Constrained ML Estimation of Structured Covariance Matrices with Applications in Radar STAP

(Invited Paper)

Bosung Kang, Vishal Monga
Department of Electrical Engineering
The Pennsylvania State University
University Park, Pennsylvania 16802

Muralidhar Rangaswamy
Air Force Research Lab.
Wright-Patterson Air Force Base
Dayton, Ohio

***Abstract*— The disturbance covariance matrix in radar space time adaptive processing (STAP) must be estimated from training sample observations. Traditional maximum likelihood (ML) estimators are effective when training is generous but lead to degraded false alarm rates and detection performance in the realistic regime of limited training. We exploit physically motivated constraints such as 1.) rank of the clutter subspace which can be inferred using existing physics based models such as the Brennan rule, and 2.) the Toeplitz constraint that applies to covariance matrices obtained from stationary random processes. We first provide a closed form solution of the rank constrained maximum likelihood (RCML) estimator and then subsequently develop an efficient approximation under joint Toeplitz and rank constraints (EASTR). Experimental results confirm that the proposed EASTR estimators outperform state-of-the-art alternatives in the sense of widely used measures such as the signal to interference and noise ratio (SINR) and probability of detection – particularly when training support is limited.**

I. INTRODUCTION

Space time adaptive processing (STAP) is widely used in modern radar signal processing since it creates an ability to suppress interfering signals while simultaneously preserving gain on the desired signal. Interference statistics, in particular the disturbance covariance matrix which plays a vital role in STAP must be estimated from training samples. In the absence of any prior knowledge about the interference environment, a challenge for STAP is that a large number of homogeneous (target free) disturbance training samples are required for STAP to be successful. Since generous homogeneous observations are generally not available, many approaches exploiting particular structure of the disturbance covariance matrix have been developed to overcome this practical issue of lack of generous training. Examples of structure include persymmetry, circulant structure, physical constraints, and so on.

The fast maximum likelihood (FML) method [1] which enforces special eigenstructure was proposed and in fact is shown to be the most competitive technique experimentally. In particular, the disturbance covariance matrix obeys the following structure

$$\mathbf{R} = \sigma^2\mathbf{I} + \mathbf{R}_c, \tag{1}$$

where $\mathbf{R}_c$ represents rank deficient clutter matrix which is positive semi-definite and $\mathbf{I}$ is an identity matrix. The FML technique enforces all eigenvalues of the estimated covariance matrix to be greater than σ^2. Recently Kang *et al.* proposed the rank constrained ML (RCML) estimation which incorporates the rank of the clutter covariance matrix, $\mathbf{R}_c$, explicitly into ML estimation of the disturbance covariance matrix [2]. Their solution is shown to be optimal for all training regimes.

Since the covariance matrix from a stationary stochastic signal is Hermitian and Toeplitz, estimating Toeplitz covariance benefits many applications such as array processing and time series analysis. The seminal work by Burg *et al.* [3] proposed an *iterative* method for estimation of structured covariance matrices using the ML method in its full generality . Li *et al.* developed the asymptotic maximum likelihood (AML) estimation for structured covariance matrices [4] using the extended invariance principle (EXIP) [5]. Approximation of arbitrary matrices by a (Hermitian) Toeplitz matrix using matrix decompositions and outer approximations has separately been pursued in applied mathematics. Of particular interest is Al-Homidan's l_1 sequential quadratic programming (SQP) method to find the nearest symmetric positive semidefinite Toeplitz matrix to given a matrix [6].

Although various methods have been proposed for estimating Toeplitz covariance matrices, it is well known that the exact ML estimation of a Hermitian Toeplitz covariance matrix has no closed-form solution that is valid for all training regimes. Li *et al.* [4] derived a closed form solution of the asymptotic maximum likelihood (AML) estimation for structured covariance matrices using the extended invariance principle (EXIP). However, it is asymptotically valid, which means it performs well only for generous homogeneous training samples which is generally not available in practice. In the regime of realistic training, methods relying on numerical optimization (often non-convex) such as ITAM [7] are computationally involved and hence are unsuitable.

