributions [83], i.e.,
\[ D_{KL}(X, Y) = \text{Tr}(XY^{-1}) - \log \det(XY^{-1}) - N \]
where \( X, Y \in S_{N_{++}} \). As shown in [84], the KL divergence \( D_{KL}(X, I_N) \) can effectively constrain the condition number of \( X \). We, thus, add the penalty terms \( a_s D_{KL}(R_s^{-1}, I_{N_s}) \) and \( a_p D_{KL}(R_p^{-1}, I_{N_p}) \) to the negative log-likelihood function in (13) to promote well-conditioned estimates \( \hat{R}_s \) and \( \hat{R}_p \), where \( a_s = ((N_p \rho_s)/(1 - \rho_s)) \) and \( a_p = ((N_s \rho_p)/(1 - \rho_p)) \) with \( \rho_s \in [0, 1) \) and \( \rho_p \in [0, 1) \). Ignoring some additive constants that are irrelevant to \( \hat{R}_s \) and \( \hat{R}_p \), the penalized negative log-likelihood function is obtained as
\[ L(\hat{R}_s, \hat{R}_p) = \frac{N_s}{1 - \rho_s} \log \det(\hat{R}_s) + \frac{N_p}{1 - \rho_p} \log \det(\hat{R}_p) + \frac{N}{L} \sum_{l=1}^{L} \log y_l^H (\hat{R}_s \otimes \hat{R}_p)^{-1} y_l + \frac{N_p \rho_s}{1 - \rho_s} \text{Tr}(\hat{R}_s^{-1}) + \frac{N_s \rho_p}{1 - \rho_p} \text{Tr}(\hat{R}_p^{-1}) \]
(14)
which reduces to \( L_0(\hat{R}_s, \hat{R}_p) \) in (13) when \( \rho_s = \rho_p = 0 \). By adding the penalty terms that are convex, the obtained objective function is also g-convex w.r.t. \( \hat{R}_s \) and \( \hat{R}_p \). This guarantees that all local minimizers of \( L(\hat{R}_s, \hat{R}_p) \) are also globally optimal, following [59, Proposition 1]. Minimizing the penalized log-likelihood function by setting \( \partial L(\hat{R}_s, \hat{R}_p)/\partial \hat{R}_s = 0 \) yields the fixed-point equations
\[ \hat{R}_s = (1 - \rho_s) \frac{N_s}{L} \sum_{l=1}^{L} \frac{y_l^H \hat{R}_p^{-1} y_l}{\hat{R}_s^{-1} \otimes \hat{R}_p^{-1}} + \rho_s I_{N_s} \]  
(15a)
\[ \hat{R}_p = (1 - \rho_p) \frac{N_p}{L} \sum_{l=1}^{L} \frac{y_l \hat{R}_s^{-1} y_l^H}{\hat{R}_s^{-1} \otimes \hat{R}_p^{-1}} + \rho_p I_{N_p} \]  
(15b)
In the above, we have defined
\[ Y_l = \text{unvec}_{N_p,N_s}(y_l) \]
(16)
where \( y_l^{(i)} \) denotes the \( i \)-th entry of \( y_l \) and \( \text{unvec}_{N_p,N_s}() \) reshapes a vector into an \( N_p \times N_s \) matrix as shown above. Therefore, the solution to (15), if exists, can be interpreted as the minimizer of the penalized negative log-likelihood function (14). These fixed-point equations interestingly have the same form as the linear shrinkage estimators for unstructured CM [45], [47]–[51]. Following these works, we refer to the resultant CM estimator as the RSKE, with shrinkage factors \( \rho_s \) and \( \rho_p \). The KMLE [60] can be obtained as a special case of RSKE by letting \( \rho_s = \rho_p = 0 \).

It should be noted that, in [72], estimators that exploit robustification and shrinkage for the unstructured CM and robust estimators for the Kronecker-structured CM have been studied via the geodesic convexity. The KL divergence penalty has also been exploited in [50] for robust estimation of unstructured CM. We here extend these studies to the estimation of Kronecker-structured CM by simultaneously exploiting robustification and shrinkage.

D. Existence of RSKE

In this section, we examine the conditions under which the RSKE exists. When \( \rho_s \) and \( \rho_p \) are small, it is possible that the cost function (14) tends to \(-\infty\) on the boundary of the set \( S_{N_s}^{N_{++}} \) and \( S_{N_p}^{N_{++}} \), i.e., (14) becomes unbounded below, and there is no solution to the fix-point equations of (15). The existence of the shrinkage Tyler’s estimator for unstructured CM has been studied in [50], where the relationship between the shrinkage factors, sample size, and dimensionality is revealed. By establishing the condition under which the cost function tends to \(+\infty\) on the boundary of the set of positive-definite, Hermitian matrix, the minimum shrinkage factor for the existence of the CM estimator is obtained [50]. This result, however, cannot directly determine the conditions of the two shrinkage factors affecting each other. In this work, we follow [50, Th. 3] and its proof to study the RSKE. We first construct auxiliary functions by which the penalized negative log-likelihood function (14) can be lower bounded. The two auxiliary functions have a similar form as [50, eq. (15)]. Thus, using the same treatment of [50], we can examine
the conditions for the auxiliary functions tending to \(+ \infty\) at the boundary. Based on the results, we can obtain the following sufficient condition for the existence of a solution to the RSKE.

**Proposition 1:** The cost function (14) has a finite lower bound over the set of positive-definite \(R_a\) and \(R_p\), i.e., a solution to (15) exists if the following conditions are satisfied.

1. None of \(r_{j,j}\) and \(c_{i,j}\) is an all-zero vector, where \(r_{j,j} \in \mathbb{C}^{N_a \times 1}\) denotes the \(j\)th row of \(Y_i\) and \(c_{i,j} \in \mathbb{C}^{N_p \times 1}\) denotes the \(i\)th column of \(Y_i\).
2. There exist \(\beta_1 \in [0,1], \beta_2 \in [0,1]\) with \(\beta_1 + \beta_2 = 1\) such that, for any proper subspace, \(S_st \subset \mathbb{C}^{N_a \times 1}\) and \(S_p \subset \mathbb{C}^{N_p \times 1}\) in the space of length-\(N_st\) and -\(N_p\) vectors, respectively,

\[
P_{L,N_st}(S_{st}) < \frac{(LN_p + \alpha_{st}L) \dim(S_{st}) - \beta_2 LN_p}{\beta_1 LN} \quad (17a)
\]

\[
P_{L,N_p}(S_p) < \frac{(LN_a + \alpha_pL) \dim(S_p) - \beta_1 LN_a}{\beta_2 LN_a} \quad (17b)
\]

where \(P_{L,N_st}(S_{st}) \triangleq \left(\sum_{i=1}^{N_st} \sum_{j=1}^{L} 1_{r_{j,i} \in S_{st}}\right)/(LN_p), P_{L,N_p}(S_p) \triangleq \left(\sum_{i=1}^{N_p} \sum_{j=1}^{L} 1_{c_{i,j} \in S_p}\right)/(LN_a)\), and \(1_x\) denotes the indicator function.

**Proof:** See Appendix A.

In general, the above conditions require that the number of samples is sufficiently large, and the samples are evenly spread out in the whole space.

**Corollary 1:** If the samples \(s_{st}\) are evenly spread out in the whole space, such that \(P_{L,N_st}(S_{st}) \leq (\dim(S_{st}))/(\min(N_st, LN_p))\) = \((\dim(S_{st}) \max(N_st, LN_p))/LN\) and \(P_{L,N_p}(S_p) \leq (\dim(S_p) \max(N_p, LN_a))/LN\), then Condition (2) in Proposition 1 is equivalent to

\[
\rho_{st} > 1 - \frac{LN_p}{\beta_1 \max(N_st, LN_p) + \beta_2 LN} \quad (18a)
\]

\[
\rho > 1 - \frac{LN_a}{\beta_2 \max(N_p, LN_a) + \beta_1 LN} \quad (18b)
\]

**Proof:** Let \(\dim(S_{st}) \triangleq d_{st}\). Recall that \(\alpha_{st} = (N_p_p \rho_{st})/(1 - \rho_{st})\) and \(\alpha_p = (N_st \rho_p)/(1 - \rho_p)\). The condition (17a) is satisfied when

\[
d_{st} \max(N_st, LN_p) < \frac{LN_p}{\beta_1 d_{st} - \beta_2 LN} \quad (19)
\]

Rearranging (19), one has \(\rho_{st} > 1 - ((LN_p)/\beta_1 \max(N_st, LN_p) + \beta_2 LN/d_{st})\) for arbitrary \(d_{st} = 1, \ldots, N_st - 1\), i.e.,

\[
\rho_{st} > \frac{\max(N_st, LN_p)}{\beta_1 \max(N_st, LN_p) + \beta_2 LN/d_{st}} = 1 - \frac{LN_p}{\beta_1 \max(N_st, LN_p) + \beta_2 LN}
\]

which is exactly (18a). Similarly, we have (18b).

