Optimal Toeplitz Completion of Covariance Matrix for Robust DOA Estimation

Carl-Inge C. Nilsen, *Member, IEEE,* Ines Hafizovic, *Student Member, IEEE*

***Abstract*—Algorithms for direction-of-arrival (DOA) estimation and beamforming will suffer from decreased performance when the estimated statistics deviate from the underlying model. This can be due to estimation from a finite number of data vectors or in case of correlation between signals. Covariance matrix estimators can attempt to counteract these effects by forcing the estimate to have a Toeplitz structure, such as redundancy averaging. It is known that redundancy averaging improves performance, but results in biased DOA estimates that are worse than what can be accomplished with non-Toeplitz constrained techniques such as e.g. spatial averaging. In this letter/correspondence we introduce an optimal, iterative Toeplitz-constrained covariance matrix estimator. We show that the estimator yields redundancy averaging as its first step, and that subsequent steps improve the DOA estimates by reducing the bias and increasing resolution beyond that of redundancy averaging.**

I. Introduction

Several adaptive beamforming and DOA estimation methods are based on examining the spatial covariance matrix. For certain array geometries, such as the uniformly spaced linear array (ULA), the covariance matrix should ideally be Toeplitz. Due to different deviations from our assumptions, such as too few data vectors and/or signal correlation (*should we mention element position errors?*) this criterion is often not satisfied. Several authors have suggested so-called Toeplitz-completion-based estimation methods that force the estimated covariance matrix to be nearly Toeplitz, e.g. adaptive spatial averaging [1], or perfectly Toeplitz, e.g. redundancy averaging [2]. We have chosen to investigate the latter method. It is known that redundancy averaging creates a covariance matrix that is inconsistent with the underlying signal model, yielding biased DOA estimates [3] that can be outperformed by ordinary spatial averaging. It is not obvious how to further develop or modify redundancy averaging to improve its performance, as the technique does not include any parameters and has not been derived using any optimality criterion.

In this correspondence we derive a novel technique for covariance matrix decorrelation through Toeplitz completion and show that a slightly modified form of redundancy averaging is the result of the first iteration. Although the covariance matrix estimates are still inconsistent, the bias induced in the DOA estimates by this inconsistency is significantly reduced for each iteration. The new technique is shown through simulations to have better threshold and resolution properties than both redundancy averaging avd spatial averaging, as well as a DOA RMSE that is superior to redundancy averaging and approaches spatial averaging as SNR increases.

C.-I.C.Nilsen is with the Department of Informatics, University of Oslo. Ines Hafizovic is with Squarehead Technology AS.

Manuscript received XX XX, 2012.

II. Background

A. Array and Signal Model

We assume an M-element uniformly spaced linear array (ULA) which captures N temporal samples of a spatial data vector, $\vec{x}[n]$, consisting of $D \leq M$ plane-wave signals of interest, $\vec{s}[n]$, and spatially white noise, $\vec{n}[n]$:

$$\vec{x}[n] = \vec{s}[n] + \vec{n}[n] \in \mathbb{C}^M, n = 0, 1, \cdots, N-1 \quad (1)$$

The signals of interest can be described as:

$$\vec{s}[n] = e^{i\omega n} \sum_{d=0}^{D-1} A_d \vec{v}(k_d) \quad (2)$$

where A_d are the complex signal amplitudes, and the propagation vectors $\vec{v}(k_d)$ for signals arriving at the ULA are:

$$\vec{v}(k_d) = \left[1, e^{ik_d}, \cdots, e^{ik_d(M-1)}\right]^T. \quad (3)$$

where the signal wavenumbers k_d are defined from the directions of arrival (DOA) θ_d as as:

$$k_d = \frac{2\pi}{\lambda} \sin\theta_d, \quad (4)$$

Our primary goal is to estimate the DOAs θ_d from the data $\vec{x}[n]$. The number of signals, D, is either known, or it must be estimated from the data as well.

