Maximum Likelihood Estimation of Toeplitz-Block-Toeplitz Covariances in the presence of Subspace Interference

F. C. Nicolls and G. de Jager
Digital Image Processing Laboratory
Department of Electrical Engineering
University of Cape Town
Rondebosch 7701, South Africa
nicolls@dip.ee.uct.ac.za, gdj@dip.ee.uct.ac.za

Abstract

The EM algorithm is a commonly cited solution in the literature for the problem of maximum likelihood estimation of covariance matrices under a Toeplitz constraint. In this paper, the solution is extended to the case of two-dimensional signals, where spatial stationarity enforces a Toeplitz-block-Toeplitz structure on the covariance matrix.

A further generalisation which is presented involves the estimation of the covariance when the observations are subject to subspace interference. It is shown that this situation is amenable to a missing data interpretation, and can be incorporated into the EM iteration with moderate ease. The solution shares all the characteristics of the 1-D Toeplitz estimate.

The need to solve this problem arises in many invariance applications, where it is required to fit a stationary multivariate normal model to data which is subject to a certain type of interference. The case of unknown DC offset is included in this class.

1. Introduction

This paper discusses a method of estimating the covariance matrix of a MVN random process when each data observation has an additive contribution which lies in a known linear subspace but is otherwise arbitrary. A constraint placed on the estimation is that the covariance have a Toeplitz-block-Toeplitz (TBT) structure, corresponding to a stationary assumption in two dimensions.

By far the most common way of calculating maximum likelihood estimates of Toeplitz-structured covariance matrices is by means of the expectation-maximisation (EM) algorithm of Dempster, Laird, and Rubin [1], as reported by Miller, et al. [4, 5]. We present a modified EM iteration which handles the case of Toeplitz-block-Toeplitz covariances, as well as subspace interference. Naturally the solution specialises to the 1-D case of the covariance matrix simply being Toeplitz.

This work has direct bearing on statistical modelling of image-like data. The problem of radar target detection in clutter is one such an application. Detection of tumours in mammography is another.

2. Subspace interference

Two-dimensional observations need to be reordered into vector form. If row or column-ordering is assumed [2, p. 23], stationarity implies that covariance matrices are TBT (block Toeplitz with Toeplitz blocks).

It is assumed that the covariance of the random process $\mathbf{X} : N[\mathbf{0}, \mathbf{T}]$, with $\mathbf{T}$ a TBT matrix, needs to be estimated. However, we cannot observe realisations of this process directly; each observation is contaminated by subspace interference. Thus if $\mathbf{x}_1, \ldots, \mathbf{x}_m$ are independent realisations of this process, then the observed data are $\mathbf{y}_1, \ldots, \mathbf{y}_m$ with

$$\mathbf{y}_j = \mathbf{x}_j + \mathbf{U}_I \mathbf{c}_j. \tag{1}$$

In this equation $\mathbf{U}_I$ is a matrix which spans the interference subspace, and for convenience we may assume that $\mathbf{U}_I^T \mathbf{U}_I = \mathbf{I}$. The i-dimensional vector $\mathbf{c}_j$ is a completely unknown constant which differs for each observation.

This problem is amenable to a missing data interpretation: the component of $\mathbf{X}$ which lies in the interference subspace is destroyed by the unknown $\mathbf{c}_j$, and is therefore useless for inferential purposes. Thus if $\mathbf{U}_H$ is a matrix with orthogonal columns which span the subspace complementary to $\mathbf{U}_I$, only the component $\mathbf{y}_j^e = \mathbf{U}_H^T \mathbf{x}_j = \mathbf{U}_H^T \mathbf{y}_j$ of the observation is valid for estimating the parameters in the distribution of $\mathbf{X}$.

3. Maximum likelihood parameter estimation

The portion of the data that is uncorrupted by interference is $\mathbf{Y}^e : N[\mathbf{0}, \mathbf{U}_H^T \mathbf{T} \mathbf{U}_H]$. It is required to estimate $\mathbf{T}$ from m samples of this quantity, under the constraint that $\mathbf{T}$ be Toeplitz-block-Toeplitz. If m is large, maximum likelihood estimation is approximately optimal.