In this paper, we deal with estimation of the structural disturbance covariance matrix under practical constraints, in particular, the knowledge of the rank of the clutter matrix and Toeplitz structure. First we introduce the rank constrained

978-1-4673-3146-3/13/$31.00 ©2013IEEE

ML (RCML) estimation and then develop an efficient approximation of structured covariance under joint Toeplitz and rank constraints (EASTR) as an extension of the RCML. For RCML estimator, we derive a closed form solution from the optimization problem which is initially not a convex problem. The EASTR is based on a cascade of two closed form solutions. The first closed form is the RCML and then we propose a new method to perturb the eigenvalues of the RCML estimator in a rank preserving manner so as to impose additional Toeplitz structure. We formulate a new quadratic programming (QP) optimization problem that solves for the optimal eigenvalues while incorporating Toeplitz constraints and demonstrate that this problem also admits a closed form solution.

II. Rank Constrained ML Estimation of Structured Covariance Matrices

Let $\mathbf{z}_i \in \mathbb{C}^N$ be the i-th realization of the target-free (stochastic) disturbance vector and K be the number of training samples. That is, $i = 1, 2, \ldots, K$ and N is the dimension of the vector. Therefore, in each training sample, $\mathbf{z}_i$, under assumption of zero mean, obeys

$$f(\mathbf{z}_i) = \frac{1}{\pi^N |\mathbf{R}|} \exp(-\mathbf{z}_i^H \mathbf{R}^{-1} \mathbf{z}_i), \tag{2}$$

which comes from a zero-mean complex circular Gaussian distribution and $\mathbf{R}$ is the $N \times N$ disturbance covariance matrix. Since each observations $\mathbf{z}_i$ are i.i.d. (independent and identically distributed), the likelihood of observing $\mathbf{Z}$ given $\mathbf{R}$ is given by

$$f_{(\mathbf{R})}(\mathbf{Z}) = \frac{1}{\pi^{NK}} |\mathbf{R}|^{-K} \exp\big(-K \cdot tr\{\mathbf{R}^{-1}\mathbf{S}\}\big), \tag{3}$$

where $\mathbf{S} = \frac{1}{K}\mathbf{Z}\mathbf{Z}^H$ is the well-known sample covariance matrix. Maximizing the likelihood function is equivalent to minimizing the following which is the cost function of the optimization problem,

$$tr\{\mathbf{S}'\mathbf{X}\} - \log(|\mathbf{X}|), \tag{4}$$

where $\mathbf{X} = \sigma^2 \mathbf{R}^{-1}$ is the inverse of normalized covariance matrix and $\mathbf{S}' = \frac{1}{\sigma^2}\mathbf{S}$ is the sample covariance matrix normalized by σ^2. Further, the cost function (4) can be simplified to the function of eigenvalues of $\mathbf{X}$ by the eigenvalue decomposition of $\mathbf{X}$ and $\mathbf{S}'$ and the fairly well known fact that the function (4) has the minimum value when both the eigenvector matrices of $\mathbf{X}$ and $\mathbf{S}'$ are identical [8],

$$\mathbf{d}^T \boldsymbol{\lambda} - \mathbf{1}^T \log \boldsymbol{\lambda}, \tag{5}$$

where $\mathbf{d}$ and $\boldsymbol{\lambda}$ are vectors with entries of eigenvalues of $\mathbf{S}'$ and $\mathbf{X}$ respectively and $\log \boldsymbol{\lambda} = [\log \lambda_1, \log \lambda_2, \cdots, \log \lambda_N]^T$.

Now we consider the constraints of the optimization problem,

$$\begin{cases} \mathbf{R} = \sigma^2 \mathbf{I} + \mathbf{R}_c \\ rank(\mathbf{R}_c) = r \\ \quad \mathbf{R}_c \succeq \mathbf{0} \end{cases}. \tag{6}$$

Since $rank(\mathbf{R}_c) = r$, $\mathbf{R}_c$ has r positive eigenvalues and the rest eigenvalues are all zero. Consequently, from Eq. (1), $\mathbf{R}$ has r eigenvalues greater than or equal to σ^2 and the rest eigenvalues equal to σ^2. Now let the i-th eigenvalue of $\mathbf{X}$ be λ_i. Since $\mathbf{X} = \sigma^2 \mathbf{R}^{-1}$, we can readily see that $\mathbf{X}$ has r positive eigenvalues less than or equal to 1 and the rest equal to 1. Finally, the constraint about the eigenvalues of $\mathbf{X}$ is

$$0 < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_r \le \lambda_{r+1} = \lambda_{r+2} = \cdots = 1. \tag{7}$$