We will make use of the wavenumber domain representation of data, which relates to the spatio-temporal domain through a Fourier transformation:

$$S(k) = \sum_{m=-\infty}^{\infty} s_m e^{-ikm} \quad (5)$$

$$s_m = \frac{1}{2\pi}\int_{-\pi}^{\pi} S(k) e^{ikm}\, dk. \quad (6)$$

Using an M-element ULA restricts our knowledge of s_m to the interval $m = 0, \cdots, M-1$.

We define the data spatial covariance matrix of the data as:

$$\mathbf{R} = E\left\{\vec{x}[n]\vec{x}^H[n]\right\} \quad (7)$$

When the signals of interest are uncorrelated, i.e. $E\left\{A_d A^*_{d+d'}\right\} = \sigma_d^2 \delta_{d'}$ or $\vec{v}_d^H \vec{v}_{d+d'} \propto \delta_{d'}$, the covariance matrix can be decomposed into the signal and noise covariance matrices:

$$\begin{aligned}\mathbf{R} &= \mathbf{R}_s + \mathbf{R}_n \\ \mathbf{R}_s &= \sum_{d=0}^{D-1} \sigma_d^2 \vec{v}_d \vec{v}_d^H \\ \mathbf{R}_n &= \sigma_n^2 \mathbf{I}.\end{aligned} \tag{8}$$

These covariance matrices have a Toeplitz structure. When $\mathbf{R}$ is not known, it must be estimated from the data. A common estimator is the sample covariance matrix:

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{n=0}^{N-1} \vec{x}[n] \vec{x}^H[n]. \tag{9}$$

B. Subspace-Based Methods for DOA Estimation

There are several non-parametric and parametric techniques for achieving our primary goal. The non-parametric methods are typically beamscan algorithms (see e.g. [4]), which form a spatial spectrum over all wavenumbers and locate the D largest peaks within. The most researched parametric methods are MUSIC [5] and ESPRIT [6], and we will concern ourselves with the former. Both these methods are subspace-based, and use the eigendecomposition of the spatial covariance matrix $\mathbf{R}$ to estimate the DOAs of the D signals:

$$\mathbf{R} = \mathbf{V} \Lambda \mathbf{V}^H. \tag{10}$$

The eigenvector matrix is divided into the orthogonal signal-plus-noise- and noise subspaces, where the former consists of the eigenvectors corresponding to the D largest eigenvalues:

$$\begin{aligned}&\Lambda = \mathrm{diag}\left\{\lambda_0, \cdots, \lambda_{M-1}\right\}, \lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{M-1} \\ &\mathbf{V} = [\vec{v}_0, \cdots, \vec{v}_{M-1}] = [\mathbf{V}_{S+N}, \mathbf{V}_N] \\ &\mathbf{V}_{S+N} = [\vec{v}_0, \cdots, \vec{v}_{D-1}], \mathbf{V}_N = [\vec{v}_D, \cdots, \vec{v}_{M-1}]\end{aligned} \tag{11}$$

MUSIC is based on forming the so-called null spectrum from the eigenvectors in the noise subspace:

$$Q_{MU}(k) = \vec{v}^H(k) \mathbf{V}_N \mathbf{V}_N^H \vec{v}(k) \tag{12}$$

The signal wavenumbers are estimated by finding the roots of the null spectrum, and the DOAs are derived from these wavenumbers using the relationship in (4).

C. Robust Covariance Matrix Estimation

In practice, the sample covariance matrix of (9) will not be identical to the ideal covariance matrix of (8). This can be due to different reasons. If $\hat{\mathbf{R}}$ is based on a finite number of temporal samples, the deviation from the ideal Toeplitz structure will typically increase as the number of samples N is decreased. If the signals of interest are correlated, the sample covariance matrix will always have a non-Toeplitz structure independent of N. An interesting problem is whether it is still possible to regain the Toeplitz covariance matrix of (8) despite these deviations. Several techniques have been suggested, the most well-known being spatial averaging [7]. Spatial averaging consists of dividing the array data vectors $\vec{x}[n]$ into K subarray data vectors $\vec{x}_k[n]$ of length $L = M - K + 1$ that overlap by all but one element:

