Title
Statistical analysis of an eigendecomposition based method for 2-D frequency estimation

Journal
Automatica, 30(1)

ISSN
0005-1098

Authors
Yang, Hua
Hua, Yingbo

Publication Date
1994

DOI
10.1016/0005-1098(94)90235-6

Peer reviewed

Automatica, Vol. 30, No. 1, pp. 157–168, 1994
Printed in Great Britain.

0005-1098/94 $6.00 + 0.00

Statistical Analysis of An Eigendecomposition Based Method for 2-D Frequency Estimation*†

HUA YANG‡ and YINGBO HUA‡

The large-sample covariances of 2-D frequency estimates using the matrix enhancement and matrix pencil method have been derived and analyzed.

Key Words—Signal processing; frequency estimation; matrix pencil; statistical analysis; eigendecomposition.

Abstract—An eigendecomposition based method for two-dimensional frequency estimation is analyzed in this paper. This method, to be referred to as matrix pencil (MP) method, computes a smoothed data covariance matrix, then its eigendecomposition, and then the two-dimensional frequencies via a MP approach. The MP method is now known to be more efficient in computation than many other methods and able to provide a near optimum performance for relatively large signal-to-noise ratio (SNR). The aim of this paper is to provide a further analysis of the MP method assuming a moderate SNR. To make the problem tractable, a large two-dimensional data set is considered. In this paper, a number of fundamental relations inherent in the MP method are revealed which lead to a general expression of the large-sample covariances of the estimated two-dimensional frequencies. The large-sample covariances are reduced to a very simple form for the single two-dimensional frequency case. The theoretical covariances are verified by the simulation results.

1. INTRODUCTION

THE STATISTICAL ANALYSIS of multi-dimensional (M-D) frequency estimation techniques is an open research field. In the past two decades, a large number of algorithms have been developed for frequency estimation (or the related problem: array processing), and an increasing attention has been received for statistically analyzing these algorithms (see Stoica and Söderström (1991) and a list of references thereof). However, relatively little analytical work has been reported for M-D frequency estimation methods. This is partially due to the fact that most one-dimensional methods can be effectively used for the M-D problem and an understanding of the one-dimensional methods can provide many insights we need to know. The second reason for a lack of analytical work for the M-D problem is the level of complexity involved in the M-D analysis which certainly has discouraged many researchers. On the other hand, however, the M-D analysis is very important in that (1) the M-D analysis can provide an insight that could not be obtained through one-dimensional analysis, and (2) the M-D analysis can allow us to see certain aspects of the M-D methods more easily than time-consuming computer simulations. Note that most M-D frequency estimation methods are searching-type which can be very expensive in computation.

In this paper, we present a statistical analysis of an M-D frequency estimation method which is to be referred to as the matrix pencil (MP) method. We will consider the two-dimensional case instead of the general M-D case. The MP method recently shown in Hua (1992) is superior to many existing two-dimensional methods in computation, and can achieve a near optimum performance for large SNR. A statistical analysis of the MP method for large SNR is available in Hua *et al.* (1993). This paper focuses on the case of relatively moderate SNR. To make the problem tractable, a large data set will be assumed. This work incorporates the skill shown in Stoica and Söderström (1991). But the novelty here includes: (a) considering the two-dimensional case instead of the one-dimensional, and (b) assuming unknown deterministic phases instead of the uniform random phases (the latter was assumed in Stoica and Söderström (1991) to make their problem simpler). As noted in Stoica and Söderström (1991), the frequency estimation problem where a single data set is available is unique from the array problem (e.g. in Ottersten *et al.*, 1991) where a large number of data sets

* Received 4 September 1992; revised 10 March 1993; received in final form 26 March 1993. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by the Guest Editors. Corresponding author: Y. Hua. Fax +61-3-344-6678; E-mail yhua@mullian.ee.mu.oz.au.
† This work has been supported by the Australian Research Council Large Grant Scheme and Australian Center for Sensor Signal and Information Processing.
‡ Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Victoria 3052, Australia.

are assumed. In fact, the former problem is more involved than the latter due to the need (in the former) to deal with correlated noise. Also unique for the two-dimensional frequency estimation is the pairing procedure required by the MP method. But due to the space limit, our work on the pairing analysis will be on a separate paper.

The paper is organized as follows. In Section 2, the data model of the two-dimensional frequency estimation problem is formulated. In Section 3, the MP method is summarized. Section 4 contains the original contributions of this paper, where a number of fundamental relations inherent in the MP method are shown in Sections 4.1–4.4, a general expression for the covariances of the estimated two-dimensional frequencies are provided in Section 4.5, an analysis of this general expression is carried out in Section 4.6, and then a simulation is illustrated to verify our theory in Section 4.7. In particular, a very simple form of the estimation covariances is given in Section 4.6 which clearly reveals how the accuracy of the MP method is affected by the data sizes, SNR, and the window sizes (explained in Section 3). Our expressions for the estimation covariances are valid for a median range of SNR.

2. DATA MODEL

We consider a two-dimensional data set $z(t, t')$ which consists of a sum of two-dimensional complex exponentials and noise:

$$z(t, t') = s(t, t') + e(t, t'), \quad t = 0, 1, \ldots, M-1, \quad t' = 0, 1, \ldots, N-1, \tag{1}$$

where

$$s(t, t') \triangleq \sum_{i=1}^{I} r_i \exp\{j(\phi_i + \omega_{1i}t + \omega_{2i}t')\}.$$

Note $j \triangleq \sqrt{-1}$, I is the number of the two-dimensional complex exponentials, $\{r_i\}$ are (positive) amplitudes, $\{\phi_i\}$ are phases, and $\{(\omega_{1i}, \omega_{2i})\}$ are the two-dimensional frequencies. All the parameters are assumed to be deterministic. Except I, they are also unknown. Let $y_i \triangleq e^{j\omega_{1i}}$ and $z_i \triangleq e^{j\omega_{2i}}$ which are called poles. $\{(y_i, z_i)\}$ are assumed to be distinct pairs of poles.

It should be noted here that in the one-dimensional analysis shown in Stoica and Söderstrom (1991), $\{\phi_i\}$ are assumed to be independent uniform random variables which has simplified that analysis. But it appears to be a more natural model for $\{\phi_i\}$ to be unknown deterministic.

The two-dimensional noise component $e(t, t')$ is complex, white and Gaussian, and satisfies

$$E[e(t, t')] = 0,$$
$$E[e(t, t')e^{c}(s, s')] = \sigma^2 \delta_{t,s}\delta_{t',s'},$$
$$E[e(t, t')e(s, s')] = 0 \quad \text{(for all } t, t', s \text{ and } s'),$$

where the superscript c denotes the complex conjugation, and $\delta_{t,s}$ is the Dirac delta function.

While only the case of single two-dimensional data set will be considered in this paper, a finite number of independent two-dimensional data sets could be treated similarly, But it should be noted that our analysis will assume a large data set which is opposed to the assumption of a large number of finite data sets. The latter assumption was treated by Ottersten *et al.* (1991).

3. TWO-DIMENSIONAL MATRIX PENCIL METHOD

The MP method was introduced in Hua (1992) for estimating the two-dimensional frequencies from the two-dimensional data $z(t, t')$, which can be summarized as follows just for the purpose of our analysis.