**Remark 1:** Condition (2) in Corollary 1 shows the relationship between the shrinkage factors, the number of samples \(L\), and the dimension of the sub-CMs \(N_st\) and \(N_p\). In general, a larger shrinkage factor \(\rho_{st}\) is required when \(L\) decreases or \(N_st\) increases. Moreover, Condition (2) can be easily checked. For example, when \(\beta_1 = 1\) and \(\beta_2 = 0\), \(\rho_{st} > \max(1 - ((LN_p)/(\max(N_st, LN_p))), 0)\), and \(\rho > \max(1 - (1/N_p), 0)\). When \(N_p = 1\) and \(N = N_st\), the Kronecker-structured CM reduces to an unstructured one. Then, Condition (2) becomes \(\rho_{st} > 1 - ((LN_p)/(\max(N, LN)))\) and \(\rho > 0\). When \(L \geq N\), the condition is \(\rho_{st} \in (0,1)\). When \(L < N\), the condition is \(\rho_{st} \in (1 - (L/N), 1)\), which agrees with the result in [49] and [50] for the case of unstructured CM.

**E. Iterative Solver and Its Convergence**

Similar to [48]–[51], we solve (15) by applying the following process, which involves two fixed-point iterations

\[
\hat{R}_{st}^{(k+1)}(\rho_{st}) = (1 - \rho_{st}) \hat{R}_{st}^{(k)} + \rho_{st} I_{N_st} \\
\hat{R}_{p}^{(k+1)}(\rho_p) = (1 - \rho_p) \hat{R}_{p}^{(k)} + \rho_p I_{N_p} \quad (20a)
\]

where

\[
\hat{C}_{st}^{(k+1)} = \frac{N_st}{L} \sum_{l=1}^{L} Y_l^H \left(\hat{R}_{st}^{(k)}\right)^{-1} Y_l \quad (21a)
\]

\[
\hat{C}_{p}^{(k+1)} = \frac{N_p}{L} \sum_{l=1}^{L} Y_l^H \left(\hat{R}_{p}^{(k)}\otimes \hat{R}_{st}^{(k)}\right)^{-1} Y_l \quad (21b)
\]

and \(\hat{R}_{st}^{(k)}\) and \(\hat{R}_{p}^{(k)}\) denote the estimates of the sub-CMs at the \(k\)th iteration. In this article, we choose the initial CM estimates \(R_{st}^{(0)} = I_{N_st}\) and \(R_{p}^{(0)} = I_{N_p}\) for simplicity.

It is useful to examine the convergence property of the above iterative estimator that generalizes Tyler’s estimator [39] and its shrinkage extension [48], [50], [51] to the case of Kronecker-structured CM. The works [39], [48], [50], [51] assume unstructured CM, and thus, their solutions can be characterized by a single fixed-point equation. The convergence of the iterative process for Tyler’s estimator is proven in [39] by examining the fixed-point iterations. For the shrinkage extension of Tyler’s estimator, the convergence is proven in [48] by applying the concave Perron–Frobenius theory, in [50] by applying the MM theorem, and in [51] by applying the monotone bounded convergence theorem. For the Kronecker-structured CM, though the case of the KMLE has been studied in [59], in this work, we incorporate shrinkage into the estimator, and the convergence has not been analyzed earlier to the best of our knowledge. Exploiting the MM framework [85], we have the following proposition that establishes the converging property of the fixed-point iterations in (20).

**Proposition 2:** The fixed-point iterations in (20) converge to the solution of (15) for arbitrary positive-definite initial matrices \(\hat{R}_{st}^{(0)}\) and \(\hat{R}_{p}^{(0)}\) when the conditions in Proposition 1 are satisfied.

**Proof:** See Appendix B.

**Remark 2:** The iterations in (20) can be terminated by using a distance metric

\[
D(\hat{R}_{st}^{(k+1)}, \hat{R}_{st}^{(k)}) = \left\| \frac{\hat{R}_{st}^{(k+1)}}{\text{Tr}(\hat{R}_{st}^{(k+1)})} - \frac{\hat{R}_{st}^{(k)}}{\text{Tr}(\hat{R}_{st}^{(k)})} \right\|
\]
where \( \hat{R}^{(k)} = \hat{R}_s^{(k)} \otimes \hat{R}_p^{(k)} \) and \( \| \cdot \| \) denotes the Frobenius norm. This metric measures the variation of the solution over iterations. Then, a stopping criterion can be set to terminate the iterations when
\[
D(\hat{R}^{(k+1)}, \hat{R}^{(k)}) < \delta
\]
or \( k > K_{\text{max}} \) is met, where \( \delta \) denotes a preset threshold and \( K_{\text{max}} \) denotes the maximum number of iterations allowed.

**III. CHOICE OF THE SHRINKAGE FACTORS**

The performance of the RSKE depends highly on the choice of the shrinkage factors \( \rho_{st} \) and \( \rho_p \). In practice, however, the optimal shrinkage factors are unavailable since the true CM is unknown. In this section, we propose two different choices, based on OAS and LOOCV, respectively, to provide solutions with different performances and complexities.

**A. KOAS Method**

In [48], an OAS strategy for choosing the shrinkage factor for unstructured CM is derived by exploiting the MMSE criterion and plug-in estimates. We can extend this strategy to the RSKE. The choice of the two shrinkage factors will be decoupled into separate problems to enable a low-complexity estimation with different performances and complexities.

\[
\min_{\rho_{st}} \mathbb{E}\left\{ \| \hat{R}_s - R_{st} \|^2 \right\}
\]

subject to
\[
\hat{R}_s = (1 - \rho_{st})C_{st} + \rho_{st}I_{N_s}
\]

and
\[
\min_{\rho_p} \mathbb{E}\left\{ \| \hat{R}_p - R_p \|^2 \right\}
\]

subject to
\[
\hat{R}_p = (1 - \rho_p)C_p + \rho_pI_{N_p}
\]

where \( \mathbb{E}\{\cdot\} \) denotes the mathematical expectation and
\[
C_{st} \triangleq \frac{N}{LN_p} \sum_{i=1}^{L} \frac{Y_i^H R_p^{-1} Y_i}{y_i^H (R_s \otimes R_p)^{-1} y_i}
\]

\[
C_p \triangleq \frac{N}{LN_s} \sum_{i=1}^{L} \frac{Y_i R_p^{-1} Y_i^H}{y_i^H (R_s \otimes R_p)^{-1} y_i}.
\]

The following proposition extends the OAS solution of [48] to the Kronecker-structured CM.

**Proposition 3:** The shrinkage factors that achieve the MMSE are given as (27a) and (27b), shown at the bottom of the page.

**Proof:** See Appendix C.

In practice, \( R_s \) and \( R_p \) in (27) are unknown. Similar to [48], we propose to replace them by their trace-normalized estimates \( \hat{R}_s \) and \( \hat{R}_p \), such as the KNSCM [69] and KMLE [60]. We will show the performance of the resulting shrinkage factors \( (\rho_{st}, \text{KOAS}, \rho_p, \text{KOAS}) \), referred to as the Kronecker OAS (KOAS) choice, in Section IV. Note that, if \( N_s = 1 \) or \( N_p = 1 \), the Kronecker-structured CM reduces to the unstructured CM and (27) agrees with [48, eq. (17)]. If \( \rho_{st} \text{KOAS} < 0 \) is produced, we then truncate it to \( \rho_{st} \text{KOAS} = 0 \). If \( \rho_p \text{KOAS} \geq 1 \), we simply set the CM estimate to be the shrinkage target matrix. The treatments are similar for \( \rho_p \text{KOAS} < 0 \) and \( \rho_p \text{KOAS} \geq 1 \), and also the LOOCV-based choices of the shrinkage factors to be introduced in Section III-B.

**B. LOOCV Method**

We next provide an alternative for choosing the shrinkage factors based on LOOCV. In order to achieve good performance and complexity tradeoff, the cost for LOOCV must be carefully chosen. In this work, we extend the quadratic cost used in [54] to obtain a data-driven, analytical solution. Note that Tong et al. [54] consider unstructured CM for Gaussian data, whereas this article considers Kronecker-structured CM estimation with elliptically distributed data for which iterative solvers are required.

Let \( \Sigma_{st} \) and \( \Sigma_p \) be two positive-definite, Hermitian matrices. Define the following cost function:
\[
\mathcal{J}_{st}(\Sigma_{st}) = \mathbb{E}\{\| \Sigma_{st} - S_{st} \|^2 \}
\]

\[
\mathcal{J}_{p}(\Sigma_{p}) = \mathbb{E}\{\| \Sigma_{p} - S_{p} \|^2 \}
\]

where the expectation is with respect to \( Y = \text{unvec}_{N_s N_p}(y) \)
\[
S_{st} \triangleq \frac{N_s Y^H R_p^{-1} Y}{Y^H (R_s \otimes R_p)^{-1} Y}, \quad S_{p} \triangleq \frac{N_p Y R_s^{-1} Y^H}{y^H (R_s \otimes R_p)^{-1} y}.
\]

**Proposition 4:** The expectations of \( S_{st} \) and \( S_p \) are, respectively, given as \( \mathbb{E}(S_{st}) = R_s \) and \( \mathbb{E}(S_p) = R_p \), and \( \mathcal{J}_{st}(\Sigma_{st}) \) and \( \mathcal{J}_{p}(\Sigma_{p}) \) are minimized by \( \Sigma_{st} = R_s \) and \( \Sigma_{p} = R_p \), respectively.