$$\vec{x}_k[n] = [x_k[n], \cdots, x_{k+L-1}[n]]^T \in \mathbb{C}^L. \tag{13}$$

The spatially averaged covariance matrix is formed by averaging the subarray data vectors:

$$\hat{\mathbf{R}}_{SA} = \frac{1}{KN} \sum_{n=0}^{N-1} \sum_{k=0}^{K-1} \vec{x}_k[n] \vec{x}_k^H[n]. \tag{14}$$

Unfortunately, this matrix only achieves Toeplitz structure asymptotically as $K \to \infty$, which is an unrealistic assumption. Therefore, the alternative method of redundancy averaging was suggested to achieve Toeplitz structure for any value of K and N [2]. Redundancy averaging consists of replacing all elements of the covariance matrix by the average across the corresponding matrix diagonal:

$$[\hat{\mathbf{R}}_{RA}]_{m,n} = \frac{1}{M-P} \sum_{p=0}^{P-1} \left[\hat{\mathbf{R}}\right]_{p,p+P}, P = |m-n| \tag{15}$$

Redundancy averaging can also be derived as a least-squares structured covariance estimate [8]. Both spatial and redundancy averaging achieves some degree of robustness against above-mentioned model deviations, and they both have their advantages and disadvantages.

The main advantage of spatial averaging is that it preserves the underlying covariance matrix structure. Therefore, it converges to the true covariance matrix, i.e. $\lim_{K\to\infty} \hat{\mathbf{R}}_{SA} = \mathbf{R}$. The disadvantage is that increasing K decreases the degrees of freedom by the same amount, which means that fewer sources can be detected.

The advantage of redundancy averaging is that it can yield a fully Toeplitz matrix with full rank from as little as a single data vector without any reduction of the degrees of freedom. The disadvantage is that it does not preserve the underlying covariance matrix structure, and can therefore not be used to perfectly regain $\mathbf{R}$ except for certain exceptional scenarios. It has been shown that the bias induced in the DOA estimates from redundancy averaging can be rather large [3].

Although redundancy averaging has its merits, it is not obvious how it could be improved or modified to reduce the associated DOA bias. In the next section we present an alternative approach to covariance matrix Toeplitz completion, which is more flexible and can be interpreted as an extension of redundancy averaging.

III. Covariance Matrix Decorrelation Through Optimal Toeplitz Completion

Instead of deriving a covariance matrix estimator in the spatio-temporal domain, it can be advantageous to look at the problem in the wavenumber domain. An element of the signal covariance matrix from (8) can be written in the wavenumber domain through a Fourier transformation:

$$[\mathbf{R}_s]_{m,n} = \frac{1}{2\pi} \int_{-\pi}^{\pi} |S(k)|^2 e^{i(m-n)k} \, dk \tag{16}$$

The wavefield magnitude at wavenumber k_s, $|S(k_s)|^2$, is generally unknown, but can be estimated using beamforming:

$$\hat{S}(k_s) = \vec{w}^H \vec{x} \Rightarrow E\left\{\left|\hat{S}(k_s)\right|^2\right\} = \vec{w}^H \mathbf{R} \vec{w}. \tag{17}$$

The quality of the estimate depends on the properties of the beamformer. A simple example is the uniformly weighted delay-and-sum beamformer, which is given as:

$$\vec{w}_{DAS} = \frac{1}{M}\vec{v}(k_s). \tag{18}$$

We note that the expected power of the beamformer estimate in (17) can be written in the wavenumber domain:

$$\vec{w}^H \mathbf{R} \vec{w} = E\left\{\left|\frac{1}{2\pi}\int_{-\pi}^{\pi} W(k) S(k)\, dk\right|^2\right\}. \tag{19}$$