The EM algorithm is commonly used for maximum likelihood estimation of Toeplitz covariances. It is an iterative method whereby a difficult parametric optimisation problem is embedded inside a higher-dimensional but computationally more tractable one [1]. This is an ideal formulation for the problem outlined here: the hypothetical complete data observations are $\mathbf{z}_j : N[\mathbf{0}, \mathbf{C}]$, with $\mathbf{C}$ a CBC matrix representing the parameters to be optimised over, and the actual useful observations $\mathbf{y}_j^e$ take the role of the incomplete data. The embedding is such that the unobserved interference-free data $\mathbf{x}_j$ (reordered from a $r \times s$ observation) is related to the complete data $\mathbf{z}_j$ (reordered from a $u \times v$ observation) by

$$\mathbf{x}_j = (\mathbf{I}_{r \times u} \otimes \mathbf{I}_{s \times v}) \mathbf{z}_j. \tag{2}$$

Here $\otimes$ represents the matrix Kronecker product, $\mathbf{I}_{j \times k}$ is a $j \times k$ identity matrix of zeros with ones along the main diagonal, $\mathbf{C}$ has $u \times u$ blocks each of dimension $v \times v$, and $\mathbf{T}$ has $r \times r$ blocks each of dimension $s \times s$. The useful observations $\mathbf{y}_j^e$ are related to the complete data $\mathbf{z}_j$ by

$$\mathbf{y}_j^e = \mathbf{U}_H^T \mathbf{x}_j = \mathbf{U}_H^T (\mathbf{I}_{r \times u} \otimes \mathbf{I}_{s \times v}) \mathbf{z}_j. \tag{3}$$

The reason for the EM algorithm being effective in this problem is because a CBC matrix is very easily diagonalised.

The method of solution redefines the problem slightly: instead of maximising the likelihood over the set of all TBT matrices, the maximisation is performed over the set of all matrices with positive definite circulant-block-circulant (CBC) extensions. This is the 2-D analogue of the the standard 1-D Toeplitz formulation, found for example in [4]. The covariance matrix $\mathbf{T}$ is obtained from the corresponding complete data circulant covariance $\mathbf{C}$ by

$$\mathbf{T} = (\mathbf{I}_{r \times u} \otimes \mathbf{I}_{s \times v}) \mathbf{C} (\mathbf{I}_{r \times u} \otimes \mathbf{I}_{s \times v})^T. \tag{4}$$

A notable feature of the EM algorithm is its use of a missing data formalism to arrive at the required solution. In the previous section is was demonstrated that subspace interference is also conducive to a missing data interpretation. This presents further justification for using the EM technique.

4. EM formulation of solution

The quantity $\mathbf{Y}^e = \mathbf{U}_H^T \mathbf{X}$ is all that is observed of the hypothetical uv-dimensional complete data $\mathbf{Z} : N[\mathbf{0}, \mathbf{C}]$, where $\mathbf{C}$ is a circulant-block-circulant matrix. It is simpler to consider the problem in a rotated coordinate system where the covariance matrix is diagonalised.

Let $\mathbf{W} = \mathbf{W}_u \otimes \mathbf{W}_v$, where $\mathbf{W}_u$ and $\mathbf{W}_v$ are the u and v-dimensional unitary DFT matrices. It can be shown that this matrix diagonalises the class of all circulant-block-circulant matrices with $u \times u$ blocks each of dimension $v \times v$ [2, p. 150]. The transformed complete data $\mathbf{D} = \mathbf{W}\mathbf{Z}$ is therefore distributed as $\mathbf{D} : N[\mathbf{0}, \mathbf{\Sigma}]$, with $\mathbf{\Sigma} = \mathbf{W}\mathbf{C}\mathbf{W}^\dagger = \mathrm{diag}[\sigma_1^2, \ldots, \sigma_{uv}^2]$ a diagonal matrix comprised of the eigenvalues of $\mathbf{C}$. The log-likelihood in this rotated coordinate system is