**Proof:** See Appendix D.

Inspired by Proposition 4, we aim to estimate the cost function in (28) and then minimize it over the shrinkage factors. This may be achieved using different strategies, e.g., [45]. In this article, we apply the LOOCV strategy [53] to estimate \( \mathcal{J}_{st}(\Sigma_{st}) \) and \( \mathcal{J}_{p}(\Sigma_{p}) \), and minimize them to determine the shrinkage factors. With the standard LOOCV, the samples \( \mathcal{Y}_l \) are repeatedly split into two sets. For the \( l \)th split, the samples in the training set \( \mathcal{Y}_l \) (with the \( l \)th sample \( y_l \) omitted from \( \mathcal{Y}_l \)) are used for producing shrinkage CM estimates \( \{S_{st}, \Sigma_p\} \), and the remaining sample \( y_l \) is used for constructing \( \{S_{st}, S_p\} \) to...
estimate \( J_{st}(\Sigma_a) \) and \( J_p(\Sigma_p) \). The standard LOOCV process requires the iterative estimator to be applied for \( L \) times for each pair of candidate shrinkage factors \((\rho_a, \rho_p)\), which can lead to significant complexity, especially when grid search of \((\rho_a, \rho_p)\) is conducted. In order to address this complexity challenge, we propose an alternative solution by using proxy estimators so that closed-form expressions can be found for the optimized shrinkage factors.

Similar to KOAS, we first assume that the CMs are “known” and consider estimates of the CMs from the samples \( Y_j = \{ Y_j, j \neq l \} \) as

\[
\begin{align*}
\hat{R}^{(l)}_{st}(\rho_a) &= (1 - \rho_a) \hat{C}^{(l)}_{st} + \rho_a I_{N_a} \\
\hat{R}^{(l)}_p(\rho_p) &= (1 - \rho_p) \hat{C}^{(l)}_p + \rho_p I_{N_p}
\end{align*}
\]  

(30a) \hspace{1cm} (30b)

where

\[
\begin{align*}
\hat{C}^{(l)}_{st} &= \frac{N_a}{L - 1} \sum_{j \neq l} \frac{Y_j^H R_p^{-1} Y_j}{y_j^H (R_s \otimes R_p)^{-1} y_j} \\
\hat{C}^{(l)}_p &= \frac{N_p}{L - 1} \sum_{j \neq l} \frac{Y_j R_s^{-1} Y_j^H}{y_j^H (R_s \otimes R_p)^{-1} y_j}
\end{align*}
\]  

(31a) \hspace{1cm} (31b)

Following Tong et al. [54], we adopt the quadratic cost functions as follows:

\[
\begin{align*}
J_{st,\text{CV}}(R_s) &= \frac{1}{L} \sum_{l=1}^{L} \| \hat{R}^{(l)}_{st}(\rho_a) - \hat{S}^{(l)}_{st} \|^2 \\
J_p,\text{CV}(R_p) &= \frac{1}{L} \sum_{l=1}^{L} \| \hat{R}^{(l)}_p(\rho_p) - \hat{S}^{(l)}_p \|^2
\end{align*}
\]  

(32a) \hspace{1cm} (32b)

where

\[
\begin{align*}
\hat{S}^{(l)}_{st} &= \frac{N_a Y_j^H R_p^{-1} Y_l}{y_j^H (R_s \otimes R_p)^{-1} y_j} \\
\hat{S}^{(l)}_p &= \frac{N_p Y_j R_s^{-1} Y_j^H}{y_j^H (R_s \otimes R_p)^{-1} y_j}
\end{align*}
\]  

(33a) \hspace{1cm} (33b)

Substituting (30a) into (32a), the cost function can be rewritten as

\[
J_{st,\text{CV}}(\rho_a) = \frac{1}{L} \sum_{l=1}^{L} \| (1 - \rho_a) \hat{C}^{(l)}_{st} + \rho_a I_{N_a} - \hat{S}^{(l)}_{st} \|^2
\]  

(34)

We treat \( J_{st,\text{CV}}(\rho_a) \) as a proxy of \( J_{st}(\Sigma_a) \) and choose the shrinkage factor \( \rho_a \) as the minimizer of (34) as

\[
\rho_{st,\text{CV}} = \text{Re} \left( \frac{\sum_{l=1}^{L} \text{Tr} \left( \left( I_{N_a} - \hat{C}^{(l)}_{st} \right) \left( \hat{S}^{(l)}_{st} - \hat{C}^{(l)}_{st} \right) \right) \right)}{\sum_{l=1}^{L} \text{Tr} \left( \left( I_{N_a} - \hat{C}^{(l)}_{st} \right)^2 \right)} 
\]  

(35)

Similarly, we choose \( \rho_p \) as

\[
\rho_{p,\text{CV}} = \frac{\text{Re} \left( \sum_{l=1}^{L} \text{Tr} \left( \left( I_{N_p} - \hat{C}^{(l)}_p \right) \left( \hat{S}^{(l)}_p - \hat{C}^{(l)}_p \right) \right) \right)}{\sum_{l=1}^{L} \text{Tr} \left( \left( I_{N_p} - \hat{C}^{(l)}_p \right)^2 \right)}
\]  

(36)

Alternative expressions can be derived for (35) and (36) to reduce the computational costs. Let

\[
\hat{C}_{st} = \frac{N_a}{L} \sum_{l=1}^{L} Y_l^H R_p^{-1} Y_l
\]  

(37)

Recalling (31a) and (33a), we have

\[
\hat{C}^{(l)}_{st} = \frac{L}{L - 1} \hat{C}_{st} - \frac{1}{L - 1} \hat{S}^{(l)}_{st}, \quad \hat{L}_{st} = \hat{C}_{st} = \frac{L}{L - 1} \sum_{l=1}^{L} \hat{C}^{(l)}_{st} = \sum_{l=1}^{L} \hat{S}^{(l)}_{st}.
\]  

(38)

Note that \( \hat{C}_{st}, \hat{C}^{(l)}_{st}, \hat{S}^{(l)}_{st} \), and \( I_{N_a} \) are all Hermitian matrices. By using (38), we have

\[
\begin{align*}
\sum_{l=1}^{L} \text{Tr} \left( \left( \hat{C}^{(l)}_{st} \right)^2 \right) &= \frac{L^2}{L - 1} \text{Tr} \left( \hat{C}^{2}_{st} \right) - \frac{\sum_{l=1}^{L} \text{Tr} \left( \left( \hat{S}^{(l)}_{st} \right)^2 \right)}{L - 1} \\
\sum_{l=1}^{L} \text{Tr} \left( \left( \hat{C}^{(l)}_{st} \right)^2 \right) &= \frac{L^2(1 - L)}{(L - 2)^2} \text{Tr} \left( \hat{C}^{2}_{st} \right) + \frac{\sum_{l=1}^{L} \text{Tr} \left( \left( \hat{S}^{(l)}_{st} \right)^2 \right)}{(L - 2)^2}
\end{align*}
\]  

(39a) \hspace{1cm} (39b)

Substituting (39) into (35), we obtain (41a), as shown at the bottom of the page, to quickly evaluate the shrinkage factors \( \rho_{st,\text{CV}} \). Similarly, we can obtain (41b), as shown at the bottom of the page, there for \( \rho_{p,\text{CV}} \), where

\[
\hat{C}_{p} = \frac{N_p}{L} \sum_{l=1}^{L} Y_l^H R_s^{-1} Y_l^H.
\]  

(40)

The shrinkage factors determined by (41) still require the true CM \( R_s \) and \( R_p \) to be known to compute (33), (37), and (40). Similar to KOAS, we propose to substitute them by their trace-normalized estimates \( \hat{R}_s \) and \( \hat{R}_p \). We refer to the resultant solutions as the CV choice.

Remark 3: The proposed methods exhibit different complexities. If the shrinkage factors are given, the computational complexity of the iterative process in (20) is about \( O(N_0 N_0^3 + N_0 N_0^2 + L(N_0 N_0^2 + N_0^2 N_0)) \), where \( N_0 \) denotes the number of iterations, and we have used the identities \((A \otimes B)^{-1} = A^{-1} \otimes B^{-1}\) and \((B^T \otimes A)\text{vec}(X) = \text{vec}(AXB)\). All the shrinkage factors
proposed are given in closed forms without the need of grid search. Their complexities are summarized in the following, where only the highest order of the complexity is counted.

1) KOAS: The computational complexity of (27) mainly arises from the computation of $\text{Tr}(\mathbf{R}_s^2)$ and $\text{Tr}(\mathbf{R}_p^2)$, which is $O(N_{st}^2 + N_p^3)$, when the plug-in CMs $\mathbf{R}_s$ and $\mathbf{R}_p$ are known.

2) CV: Given $\mathbf{R}_s$ and $\mathbf{R}_p$, (41) can be evaluated at a complexity of $O(N_{st}^3 + N_p^3 + L(N_{st}^2N_p + N_{st}N_p^2))$.

It can be seen that, ignoring the cost for finding the plug-in CMs, the complexity of finding the shrinkage factors is dominated by that of iteratively updating the CMs in (20).

IV. SIMULATION RESULTS

In this section, we show the performance of the proposed RSKE estimators. We compare the proposed estimators with the following CM estimators: KMLE [60], [69] and KNSCM [69]. We will then demonstrate the superiority of our proposed methods over these existing methods with the true data and generated simulation data.

