AN EM-ALGORITHM FOR BAND-TOEPLITZ COVARIANCE MATRIX ESTIMATION

Lars P. B. Christensen

Technical University of Denmark
Informatics and Mathematical Modelling
email: lc@imm.dtu.dk

ABSTRACT

Toeplitz covariance matrix estimation has many uses in statistical signal processing due to the stationarity assumption of many signals. For some applications, further constraints may exist on the maximum lag at which the correlation function is non-zero and thereby giving rise to a band-Toeplitz covariance matrix. In this paper, an existing EM-algorithm for Toeplitz estimation is generalized to the case of band-Toeplitz estimation. In addition, the Cramer-Rao lower-bound for unbiased band-Toeplitz covariance matrix estimation is derived and through simulations it is shown that the proposed estimator achieves the bound for medium and large sample-sizes.

***Index Terms*—** Structured covariance matrix estimation, banded Toeplitz, EM-algorithm, Cramer-Rao lower-bound.

1. INTRODUCTION

Estimation of Toeplitz covariance matrices is inherently connected to signal processing of stationary processes and applications are numerous, e.g. communications and radar systems. However, the constraint of stationarity and its resulting requirement for a Toeplitz structure in the covariance matrix, makes Maximum-Likelihood (ML) estimation challenging and no general closed-form solution is known [1, 2]. In [1], an EM-algorithm for Toeplitz covariance matrix estimation is constructed by exploiting a circulant extension of the Toeplitz matrix and the idea is further generalized to Block-Toeplitz in [3].

The contribution of this paper is to generalize the idea of using an EM-algorithm based on a circulant extension to covariance matrices that are not only Toeplitz, but band-Toeplitz with bandwidth B, i.e. having non-zero correlations only up to and including lag B. Such estimates are important in many practical applications as it is often reasonable to set an upper limit on the maximum lag of the estimate due to properties of the system considered. The proposed method therefore bridges the gap in correlation estimation from simple power estimation ($B = 0$) to full Toeplitz covariance matrix estimation. A natural generalization to block-banded block-Toeplitz matrices exists following [3], but this is outside the scope of this paper.

The author would like to thank Nokia Denmark for funding

Section 2 presents the system model and preliminaries and in Section 3 the EM-algorithm for Toeplitz covariance matrix estimation is outlined. Next, Section 4 modifies the existing M-step of the EM-algorithm to allow for constrained band-Toeplitz estimation. Section 5 derives the Cramer-Rao lower-bound for band-Toeplitz estimation and Section 6 outlines a traditional linear estimator used for comparison. Finally, Section 7 presents a numerical example and conclusion.

In the following, bold letters such as $\mathbf{x}$ and $\mathbf{X}$ are respectively column vectors and matrices with x_i and $[\mathbf{X}]_{i,j}$ being a specific scalar element with the indices starting at zero. Further, $(\cdot)^T$ indicates matrix transpose, $(\cdot)^*$ matrix conjugation, $(\cdot)^H \triangleq \left((\cdot)^*\right)^T$ Hermitian transpose and $tr\left\{\cdot\right\}$ the trace operator. The notation $|\cdot|$ indicates the determinant of a matrix or the absolute value of a scalar. Finally, $diag\left(\cdot\right)$ constructs a diagonal matrix from a vector or, if operating on a matrix, produces a vector from the diagonal elements of the matrix.

2. SYSTEM MODEL AND PRELIMINARIES

Let $\mathcal{X} \triangleq \{\mathbf{x}_k\}_{k=1}^K$ be a collection of independent realizations of a zero-mean circular complex Gaussian distribution $\mathbf{x}_k \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{\Sigma}_x\right)$ with $\mathbf{x}_k \in \mathbb{C}^{N_x}$. The distribution of the observations is therefore given by

$$\begin{aligned} -ln\left(p\left(\mathcal{X} \mid \mathbf{\Sigma}_x\right)\right) - Kln\left|\pi\mathbf{\Sigma}_x\right| &= \sum_{k=1}^{K} \mathbf{x}_k^H \mathbf{\Sigma}_x^{-1} \mathbf{x}_k \\ &= Ktr\left\{\mathbf{\Sigma}_x^{-1}\mathbf{S}_x\right\} \end{aligned} \quad (1)$$

with $\mathbf{S}_x \in \mathbb{C}^{N_x \times N_x}$ being the sample covariance matrix

$$\mathbf{S}_x \triangleq \frac{1}{K} \sum_{k=1}^{K} \mathbf{x}_k \mathbf{x}_k^H \quad (2)$$