Step 1. Choose two integers (window sizes) K and L such that

$$M - I + 1 \geq K \geq I + 1,$$
$$N - I + 1 \geq L \geq I + 1.$$

Step 2. Compute the data covariance matrix

$$\hat{R}_c \triangleq \frac{1}{c}\mathbf{Z}_c\mathbf{Z}_c^{H} \quad \text{with } c \triangleq (M-K+1)(N-L+1),$$

where the superscript H denotes the conjugate transpose and

$$\mathbf{Z}_c \triangleq \begin{pmatrix} Z_0 & Z_1 & \cdots & Z_{M-K} \\ Z_1 & Z_2 & \cdots & Z_{M-K+1} \\ \cdots & \cdots & \cdots & \cdots \\ Z_{K-1} & Z_K & \cdots & Z_{M-1} \end{pmatrix}$$

($K \times (M-K+1)$ Hankel block matrix)

with

$$Z_t \triangleq \begin{pmatrix} z(t,0), & z(t,1) & \cdots & z(t,N-L) \\ z(t,1), & z(t,2) & \cdots & z(t,N-L+1) \\ \cdots & \cdots & \cdots & \cdots \\ z(t,L-1), & z(t,L) & \cdots & z(t,N-1) \end{pmatrix}$$

($L \times (N-L+1)$ Hankel matrix).

Step 3. Compute the eigendecomposition of $\hat{R}_c$, i.e. $\hat{R}_c = \sum_{i=1}^{KL} \hat{\lambda}_i \hat{s}_i \hat{s}_i^{H}$. Estimate I the number of the two-dimensional frequencies from the eigenvalues $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_{KL}$ by following a detection algorithm such as the one in Wax and Kailath (1985). But I will be assumed to be known for our analysis. Then define

$\hat{U}_s \triangleq (\hat{s}_1, \ldots, \hat{s}_I)$ (matrix of principal eigenvectors),
$\hat{U}_1 \triangleq \hat{U}_s$ with the last L rows deleted,
$\hat{U}_2 \triangleq \hat{U}_s$ with the first L rows deleted,
$P \triangleq (e_1, e_{1+L}, \ldots, e_{1+(K-1)L}, \ldots, e_L, e_{L+L}, \ldots, e_{L+(K-1)L})^{\mathrm{T}}$ (shuffling matrix), where e_i is the ith column of $KL \times KL$ identity matrix and the superscript T denotes the transpose,
$\hat{U}'_s \triangleq P\hat{U}_s$ (shuffled version of $\hat{U}_s$),
$\hat{U}'_1 \triangleq \hat{U}'_s$ with the last K rows deleted, and
$\hat{U}'_2 \triangleq \hat{U}'_s$ with the first K rows deleted.

Step 4. Find $\{\hat{y}_i\}$ by computing the generalized eigenvalues (GEs) of $\hat{U}_2 - \lambda\hat{U}_1$ (matrix pencil), and $\{\hat{z}_i\}$ the GEs of $\hat{U}'_2 - \lambda\hat{U}'_1$ (matrix pencil).

Step 5. Match $\{\hat{y}_i\}$ with $\{\hat{z}_i\}$ into I pairs by minimizing the null spectrum function:

$$D(p, q, \hat{U}_n) \triangleq (a^{\mathrm{H}}(p) \otimes b^{\mathrm{H}}(q))\hat{U}_n\hat{U}_n^{\mathrm{H}} \times (a(p) \otimes b(q)),$$

where $\hat{U}_n \triangleq (\hat{s}_{I+1}, \ldots, \hat{s}_{KL})$, $a(p) \triangleq (1, p, \ldots, p^{K-1})^{\mathrm{T}}$, $b(q) \triangleq (1, q, \ldots, q^{L-1})^{\mathrm{T}}$, and p and q are two complex variables associated with y_i and z_i, respectively. $\otimes$ denotes the Kronecker product.

Step 6. Compute $\hat{\omega}_{1i} \triangleq \mathrm{Im}[\log \hat{y}_i]$ and $\hat{\omega}_{2i} \triangleq \mathrm{Im}[\log \hat{z}_i]$ which are the MP estimates of ω_{1i} and ω_{2i}.

4. STATISTICAL ANALYSIS OF MATRIX PENCIL METHOD

4.1. *Preparation*

In this section, we first rewrite the data model into a matrix form to facilitate our analysis. A number of essential covariance matrices are then introduced.

The data model (1) can be rewritten as

$$z(t+u, t'+v) = \sum_{i=1}^{I} x_i(t, t') y_i^u z_i^v + e(t+u, t'+v), \quad (2)$$

where $x_i(t, t') \triangleq r_i \exp\{j(\phi_i + \omega_{1i}t + \omega_{2i}t')\}$.

We define a sequence of data submatrices $\{Z(t, t')\}$ by moving a window of the size $K \times L$ in the t–t' plane. Each submatrix is given by

$$Z(t, t') \triangleq \begin{pmatrix} z(t, t') & z(t, t'+1) & \cdots & z(t, t'+L-1) \\ z(t+1, t') & z(t+1, t'+1) & \cdots & z(t+1, t'+L-1) \\ \cdots & \cdots & \cdots & \cdots \\ z(t+K-1, t') & z(t+K-1, t'+1) & \cdots & z(t+K-1, t'+L-1) \end{pmatrix}.$$

In the same way, we define a sequence of noise submatrices $\{\varepsilon(t, t')\}$ from $e(t, t')$.

Let $V(v, m)$ be a Vandermond matrix

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ v_1 & v_2 & \cdots & v_I \\ \cdots & \cdots & \cdots & \cdots \\ v_1^{m-1} & v_2^{m-1} & \cdots & v_I^{m-1} \end{pmatrix},$$

where m is an integer and $v = (v_1, v_2, \ldots, v_I)^{\mathrm{T}}$ is a vector.

It can be shown that $Z(t, t')$ can be decomposed as

$$Z(t, t') = V(y, K)\,\mathrm{diag}\,(x_1(t, t'), \ldots, x_I(t, t'))V^{\mathrm{T}} \times (z, L) + \varepsilon(t, t'), \quad (3)$$

for $t = 0, 1, \ldots, M-K$, $t' = 0, 1, \ldots, N-L$, where $y \triangleq (y_1, y_2, \ldots, y_I)^{\mathrm{T}}$ and $z \triangleq (z_1, z_2, \ldots, z_I)^{\mathrm{T}}$.