A. Target Detection

In this section, we show simulation results to demonstrate the performance of the RSKE for the polarization target detection in the context of real heterogeneous sea clutter data. The Ice Multiparameter Imaging X-Band (IPIX) 1998 is collected using the McMaster IPIX radar with one single antenna from Grimsby, Canada [86]. One dataset that we use is IPIX 1998 file “19980223_171533.” In Fig. 1, we show the normalized logarithmic amplitude of the clutter in this file. Key parameters of the dataset include the carrier frequency of 9.39GHz, PRF of 1000 Hz, a pulseslength of 20 ns, and a range resolution of 3 m. We refer the reader to the official website [86] for more details. From Fig. 1, we can see that there are many strong scattering points whose echo amplitude is significantly large. This indicates that the data fit the compound Gaussian distribution better due to its heavy tail in contrast to the Gaussian one.

In order to illustrate this, we use the compound Gaussian distribution to fit the pdf of the amplitude of the sea clutter in files “19980223_171533” and “19980226_215015” under different polarizations. Note that the data correspond to different polarizations. The fitting errors for the VV data of the real sea clutter data have a heavier tail than the Gaussian model. The fitting errors for the VV data of “19980223_171533” are given in Table I, which demonstrates that the fitting error of the Gaussian distribution is larger than that of the CG distributions. This shows the suitability of the CG model for fitting the real sea clutter. Note that, under different sea states, the different types of CG distribution may provide different accuracies for fitting the clutter data. However, the CG model always fits the data better than the Gaussian one. Meanwhile, the proposed RSKE is effective for various CG data, regardless of the specific type.

To assess the detection performance, we consider the well-known normalized matched filter (NMF) detector [51], i.e.,

$$\Lambda = \frac{|s^H\mathbf{R}^{-1}y|^2}{\left(s^H\mathbf{R}^{-1}s\right)\left(y^H\mathbf{R}^{-1}y\right)}$$

Recall that $\mathbf{s}$ denotes the steering vector of the desired signal, $\mathbf{R}$ denotes the estimated CM, $\mathbf{y}$ denotes the received echo, and $\delta$ denotes the detection threshold.

In order to obtain $\delta$, we first implement 100/Pfa Monte Carlo trials to ensure a preassigned value of the probability of false alarm $P_{fa}$. In this section, we set $N_s = 1$, $N_t = 8$, and $L = 8$. The normalized Doppler frequency of the target is 0.25, and its azimuth and elevation angles are 0° and 3.6°, respectively. We use three different polarization channels, i.e., HH, HV, and VV. Note that the SCR is computed as $\text{SCR} = \sigma_s^2/\sigma_c^2$, where $\sigma_s^2$ and $\sigma_c^2$ are the powers of the target and clutter, respectively.

Fig. 3 shows the detection performance for the NMF versus the input SCR. For each abscissa, 10000 Monte Carlo experiments are performed. It is seen that the proposed methods can achieve the best detection performance among several estimators under different $P_{fa}$’s. For example, when the SCR is $-10$ dB, the detection probability with the proposed estimators is about 62%, while those with KMLE and KNSCM are 49% and 31%, respectively. This shows that the RSKE is effective for the target detection application with similar computational complexity as that of the KMLE.

| Distribution | Error ($\times 10^{-4}$) |
|--------------|--------------------------|
| Gaussian     | 12.1012                  |
| Weibull      | 9.5881                   |
| IG-CG        | 3.0916                   |
| K            | 2.4125                   |

$\text{FITTING ERROR}$

$\text{TABLE I}$

$\text{FITTING ERROR}$
Fig. 2. Fitting the real sea clutter by Weibull, IG-CG, and K distributions with different polarization. (a) HH in 19980223_171533. (b) VH in 19980223_171533. (c) VV in 19980223_171533. (d) HH in 19980226_215015. (e) VH in 19980226_215015. (f) VV in 19980226_215015.

Fig. 3. Detection performance versus the SCR. (a) \( P_{fa} = 10^{-2} \). (b) \( P_{fa} = 10^{-4} \).

**B. CM Estimation Accuracy**

In order to evaluate the CM estimation accuracy, we use the following normalized mse (NMSE) as the performance metric [87]:

\[
\text{NMSE} \triangleq \mathbb{E}\left\{ \frac{\| \tilde{R}/\text{Tr}(\tilde{R}) - R/\text{Tr}(R) \|^2}{\| R/\text{Tr}(R) \|^2} \right\}.
\]  

(43)

Since the true CM of the real data is unknown, we use synthetic data here. Considering the model in Section II, the samples are generated according to \( y_l = \sqrt{\tau_l} u_l + n_l, l = 1, 2, \ldots, L \), where \( u_l \) is generated by (3) and \( n_l \) denotes the additive white Gaussian noise. Then, the corresponding true CM is given by (8). According to Fig. 2, the sea clutter fits the CG distribution well. Therefore, we assume that the texture \( \tau_l \) follows a Gamma distribution [75] of shape parameter \( \nu \) and scale parameter \( 1/\nu \), i.e., \( \tau_l \sim \Gamma(v, 1/\nu) \), \( u_l \sim \mathcal{CN}(0, R) \). The generated samples \( \{y_l\} \) follow a zero-mean CES distribution. The estimated sub-CMs \( \tilde{R}^{(k)} \) and \( \tilde{R}^{(k)} \) in (21) are initialized as identity matrices for simplicity, but other initialization can produce similar results.

Here, we set \( N_s = 1, N_t = 8, \) and \( N_p = 3 \). The polarization parameters in (10) are set as \( \rho_c = 0.89, \gamma_c = 0.61, \) and \( \delta_c = 0.16 \). Other radar parameters include the carrier frequency of 1.2 GHz, a wavelength of 0.25 m, PRF of 2000 Hz, a platform velocity of 125 m/s, and CNR of 30 dB. In the rest of this section, for terminating the iterations, we choose the threshold \( \delta \) in (23) as \( 10^{-3} \) and \( K_{\text{max}} = 15 \). For the RSKE, in addition to the KOAS and CV choices of the shrinkage factors, the oracle choice of the shrinkage factors is also considered, which minimizes the NMSE defined in (43) at each iteration under the assumption that the true CM is known.
Fig. 4 shows the NMSE performance under different numbers of samples $L$. For each abscissa, 2000 Monte Carlo experiments are performed. Note that even a small numerical gap in the NMSE performance may lead to a large error between the estimated result and the true CM since the NMSE is normalized. We can see that the proposed RSKE can improve the estimation accuracy as compared with several existing estimators in different cases. The CV choices of the shrinkage factors can produce near-oracle performance. The performance with KOAS and CV depends on the choice of the plug-in estimates used, and CV performs slightly better than KOAS.

Fig. 5 shows the NMSE versus the space–time number. Here, we fix $N_t = 2$, $L = (1/2)N_s N_t$, and varies $N_s$ from 4 to 8. As the dimension and the number of samples increase with a constant ratio, the estimation accuracy is also improved.

Fig. 6 shows the NMSE versus $\rho_{st}$ and $\rho_p$. Here, we fix $L = 12$, and other parameters are the same as in Fig. 4. 100 Monte Carlo experiments are performed. The average NMSE achieved by RSKE with different $\rho_{st}$’s and $\rho_p$’s is demonstrated in Fig. 6 where the averages of the shrinkage factors chosen by KOAS and CV are also marked. Each line shows the contour of NMSE. It confirms that the different plug-in estimators used lead to different shrinkage factors. Moreover, CV yields solutions closer to the oracle ones compared to KOAS. The selected shrinkage coefficients are also listed in Table II.

Fig. 7 shows the condition number of the estimated CM of RSKE (with CV and KOAS), KMLE, and KNSCM. We set the plug-in estimator for CV and KOAS as KNSCM. One can see that the proposed CV and KOAS algorithms yield CM
estimates that are better-conditioned than those with KNSCM and KMLE, especially when the number of samples is small. As they also improve the NMSE, it is expected that the RSKE with the proposed shrinkage factor choices can improve the performance for applications where the inverse of the CM is required, such as beamforming and spectral estimation applications.

The performance of clutter suppression in PSTAP is often evaluated via the normalized SCNR loss [21], [80], [81]

\[
\text{SCNR}_{\text{loss}} = \frac{\left(\mathbf{s}^H \mathbf{R}^{-1} \mathbf{s}\right)^2}{\left(\mathbf{s}^H \mathbf{R}^{-1} \mathbf{R} \mathbf{R}^{-1} \mathbf{s}\right)}.
\]  

(44)

It is clear that its maximum $\text{SCNR}_{\text{loss}} = 1$ is achieved when the CM is perfectly estimated, and a larger value indicates better performance. Parameters are the same as those in Fig. 4. For each abscissa, 2000 Monte Carlo experiments are performed. Fig. 8 shows the SCNR loss resulted from different covariance estimators. We can see that the proposed RSKE with KOAS and CV can also outperform KNSCM, KMLE, and SCM.

V. CONCLUSION

In this article, we investigate a robust, iterative shrinkage estimator for Kronecker-structured CMs of compound Gaussian data, which is referred to as RSKE. The RSKE can be obtained by minimizing a negative log-likelihood function penalized by Kullback–Leibler divergence and interpreted by integrating linear shrinkage into the fixed-point iterations. The conditions for the existence of the RSKE are investigated, and the convergence of the iterative solver is investigated. We also introduce two methods for choosing the shrinkage factors by exploiting OAS and CV, respectively. The proposed estimators are then applied to polarization radar detection in the real sea clutter context. Compared with the state-of-the-art estimators, the RSKE achieves better detection performance and more accurate CM estimation, and improves the condition number by significantly reducing the number of unknown parameters and integrating shrinkage into the robust estimation.