Also define $\mathbf{c} \in \mathbb{C}^{B+1}$ as the top $B+1$ terms of the first column in $\mathbf{\Sigma}_x$ and let $\mathbf{r} \in \mathbb{R}^{2B+1}$ be the stacking of the real and imaginary part of $\mathbf{c}$ with the real parts in the top of the vector. The length of $\mathbf{r}$ is only $2B+1$ as the imaginary part of c_0 must be zero.

Given the observations $\mathcal{X}$, the task is to provide the constrained ML estimate of $\mathbf{\Sigma}_x$. If no such constraints existed, the ML solution is simply given by $\mathbf{S}_x$.

Let $\hat{\mathbf{\Sigma}}_x^{(k)} \in \mathbb{C}^{N_x \times N_x}$ be an estimate of $\mathbf{\Sigma}_x$ at iteration k that obeys the Toeplitz constraint. Now, let $\hat{\mathbf{\Sigma}}_y^{(k)} \in \mathbb{C}^{N_y \times N_y}$ be a circularly extended version of $\hat{\mathbf{\Sigma}}_x^{(k)}$ where $N_y \geq 2N_x - 1$ makes a circulant extension possible. It is well-known that the Discrete Fourier Transform (DFT) diagonalizes any circulant matrix, i.e.

$$\hat{\mathbf{\Sigma}}_y^{(k)} = \mathbf{D}^H \hat{\mathbf{\Lambda}}^{(k)} \mathbf{D} \tag{3}$$

with $[\mathbf{D}]_{i,j} = N_y^{-\frac{1}{2}} e^{-\frac{2\pi\sqrt{-1}}{N_y} ij}$ being the normalized DFT matrix of size N_y. The diagonal matrix $\hat{\mathbf{\Lambda}}^{(k)} \triangleq diag\left(\hat{\boldsymbol{\lambda}}^{(k)}\right)$ holds the eigenvalues given by

$$\hat{\boldsymbol{\lambda}}^{(k)} \triangleq N_y^{\frac{1}{2}} \mathbf{D} \hat{\boldsymbol{\sigma}}_y^{(k)} \tag{4}$$

where $\hat{\boldsymbol{\sigma}}_y^{(k)} \triangleq \left[\hat{\mathbf{\Sigma}}_y^{(k)}\right]_{:,0}$ is the first column of the circulant matrix. The eigenvalues of (4) contain all the information about the underlying Toeplitz matrix and we therefore have

$$\hat{\mathbf{\Sigma}}_x^{(k)} = \tilde{\mathbf{D}}^H \hat{\mathbf{\Lambda}}^{(k)} \tilde{\mathbf{D}} \tag{5}$$

with $\tilde{\mathbf{D}} \triangleq \mathbf{D}\left[\mathbf{I}_{N_x} \quad \mathbf{0}\right]^T$. If the desired covariance estimate is of size $M \leq N_x$, the result is given by the upper left submatrix of $\hat{\mathbf{\Sigma}}_x^{(k)}$.

3. EM-ALGORITHM FOR TOEPLITZ ESTIMATION

Here, the EM-algorithm applied to the problem of Toeplitz covariance matrix estimation is briefly outlined as described in [1] with [2] providing an efficient implementation. The E-step can be expressed as

$$E: \quad \mathbf{\Delta}^{(k)} \triangleq \hat{\mathbf{\Lambda}}^{(k)} \tilde{\mathbf{D}} \left(\mathbf{W}^{(k)} \mathbf{S}_x \mathbf{W}^{(k)} - \mathbf{W}^{(k)}\right) \tilde{\mathbf{D}}^H \hat{\mathbf{\Lambda}}^{(k)} \tag{6}$$

with $\mathbf{W}^{(k)} \triangleq \left(\hat{\mathbf{\Sigma}}_x^{(k)}\right)^{-1}$ and $\mathbf{\Delta}^{(k)}$ being the unconstrained update to the complete-data sample covariance matrix $\mathbf{S}_y$. The M-step should now choose the complete-data ML covariance estimate fulfilling the structural constraints based on the sufficient statistic $\mathbf{S}_y$. As the estimate is known to be Toeplitz, meaning that the update must be a diagonal matrix, it is straight-forward to show that the constrained ML update is exactly the diagonal of $\mathbf{\Delta}^{(k)}$, i.e.