However, we want to select $\vec{w}$ to optimize our estimate of the wavefield. An optimum estimator of $|S(k)|^2$ is the Minimum Variance Distortionless Response (MVDR) beamformer, also known as the Capon beamformer, which is defined as:

$$\left|\hat{S}_{MV}(k_s)\right|^2 = \vec{w}_{MV}^H \mathbf{R} \vec{w}_{MV} \tag{20}$$
$$\text{where } \vec{w}_{MV} = \text{argmin}_{\vec{w}} \text{VAR}\left\{\vec{w}^H \vec{x}\right\} \text{ s.t. } \vec{w}^H \vec{v}(k_s) = 1.$$

Given our underlying assumption of an uncorrelated wavefield, (20) can be written in the wavenumber domain as in (19):

$$\vec{w}_{MV} = \text{argmin}_{\vec{w}} \frac{1}{2\pi}\int_{-\pi}^{\pi} |W(k)|^2 E\left\{|S(k)|^2\right\} dk \tag{21}$$
$$\text{subject to } W(k_s) = 1 \text{ and } w_m = 0 \text{ for } m < 0 \text{ or } m \geq M.$$

Since this expression depends on $|S(k)|^2$, the very quantity we are trying to estimate, we can define an iterative estimator of the uncorrelated wavefield (for steps $e = 1, 2, \cdots, \infty$):

$$\left|\hat{S}^{(e+1)}(k_s)\right|^2 = \vec{w}_{MV}^{(e+1)H} \hat{\mathbf{R}} \vec{w}_{MV}^{(e+1)} \tag{22}$$
$$\text{where } \vec{w}_{MV}^{(e+1)} = \text{argmin}_{\vec{w}} \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|W(k)\hat{S}^{(e)}(k)\right|^2 dk$$
$$\text{subject to } W(k_s) = 1 \text{ and } w_m = 0 \text{ for } m < 0 \text{ or } m \geq M.$$

It is well known that the solution of this MVDR optimization problem is:

$$\vec{w}_{MV}^{(e+1)} = \frac{\hat{\mathbf{R}}^{-(e)} \vec{v}(k)}{\vec{v}^H(k) \hat{\mathbf{R}}^{-(e)} \vec{v}(k)}, \tag{23}$$

where $\hat{\mathbf{R}}^{-(e)}$ is shorthand for $\left(\hat{\mathbf{R}}^{(e)}\right)^{-1}$.This estimate can be used with (16) to specify an iterative Toeplitz covariance matrix estimator of the associated uncorrelated wavefield:

$$\left[\hat{\mathbf{R}}^{(e+1)}\right]_{m,n} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\vec{v}^H(k)\hat{\mathbf{R}}^{-(e)}\hat{\mathbf{R}}\hat{\mathbf{R}}^{-(e)}\vec{v}(k)}{\left|\vec{v}^H(k)\hat{\mathbf{R}}^{-(e)}\vec{v}(k)\right|^2} e^{i(m-n)k}\, dk. \tag{24}$$

As for any other iterative algorithm, there is the question of initialization. A plausible initial estimate $\left|\hat{S}^{(0)}(k)\right|^2$ requiring no knowledge about the situation at hand could be the positive, constant spectrum:

$$\left|\hat{S}^{(0)}(k)\right|^2 = \alpha > 0. \tag{25}$$

However, this automatically yields a closed-form solution for the next step, namely the delay-and-sum beamformer of (18), as is shown in Appendix A. Therefore, it makes more sense to initialize with the delay-and-sum beamformer directly:

$$\left|\hat{S}^{(0)}(k)\right|^2 = \frac{1}{M^2}\vec{v}^H(k)\hat{\mathbf{R}}\vec{v}(k) \tag{26}$$

An interesting observation comes when inserting (26) into (22) and, subsequently, the result into (24) to estimate $\hat{\mathbf{R}}^{(1)}$, namely that the first covariance matrix estimate becomes (see Appendix B for a proof):

$$[\hat{\mathbf{R}}^{(1)}]_{m,n} = \frac{1}{M^2}\sum_{p=0}^{P-1}\left[\hat{\mathbf{R}}\right]_{p,p+P}, P = |m-n|. \tag{27}$$

This is equal to the redundancy averaged estimator of (15) except for the denominator of the leading fraction. As we will show in the next section, both this scaling and iterating further will improve the performance beyond that of redundancy averaging. However, since step 1 of the iteration still yields a closed form solution, it ultimately makes more sense to start with (27) instead of the conventional beamformer estimate.