APPENDIX A

PROOF OF PROPOSITION 1

In this appendix, we examine the conditions under which a solution to (15) exists by constructing two auxiliary functions to lower bound the cost function in (14).

Let $\lambda^{(1)}_{st} \geq \lambda^{(2)}_{st} \geq \cdots \geq \lambda^{(N_s)}_{st}$ and $\lambda^{(1)}_{p} \geq \lambda^{(2)}_{p} \geq \cdots \geq \lambda^{(N_p)}_{p}$ be the eigenvalues of $\mathbf{R}_{st}$ and $\mathbf{R}_{p}$, respectively. Then, we have

\[
\log y_i^H \left(\mathbf{R}_{st} \otimes \mathbf{R}_{p}\right)^{-1} y_i \\
\geq \log \frac{1}{N_p} \sum_{j=1}^{N_p} \log r_{j,j}^H \mathbf{R}_{st}^{-1} r_{j,j} - \log \lambda^{(1)}_{p} + \log N_p 
\]

(45)

where we have utilized Jensen’s inequality in the last step. Similarly, we have

\[
\log y_i^H \left(\mathbf{R}_{st} \otimes \mathbf{R}_{p}\right)^{-1} y_i \\
\geq \frac{1}{N_s} \sum_{i=1}^{N_s} \log c_{i,i}^H \mathbf{R}_{p}^{-1} c_{i,i} - \log \lambda^{(1)}_{st} + \log N_s.
\]

(46)

Here, we have assumed that none of $r_{j,j}$ and $c_{i,i}$ is an all-zero vector such that $r_{j,j}^H \mathbf{R}_{st}^{-1} r_{j,j} \neq 0$ and $c_{i,i}^H \mathbf{R}_{p}^{-1} c_{i,i} \neq 0, \forall i, \forall j, \forall l$. Then, let us define the following auxiliary functions:

\[
\mathcal{F}_1\left(\mathbf{R}_{st}\right) = \frac{N_p L}{2} \log \det \left(\mathbf{R}_{st}\right) + \frac{\beta_1 N_s}{2} \sum_{i=1}^{N_s} \log \lambda^{(1)}_{st}
\]

\[
+ \frac{\alpha_1 L}{2} \text{Tr} \left(\mathbf{R}_{st}^{-1}\right) + \frac{\alpha_2 L}{2} \log \det \left(\mathbf{R}_{st}\right) + \frac{\beta_2 L N}{2} \log \lambda^{(1)}_{st}
\]

\[
\mathcal{F}_2\left(\mathbf{R}_{p}\right) = \frac{N_s L}{2} \log \det \left(\mathbf{R}_{p}\right) + \frac{\beta_1 N_p}{2} \sum_{i=1}^{N_p} \log \lambda^{(1)}_{p}
\]

\[
+ \frac{\alpha_1 L}{2} \text{Tr} \left(\mathbf{R}_{p}^{-1}\right) + \frac{\alpha_2 L}{2} \log \det \left(\mathbf{R}_{p}\right) + \frac{\beta_2 L N}{2} \log \lambda^{(1)}_{p}
\]

where $\beta_1 + \beta_2 = 1$ and $\beta_1, \beta_2 \in [0, 1]$. From (45) and (46), we have

\[
\mathcal{L}\left(\mathbf{R}_{st}, \mathbf{R}_{p}\right) \\
\geq \frac{2}{L} \left(\mathcal{F}_1\left(\mathbf{R}_{st}\right) + \mathcal{F}_2\left(\mathbf{R}_{p}\right)\right) + N (\beta_1 \log N_p + \beta_2 \log N_s).
\]

Since $L, N_s, N_p$, and $N_s$ are finite, if $\mathcal{F}_1\left(\mathbf{R}_{st}\right) \rightarrow +\infty$ and $\mathcal{F}_2\left(\mathbf{R}_{p}\right) \rightarrow +\infty$, then $\mathcal{L}\left(\mathbf{R}_{st}, \mathbf{R}_{p}\right) \rightarrow +\infty$. In the following, we check the conditions under which $\mathcal{F}_1\left(\mathbf{R}_{st}\right) \rightarrow +\infty$ and $\mathcal{F}_2\left(\mathbf{R}_{p}\right) \rightarrow +\infty$ on the boundary of the set of positive-definite, Hermitian matrices. Note that $\mathcal{F}_1$ and $\mathcal{F}_2$ are similar to the first equation of [50, Appendix A].

Denote the eigenvectors corresponding to $\lambda^{(i)}_{st}$ and $\lambda^{(j)}_{p}$ by $\mathbf{v}^{(i)}_{st}$ and $\mathbf{v}^{(j)}_{p}$, respectively, for $\mathbf{R}_{st}$ and $\mathbf{R}_{p}$. Then, denote the subspace spanned by $\{\mathbf{v}^{(i)}_{st}, \ldots, \mathbf{v}^{(i)}_{st}\}$ and $\{\mathbf{v}^{(j)}_{p}, \ldots, \mathbf{v}^{(j)}_{p}\}$ as $S^{(i)}$ and $S^{(j)}$, respectively. Formally, define $[r_{st}, s_{st}]$ with $1 \leq r_{st} \leq s_{st} \leq N_s$ such that $\lambda^{(i)}_{st} \rightarrow \infty$ for $i \in [1, r_{st}]$, and $\lambda^{(i)}_{st}$ is bounded for $i \in (s_{st}, N_s]$ and $\lambda^{(i)}_{s} \rightarrow 0$ for $i \in (s_{st}, N_s]$. Similarly, define $[r_p, s_p]$ for $\lambda^{(j)}_{p}$. Here, we consider the case with $r_{st} \geq 1$, i.e., there exists at least one eigenvalue diverging.
following Sun et al. [50], in order to examine the condition for \( \mathcal{F}_1(\mathbf{R}_st) \to +\infty \) at the boundary of feasible set for \( \mathbf{R}_st \).

Define \( \mathcal{G}_1(\mathbf{R}_st) = \exp(-\mathcal{F}_1(\mathbf{R}_st)) \) and \( \mathcal{G}_2(\mathbf{R}_p) = \exp(-\mathcal{F}_2(\mathbf{R}_p)) \).

It is clear that \( \mathcal{F}_1(\mathbf{R}_st) \to +\infty \) is equivalent to \( \mathcal{G}_1(\mathbf{R}_st) \to 0 \). From [50, Appendix A], the condition for \( \mathcal{G}_1(\mathbf{R}_st) \to 0 \) can be checked by examining the infinitesimal equivalence of \( \mathcal{G}_1(\mathbf{R}_st) \) in terms of the eigenvalues \( \lambda^{(i)}_st \) of \( \mathbf{R}_st \). From (36) in [50, Appendix A], \( \mathcal{G}_1(\mathbf{R}_st) \to 0 \) if the orders of all the eigenvalues \( \lambda^{(i)}_st \to \infty \) in the infinitesimal equivalence are negative and those of \( \lambda^{(i)}_st \to 0 \) are positive. Following this argument, we invoke (36) in [50, Appendix A] by letting \( N = LN_p, K = Nst, \rho(s) = ((\beta_1 Nst)/2)\log(s), h_1(s) = s, \alpha = \alpha_1 = (\alpha st L)/2, \) and \( \mathbf{A}_1 = \mathbf{I}_{Nst} \), and hence, \( \alpha_0 = \alpha' = \beta_1 Nst \) and \( \alpha_1 = +\infty, \alpha_1' = 0 \).\footnote{Note also that, for any \( \epsilon > 0 \)}

\[
\left( \lambda^{(1)}_st \right)^{\epsilon} = o \left( \left( \phi^{(r)}_st \right)^{\epsilon} \right) = o \left( \phi^{(r)}_st \right)^{-\epsilon}
\]

where \( o() \) denotes the higher order infinitesimal and \( \phi^{(r)}_st \triangleq \left( \lambda^{(r)}_st \right)^{-1} \). Then, we impose the same condition as the first line\footnote{The second line of (36) in [50, Appendix A] is always met since \( a_1 = +\infty \) in this article.} of (36) in [50, Appendix A], i.e.,

\[
\left( \frac{LN_p}{2} + \frac{\alpha st L}{2} - \epsilon \right) - \frac{\beta_1 Nst + \epsilon}{2} LN_p P_{LNp} \left( S^{(d)}_st \right) - \frac{\beta_2 NL}{2} - \epsilon \geq 0, \quad d = 1, \ldots, Nst - 1.
\]

Under this condition, \( \mathcal{G}_1(\mathbf{R}_st) \) goes to zero, i.e., \( \mathcal{F}_1(\mathbf{R}_st) \to +\infty \) on the boundary of positive-definite and Hermitian \( \mathbf{R}_st \) [50]. Letting \( \epsilon \to 0 \) and rearranging the terms, one has

\[
P_{LNp} \left( S^{(d)}_st \right) \leq \frac{\left( LN_p + \alpha st L \right) d - \beta_2 LN}{\beta_1 LN}
\]

for arbitrary \( d = 1, \ldots, Nst - 1 \). Intuitively, this requires that the samples are evenly spread in the subspace spanned by the eigenvectors of \( \mathbf{R}_st \). The condition (48) is then rewritten in a general form as (17a). Similarly, we have (17b).