The following notations and names will be used:

parameter matrix:

$$A \triangleq (a(y_1) \otimes b(z_1), a(y_2) \otimes b(z_2), \ldots, a(y_I) \otimes b(z_I)), \quad (4)$$

source covariance matrix (spatial varying):

$$B_{n,n'}^{t,t'} \triangleq \begin{pmatrix} x_1(t, t') \\ \vdots \\ x_I(t, t') \end{pmatrix} \times (x_1^c(t-n, t'-n'), \ldots, x_I^c(t-n, t'-n')),$$

averaged source covariance matrix:

$$B_{n,n'} \triangleq \frac{1}{C} \sum_{t=0}^{M-K} \sum_{t'=0}^{N-L} B_{n,n'}^{t,t'},$$

data covariance matrix (spatial varying):

$$R_{n,n'}^{t,t'} \triangleq E[Vec(Z(t, t'))Vec^{\mathrm{H}}(Z(t-n, t'-n'))],$$

averaged data covariance matrix:

$$R_{n,n'} \triangleq \frac{1}{C} \sum_{t=0}^{M-K} \sum_{t'=0}^{N-L} R_{n,n'}^{t,t'},$$

averaged sample data covariance matrix:

$$\hat{R}_{n,n'} \triangleq \frac{1}{c}\sum_{t=0}^{M-K}\sum_{t'=0}^{N-L} Vec(Z(t,t'))Vec^{H} \times (Z(t-n,t'-n')), \quad (5)$$

noise covariance matrix:

$$Q_{n,n'} \triangleq E[Vec(\varepsilon(t,t'))Vec^{H}(\varepsilon(t-n,t'-n'))], \quad (6)$$

where $c=(M-K+1)(N-L+1)$ as defined before and $Vec(\cdot)$ is the stretch operator which cascades the rows of the matrix under operation into one column vector. From the above definitions, the averaged data covariance is the expectation of the averaged sample data covariance, i.e. $R_{n,n'}=E[\hat{R}_{n,n'}]$.

For zero lags ($n=0$ and $n'=0$), the three matrices $B_{0,0}$, $R_{0,0}$ and $\hat{R}_{0,0}$ will be referred to as B, R and $\hat{R}$. By the definition of $\mathbf{Z}_c$ and $\hat{R}_c$ in Step 2 of the MP method, it can be verified that $\hat{R}=\hat{R}_c$. So $R=E[\hat{R}]=E[\hat{R}_c]$. The structure of the data covariance $R_{n,n'}$, the noise covariance $Q_{n,n'}$ and the source covariance $B_{n,n'}$ are essential in our analysis and discussed next.

4.2. *Structure of data, source and noise covariances*

The following result gives the structure of $R^{t,t'}_{n,n'}$ and $R_{n,n'}$.

Theorem 1.

$$R^{t,t'}_{n,n'}=AB^{t,t'}_{n,n'}A^{H}+Q_{n,n'}. \quad (7)$$

$$R_{n,n'}=AB_{n,n'}A^{H}+Q_{n,n'}. \quad (8)$$

Proof. See Appendix A.

The noise covariances $\{Q_{n,n'}\}$ are a sequence of $KL\times KL$ matrices with the properties (a) $Q_{n,n'}=0$ for $|n|\geq K$ or $|n'|\geq L$ and (b) $Q^{H}_{n,n'}=Q_{-n,-n'}$.

Define $J_{l,l'}\triangleq\frac{1}{\sigma^2}Q_{l,l'}$. Then $J_{l,l'}$ is a sparse matrix and its elements are zero or one. The following lemma shows its exact structure.

Lemma 1. For all l and l', we have the equality

$$J_{l,l'}=J^{(K)}_{l}\otimes J^{(L)}_{l'}, \quad (9)$$

where $\{J^{(K)}_l, l=0,\pm1,\pm2,\ldots\}$ is a sequence of $K\times K$ matrices. For $|l|\geq K$, $J^{(K)}_l=0$, and for $-K<l<K$, the (u,v)-element of

$$J^{(K)}_l=\begin{cases}1, & \text{if } v-u=l,\\ 0, & \text{otherwise.}\end{cases}$$

The matrices $\{J^{(L)}_{l'}\}$ are defined similarly.

Proof. See Appendix B.

Note each source covariance $B_{n,n'}$ is not diagonal for finite data sizes. The asymptotic structure of $B_{n,n'}$ is given by the following Lemma 2 which tells us that the off-diagonal elements in $B_{n,n'}$ are approximately equal to zero when the data sizes M and N are large.

We need the notation $O(\cdot)$. Let $O(1)$ denote a bounded quantity $|O(1)|\leq C_0$, where the constant C_0 does not depend on the data sizes M and N. Also define

$$O\left(\frac{1}{M}\right)=\frac{1}{M}O(1) \quad\text{and}\quad O\left(\frac{1}{MN}\right)=\frac{1}{MN}O(1).$$

Using the assumption that $\{(y_k,z_k)\}$ are distinct two-dimensional poles and the fact that

$$\sum_{m=0}^{M-1}z^m=\frac{1-z^M}{1-z},$$

when $z\neq1$, it is easy to prove the following lemma.

Lemma 2. $B_{n,n'}$ is an $I\times I$ matrix. When the data sizes, M and N, are large,

$$B_{n,n'}=\operatorname{diag}(r_1^2y_1^nz_1^{n'},\ldots,r_I^2y_I^nz_I^{n'}) +O\left(\frac{1}{M}\right)+O\left(\frac{1}{N}\right)+O\left(\frac{1}{MN}\right). \quad (10)$$

4.3. *Results easy to get*

From (8) in Theorem 1 and (9) in Lemma 1, we know that for $n=0$ and $n'=0$

$$R=ABA^{H}+\sigma^2I_{KL}, \quad (11)$$

where I_{KL} denotes the $KL\times KL$ identity matrix. The properties of the eigendecomposition of the data covariance matrix R are known in Stoica and Söderström (1991). But for easy reference in later analysis, they are reproduced below.

Define $U_s\triangleq(s_1,\ldots,s_I)$, consisting of the I (orthonormal) principal eigenvectors of R; $U_n\triangleq(s_{I+1},\ldots,s_{KL})$, consisting of the rest (orthonormal) eigenvectors of R; $\Lambda\triangleq\operatorname{diag}(\lambda_1,\ldots,\lambda_I)$, consisting of the I largest eigenvalues of R; and $\bar{\Lambda}\triangleq\Lambda-\sigma^2I_I$. Note that given A and B which are of the rank I, the decomposition (11) implies $\lambda_k>\sigma^2$ for $k=1,\ldots,I$, and $\lambda_{I+1}=\cdots=\lambda_{KL}=\sigma^2$. The conditions for A to be of the rank I are discussed in Hua (1992) and should be satisfied by *Step* 1 of the MP method. The full rankness of B is easily understood.

Based on the equation $U_s\bar{\Lambda}U_s^{H}=ABA^{H}$ which follows from (11), we can show that

$$U_s=AC, \quad (12)$$

where C is an $I \times I$ nonsingular matrix and given by $C = BA^{\mathrm{H}}U_s\bar{\Lambda}^{-1}$, and its inverse is

$$C^{-1} = U_s^{\mathrm{H}}A. \tag{13}$$

Also from $U_s\bar{\Lambda}U_s^{\mathrm{H}} = ABA^{\mathrm{H}}$, we have

Lemma 3. Let $X \triangleq U_s\bar{\Lambda}^{-1}U_s^{\mathrm{H}}A$. Then

$$X = A(A^{\mathrm{H}}A)^{-1}B^{-1}, \tag{14}$$

It follows from (10) and (14) that

$$x_k \to \frac{1}{r_k^2}a_k, \quad k = 1, \dots, I,$$

as

$$M \to \infty \text{ and } N \to \infty. \tag{15}$$

where x_k and a_k are the kth column of the matrix X and the kth column of the matrix $A(A^{\mathrm{H}}A)^{-1}$, respectively.

The following equality will be used to simplify some expressions in later sections.