Unfortunately, it is not trivial to prove the convergence of the iterative estimator in (24). In the next section, we will demonstrate convergence for a range of cases through simulations.

It is worth noting that the suggested method is more computationally demanding than both spatial and redundancy averaging. Each iteration requires the formation of an $M \times M$ Toeplitz matrix, where the value along each of the M diagonals is calculated from the integral in (24). As there is no closed form solution to this integral, it must be done numerically. The resulting covariance matrix must be then be inverted, which has complexity of $O(M^2)$ or less due to its Toeplitz structure.

IV. Simulations and Discussion

In this section we will compare the performance of redundancy averaging, spatial averaging, and the suggested method for DOA estimation using the Root-MUSIC algorithm.

We will first investigate the performance with respect to DOA estimation. We simulate a scenario with a ULA of $M = 10$ elements in the presence of two equally strong sources located at broadside (i.e. $\theta = 0$) and the half-power beamwidth angle (i.e. $\theta = 2\sin^{-1}\left(\frac{0.891}{M}\right)$). In Fig. 1 we show the DOA RMSE normalized by the null-to-null beamwidth of the full array for $N = 100$. In Fig. 2 we see the single-snapshot behavior.

We will also investigate the probability of resolution (POR) for the different methods. Resolution can be a tricky thing to define, especially for parametric, non-spectral methods like Root-MUSIC because the number of sources is defined beforehand and there is no continuous spectral curve to evaluate with respect to peaks and variations. There are several existing interpretations, but we have chosen to define that two sources are resolved if they are both detected within one (null-to-null) beamwidth from their actual positions. In Fig. 3 we show the POR as a function of SNR for $N = 100$.

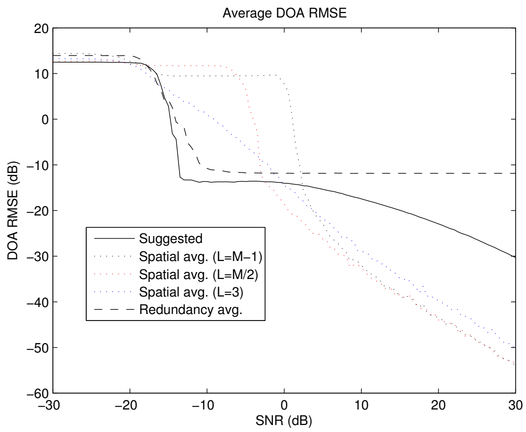

Fig. 1. DOA RMSE as a function of SNR. *Parameters:* $\theta_0 = 0$, $\theta_1 = 2\sin^{-1}(0.0891)$, $N = 100$, $M = 10$, $K = 2$, $E = 20$.