In summary, we have obtained conditions (17a) and (17b) under which the cost function (14) tends to positive infinity at the boundary of the set of positive definite and Hermitian matrix. By [50, Lemma 1], these also give a sufficient condition that a solution to (15) exists.

APPENDIX B

PROOF OF PROPOSITION 2

In this appendix, we prove the convergence of the proposed iteration process, following the methodology of [50] and [59].

By the concavity of the logarithm function, one has \( \log x \leq \log a + \frac{(x/a) - 1}{2} \). The equality holds when \( x = a \). Then, we have

\[
\log \left[ y_i^H \left( \mathbf{R}_{st}^{(k)} \otimes \mathbf{R}_p^{(k)} \right)^{-1} y_i \right] \leq \frac{y_i^H \left( \mathbf{R}_{st}^{(k)} \otimes \mathbf{R}_p^{(k)} \right)^{-1} y_i}{y_i^H \left( \mathbf{R}_{st}^{(k)} \otimes \mathbf{R}_p^{(k)} \right)^{-1} y_i} - 1
\]

where the equality holds when \( \mathbf{R}_{st} = \mathbf{R}_p^{(k)} \). We then construct the surrogate function

\[
\mathcal{G}_1 \left( \mathbf{R}_{st}^{(k)}, \mathbf{R}_p^{(k)} \right) = \frac{Np}{1 - \rho_p} \log \det \left( \mathbf{R}_{st}^{(k)} \right) + \frac{Nst}{1 - \rho_p} \log \det \left( \mathbf{R}_p^{(k)} \right)
\]

\[
+ \frac{N}{L} \sum_{i=1}^{L} y_i^H \left( \mathbf{R}_{st}^{(k)} \otimes \mathbf{R}_p^{(k)} \right)^{-1} y_i
\]

\[
+ \frac{N}{L} \sum_{i=1}^{L} \log \left[ y_i^H \left( \mathbf{R}_{st}^{(k)} \otimes \mathbf{R}_p^{(k)} \right)^{-1} y_i \right] - N
\]

\[
= \frac{Np}{1 - \rho_p} \text{Tr} \left( \mathbf{R}_{st}^{(k)} \right) + \frac{Nst \rho_p}{1 - \rho_p} \text{Tr} \left( \mathbf{R}_p^{(k)} \right).
\]

Recalling (49), we have

\[
\mathcal{L} \left( \mathbf{R}_{st}^{(k)}, \mathbf{R}_p^{(k)} \right) \leq \mathcal{G}_1 \left( \mathbf{R}_{st}^{(k)}, \mathbf{R}_p^{(k)} \right) \leq \mathcal{G}_1 \left( \mathbf{R}_{st}^{(k)}, \mathbf{R}_p^{(k)} \right)
\]

and the equality holds when \( \mathbf{R}_{st} = \mathbf{R}_p^{(k)} \), i.e.,

\[
\mathcal{L} \left( \mathbf{R}_{st}^{(k)}, \mathbf{R}_p^{(k)} \right) = \mathcal{G}_1 \left( \mathbf{R}_{st}^{(k)}, \mathbf{R}_p^{(k)} \right).
\]

It is easy to verify that the minimizer of (50) is exactly (20a) by setting the gradient of (50) with respect to \( \mathbf{R}_{st} \) to zero. It follows that

\[
\mathbf{R}_{st}^{(k+1)} = \text{arg min} \mathcal{G}_1 \left( \mathbf{R}_{st}, \mathbf{R}_{st}^{(k)}, \mathbf{R}_p^{(k)} \right).
\]

Therefore,

\[
\mathcal{L} \left( \mathbf{R}_{st}^{(k+1)}, \mathbf{R}_p^{(k)} \right) \leq \mathcal{G}_1 \left( \mathbf{R}_{st}^{(k+1)}, \mathbf{R}_p^{(k)} \right)
\]

\[
= \text{min} \mathcal{G}_1 \left( \mathbf{R}_{st}, \mathbf{R}_{st}^{(k)}, \mathbf{R}_p^{(k)} \right)
\]

\[
\leq \mathcal{G}_1 \left( \mathbf{R}_{st}^{(k+1)}, \mathbf{R}_p^{(k)} \right)
\]

\[
= \mathcal{L} \left( \mathbf{R}_{st}^{(k+1)}, \mathbf{R}_p^{(k)} \right).
\]

Then, define

\[
\mathcal{G}_2 \left( \mathbf{R}_p, \mathbf{R}_{st}^{(k+1)}, \mathbf{R}_p^{(k)} \right) = \frac{Np}{1 - \rho_p} \log \det \left( \mathbf{R}_{st}^{(k+1)} \right) + \frac{Nst}{1 - \rho_p} \log \det \left( \mathbf{R}_p^{(k)} \right)
\]

\[
+ \frac{N}{L} \sum_{i=1}^{L} y_i^H \left( \mathbf{R}_{st}^{(k+1)} \otimes \mathbf{R}_p^{(k)} \right)^{-1} y_i
\]

\[
+ \frac{N}{L} \sum_{i=1}^{L} \log \left[ y_i^H \left( \mathbf{R}_{st}^{(k+1)} \otimes \mathbf{R}_p^{(k)} \right)^{-1} y_i \right] - N
\]

\[
+ \frac{Np \rho_p}{1 - \rho_p} \text{Tr} \left( \mathbf{R}_{st}^{(k+1)} \right) + \frac{Nst \rho_p}{1 - \rho_p} \text{Tr} \left( \mathbf{R}_p^{(k)} \right).
\]
Similarly, we can verify that the minimizer of (55) is exactly (20b), and
\[
\mathcal{L}'\left(\hat{R}_{p}^{(k+1)}, \hat{R}_{p}\right) \leq \mathcal{G}_{2}\left(\hat{R}_{p}\bigg|\hat{R}_{p}^{(k+1)}, \hat{R}_{p}^{(k)}\right) \tag{56}
\]
where the equality holds when \(\hat{R}_{p} = \hat{R}_{p}^{(k)}\), i.e.,
\[
\mathcal{L}'\left(\hat{R}_{p}^{(k+1)}, \hat{R}_{p}\right) = \mathcal{G}_{2}\left(\hat{R}_{p}\bigg|\hat{R}_{p}^{(k+1)}, \hat{R}_{p}^{(k)}\right) \tag{57}
\]
It follows that
\[
\mathcal{L}'\left(\hat{R}_{p}^{(k+1)}, \hat{R}_{p}\right) \leq \mathcal{G}_{2}\left(\hat{R}_{p}\bigg|\hat{R}_{p}^{(k+1)}, \hat{R}_{p}^{(k)}\right) = \min_{R_{p}} \mathcal{G}_{2}\left(\hat{R}_{p}\bigg|\hat{R}_{p}^{(k+1)}, \hat{R}_{p}^{(k)}\right) \leq \mathcal{L}'\left(\hat{R}_{p}^{(k+1)}, \hat{R}_{p}\right) = \mathcal{L}\left(\hat{R}_{p}^{(k+1)}, \hat{R}_{p}\right) \tag{58}
\]
Combining (54) and (58), we have
\[
\mathcal{L}\left(\hat{R}_{p}^{(k+1)}, \hat{R}_{p}\right) \leq \mathcal{L}\left(\hat{R}_{p}^{(k)}, \hat{R}_{p}\right) \tag{59}
\]
i.e., the penalized log-likelihood function \(\mathcal{L}(\hat{R}_{p}, \hat{R}_{p})\) in (14) is decreasing with iterations.

Since \(\mathcal{L}(\hat{R}_{p}, \hat{R}_{p})\) is g-convex, its minimizer exists, and denote it by \((\hat{R}_{p}^{\infty}, \hat{R}_{p}^{\infty})\). Then, \(\mathcal{L}(\hat{R}_{p}^{\infty}, \hat{R}_{p}^{\infty})\) lower bounds the sequence \(\{\mathcal{L}(\hat{R}_{p}^{k}, \hat{R}_{p}^{k})\}, k = 1, 2, \cdots\). This indicates that the decreasing sequence \(\{\mathcal{L}(\hat{R}_{p}^{(k)}, \hat{R}_{p}^{(k)})\}\) is bounded by an infimum. Then, according to the monotone convergence theorem [88], the sequence will converge to the infimum as \(k\) increases, i.e., \((\hat{R}_{p}^{(k)}, \hat{R}_{p}^{(k)})\) will converge to the minimizer of \(\mathcal{L}(\hat{R}_{p}, \hat{R}_{p})\), i.e., the solution to (15).