$$M: \quad \hat{\mathbf{\Sigma}}_x^{(k+1)} = \tilde{\mathbf{D}}^H diag\left(\underbrace{\hat{\boldsymbol{\lambda}}^{(k)} + diag\left(\mathbf{\Delta}^{(k)}\right)}_{\hat{\boldsymbol{\lambda}}^{(k+1)}}\right) \tilde{\mathbf{D}} \tag{7}$$

As the EM-algorithm is only guaranteed to converge to a local maximum in the complete-data likelihood function, initialization is important. A reasonable choice of initialization, which is used throughout this paper, is $\hat{\mathbf{\Sigma}}_x^{(0)} = N_x^{-1} tr\left\{\mathbf{S}_x\right\} \mathbf{I}_{N_x}$.

4. A MODIFIED M-STEP FOR BAND-TOEPLITZ ESTIMATION

Assuming it is known apriori that the covariance matrix $\mathbf{\Sigma}_x$ is band-Toeplitz with bandwidth B, the idea is now to constrain the covariance estimate by requiring a functional form of $\hat{\boldsymbol{\lambda}}^{(k)}$ that guarantee this constraint. From (4) we have

$$\hat{\boldsymbol{\sigma}}_y^{(k)} = N_y^{-\frac{1}{2}} \mathbf{D}^H \hat{\boldsymbol{\lambda}}^{(k)} \tag{8}$$

so a set of eigenvalues $\hat{\boldsymbol{\lambda}}_{BT}^{(k)}$ fulfilling the structural constraint, must decompose as

$$\begin{aligned} \hat{\boldsymbol{\lambda}}_{BT}^{(k)} &= N_y^{\frac{1}{2}} \sum_{b=-B}^{B} \hat{c}_b^{(k)} \mathbf{d}_b \qquad s.t. \quad \hat{c}_b^{(k)} = \left(\hat{c}_{-b}^{(k)}\right)^* \\ &= N_y^{\frac{1}{2}} \left(\hat{c}_0^{(k)} \mathbf{d}_0 + 2\sum_{b=1}^{B} Re\left\{\hat{c}_b^{(k)} \mathbf{d}_b\right\}\right) \end{aligned} \tag{9}$$

where $\mathbf{d}_b = \mathbf{d}_{-b}^* \triangleq [\mathbf{D}]_{:,mod(b,N_y)}$ with $mod\,(x,y)$ meaning x modulo y. The functional form of the eigenvalues in (9) effectively forces the covariance estimate to be $\hat{\mathbf{c}}^{(k)} \triangleq \left[\hat{c}_0^{(k)}, \ldots, \hat{c}_B^{(k)}\right]^T$ for the non-zero band and zero elsewhere. However, as the eigenvalues must be real-valued and there are only $N_b \triangleq 2B+1$ real-valued degrees-of-freedom in the decomposition, we choose to reformulate the constraint as a real-valued decomposition, i.e.

$$\hat{\boldsymbol{\lambda}}_{BT}^{(k)} = \mathbf{T}\hat{\mathbf{r}}^{(k)} \tag{10}$$

Here $\hat{\mathbf{r}}^{(k)} \in \mathbb{R}^{N_b}$ are the unknowns and $\mathbf{T} \triangleq [\mathbf{t}_0, \ldots, \mathbf{t}_{N_b-1}]$ is defined by

$$\mathbf{t}_b \triangleq \begin{cases} N_y^{\frac{1}{2}} Re\left\{\mathbf{d}_b\right\} & , b = 0 \\ 2N_y^{\frac{1}{2}} Re\left\{\mathbf{d}_b\right\} & , 1 \leq b \leq B \\ -2N_y^{\frac{1}{2}} Im\left\{\mathbf{d}_{b-B}\right\} & , B < b < N_b \end{cases} \tag{11}$$

The upper $B+1$ coefficients of $\hat{\mathbf{r}}^{(k)}$ will therefore hold the real part of $\hat{\mathbf{c}}^{(k)}$ while the lower B coefficients are the imaginary part. In the case of real-valued covariance estimation, having $N_b \triangleq B+1$ is therefore sufficient to parameterize the constrained estimate.