Lemma 4. For $u \in \mathrm{Range}(U_n)$, we have

$$\sum_{|l|<K}\sum_{|l'|<L} J_{l,l'}u\mathrm{e}^{-jl\omega_{1k}-jl'\omega_{2k}} = 0, \quad k = 1, \dots, I. \tag{16}$$

Proof. Following (12), Range (A) = Range (U_s). Then

$A^{\mathrm{H}}U_n = 0$ and $(a(\omega_{1k}) \otimes b(\omega_{2k}))^{\mathrm{H}}u = 0$. So

$$\begin{aligned}0 &= (a(y_k) \otimes b(z_k))(a(y_k) \otimes b(z_k))^{\mathrm{H}}u \\ &= (a(y_k)a^{\mathrm{H}}(y_k) \otimes b(z_k)b^{\mathrm{H}}(z_k))u \\ &= \left(\left(\sum_{|l|<K} J_l^{(K)}\mathrm{e}^{-jl\omega_{1k}}\right) \otimes \left(\sum_{|l'|<L} J_{l'}^{(L)}\mathrm{e}^{-jl'\omega_{2k}}\right)\right)u \\ &= \sum_{|l|<K}\sum_{|l'|<L} (J_l^{(K)} \otimes J_{l'}^{(L)})u\mathrm{e}^{-jl\omega_{1k}-jl'\omega_{2k}} \\ &= \sum_{|l|<K}\sum_{|l'|<L} J_{l,l'}u\mathrm{e}^{-jl\omega_{1k}-jl'\omega_{2k}}. \quad \text{Q.E.D.}\end{aligned}$$

4.4. *Results hard to get*

To calculate the large sample covariance of the estimated frequencies, we need the following lemma to compute $G(u_1, u_2) \triangleq E[v_1v_2^{\mathrm{H}}]$, and $\bar{G}(u_1, u_2) \triangleq E[v_1v_2^{\mathrm{T}}]$, where u_1 and u_2 belong to Range (U_n) and $v_i \triangleq (u_i^{\mathrm{H}}(\hat{s}_1 - s_1), \dots, u_i^{\mathrm{H}}(\hat{s}_I - s_I))^{\mathrm{T}}$, $i = 1, 2$.

Lemma 5. For large M and N, we have

$$G(u_1, u_2) = \frac{1}{MN}\sum_{|k|<K}\sum_{|k'|<L} u_1^{\mathrm{H}}Q_{k,k'}u_2\bar{\Lambda}^{-1}U_s^{\mathrm{T}}R_{k,k'}^{c}U_s^{c}\bar{\Lambda}^{-1}, \tag{17}$$

$$\bar{G}(u_1, u_2) = \frac{1}{MN}\sum_{|k|<K}\sum_{|k'|<L} \bar{\Lambda}^{-1}U_s^{\mathrm{T}}Q_{k,k'}u_2^{c}u_1^{\mathrm{H}}Q_{k,k'}U_s\bar{\Lambda}^{-1}. \tag{18}$$

Proof. See Appendix C.

The two expressions above are the two-dimensional version of the one-dimensional expressions (3.22) and (3.23) in Stoica and Söderström (1991). It is important to note that (17) and (18) hold for both the random (uniformly distributed) phase model and the deterministic phase model. The derivation of these expressions for the deterministic phase model is more involved because the data covariance matrix $R_{n,n'}^{t,t'}$ depends on t and t' (i.e. spatial or time dependent) and the sequence of the noise submatrices $\{\varepsilon(t, t')\}$ are correlated. The derivation for the two-dimensional random phase model is much simpler because the data covariance matrix $R_{n,n'}^{t,t'}$ is independent of t and t'.

4.5. *Large sample covariance of the MP estimator*

Define

$A_1 \triangleq A$ with the last L rows deleted,
$A_2 \triangleq A$ with the first L rows deleted,
$A' \triangleq PA$(shuffled version of A),
$U_s' \triangleq PU_s$(shuffled version of U_s),
$A_1' \triangleq A'$ with the last K rows deleted, and
$A_2' \triangleq A'$ with the first K rows deleted

In the same way that $\hat{U}_1$, $\hat{U}_2$, $\hat{U}_1'$ and $\hat{U}_2'$ are defined from $\hat{U}_s$ and $\hat{U}_s'$ in Section 3, we define U_1, U_2, U_1' and U_2' from U_s and U_s'. Also define

$$F_k \triangleq (0 \mid I_{(K-1)L}) - \mathrm{e}^{j\omega_{1k}}(I_{(K-1)L} \mid 0),$$
$$F_k' \triangleq (0 \mid I_{(L-1)K}) - \mathrm{e}^{j\omega_{2k}}(I_{(L-1)K} \mid 0),$$
$$\eta_k^{\mathrm{H}} \triangleq \text{the } k\text{th row of the matrix } \times (A_1^{\mathrm{H}}A_1)^{-1}A_1^{\mathrm{H}}F_k,$$

and

$$\eta_k'^{\mathrm{H}} \triangleq \text{the } k\text{th row of the matrix } \times (A_1'^{\mathrm{H}}A_1')^{-1}A_1'^{\mathrm{H}}F_k'P.$$

Then we have:

Theorem 2. The two-dimensional MP estimators $\{(\hat{\omega}_{1k}, \hat{\omega}_{2k}),\ k = 1, \dots, I\}$ of the frequencies $\{(\omega_{1k}, \omega_{2k}),\ k = 1, \dots, I\}$ have the large sample covariance matrix:

$$\begin{pmatrix} \operatorname{cov}(\hat{\omega}_{1i}, \hat{\omega}_{1k}) & \operatorname{cov}(\hat{\omega}_{1i}, \hat{\omega}_{2k}) \\ \operatorname{cov}(\hat{\omega}_{2i}, \hat{\omega}_{1k}) & \operatorname{cov}(\hat{\omega}_{2i}, \hat{\omega}_{2k}) \end{pmatrix}$$
$$= \frac{\sigma^4}{2MNr_i^2r_k^2} \sum_{|l|<K} \sum_{|l'|<L} \operatorname{Re}\left[\begin{pmatrix} y_i^c\eta_i^{\mathrm{H}} & 0 \\ 0 & z_i^c\eta_i'^{\mathrm{H}} \end{pmatrix} \times \begin{pmatrix} J_{l,l'} & 0 \\ 0 & J_{l,l'} \end{pmatrix}\begin{pmatrix} \eta_k a_i^{\mathrm{T}} & \eta_k' a_i^{\mathrm{T}} \\ \eta_k a_i^{\mathrm{T}} & \eta_k' a_i^{\mathrm{T}} \end{pmatrix}\begin{pmatrix} J_{l,l'} & 0 \\ 0 & J_{l,l'} \end{pmatrix} \times \begin{pmatrix} y_k a_k^c & 0 \\ 0 & z_k a_k^c \end{pmatrix} - \begin{pmatrix} y_i^c\eta_i^{\mathrm{H}} & 0 \\ 0 & z_i^c\eta_i'^{\mathrm{H}} \end{pmatrix}\begin{pmatrix} J_{l,l'} & 0 \\ 0 & J_{l,l'} \end{pmatrix} \times \begin{pmatrix} a_k a_i^T & a_k a_i^{\mathrm{T}} \\ a_k a_i^T & a_k a_i^{\mathrm{T}} \end{pmatrix}\begin{pmatrix} J_{l,l'} & 0 \\ 0 & J_{l,l'} \end{pmatrix}\begin{pmatrix} y_k^c\eta_k^c & 0 \\ 0 & z_k^c\eta_k'^c \end{pmatrix}\right]. \tag{19}$$