$$\begin{aligned} L(\mathbf{\Sigma}, \mathbf{d}_1, \ldots, \mathbf{d}_m) &= K - \frac{m}{2} \log |\mathbf{\Sigma}| - \frac{1}{2} \sum_{j=1}^{m} \mathbf{d}_j^\dagger \mathbf{\Sigma}^{-1} \mathbf{d}_j \\ &= K - \frac{m}{2} \sum_{k=1}^{uv} \log \sigma_k^2 - \frac{1}{2} \sum_{k=1}^{uv} \sum_{j=1}^{m} \frac{|\mathbf{d}_j(k)|^2}{\sigma_k^2}, \end{aligned} \tag{5}$$

where $\mathbf{d}_j = [\mathbf{d}_j(1), \cdots, \mathbf{d}_j(uv)]^T$.

Consider the parameter to be estimated to be the diagonalised covariance matrix $\mathbf{\Sigma}$, which uniquely specifies the complete data CBC covariance. The EM algorithm proceeds as follows: for the E (expectation) step, the current best estimate $\mathbf{\Sigma}^{(p)}$ of the parameter is used to find the expected log-likelihood function $L(\mathbf{\Sigma}, \mathbf{d}_1, \ldots, \mathbf{d}_m)$, conditioned on the observations $\mathbf{y}_1^e, \ldots, \mathbf{y}_m^e$. In the M (maximisation) step, this conditional expectation is maximised with respect to the parameters to yield the next iterate $\mathbf{\Sigma}^{(p+1)}$. For the problem addressed in this paper, these steps will now be formalised.

4.1. Expectation step

Given the previous best estimate $\mathbf{\Sigma}^{(p)}$ of the parameters as well as the incomplete data $\mathbf{y}_1^e, \ldots, \mathbf{y}_m^e$, the expected value of the complete data log likelihood is

$$\begin{aligned} E\{L|\mathbf{\Sigma}^{(p)}, \mathbf{y}_1^e, \ldots, \mathbf{y}_m^e\} = K - \frac{m}{2} \sum_{k=1}^{uv} \log \sigma_k^{2(p)} - \\ \frac{1}{2} \sum_{k=1}^{uv} \sum_{j=1}^{m} \frac{E\{|d_j(k)|^2 | \mathbf{\Sigma}^{(p)}, \mathbf{y}_1^e, \ldots, \mathbf{y}_m^e\}}{\sigma_k^2}. \end{aligned} \tag{6}$$

4.2. Maximisation step

This involves finding the new parameters $\mathbf{\Sigma}^{(p+1)}$ which maximise the conditional expected log-likelihood in equation 6. Taking the derivative with respect to σ_l^2 and setting to zero yields a necessary condition for a maximum:

$$\frac{\partial E\{L|\mathbf{\Sigma}^{(p)}, \mathbf{y}_1^e, \dots, \mathbf{y}_m^e\}}{\partial \sigma_l^2} = -\frac{m}{2}\frac{1}{\sigma_l} - \frac{1}{2}\sum_{j=1}^{m} -\frac{E\{|d_j(l)|^2|\mathbf{\Sigma}^{(p)}, \mathbf{y}_1^e, \dots, \mathbf{y}_m^e\}}{(\sigma_l^2)^2} = 0 \quad (\forall l), \quad (7)$$

so

$$\sigma_l^{2(p+1)} = \frac{1}{m}\sum_{j=1}^{m} E\{|d_j(l)|^2|\mathbf{\Sigma}^{(p)}, \mathbf{y}_1^e, \dots, \mathbf{y}_m^e\}. \quad (8)$$

Given the values $\sigma_l^{2(p+1)}$ for each l, the new estimate of the parameter is $\mathbf{\Sigma}^{(p+1)} = \text{diag}(\sigma_1^{2(p+1)}, \dots, \sigma_{uv}^{2(p+1)})$. Since $\mathbf{C}^{(p+1)} = \mathbf{W}^\dagger\mathbf{\Sigma}^{(p+1)}\mathbf{W}$, the improved covariance matrix estimate $\mathbf{T}^{(p+1)}$ can be obtained from this using equation 4.