Now the constraint (7) can be expressed in vector and matrix forms. First, $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_N$ is $\mathbf{U}\boldsymbol{\lambda} \preceq \mathbf{0}$ where

$$\mathbf{U} = \begin{bmatrix} 1 & -1 & \cdots & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & \cdots & 1 & -1 \\ 0 & \cdots & 0 & -1 \end{bmatrix} \in \mathbb{R}^{N \times N}. \tag{8}$$

Second, $0 < \lambda_i \le 1$ which can be expressed by

$$\boldsymbol{\varepsilon} \preceq \boldsymbol{\lambda} \preceq \mathbf{1}, \tag{9}$$

where $\boldsymbol{\varepsilon}$ is a vector with all entries equal to the same constant ϵ such that ϵ is picked close to zero. The final constraint is $\lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_N = 1$, and is expressed as

$$\mathbf{E}\boldsymbol{\lambda} = \mathbf{h}, \tag{10}$$

where

$$\mathbf{E} = \begin{bmatrix} \mathbf{0}_{r \times r} & \mathbf{0}_{r \times (N-r)} \\ \mathbf{0}_{(N-r) \times r} & \mathbf{I}_{N-r} \end{bmatrix} \in \mathbb{R}^{N \times N} \tag{11}$$

and $\mathbf{h} = [0, 0, \cdots, 0_r, 1, 1, \cdots, 1]^T$.

We therefore have the following optimization problem.

$$\begin{cases} \min\limits_{\boldsymbol{\lambda}} & \mathbf{d}^T \boldsymbol{\lambda} - \mathbf{1}^T \log \boldsymbol{\lambda} \\ s.t. & \mathbf{F}\boldsymbol{\lambda} \preceq \mathbf{g} \\ & \mathbf{E}\boldsymbol{\lambda} = \mathbf{h} \end{cases}, \tag{12}$$

where $\mathbf{F} = \begin{bmatrix} \mathbf{U}^T & -\mathbf{I} & \mathbf{I} \end{bmatrix}^T$, $\mathbf{g} = \begin{bmatrix} \mathbf{0}^T & -\boldsymbol{\varepsilon}^T & \mathbf{1}^T \end{bmatrix}^T$. The optimization problem (12) is obviously a convex optimization problem because the cost function is a convex function and feasible constraint sets are convex as well. A closed form solution for (12) can in fact be derived using KKT conditions [9] in constrained optimization and shown in [2]. It should be noted that this is a generalization of the FML solution in [1] with the rank-information incorporated.

III. Efficient Approximation of Structured Covariance under Joint Toeplitz and Rank Constraints

The EASTR now involves enforcing the Toeplitz structure on top of the RCML estimator. The EASTR is based on a cascade of two closed form solutions that capture the rank and Toeplitz constraints respectively. We modify the RCML estimate in way that the Toeplitz structure is captured but without compromising the rank constraint. To do that, we optimize the eigenvalues of the clutter matrix $\mathbf{R}_c$ so that

$\mathbf{R}_c$ be Toeplitz. Let the eigenvector matrix of $\mathbf{S}$ be $\mathbf{\Phi}$[1] and the eigenvalues of $\mathbf{R}_c$ be $\lambda_1, \lambda_2, \ldots, \lambda_r, \ldots, \lambda_N$. Since $rank(\mathbf{R}_c) = r$, $\lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_N = 0$. Therefore, $\mathbf{R}_c$ can be expressed as

$$\mathbf{R}_c = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^H, \tag{13}$$

where $\mathbf{\Lambda}$ is a diagonal matrix with r positive eigenvalues and $N-r$ zero's as diagonal entries. Therefore, ijth component of $\mathbf{R}_c$ is given by

$$(\mathbf{R}_c)_{ij} = \sum_{k=1}^{r} \lambda_k \phi_{ik} \phi_{jk}^*. \tag{14}$$