Proof. We first derive the following expression:

$$\operatorname{cov}(\hat{\omega}_{1i}, \hat{\omega}_{1k}) \triangleq E[(\hat{\omega}_{1i} - \omega_{1i})(\hat{\omega}_{1k} - \omega_{1k})] \approx \frac{\sigma^4}{2MNr_i^2r_k^2} \sum_{|l|<K} \sum_{|l'|<L} \times \operatorname{Re}[e^{j(\omega_{1k}-\omega_{1i})}\eta_i^{\mathrm{H}}J_{l,l'}\eta_k a_k^{\mathrm{H}}J_{l,l'}^{\mathrm{T}}a_i - e^{-j(\omega_{1k}+\omega_{1i})}a_i^{\mathrm{T}}J_{l,l'}\eta_k^c\eta_i^{\mathrm{H}}J_{l,l'}a_k]. \tag{20}$$

From the definition of U_1, U_2, A_1 and A_2, we have the following relations: $U_1 = A_1C$, $U_2 = A_2C$ and $A_2 = A_1Y_d$, where $Y_d = \operatorname{diag}(y_1, \ldots, y_I)$. Given $\operatorname{rank}(U_1) = I$ as guaranteed by the *Step* 1 of the MP method, the GEs of $U_2 - \lambda U_1$ are also the eigenvalues of the matrix $\Phi \triangleq (U_1^{\mathrm{H}}U_1)^{-1}U_1^{\mathrm{H}}U_2$. By the above relations, $U_2 = A_2C = A_1Y_dC = U_1C^{-1}Y_dC$, and so $\Phi = C^{-1}Y_dC$, i.e. $\{y_k\}$ are eigenvalues of Φ, and $\gamma_k^{\mathrm{H}} \triangleq e_k^TC$ (the kth row of C) and $\beta_k \triangleq C^{-1}e_k$ (the kth column of C^{-1}), $k = 1, \ldots, I$, are the right and left eigenvectors of Φ, respectively.

Note the GEs $\{\hat{y}_k\}$ of $\hat{U}_2 - \lambda\hat{U}_1$ are the eigenvalues of the matrix $\hat{\Phi} \triangleq (\hat{U}_1^{\mathrm{H}}\hat{U}_1)^{-1}\hat{U}_1^{\mathrm{H}}\hat{U}_2$. Similarly to the proof of Theorem 2 in Stoica and Söderström (1991), we can get the following expressions:

$$\hat{\omega}_{1k} - \omega_{1k} \approx \operatorname{Im}(e^{-j\omega_{1k}}\eta_k^{\mathrm{H}}(\hat{U}_s - U_s)\beta_k) = \operatorname{Im}(e^{-j\omega_{1k}}v_k^{\mathrm{T}}\beta_k) = \operatorname{Im}(e^{-j\omega_{1k}}\beta_k^{\mathrm{T}}v_k), \tag{21}$$

$$\operatorname{cov}(\hat{\omega}_{1i}, \hat{\omega}_{1k}) \approx \tfrac{1}{2}\operatorname{Re}[e^{j(\omega_{1k}-\omega_{1i})}\beta_i^{\mathrm{T}}G(\eta_i, \eta_k)\beta_k^c - e^{-j(\omega_{1k}+\omega_{1i})}\beta_i^{\mathrm{T}}\bar{G}(\eta_i, \eta_k)\beta_k], \tag{22}$$

where $v_i \triangleq (\eta_i^{\mathrm{H}}(\hat{s}_1 - s_1), \ldots, \eta_i^{\mathrm{H}}(\hat{s}_I - s_I))^{\mathrm{T}}$, $i = 1, \ldots, I$.

Recalling (13) and the definition of x_k following (15), we know that $U_s\bar{\Lambda}^{-1}\beta_k = U_s\bar{\Lambda}^{-1}C^{-1}e_k = U_s\bar{\Lambda}^{-1}U_s^{\mathrm{H}}Ae_k = x_k$, i.e. $U_s\bar{\Lambda}^{-1}\beta_k = x_k$. Applying this relation and Lemma 5, for large M and N, we have

$$MN\beta_i^{\mathrm{T}}G(\eta_i, \eta_k)\beta_k^c = \sum_{|l|<K} \sum_{|l'|<L} \eta_i^{\mathrm{H}}Q_{l,l'}\eta_k\beta_i^{\mathrm{T}}\bar{\Lambda}^{-1}U_s^{\mathrm{T}}R_{l,l'}^cU_s^c\bar{\Lambda}^{-1}\beta_k^c$$
$$= \sum_{l,l'} \eta_i^{\mathrm{H}}Q_{l,l'}\eta_k x_i^{\mathrm{T}}R_{l,l'}^c x_k^c = \sum_{l,l'} \eta_i^{\mathrm{H}}Q_{l,l'}\eta_k x_k^{\mathrm{H}}R_{l,l'}^{\mathrm{H}}x_i$$
$$= \sum_{l,l'} \eta_i^{\mathrm{H}}Q_{l,l'}\eta_k x_k^{\mathrm{H}}AB_{l,l'}^{\mathrm{H}}A^{\mathrm{H}}x_i + \sum_{l,l'} \eta_i^{\mathrm{H}}Q_{l,l'}\eta_k x_k^{\mathrm{H}}Q_{l,l'}^{\mathrm{H}}x_i, \tag{23}$$

where (8) of Theorem 1 is applied to get the last equality.

By Lemma 3,

$$x_k^{\mathrm{H}}AB_{l,l'}^{\mathrm{H}}A^{\mathrm{H}}x_i = e_k^{\mathrm{T}}X^{\mathrm{H}}AB_{l,l'}^{\mathrm{H}}A^{\mathrm{H}}Xe_i = e_k^{\mathrm{T}}B^{-1}B_{l,l'}^{\mathrm{H}}B^{-1}e_i.$$

Using this and Lemma 2, for large M and N the first term in (23) is approximately equal to

$$\frac{1}{r_k^2}\sum_{|l|<K} \sum_{|l'|<L} \eta_i^{\mathrm{H}}Q_{l,l'}\eta_k y_k^{-l}z_k^{-l'}\delta_{ki} = \frac{\delta_{ki}}{r_k^2}\sigma^2 \sum_{|l|<K} \sum_{|l'|<L} \eta_i^{\mathrm{H}}J_{l,l'}\eta_k e^{-jl\omega_{1k}-jl'\omega_{2k}},$$

which is zero by Lemma 4, where δ_{ki} is the Dirac delta function.