5. Calculating the iteration

The new estimate $\sigma_l^{2(p+1)}$ in equation 8 is expressed in terms of the expectations $E\{|d_j(l)|^2|\mathbf{\Sigma}^{(p)}, \mathbf{y}_1^e, \dots, \mathbf{y}_m^e\}$, which have yet to be calculated. Taking the same approach as Miller et al. [4], we note that $\sigma_l^{2(p+1)}$ in that equation is identical to the lth diagonal element of the matrix

$$\begin{aligned}\mathbf{\Sigma}_{dd}^{(p+1)} &= \frac{1}{m}\sum_{j=1}^{m} E\{\mathbf{d}_j\mathbf{d}_j^\dagger|\mathbf{\Sigma}^{(p)}, \mathbf{y}_1^e, \dots, \mathbf{y}_m^e\} \\ &= \frac{1}{m}\sum_{j=1}^{m} E\{\mathbf{d}_j\mathbf{d}_j^\dagger|\mathbf{\Sigma}^{(p)}, \mathbf{y}_j^e\} \quad (9)\end{aligned}$$

(since the observations are independent). To calculate this expectation, the joint distribution of $\mathbf{d}_j$ and $\mathbf{y}_j^e$ is required: with $\mathbf{K}_{yy} = \mathbf{U}_H^T\mathbf{T}^{(p)}\mathbf{U}_H$ and $\mathbf{K}_{dd} = \mathbf{\Sigma}^{(p)}$ we have

$$\begin{pmatrix}\mathbf{y}_j^e \\ \mathbf{d}_j\end{pmatrix} \Big| \mathbf{\Sigma}^{(p)} : N\left[\begin{pmatrix}\mathbf{0} \\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\mathbf{K}_{yy} & \mathbf{K}_{yd} \\ \mathbf{K}_{dy} & \mathbf{K}_{dd}\end{pmatrix}\right]. \quad (10)$$

The expectation $\mathbf{K}_{dy} = \mathbf{K}_{yd}^\dagger$ can be calculated as follows:

$$\begin{aligned}\mathbf{K}_{dy} &= E\{\mathbf{d}_j\mathbf{y}_j^{e\dagger}\} = E\{\mathbf{d}_j\mathbf{z}_j^\dagger\}(\mathbf{I}_{r\times u} \otimes \mathbf{I}_{s\times v})^T\mathbf{U}_H \\ &= E\{\mathbf{d}_j\mathbf{d}_j^\dagger\}(\mathbf{W}_u \otimes \mathbf{W}_v)(\mathbf{I}_{u\times r} \otimes \mathbf{I}_{v\times s})\mathbf{U}_H \\ &= \mathbf{\Sigma}^{(p)}(\mathbf{W}_u^{(r)} \otimes \mathbf{W}_v^{(s)})\mathbf{U}_H, \quad (11)\end{aligned}$$

where $\mathbf{W}_u^{(r)}$ contains the first r columns of $\mathbf{W}_u$, and $\mathbf{W}_v^{(s)}$ the first s columns of $\mathbf{W}_v$. The conditional distribution of $\mathbf{d}_j$ given $\mathbf{y}_j^e$ and $\mathbf{\Sigma}^{(p)}$ is [6]

$$\mathbf{d}_j|\mathbf{\Sigma}^{(p)}, \mathbf{y}_j^e : N[\mathbf{K}_{dy}\mathbf{K}_{yy}^{-1}\mathbf{y}_j^e, \mathbf{\Sigma}^{(p)} - \mathbf{K}_{dy}\mathbf{K}_{yy}^{-1}\mathbf{K}_{yd}], \quad (12)$$

from which it can be shown that

$$E\{\mathbf{d}_j\mathbf{d}_j^\dagger|\mathbf{\Sigma}^{(p)}, \mathbf{y}_j^e\} = \mathbf{K}_{dy}\mathbf{K}_{yy}^{-1}\mathbf{y}_j^e\mathbf{y}_j^{e\dagger}\mathbf{K}_{yy}^{-1}\mathbf{K}_{yd} + \mathbf{\Sigma}^{(p)} - \mathbf{K}_{dy}\mathbf{K}_{yy}^{-1}\mathbf{K}_{yd}. \quad (13)$$