Note that $\mathbf{R}_c$ is already Hermitian, that is, $(\mathbf{R}_c)_{ij} = (\mathbf{R}_c)_{ji}^*$. Now in order for $\mathbf{R}_c$ to be Toeplitz matrix, all elements in each diagonal must be same. For example, in the main diagonal,

$$(\mathbf{R}_c)_{11} = (\mathbf{R}_c)_{22} = \cdots = (\mathbf{R}_c)_{NN} \tag{15}$$

must hold. Let's see the first equation, $(\mathbf{R}_c)_{11} = (\mathbf{R}_c)_{22}$. It can be expressed as in vector form,

$$\begin{bmatrix} \phi_{11}\phi_{11}^* - \phi_{21}\phi_{21}^* & \cdots & \phi_{1r}\phi_{1r}^* - \phi_{2r}\phi_{2r}^* \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_r \end{bmatrix} = 0. \tag{16}$$

The other equations also can be expressed as the vector form like Eq. (16). Consequentially we have totally $N(N-1)/2$ equations and finally get the following equation which is a constraint for Toeplitz matrix.

$$\mathbf{\Psi}\boldsymbol{\lambda} = \mathbf{0}, \tag{17}$$

where each row of $\mathbf{\Psi} \in \mathbb{C}^{N(N-1)/2 \times r}$ comes from identicalness of the diagonal elements and $\boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_r \end{bmatrix}^T$.

Eq. (17) is a homogeneous overdetermined linear system, that is, we have more equations than unknowns. Therefore we consider two cases here. The first case is that we have an infinite set of solutions when the column rank of $\mathbf{\Psi}$ is less than r. On the other hand, when $\mathbf{\Psi}$ has a full column rank, we have the only one trivial solution, $\boldsymbol{\lambda} = \mathbf{0}$, which results in the covariance matrix estimate, $\hat{\mathbf{R}} = \sigma^2\mathbf{I}$, which we do not desire. Now we derive the optimal eigenvalues for each case.

A. $\mathrm{rank}(\mathbf{\Psi}) < r$ – *Exact Toeplitz Solution*

Let $\boldsymbol{\lambda}_{\text{RCML}}$ be the eigenvalues obtained from the RCML estimator. We already know the eigenvalues $\boldsymbol{\lambda}$ from the RCML estimate are the optimal ML estimate of the eigenvalues of the true covariance matrix under only the rank constraint. We want the eigenvalues of the clutter matrix to satisfy Eq. (17). Since Eq. (17) has the infinite number of solutions, we find the closest vector of the eigenvalues to $\boldsymbol{\lambda}_{\text{RCML}}$ by solving the following convex optimization problem.

$$\begin{array}{cc} \min\limits_{\boldsymbol{\lambda}} & ||\boldsymbol{\lambda}_{\text{RCML}} - \boldsymbol{\lambda}||^2 \\ \text{subject to :} & \mathbf{\Psi}\boldsymbol{\lambda} = \mathbf{0} \end{array}. \tag{18}$$

The optimization problem (18) is a quadratic programming (QP) problem with a equality constraint and therefore the closed form solution is available using KKT condition [9] and it is given by solving the following equations.

$$\begin{bmatrix} 2\mathbf{I} & \mathbf{\Psi}^T \\ \mathbf{\Psi} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda}^\star \\ \boldsymbol{\nu}^\star \end{bmatrix} = \begin{bmatrix} 2\boldsymbol{\lambda}_{\text{RCML}} \\ \mathbf{0} \end{bmatrix}, \tag{19}$$

where $\boldsymbol{\nu}^\star$ is a vector of Lagrange multipliers.

B. $\mathrm{rank}(\mathbf{\Psi}) = r$ – *Toeplitz Approximation*

In this case, Eq. (17) has the only one solution, $\boldsymbol{\lambda} = \mathbf{0}$, which does not give us any useful information about the covariance matrix at all. Therefore, we take a little perturbation on $\mathbf{\Psi}$ so that the constraint can have infinite number of solution. We want to find the closest matrix to $\mathbf{\Psi}$ with the column rank less than r. By the well-known theorem, Eckart-Young theorem [10], we know that the closest matrix $\tilde{\mathbf{\Psi}}$ to $\mathbf{\Psi}$ in the sense of Frobenius norm is given by

$$\tilde{\mathbf{\Psi}} = \mathbf{U}\tilde{\mathbf{\Sigma}}\mathbf{V}^H, \tag{20}$$

where $\tilde{\mathbf{\Sigma}}$ is a diagonal matrix with $r-1$ largest positive singular values of $\mathbf{\Psi}$. By substituting $\mathbf{\Psi}$ with $\tilde{\mathbf{\Psi}}$ in Eq. (17), we obtain the infinite number of solutions for $\boldsymbol{\lambda}$.