By (15), the second term in (23) is approximately equal to $\frac{\sigma^4}{r_i^2r_k^2} \sum_{|l|<K, |l'|<L} \eta_i^{\mathrm{H}}J_{l,l'}\eta_k a_k^{\mathrm{H}}J_{l,l'}^{\mathrm{T}}a_i$. So (23) leads to

$$\beta_i^{\mathrm{T}}G(\eta_i, \eta_k)\beta_k^c \approx \frac{\sigma^4}{MNr_i^2r_k^2} \sum_{|l|<K} \sum_{|l'|<L} \eta_i^{\mathrm{H}}J_{l,l'}\eta_k a_k^{\mathrm{H}}J_{l,l'}^{\mathrm{T}}a_i. \tag{24}$$

Note when σ is too small or SNR is too high such that it is comparable to the data sizes M and N, the above approximation is inaccurate since the first term in (23) is not negligible as compared with the second term in (23). In this case, the contribution from the off-diagonal elements in $B_{l,l'}$ should be included in the approximation (24). But in this paper we are only interested in the median range of SNR.

In the same way deriving (24), we can get

$$\beta_i^{\mathrm{T}}\bar{G}(\eta_i, \eta_k)\beta_k \approx \frac{\sigma^4}{MNr_i^2r_k^2} \sum_{|l|<K} \sum_{|l'|<L} \eta_i^{\mathrm{H}}J_{l,l'}a_k a_i^TJ_{l,l'}\eta_k^c. \tag{25}$$

Therefore, using (24) and (25) in (22) yields (20). We can repeat the above procedures and get the large sample covariance

$$\operatorname{cov}(\hat{\omega}_{2i}, \hat{\omega}_{2k}) \triangleq E[(\hat{\omega}_{2i} - \omega_{2i})(\hat{\omega}_{2k} - \omega_{2k})]$$
$$\approx \frac{\sigma^4}{2MNr_i^2 r_k^2} \sum_{|l|<K} \sum_{|l'|<L} \times \operatorname{Re}[e^{j(\omega_{2k}-\omega_{2i})}\eta_i'^{\mathrm{H}} J_{l,l'} \eta_k' a_k^{\mathrm{H}} J_{l,l'}^{\mathrm{T}} a_i - e^{-j(\omega_{2k}+\omega_{2i})} a_i^{T} J_{l,l'} \eta_k'^{c} \eta_i'^{\mathrm{H}} J_{l,l'} a_k]. \quad (26)$$

One step used to yield the above expression is

$$\hat{\omega}_{2k} - \omega_{2k} \approx \operatorname{Im}(e^{-j\omega_{2k}} \eta_k'^{\mathrm{H}} (\hat{U}_s - U_s)\beta_k)$$
$$= \operatorname{Im}(e^{-j\omega_{2k}} v_k'^{\mathrm{T}} \beta_k)$$
$$= \operatorname{Im}(e^{-j\omega_{2k}} \beta_k^{\mathrm{T}} v_k'). \quad (27)$$

Combining (21) and (27), we can get the large sample covariances $\operatorname{cov}(\hat{\omega}_{1i}, \hat{\omega}_{2k})$ and $\operatorname{cov}(\hat{\omega}_{2i}, \hat{\omega}_{1k})$. These results combined with (20) and (26) are summarized in (19). □

4.6. *Analysis of the estimation covariance*

In the case of single two-dimensional sinusoid, the right-hand side of (19) can be simplified, through a very tedious calculation, into the following:

$$\operatorname{cov}(\hat{\omega}_{11}, \hat{\omega}_{11}) = \frac{1}{MN\mathrm{SNR}_1^2} \frac{2(2L^2+1)}{3(K-1)K^2L^3},$$
$$\operatorname{cov}(\hat{\omega}_{21}, \hat{\omega}_{21}) = \frac{1}{MN\mathrm{SNR}_1^2} \frac{2(2K^2+1)}{3(L-1)L^2K^3},$$
$$\operatorname{cov}(\hat{\omega}_{11}, \hat{\omega}_{21}) = 0, \quad \text{where } \mathrm{SNR}_1 \triangleq \frac{r_1^2}{\sigma^2}.$$

These expressions have been verified through numerical computation of (19).

The first and second expressions are consistent with the expression (6.3) in Stoica and Söderström (1991), when $L = 1$ is chosen in the first expression and $K = 1$ in the second. The first expression implies that the first frequency component is more affected by the window size $K\left(\sim \frac{1}{K^3}\right)$ than the other window size $L\left(\sim \frac{1}{L}\right)$. The second expression implies a similar property for the second frequency component. Note that in the derivation of the estimation covariances, the window sizes K and L are assumed to be fixed as the data sizes M and N are allowed to become larger, i.e. much larger than K and L. The above expressions imply the estimation variances of ω_{11} and ω_{21} are always decreasing as K and L increase (provided K and L are much smaller than M and N of course). The third expression tells us that the estimates of the two components of the single two-dimensional frequency are uncorrelated. Also note that all covariances here are independent of the value of the single two-dimensional frequency.

In general multiple-frequency case, Theorem 2 clearly shows a simple relationship between the estimation covariances and the important parameters: the data sizes M and N and the signal-to-noise ratios $\mathrm{SNR}_k \triangleq \frac{r_k^2}{\sigma^2}$, $k = 1, \ldots, I$. For the following discussion we define

$$\operatorname{var}_0(\hat{\omega}_{li}) \triangleq 2MN\mathrm{SNR}_i^2 E[(\hat{\omega}_{li} - \omega_{li})^2],$$
$$l = 1, 2, \quad i = 1, \ldots, I,$$

which are normalized estimation covariances, independent of the data sizes and SNR but dependent upon the data covariance R, the number of the two-dimensional frequencies and the window sizes K and L. The following numerical example shows how the normalized covariances are affected by the window sizes K and L.

Assume that the frequency matrix is

$$\begin{pmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{21} & \omega_{22} & \omega_{23} \end{pmatrix} = 2\pi \begin{pmatrix} 0.36 & 0.24 & 0.24 \\ 0.24 & 0.24 & 0.36 \end{pmatrix}, \quad (28)$$

which specifies three two-dimensional frequency pairs. Figures 1–2 illustrate $\operatorname{var}_0(\hat{\omega}_{i1})$, $i = 1, 2$, vs K and L. The plots for other frequencies are similar and omitted. As shown in the figures, the estimation covariances are reduced by two orders of magnitude by increasing the window sizes from $(3, 3)$ to $(7, 7)$. (The corresponding computations increase is also dramatic (Hua, 1992).) Being consistent with the single frequency case, these figures also suggest that the variance of a frequency estimate in each dimension is more affected by the window size in that dimension than the one in the other.