**APPENDIX C**

**PROOF OF PROPOSITION 3**

We here complete the proof by exploiting results from random matrix theory. Following LW [89], when the true CM \(R_{st}\) and \(R_{p}\) are known, the oracle shrinkage factor \(\rho'_{p}\), i.e., the solution to (25), is given by

\[
\rho'_{p} = \frac{\mathbb{E}\{\text{Re}\left(\text{Tr}\left((\mathcal{I}_{N_{p}} - C_{p}) (R_{p} - C_{p})^{H}\right)\right)\}}{\mathbb{E}\{\text{Re}\left(\text{Tr}\left(C_{p}\right)\right)^{2}\}} = \frac{E_{1} - E_{2} - E_{3} + \text{Tr}(R_{p})}{E_{1} - 2E_{2} + N_{p}} \tag{60}
\]
where \(\text{Re}(\cdot)\) denotes the real part and
\[
E_{1} = \mathbb{E}\{\text{Tr}(C_{p})\}, \quad E_{2} = \mathbb{E}\{\text{Re}(\text{Tr}(C_{p}))\}
\]
\[
E_{3} = \mathbb{E}\{\text{Re}(\text{Tr}(C_{p}R_{p}^{H}))\} \tag{61}
\]
and \(C_{p}\) is defined by (40). The resulting optimal shrinkage estimate can be interpreted as the projection of the true CM onto the linear space spanned by \(C_{p}\) and \(\mathcal{I}_{N_{p}}\).

Let the eigen-decomposition of \(R, R_{st}\) and \(R_{p}\) be \(R = VA\Lambda^{H}, R_{st} = V_{st}\Lambda_{V_{st}}^{H}\), and \(R_{p} = V_{p}\Lambda_{p}^{H}V_{p}^{H}\), respectively. Then, we define \(z_{l} = ((D^{-1}y_{l})/(\|D^{-1}y_{l}\|_{2}))\), where \(D = VA^{1/2}\). It is easy to see that \(\|z_{l}\|_{2} = 1\) and \(\{z_{l}\}\) are independent of each other. Moreover, the whiten vectors \(\{z_{l}\}\) are isotropically distributed [90] and satisfy [47], [48]

\[
\mathbb{E}\{z_{l}z_{l}^{H}\} = \frac{1}{N}I_{N}
\]
\[
\mathbb{E}\{z_{l}^{H}z_{q}\} = \frac{1}{N(N+1)}\text{Tr}(R^{2}) + \frac{1}{N+1}\text{Tr}(\mathcal{I}_{N}) \tag{62}
\]
Note that \(D = D_{st} \otimes D_{p}\), where \(D_{st} = V_{st}\Lambda_{V_{st}}^{1/2}\) and \(D_{p} = V_{p}\Lambda_{p}^{1/2}\). We then reshape \(z_{l}\) into a matrix satisfying

\[
Z_{l} = \text{unvec}_{N_{p}N_{q}}(z_{l}) = \frac{D_{p}^{-1}Y_{l}(D_{st}^{-1})^{H}}{\|D_{st}^{-1}y_{l}\|_{2}} \tag{63}
\]
which can be easily verified by vectorizing both sides of (63).

In order to determine the shrinkage factor for the robust shrinkage estimator of unstructured CM, Chen et al. [48] analyzed the feature of \(Z_{l}\) where it reduces to a vector. We here extend the analysis to the more general case of matrix-valued \(Z_{l}\) by exploiting random matrix theory and properties of the Kronecker product. Let \(z_{l}^{(i)}\) be the \(i\)th entry of \(z_{l}\). From (62), one has

\[
\mathbb{E}\{z_{l}^{(i)}(z_{l}^{(j)})^{*}\} = \begin{cases} 1/N, & i = j \\ 0, & i \neq j. \end{cases} \tag{64}
\]
This indicates that \(\{z_{l}^{(i)}\}_{i=1}^{N_{p}}\) are i.i.d. with zero mean and variance \(1/N\). Consequently, we have

\[
\mathbb{E}\{Z_{l}Z_{l}^{H}\} = \frac{N_{p}}{N}I_{N_{p}}, \quad \mathbb{E}\{Z_{l}^{H}Z_{l}\} = \frac{N_{q}}{N}I_{N_{q}}. \tag{65}
\]
Note that \(\|D_{st}^{-1}y_{l}\|_{2}^{2} = y_{l}^{H}(R_{st} \otimes R_{p})^{-1}y_{l}\), and we have

\[
\frac{Y_{l}^{H}R_{st}^{-1}Y_{l}^{H}}{y_{l}^{H}(R_{st} \otimes R_{p})^{-1}y_{l}} = D_{st}Z_{l}Z_{l}^{H}D_{p}^{H} \tag{66}
\]
Note that \(\{\}, \mathbb{E}(\cdot), \text{Re}(\cdot), \) and \(\text{Tr}(\cdot)\) are exchangeable to each other. Substituting (66) into (61), one has

\[
E_{2} = \text{Tr}\left(\frac{N}{L N_{st}} \sum_{l=1}^{L} D_{p} \mathbb{E}(Z_{l}Z_{l}^{H})D_{p}^{H}\right) = \text{Tr}(R_{p})
\]
\[
E_{3} = \text{Tr}(R_{p}^{2}) \tag{67}
\]
From [91] and [92], we have

\[
\mathbb{E}\{\|z_{l}^{(i)}\|_{2}^{4}\} = \frac{2}{N(N+1)}, \quad \mathbb{E}\{\|z_{l}^{(i)}\|_{2}^{2}\} = \frac{1}{N(N+1)} \tag{68}
\]
Since \(\{z_{l}^{(i)}\}_{i=1}^{N_{p}}\) are i.i.d., we have

\[
\mathbb{E}\{\|z_{l}^{(i)}\|_{2}^{2}\} = \frac{1}{N}, \quad \mathbb{E}\{z_{l}^{(i)}(z_{l}^{(j)})^{*}\} = 0. \tag{69}
\]
Therefore, (61) can be rewritten as

\[ E_1 = \left( \frac{N}{LN_{st}} \right)^2 \mathbb{E} \left\{ \sum_{i=1}^{L} \sum_{q=1}^{L} L_p Z_i^H p D_p Z_q^H p \right\} \]

Utilizing [91, Lemma 1.1] and substituting (68) and (69) into (70), \( E_1 \) is obtained as (71), shown at the top of the page. Substituting (71) and (67) into (60), (27b) is obtained. Similarly, we can have the optimal \( \rho_{st}^* \), i.e., (27a). The resulting expressions of \( \rho_{st}^* \) and \( \rho^* \) can be used to produce the KOAS choice \( \rho_{st, KOAS} \) and \( \rho_{p, KOAS} \) by plugging estimates of \( \mathbf{R}_d \) and \( \mathbf{R}_p \) into (27).

**Appendix D**

**Proof of Proposition 4**

This proposition can be proven by combining the results in Appendix C. Recalling (29), (65), and (66), we have

\[ \mathbb{E}(S_d) = N_d D_d \mathbb{E}(Z_i Z_i^H) D_i^H = \mathbf{R}_{st} \]

\[ \mathbb{E}(S_p) = N_p D_p \mathbb{E}(Z_i Z_i^H) D_i^H = \mathbf{R}_p. \]  

Moreover, (28) can be rewritten as

\[ J_{st}(\Sigma_{st}) = \text{Tr}(\Sigma_{st}^2 - 2\text{Re}(\Sigma_{st} \mathbb{E}(S_d)) + \mathbb{E}(S_d^2)) \]

\[ J_p(\Sigma_p) = \text{Tr}(\Sigma_p^2 - 2\text{Re}(\Sigma_p \mathbb{E}(S_p)) + \mathbb{E}(S_p^2)). \]

By setting the derivative of (73a) and (73b) with respect to \( \Sigma_{st} \) and \( \Sigma_p \) to zero, we have the minimizer of (28) as \( \Sigma_{st} = \mathbf{R}_{st} \) and \( \Sigma_p = \mathbf{R}_p \).

**References**

[1] P. Huang, Z. Zou, X.-G. Xia, X. Liu, G. Liao, and Z. Xin, “Multichannel sea clutter modeling for spaceborne early warning radar and clutter suppression performance analysis,” IEEE Trans. Geosci. Remote Sens., vol. 59, no. 10, pp. 1–18, Oct. 2020.

[2] J. Yin, C. Unal, M. Schleiss, and H. Russchenberg, “Radar target and moving clutter separation based on the low-rank matrix optimization,” IEEE Trans. Geosci. Remote Sens., vol. 56, no. 8, pp. 4765–4780, Aug. 2018.

[3] S. Allabaksh, S. Lim, P. Yasodha, H. Kim, and G. Lee, “Intermittent clutter suppression method based on adaptive harmonic wavelet transform for L-band radar wind profiler,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 11, pp. 8546–8556, Nov. 2019.

[4] E. Makhoul, C. López-Martínez, and A. Broquetas, “Exploiting polarimetric TerraSAR-X data for sea clutter characterization,” IEEE Trans. Geosci. Remote Sens., vol. 54, no. 1, pp. 358–372, Jan. 2016.

[5] J. Carretero-Moya, J. Gismero-Menoyo, Á. Blanco-del-Campo, and A. Asensio-Lopez, “Statistical analysis of a high-resolution sea-clutter database,” IEEE Trans. Geosci. Remote Sens., vol. 48, no. 4, pp. 2024–2037, Apr. 2010.

[6] T. Zhang, L. Jiang, D. Xiang, Y. Ban, L. Pei, and H. Xiong, “Ship detection from PolSAR imagery using the ambiguity removal polarimetric notches filter,” ISPRS J. Photogramm. Remote Sens., vol. 157, pp. 41–58, Nov. 2019.