Now we have the final optimization problem which is given by

$$\begin{array}{cc} \min\limits_{\boldsymbol{\lambda}} & ||\boldsymbol{\lambda}_{\text{RCML}} - \boldsymbol{\lambda}||^2 + \gamma||\mathbf{\Psi}\boldsymbol{\lambda}||^2 \\ \text{subject to :} & \tilde{\mathbf{\Psi}}\boldsymbol{\lambda} = \mathbf{0} \end{array}. \tag{21}$$

Note that a regularization term in the cost function is added to keep faithfulness to Toeplitz structure of the estimated covariance matrix and the regularization parameter γ is determined heuristically in this paper. Eq. (21) is also a QP problem with an equality constraint and the closed form solution is available as before.

Remark: It should be noted that the actual rank of $\mathbf{\Psi}$ which is derived from $\mathbf{\Phi}$ depends on the training data. If the true covariance is indeed Toeplitz, we expect training samples to reflect that particularly in the regime of $K \geq N$ training samples (reasonably high training), this is indeed what we observe in practice.

IV. Experimental Investigation

In this section, we compare the performance of the proposed EASTR with alternative algorithms. We use the normalized signal to interference and noise ratio (SINR) and probability of detection vs. signal to noise ratio (SNR) as measurements. The normalized SINR measure [11] is commonly used in the radar literature and the detection probability is defined as the probability that the value of test statistic is greater than a threshold conditioned on the hypothesis that the received data

[1]Note that $\mathbf{\Phi}$ is the optimal eigenvector matrix of unconstrained (other than normality) ML estimation of non-singular $\mathbf{R}$. However, finding the optimal solutions of the eigenvalues and the eigenvectors for Toeplitz matrix simultaneously is very difficult and involves a computationally expensive iterative procedure. In this paper, we focus on efficient approximation and analytically tractable framework for exploiting both Toeplitz structure and clutter rank though the solution can be suboptimal.

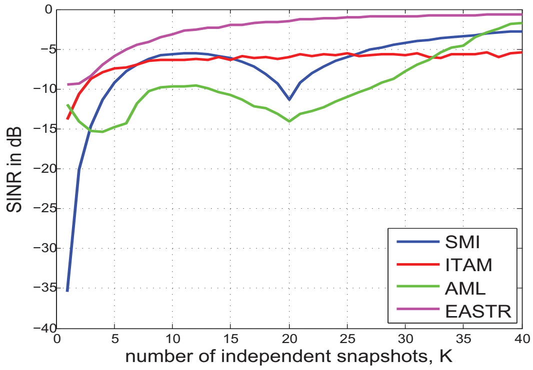

Fig. 1. Normalized SINR versus number of training samples, $N = 20$

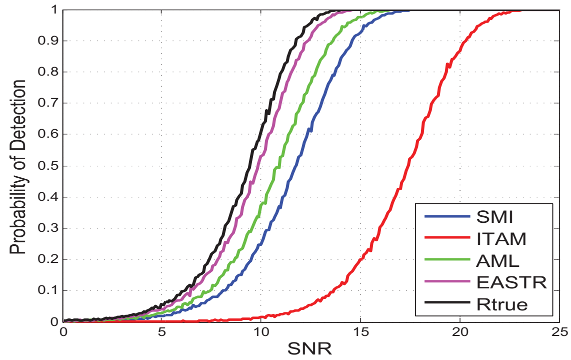

Fig. 2. Probability of detection vs. SNR via normalized matched filter (NMF) test. $K = 2N = 40$ is used.

includes target information. We apply the normalized matched filter (NMF) as the test statistics which is given by

$$\frac{|\mathbf{s}^H\mathbf{R}^{-1}\mathbf{x}|^2}{[\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}][\mathbf{x}^H\mathbf{R}^{-1}\mathbf{x}]} \underset{H_0}{\overset{H_1}{\gtrless}} \lambda_{\text{NMF}}, \tag{22}$$

where $\mathbf{x}$ and K are the observation vector and the number of training samples, respectively.