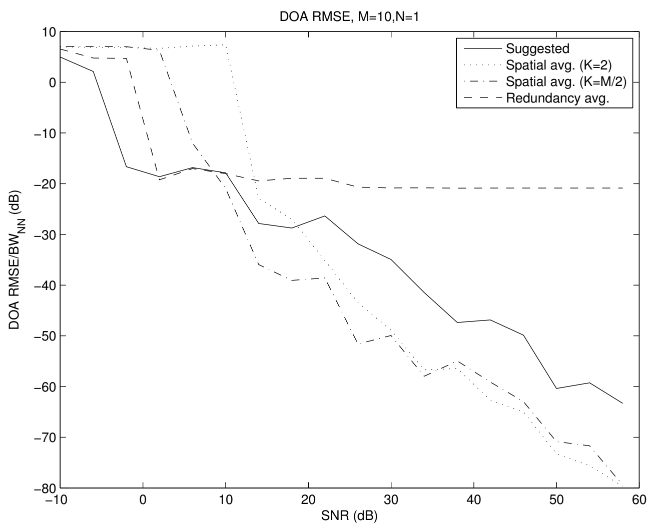

Fig. 2. DOA RMSE as a function of SNR, the single-snapshot case. *Parameters:* $\theta_0 = 0$, $\theta_1 = 2\sin^{-1}(0.0891)$, $N = 1$, $M = 10$, $K = 2$ and $M/2$, $E = 20$.

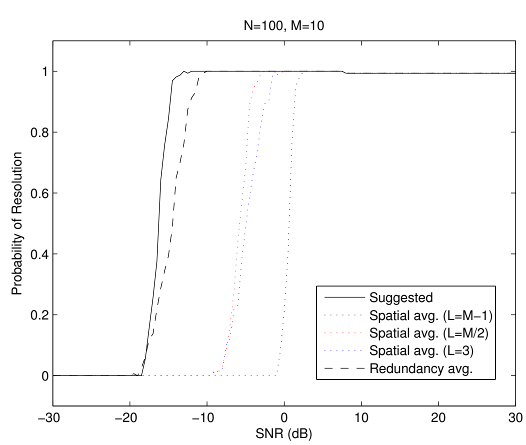

Fig. 3. Probability of resolution as a function of SNR. $N = 100, M = 10$.

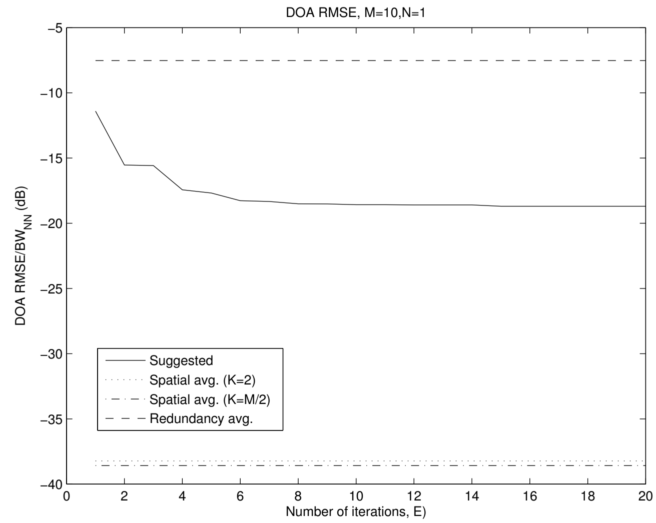

Fig. 4. DOA RMSE as a function of E. *Parameters:* $\theta_0 = 0$, $\theta_1 = 2\sin^{-1}(0.0891)$, $N = 100$, $M = 10$, $K = 2$ and $M/2$, $SNR = 20$dB.

An interesting observation is that the bias introduced by redundancy averaging has a more severe effect; the DOA RMSE is almost binary in nature and does not decrease in any significant manner after the threshold has been met. The threshold of spatial averaging occurs at a high SNR, and its performance improves very rapidly from here on. The threshold of the suggested method occurs first of all, and while its performance does not increase as quickly as spatial averaging, it is a significant improvement from redundancy averaging.

In Fig. X we show the DOA RMSE of the suggested method, compared to the others, as a function of the number of iterations E for an SNR of 20dB with $N = 100$. By observing the behavior of the suggested method as a function of E, we see that convergence is achieved approximately around $E = 10$. We also see that even for $e = 1$, a significant improvement over redundancy averaging can be achieved.

Summary of conclusions:

- The suggested method is always better than redundancy averaging, except for when there is both low correlation ($\rho \leq 0.4$) and small angles.
- Redundancy averaging and the suggested method are better than spatial averaging for small separation angles or or low-to-medium SNR; i.e. both redundancy averaging and the suggested method have lower performance thresholds and better resolution than spatial averaging.
- Spatial averaging is always superior for high SNR and large separation angles.