4.7. *Simulation and discussion*

For simulation, we assume that there are two two-dimensional frequency pairs

$$\begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix} = 2\pi \begin{pmatrix} 0.26 & 0.24 \\ 0.24 & 0.26 \end{pmatrix}$$

and the amplitudes $\{r_i\}$ are equal to one and the phases $\{\phi_i\}$ are zero. The choice of the frequency pairs as shown in (28) would lead to the same conclusion to be drawn. We choose the data sizes $M = N = 200$ which are large compared to the window sizes $K = L = 7$. The sample variances of $\hat{\omega}_{11}$ and $\hat{\omega}_{21}$ obtained from 80 independent runs (with Gaussian white noise)

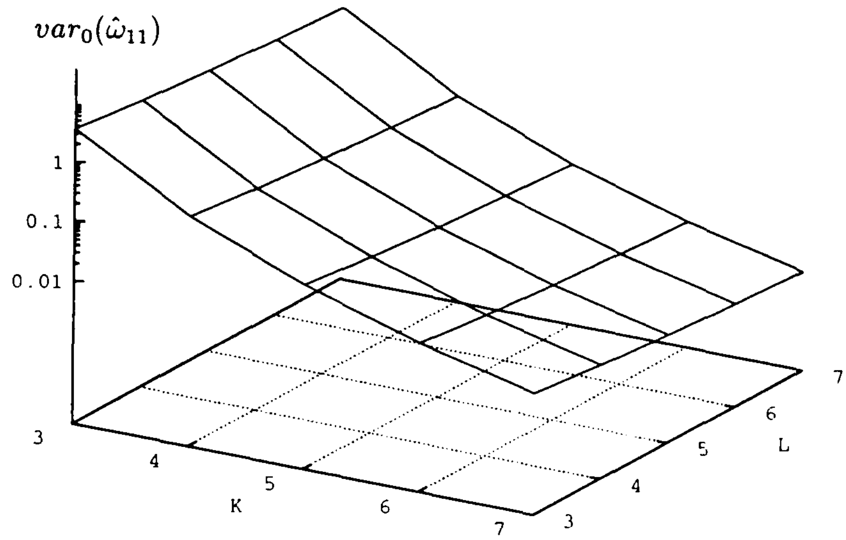

FIG. 1. Normalized variance $\mathrm{var}_0(\hat{\omega}_{11})$ vs window sizes.

are shown in Figs 3 and 4 against SNR in dB ($SNR = SNR_1 = SNR_2$). Also shown in Figs 3 and 4 are the theoretical results computed from (19) and the Cramer–Rao Bound (CRB). The plots for the second two-dimensional frequency pair are similar to Figs 3 and 4 and omitted.

It can be seen from these figures that the theory is consistent with the simulation for a median range of SNR, which is $[-12, 10]$ dB in this example. The reason why the theory is not accurate beyond a median range of SNR can be explained as follows. Our analysis is based on a first order approximation of the perturbations in the eigendecomposition of the noisy data covarance matrix $\hat{R}_e$. When SNR is too low, the first order approximation (namely (35)) breaks down and hence the result becomes inaccurate. On the other hand, when SNR is too high so that it is comparable to the data sizes M and N, then the first term in (23) is not negligible and so the approximation (24) is inaccurate. The high SNR phenomenon is unique for the single data set problem (although it is not pointed out in Stoica and Söderström (1991)) and is still under investigation.

The relationship between the estimation variance and the CRB is also worth mentioning. As suggested by the figures, the MP method is not statistically efficient for large data sizes particularly in the median range of SNR while the MP method was shown in Hua (1992) and Hua *et al.* (1993) to be near efficient for moderate data sizes and high SNR. In fact, we know that the CRB for the data model considered in this paper is proportional to $\frac{1}{\mathrm{SNR}}$ (Hua, 1992) while (19)

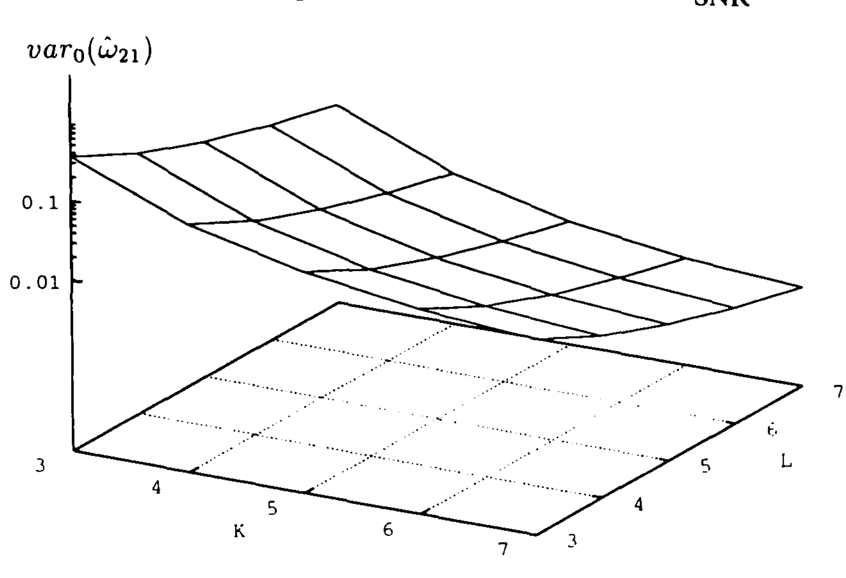

FIG. 2. Normalized variance $\mathrm{var}_0(\hat{\omega}_{21})$ vs window sizes.

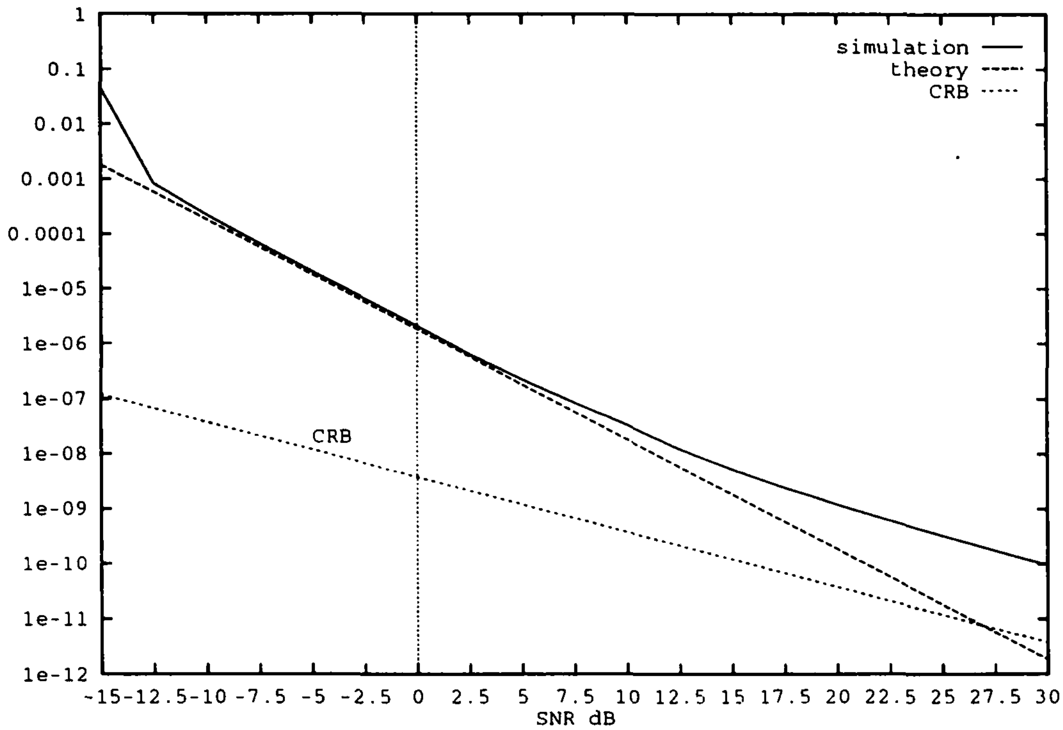

FIG. 3. Estimation variance and CRB for ω_{11}.

implies that the variance of the MP method is proportional to $\frac{1}{\mathrm{SNR}^2}$. So given the best case that the MP method approaches the CRB at high SNR, it must deviate from the CRB in the median range of SNR.