We employ a radar covariance simulation model which satisfies Toeplitz and low rank property of the true clutter covariance matrix and was successfully used in previous literature [12] for the experiments. In addition, the true clutter covariance matrix generally has rank r less than N. Therefore, this model can not only be used to simulate radar disturbance samples but also makes ground truth covariance and the clutter rank available. We compare four different Toeplitz covariance estimation techniques, SMI, iterated Toeplitz approximation method (ITAM) [7], the asymptotic maximum likelihood (AML) [4], and EASTR.

We plot the normalized average SINR versus the number of independent snapshots in Fig. 1. When $K < N$ the sample covariance is singular, therefore we used its pseudo-inverse instead of inverse itself.[2] The AML does particularly well when training is generous $K \gg N$ because AML is asymptotically based. ITAM is effective in very low training as expected because its exploits both rank and Toeplitz constraints - ITAM does not exhibit scalable improvements as training support is increased. By virtue of incorporating the rank information and Toeplitz property, the proposed EASTR outperforms the competing methods.

Fig. 2 shows the detection probability P_d plotted as a function of SNR for different estimators. We use $K = 2N = 40$ training samples to estimate the covariance matrix. The proposed EASTR is also the closest to the P_d achieved by using the true covariance matrix (upper bound) and AML follows EASTR.

[2]Note that SMI and AML have a dip when $K = 20$ due to numerical instabilities in the $K = N$ training regime. In contrast, ITAM and EASTR guarantee nonsingularity in all training regimes.

V. Conclusion

Our work focuses on estimation of structured covariance matrices for radar STAP under practical constraints, 1.) the rank constraint and 2.) Toeplitz structure. We introduce a closed form solution of the RCML estimator and then develop a new estimator that is based on a cascade of two closed forms. Crucially, this optimization also has a closed form making the overall estimator very friendly from a computational standpoint. Via evaluating probability of detection and normalized SINR, our estimators are shown to outperform traditional efforts in Toeplitz and low rank covariance estimation including those based on expensive numerical solutions.

References

[1] M. Steiner and K. Gerlach, "Fast Converging Adaptive Processors or a Structured Covariance Matrix," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 4, pp. 1115–1126, October 2000.

[2] B. Kang, V. Monga, and M. Rangaswamy, "Rank Constrained ML Estimation of Structured Covariance Matrices," *IEEE Trans. Aerosp. Electron. Syst.*, 2013, accepted.

[3] J. P. Burg, D. G. Luenberger, and D. L. Wenger, "Estimation of Structured Covariance Matrices," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 963–974, September 1982.

[4] H. Li, P. Stoica, and J. Li, "Computationally Efficient Maximum Likelihood Estimation of Structured Covariance Matrices," *IEEE Trans. Signal Processing*, vol. 47, no. 5, pp. 1314–1323, May 1999.

[5] P. Stoica and T. Söderström, "On reparameterization of loss functions used in estimation and the invariance principle," *Signal Processing*, vol. 17, no. 4, pp. 383–387, August 1989.

[6] S. S. Al-Homidan, "SQP Algorithms for Solving Toeplitz Matrix Approximation Problem," *Numerical Linear Algebra with Applications*, vol. 9, no. 8, pp. 619–627, 2002.

[7] D. M. Wilkes and M. H. Hayes, "Iterated Toeplitz Approximation of Covariance Matrices," in *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on*, April 1988, vol. 3, pp. 1663–1666.

[8] A. Haimovich, "The Eigencanceler: Adaptive Radar by Eigenanalysis Methods," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 2, pp. 532–542, April 1996.

[9] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2nd edition, 2004.

[10] C. Eckart and G. Young, "The Approximation of One Matrix by Another of Lower Rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, September 1936.

[11] R. A. Monzingo and T. Miller, *Introduction to Adaptive Arrays*, SciTech Publishing, 1st edition, 2004.

[12] A. De Maio, S. De Nicola, L. Landi, and A. Farina, "Knowledge-Aided Covariance Matrix Estimation: a MAXDET Approach," *IET Radar, Sonar and Navigation*, vol. 3, no. 4, pp. 341–356, August 2009.