V. Conclusion

We have introduced a covariance matrix estimator for use in parametric DOA estimation, which achieves robustness against low sample count and correlation through optimal Toeplitz completion. The estimator is iterative, and we have shown that its first step yields a slightly modified redundancy average estimator, while the subsequent steps result in improved DOA estimates and resolution.

APPENDIX A: PROOF OF EQUATION (26)

Because the covariance matrix of (16) is Toeplitz, we will investigate the value along the l^{th} diagonal:

$$\begin{aligned}\left[\hat{\mathbf{R}}^{(1)}\right]_{m,n} &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\hat{S}^{(0)}(k)\right|^2 e^{ilk}\,dk,\\ &= \frac{\alpha}{2\pi}\int_{-\pi}^{\pi} e^{ilk}\,dk = \alpha\delta_l \Rightarrow \hat{\mathbf{R}}^{-(1)} = \alpha^{-1}\mathbf{I}\end{aligned} \tag{28}$$

Inserting (28) into (23) yields:

$$\vec{w}^{(1)} = \frac{1}{M}\vec{v}(k_s) \tag{29}$$

APPENDIX B: PROOF OF EQUATION (27)

Because the covariance matrix of (16) is Toeplitz, we will investigate the value along the l^{th} diagonal:

$$\begin{aligned}\left[\hat{\mathbf{R}}^{(1)}\right]_{m,n} &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\hat{S}^{(0)}(k)\right|^2 e^{ilk}\,dk,\\ &\text{for } l \triangleq m-n\\ &= \frac{1}{2\pi M^2}\int_{-\pi}^{\pi}\vec{v}^H(k)\hat{\mathbf{R}}\vec{v}(k)e^{ilk}\,dk\\ &= \frac{1}{2\pi T M^2}\sum_{n=0}^{T-1}\sum_{m'=0}^{M-1}\sum_{n'=0}^{M-1} x_{m'}[t]x^*_{n'}[t]\int_{-\pi}^{\pi} e^{i(l-l')k}\,dk,\\ &\text{for } l' \triangleq m'-n'\\ &= \frac{1}{M^2}\sum_{m'=0}^{M-1}\sum_{n'=0}^{M-1}\delta_{l-l'}\frac{1}{T}\sum_{t=0}^{T-1} x_{m'}[t]x^*_{n'}[t]\\ &= \frac{1}{M^2}\sum_{p=0}^{l-1}\left[\hat{\mathbf{R}}\right]_{p,p+l}.\end{aligned} \tag{30}$$

The resulting expression is equal to (27) and nearly equal to the redundancy averaged estimate of the covariance matrix as given in (15), except for the denominator in the leading fraction.

REFERENCES

[1] K. Takao and N. Kikuma, "An adaptive array utilizing an adaptive spatial averaging technique for multipath environments," *Antennas and Propagation, IEEE Transactions on*, vol. 35, no. 12, pp. 1389 – 1396, dec 1987.

[2] L. Godara, "Beamforming in the presence of correlated arrivals using structured correlation matrix," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 38, no. 1, pp. 1 –15, jan 1990.

[3] K. Indukumar and V. Reddy, "A note on redundancy averaging," *Signal Processing, IEEE Transactions on*, vol. 40, no. 2, pp. 466 –469, feb 1992.

[4] H. V. Trees, *Optimum Array Processing*. Wiley-Interscience, 2002.

[5] R.O.Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation, Vol. AP-34, pp. 276–280*, 1986.

[6] R. Roy, "Esprit: Estimation of signal parameters via rotational invariance techniques)," *Ph.D. Dissertation, Stanford University, Stanford,California*, 1987.

[7] T.-J. Shan and T. Kailath, "Adaptive beamforming for coherent signals and interference," *IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. ASSP-33, pp. 527–534*, 1985.

[8] L. L. Scharf, *Statistical signal processing: detection, estimation, and time series analysis*.