The reader should now be familiar with the overall structure of the EM-algorithm for band-Toeplitz estimation and the iteration index is therefore dropped for notational ease in the following. The challenge is now, given the current ML estimate over the space of circulant matrices $\hat{\boldsymbol{\lambda}}$, to minimize some distance measure $f\left(\hat{\boldsymbol{\lambda}}, \hat{\mathbf{r}}\right)$ between the Toeplitz and the band-Toeplitz estimate subject to the constraint that the resulting eigenvalues must all be non-negative, i.e. $\mathbf{Tr} \geq \mathbf{0}$. As the EM-algorithm proceeds, this will result in a successive tightening of the lower-bound on the marginal log-likelihood determined in the E-step.

4.1. ML Estimation

Using the complete-data negative log-likelihood as a distance measure, we maximize the lower-bound over the space of valid band-Toeplitz matrices. Letting $\left[\tilde{\mathbf{t}}_0, \ldots, \tilde{\mathbf{t}}_{N_y-1}\right] \triangleq \mathbf{T}^T$, the desired distance measure can be written as

$$f\left(\hat{\boldsymbol{\lambda}}, \hat{\mathbf{r}}\right) = \sum_{i=0}^{N_y-1} ln\left(\tilde{\mathbf{t}}_i^T \hat{\mathbf{r}}\right) + \left(\tilde{\mathbf{t}}_i^T \hat{\mathbf{r}}\right)^{-1} \hat{\lambda}_i \qquad (12)$$

As the distance measure consists of a sum of a concave and a convex term in the unknowns, the overall function is non-convex and thereby making global minimization unfeasible. Instead, the first- and second-order derivatives

$$\frac{\partial f\left(\hat{\boldsymbol{\lambda}}, \hat{\mathbf{r}}\right)}{\partial \hat{\mathbf{r}}} = \sum_{i=0}^{N_y-1} \frac{1}{\tilde{\mathbf{t}}_i^T \hat{\mathbf{r}}} \left(1 - \frac{\hat{\lambda}_i}{\tilde{\mathbf{t}}_i^T \hat{\mathbf{r}}}\right) \tilde{\mathbf{t}}_i \qquad (13)$$

$$\frac{\partial^2 f\left(\hat{\boldsymbol{\lambda}}, \hat{\mathbf{r}}\right)}{\partial \hat{\mathbf{r}} \partial \hat{\mathbf{r}}^T} = -\sum_{i=0}^{N_y-1} \frac{1}{\left(\tilde{\mathbf{t}}_i^T \hat{\mathbf{r}}\right)^2} \left(1 - \frac{2\hat{\lambda}_i}{\tilde{\mathbf{t}}_i^T \hat{\mathbf{r}}}\right) \tilde{\mathbf{t}}_i \tilde{\mathbf{t}}_i^T \qquad (14)$$

can be used in any favorite optimization scheme to determine a local minimum of the distance measure. As the distance measure may have multiple minima, the search should be started at the previous value of $\hat{\mathbf{r}}$ to make sure that the update cannot increase the distance measure.

4.2. Other Distance Measures

Instead of minimizing the negative log-likelihood function directly other criterions can also be used. However, for the EM-algorithm to converge an update must not increase (12), but it is not required to minimize it either. In this manner, it is possible to formulate an entire family of Generalized EM-algorithms for band-Toeplitz covariance estimation. An example of this strategy would be to not minimize (12), but only find an update that lowers it and thereby trade convergence speed for reduced computational complexity in the M-step.

5. LOWER-BOUND FOR BAND-TOEPLITZ COVARIANCE ESTIMATION

The Cramer-Rao Lower-Bound (CRLB) provides the lowest possible error variance of any estimator and is therefore a natural performance benchmark. Determining the bound for a biased estimator involves computing the bias-function of the estimator, which in general appears unfeasible. Instead, the Unbiased CRLB (U-CRLB) is derived and used for comparison, as it is well-known that the ML estimate is asymptotically unbiased.