Using this result with $\mathbf{y}_j^e = \mathbf{U}_H^T\mathbf{y}_j$ in equation 9 yields

$$\mathbf{\Sigma}_{dd}^{(p+1)} = \mathbf{K}_{dy}\mathbf{K}_{yy}^{-1}\mathbf{U}_H^T\mathbf{S}_{yy}\mathbf{U}_H\mathbf{K}_{yy}^{-1}\mathbf{K}_{yd} + \mathbf{\Sigma}^{(p)} - \mathbf{K}_{dy}\mathbf{K}_{yy}^{-1}\mathbf{K}_{yd}, \quad (14)$$

where $\mathbf{S}_{yy} = \frac{1}{m}\sum_{j=1}^{m}\mathbf{y}_j\mathbf{y}_j^\dagger$ is the sample covariance of the observation. Defining $\mathbf{W}_G = \mathbf{W}_u^{(r)} \otimes \mathbf{W}_v^{(s)}$, the parameters $\sigma_l^{2(p+1)}$ are the diagonal elements of

$$\begin{aligned}\mathbf{\Sigma}_{dd}^{(p+1)} &= \mathbf{\Sigma}^{(p)}\mathbf{W}_G\mathbf{U}_H(\mathbf{U}_H^T\mathbf{T}^{(p)}\mathbf{U}_H)^{-1}\mathbf{U}_H^T\mathbf{S}_{yy}\mathbf{U}_H \\ &(\mathbf{U}_H^T\mathbf{T}^{(p)}\mathbf{U}_H)^{-1}\mathbf{U}_H^T\mathbf{W}_G^\dagger\mathbf{\Sigma}^{(p)} + \mathbf{\Sigma}^{(p)} - \\ &\mathbf{\Sigma}^{(p)}\mathbf{W}_G\mathbf{U}_H(\mathbf{U}_H^T\mathbf{T}^{(p)}\mathbf{U}_H)^{-1}\mathbf{U}_H^T\mathbf{W}_G^\dagger\mathbf{\Sigma}^{(p)}. \quad (15)\end{aligned}$$

With inventive use of discrete Fourier transforms and TBT system solvers [3], it is possible to calculate the required elements even for moderately large matrices $\mathbf{T}$.

6. Conclusion

An algorithm has been presented for the constrained estimation of Toeplitz-Block-Toeplitz covariance matrices in the presence of subspace interference. The algorithm used is a generalisation of the standard method for estimating covariances under the Toeplitz constraint. It is expected that the convergence properties of the algorithm are the same as for the standard method.

References

[1] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: B*, 39:1–38, 1977.

[2] A. K. Jain. *Fundamentals of Digital Image Processing*. Prentice-Hall, 1989.

[3] J. H. Justice. A Levinson-type algorithm for two-dimensional Weiner filtering using bivariate Szegö polynomials. *Proceedings of the IEEE*, 65(6):882–886, June 1977.

[4] M. I. Miller, D. R. Fuhrmann, J. A. O'Sullivan, and D. L. Snyder. Maximum-likelihood methods for Toeplitz covariance estimation and radar imaging. In S. Haykin, editor, *Advances in Spectrum Analysis and Array Processing*, volume II, pages 145–172. Prentice-Hall, 1991.

[5] 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. *Proceedings of the IEEE*, 75(7):892–907, July 1987.

[6] I. B. Rhodes. A tutorial introduction to estimation and filtering. *IEEE Transactions on Automatic Control*, AC-16(6):688–706, Dec. 1971.