[7] T. Zhang, Z. Yang, H. Gan, D. Xiang, S. Zhu, and J. Yang, “PolSAR ship detection using the joint polarimetric information,” IEEE Trans. Geosci. Remote Sens., vol. 58, no. 11, pp. 8225–8241, Nov. 2020.

[8] Z. Xin, G. Liao, Z. Yang, Y. Zhang, and H. Dang, “A deterministic sea-clutter space–time model based on physical sea surface,” IEEE Trans. Geosci. Remote Sens., vol. 54, no. 10, pp. 6659–6673, Nov. 2016.

[9] Y. Yang, S.-P. Xiao, and X.-S. Wang, “Radar detection of small target in sea clutter using orthogonal projection,” IEEE Geosci. Remote Sens. Lett., vol. 16, no. 3, pp. 382–386, Mar. 2019.

[10] H. Ding, J. Guan, N. Liu, and G. Wang, “New spatial correlation models for sea clutter,” IEEE Geosci. Remote Sens. Lett., vol. 12, no. 9, pp. 1833–1837, Sep. 2015.

[11] H. W. Melief, H. Greidanus, P. van Genderen, and P. Hoogeboom, “Analysis of sea spikes in radar sea clutter data,” IEEE Trans. Geosci. Remote Sens., vol. 44, no. 4, pp. 985–993, Apr. 2006.

[12] A. R. Monteith and L. M. H. Ulander, “A tower-based radar study of temporal coherence of a boreal forest at P-, L-, and C-bands and linear cross polarization,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–15, 2022.

[13] N. Longepe, A. A. Mouche, L. Ferro-Famil, and R. Husson, “Co-cross-polarization coherence over the sea surface from Sentinel-1 SAR data: Perspectives for mission calibration and wind field retrieval,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–16, 2022.

[14] Y. Wang and V. Chandrasekar, “Polarization isolation requirements for linear dual-polarization weather radar in simultaneous transmission mode of operation,” IEEE Trans. Geosci. Remote Sens., vol. 44, no. 8, pp. 2019–2028, Aug. 2006.

[15] C. Lukashin, Z. Jin, G. Kopp, D. G. MacDonnell, and K. Thome, “CLARREO reflected solar spectrometer: Restrictions for instrument sensitivity to polarization,” IEEE Trans. Geosci. Remote Sens., vol. 53, no. 12, pp. 6703–6709, Dec. 2015.

[16] M. Galletti, D. Huang, and P. Kollias, “Zenith/nadir pointing mm-Wave radars: Linear or circular polarization?” IEEE Trans. Geosci. Remote Sens., vol. 52, no. 1, pp. 628–639, Jan. 2014.

[17] A. D. Maio, G. Alfano, and E. Conte, “Polarization diversity detection in compound-Gaussian clutter,” IEEE Trans. Aerosp. Electron. Syst., vol. 40, no. 1, pp. 114–131, Jan. 2004.

[18] L. Xie, Z. He, J. Tong, J. Li, and H. Li, “Transmitter polarization optimization for space-time adaptive processing with diversely polarized antenna array,” Signal Process., vol. 169, Apr. 2020, Art. no. 107401.

[19] G. Noriega and S. Pasupathy, “Adaptive estimation of noise covariance matrices in real-time preprocessing of geophysical data,” IEEE Trans. Geosci. Remote Sens., vol. 35, no. 5, pp. 1146–1159, Sep. 1997.

[20] S. Tadjudin and D. A. Landgrebe, “Covariance estimation with limited training samples,” IEEE Trans. Geosci. Remote Sens., vol. 37, no. 4, pp. 2113–2118, Jul. 1999.

[21] I. S. Reed, J. D. Mallett, and L. E. Brennan, “Rapid convergence rate in adaptive arrays,” IEEE Trans. Aerosp. Electron. Syst., vol. AES-10, no. 6, pp. 853–863, Nov. 1974.
[75] L. M. Novak, M. C. Burl, and W. W. Irving, “Optimal polarimetric processing for enhanced target detection,” IEEE Trans. Aerosp. Electron. Syst., vol. 29, no. 1, pp. 234–244, Jan. 1993.

[76] L. C. Godara, “Application of antenna arrays to mobile communications. II. Beam-forming and direction-of-arrival considerations,” Proc. IEEE, vol. 85, no. 8, pp. 1195–1245, Aug. 1997.

[77] E. A. P. Habets, J. Benesty, I. Cohen, S. Gannot, and J. Dmochowski, “New insights into the MVDR beamformer in room acoustics,” IEEE Trans. Audio, Speech, Language Process., vol. 18, no. 1, pp. 158–170, Jan. 2010.

[78] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, “Information-theoretic metric learning,” in Proc. 24th Int. Conf. Mach. Learn., vol. 25, no. 9, pp. 3080–3084, Sep. 2011.

[79] D. R. Hunter and K. Lange, “A tutorial on MM algorithms,” J. Comput. Graph. Stat., vol. 10, no. 3, pp. 603–621, 2001.

[80] O. Ledoit and M. Wolf, “Improved estimation of the covariance matrix of stock returns with an application to portfolio selection,” J. Empirical Finance, vol. 10, no. 5, pp. 603–621, 2003.

[81] A. Breloz, “Stable numerical integration of the free-electron laser,” J. Computational Phys. Rev., vol. 223, no. 2, pp. 259–280, 2006.

[82] J. Shi, L. Xie, Z. Cheng, Z. He, and W. Zhang, “Angle-Doppler channel selection method for reduced-dimension STAP based on sequential convex programming,” IEEE Commun. Lett., vol. 25, no. 9, pp. 3080–3084, Sep. 2021.

[83] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, “Information-theoretic metric learning,” in Proc. 24th Int. Conf. Mach. Learn., 2007, pp. 209–216.

[84] I. S. Dhillon, “The log-determinant divergence and its applications,” in Proc. Householder Symp., Zeuthen, Germany, 2008, pp. 1–39.

[85] R. Klemm, Principles of Space—Time Adaptive Processing. London, U.K.: IET, 2006.

[86] W. L. Melvin, “A STAP overview,” IEEE Aerosop. Electron. Syst. Mag., vol. 19, no. 1, pp. 19–35, Jan. 2004.

[87] L. Xie, Z. He, J. Tong, and W. Zhang, “A recursive angle-Doppler channel selection method for reduced-dimension space-time adaptive processing,” IEEE Trans. Aerosp. Electron. Syst., vol. 56, no. 5, pp. 3985–4000, Oct. 2020.

[88] J. Shi, L. Xie, Z. Cheng, Z. He, and W. Zhang, “Angle-Doppler channel selection method for reduced-dimension STAP based on sequential convex programming,” IEEE Commun. Lett., vol. 25, no. 9, pp. 3080–3084, Sep. 2021.

[89] F. Hiai and D. Petz, “Asymptotic freeness almost everywhere for random matrices,” Acta Sci. Math. Szeged, vol. 66, no. 1, pp. 801–826, 2000.

[90] A. M. Tulino and S. Verdú, “Random matrix theory and wireless communications,” Found. Trends Commun. Inf. Theory, vol. 1, no. 1, pp. 1–182, 2004.

Lei Xie (Member, IEEE) received the B.E. and Ph.D. degrees from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2016 and 2021, respectively. He is currently a Post-Doctoral Fellow with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong. His research interests include integrated sensing and communications, mobile networks, statistical signal processing, and optimization theory.

Zishu He was born in Chengdu, Sichuan, China, in 1962. He received the B.S., M.S., and Ph.D. degrees in signal and information processing from the University of Electronic Science and Technology of China (UESTC), Chengdu, in 1984, 1988, and 2000, respectively.

He is currently a Professor of signal and information processing with the School of Information and Communication Engineering, UESTC. He has finished more than 400 papers and has written two books: Signals and Systems and multi-input multi-output (MIMO) Radar. His research interests are involved in array signal processing, the theory on MIMO radar and cognitive radar, space and time adaptive signal processing (STAP), and clutter and interference cancellation based on knowledge.

Jun Li was born in Sichuan, China. He received the B.S., M.S., and Ph.D. degrees in signal and information processing from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2000, 2003, and 2009, respectively.

He is currently an Associate Professor with the School of Information and Communication Engineering, UESTC. His research interests include adaptive signal processing, cognitive radar, and multi-input multi-output (MIMO) radar.

Jiangtao Xi (Senior Member, IEEE) received the B.E. degree from the Beijing Institute of Technology, Beijing, China, in 1982, the M.E. degree from Tsinghua University, Beijing, in 1985, and the Ph.D. degree from the University of Wollongong, Wollongong, NSW, Australia, in 1996, all in electrical engineering.

He was a Post-Doctoral Fellow with the Communications Research Laboratory, McMaster University, Hamilton, ON, Canada, from 1995 to 1996, and a member of Technical Staff with Bell Laboratories, Lucent Technologies Inc., Murray Hill, NJ, USA, from 1996 to 1998. From 2000 to 2002, he was the Chief Technical Officer of TCL IT Group Company, Huizhou, China. In 2003, he rejoined the University of Wollongong as a Senior Lecturer, where he is currently a Professor and also the Head of the School of Electronic, Computer, and Telecommunications Engineering. His research interests include signal processing and its applications in various areas, such as optoelectronics, instrumentation and measurement, and telecommunications.