Following the derivation in [4] and modifying it to include the complex-valued observations, the Fisher information matrix $\mathbf{J} \in \mathcal{R}^{N_b \times N_b}$ for the constrained covariance estimate can be found to be

$$[\mathbf{J}]_{i,j} = Ktr\left\{\boldsymbol{\Sigma}_x^{-1} \frac{\partial \boldsymbol{\Sigma}_x}{\partial r_i} \boldsymbol{\Sigma}_x^{-1} \frac{\partial \boldsymbol{\Sigma}_x}{\partial r_j}\right\} \qquad (15)$$

and using (5) and (10) we readily get

$$\frac{\partial \boldsymbol{\Sigma}_x}{\partial r_i} = \tilde{\mathbf{D}}^H diag\left(\mathbf{t}_i\right) \tilde{\mathbf{D}} \qquad (16)$$

As the focus is on the U-CRLB, the desired lower-bound is

$$E\left[\left|r_i - \hat{r}_i\right|^2\right] \geq \left[\mathbf{J}^{-1}\right]_{i,i} \qquad (17)$$

6. WEIGHTED PROJECTED COVARIANCE ESTIMATION

This section outlines a simple method of performing Toeplitz covariance matrix estimation based on the idea of [5] in order to better understand the EM-based approach and provide a benchmark. The idea is to simply average along the diagonals of the sample covariance to estimate the correlations for the desired lags. However, to guarantee a positive definite matrix, the lag m correlation estimate is weighted by $\frac{N_x - m}{N_x}$ with $m \geq 0$. A valid Toeplitz covariance matrix of size M can now be constructed from the weighted correlation coefficients $\left[\hat{c}_0, \ldots, \hat{c}_{M-1}\right]^T$ resulting in a bias given by

$$E\left[c_m - \hat{c}_m\right] = \frac{m}{N_x} c_m \qquad (18)$$

The Mean-Squared Error (MSE) of the correlation estimates can be found by expressing fourth-order moments of the Gaussian observations by their second-order moments as

$$E\left[\left|c_m - \hat{c}_m\right|^2\right] = \frac{m^2}{N_x^2}\left|c_m\right|^2 + \frac{1}{KN_x^2} \sum_{b=-B}^{B} \left(N_x - m - |b|\right) \left|c_{|b|}\right|^2 \qquad (19)$$

The result of (19) is only valid for the complex-valued domain, but a similar result can be obtained for the real-valued domain by following the same principle. However, the proof of (19) and its real-valued equivalent are left out due to lack of space.

Although this sub-optimal method of covariance estimation can only provide full Toeplitz matrix estimates, it is related to band-Toeplitz estimation in the sense that one can choose $M = B + 1$ to produce an estimate of the non-zero correlation coefficients. Using the resulting estimate to produce a larger band-Toeplitz matrix, e.g. of size N_x, is however not guaranteed to be positive definite.

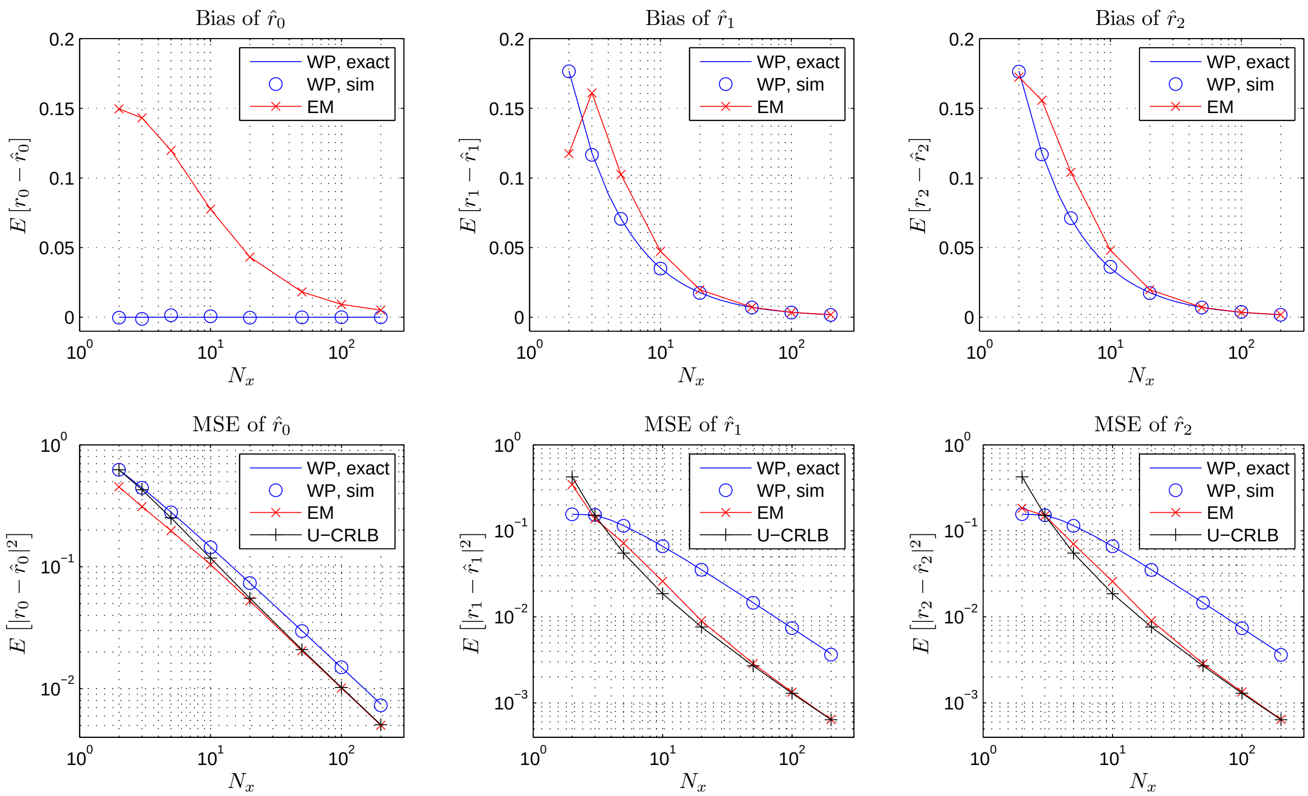

Fig. 1. Example of band-Toeplitz covariance matrix estimation as a function of the sample-size N_x for $B = 1$, $K = 1$.

7. NUMERICAL EXAMPLE AND CONCLUSION

To demonstrate the proposed method, complex-valued zero-mean white Gaussian noise with unit power is filtered by a first-order FIR filter having coefficients $\left[\frac{1}{\sqrt{2}}, \frac{1+\sqrt{-1}}{2}\right]^T$. This results in a band-Toeplitz covariance matrix $(B = 1)$ having the first column specified by $\mathbf{r} = \left[1, \frac{1}{\sqrt{8}}, \frac{1}{\sqrt{8}}\right]^T$. In figure 1, the bias and MSE of the Weighted Projected (WP) estimator and the proposed EM-based method (EM) is depicted as a function of the sample-size N_x. For WP, both exact results found by (19) and simulations are shown. The lower part of the figure depicts the MSE of the estimates and the U-CRLB is also shown for reference. It can be seen that a significant reduction in MSE is achievable by incurring a moderate bias increase with the exception of lag zero where WP is unbiased. Furthermore, as the sample-size increases, the proposed estimator becomes unbiased and tends to the U-CRLB. In conclusion, the proposed EM-based estimator provides near-optimal performance with a reasonable complexity due to its effective implementation exploiting the band-Toeplitz structure [2]. These properties make the proposed method an interesting candidate for many applications where accurate band-Toeplitz covariance matrix estimation is of great importance.

8. REFERENCES

[1] M. I. Miller and D. L. Snyder, "The role of likelihood and entropy in incomplete-data problems: Applications to estimating point-process intensities and toeplitz constrained covariances," *Procedings of the IEEE*, vol. 75, pp. 892–907, 1987.

[2] A. Dembo, C. L. Mallows, and L. A. Shepp, "Embedding nonnegative definite Toeplitz Matrices in nonnegative definite circulant matrices, with application to covariance estimation," *IEEE Trans. on Information Theory*, vol. 35, no. 6, pp. 1206–1212, Nov. 1989.

[3] D. R. Fuhrmann and T. A. Barton, "Estimation of block-toeplitz covariance matrices," in *24th IEEE Asim. Conf. on Signals, Systems and Computers.*, 1990, pp. 779–783.

[4] B. Porat and B. Friedlander, "Computation of the exact information matrix of Gaussian time series with stationary random components," *IEEE Trans. on Accoustics, Speech and Signal Proc.*, vol. 34, pp. 118–130, 1986.

[5] T. A. Barton and S. T. Smith, "Structured covariance estimation for space-time adaptive processing," in *Procedings of the IEEE ICASSP*, 1997, pp. 3493–3496.