At this point, one may ask whether other methods can achieve the CRB in the median range of SNR for large data sizes. Based on the relationship between the maximum likelihood (ML) method and the CRB, one expects the ML should be efficient in the median range of SNR for large data sizes. But the exact median range of SNR vs data sizes remains to be found. An

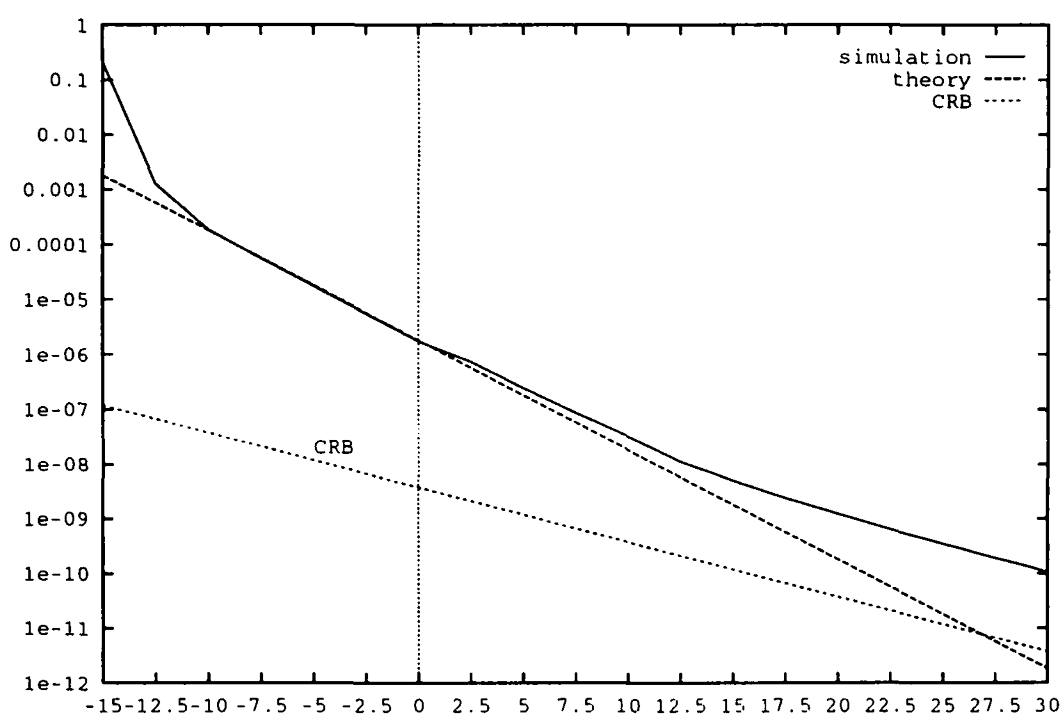

FIG. 4. Estimation variance and CRB for ω_{21}.

asymptotical version of the ML method (for large data sizes) is the periodogram. The analysis of the periodogram along the lines shown in this paper needs to be carried out. Another method that should come into one's mind in this context is (two-dimensional) MUSIC. In fact, the analysis of the MUSIC can be similarly carried out as for the MP method. Due to the amount of work involved, the properties of the MUSIC are separately shown in Yang and Hua (1993).

5. CONCLUSION

We have derived the covariances of the two-dimensional frequency estimates obtained by the MP method, assuming large data sizes, single data set and deterministic phases. The derivation has been very involved due to the nature of this problem. The problem with deterministic phases is more difficult than the one with random phases used in Stoica and Söderström (1991). A single and large data set is also more challenging than a large number of data sets as dealt with in Ottersten *et al.* (1991). The hard found general expression of the estimation covariance has also been analyzed. In particular, for the case of a single two-dimensional frequency, the general expression is reduced to a very simple form which reveals clearly how the accuracy of the MP method is affected by the data sizes, SNR and the window sizes. While the general expression remains to be studied analytically for multiple two-dimensional frequencies, it has been numerically compared to the simulation. The numerical results have confirmed the accuracy of the theoretical expression. It is important to stress that the analysis of two-dimensional frequency estimation methods is still an open field where simulation has been widely used in place of analysis. As has been demonstrated in this paper, analysis can provide important insights that could not be found through simulation.

Acknowledgment—The authors would like to thank the reviewers for their valuable comments.

REFERENCE

Hua, Y. (1992). Estimating two-dimensional frequencies by matrix enhancement and matrix pencil. *IEEE Trans. on Signal Processing*, **40**, 2267–2280.

Hua, Y., F. Baqai, Y. Zhu and D. Heilbronn (1993). Imaging of point scatterers from step-frequency ISAR data. *IEEE Trans. on Aerospace and Electronic Systems*, **29**, 195–205.

Ottersten, B., M. Viberg and T. Kailath (1991). Performance analysis of the total least squares ESPRIT algorithm. *IEEE Trans. on Signal Processing*, **39**, 1122–1135.

Stoica, P. and T. Söderström (1991). Statistical analysis of MUSIC and subspace rotation estimates of sinusoidal frequencies. *IEEE Trans. Signal Processing*, **39**, 1836–1847.

Wax, M. and T. Kailath (1985). Detection of signals by information theoretic criteria. *IEEE Trans. on Acoustics, Speech and Signal Processing*, **33**, 387–392.

Yang, H. and Y. Hua (1993). Asymptotic properties of 2-D MUSIC estimator with comparison to MP estimator, to be submitted.

APPENDIX A

Proof of Theorem 1.

Applying the stretch operator $Vec(\cdot)$ to the both sides of (3), we get

$$Vec(Z(t,t')) = (V(y,K)\otimes V(z,L)) \times Vec(\mathrm{diag}\,(x_1(t,t'),\dots,x_l(t,t'))) + Vec(\varepsilon(t,t')). \quad \text{(A.1)}$$

Let $X^d(t,t') \triangleq \mathrm{diag}\,(x_1(t,t'),\dots,x_l(t,t'))$ and

$$P^{t,t'}_{n,n'} \triangleq Vec(X^d(t,t'))Vec^{H}(X^d(t-n,t'-n')).$$

Since

$$Vec(X^d(t,t')) = (\underset{\substack{\uparrow\\ \text{1st}}}{x_1(t,t')}\ 0\ \cdots\ 0\ \underset{\substack{\uparrow\\ (l+2)\text{th}}}{x_2(t,t')}\ 0\ \cdots\ 0\ \underset{\substack{\uparrow\\ l^2\text{th}}}{x_l(t,t')})^T,$$

$$P^{t,t'}_{n,n'} = \begin{pmatrix} x_1x_1'^c & 0 & \cdots & 0 & x_1x_2'^c & 0 & \cdots & 0 & x_1x_l'^c \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & & & & \vdots & & & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ x_2x_1'^c & 0 & \cdots & 0 & x_2x_2'^c & 0 & \cdots & 0 & x_2x_l'^c \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & & & & \vdots & & & & \vdots \\ x_lx_1'^c & 0 & \cdots & 0 & x_lx_2'^c & 0 & \cdots & 0 & x_lx_l'^c \end{pmatrix}, \quad \text{(A.2)}$$

where $x_i = x_i(t,t')$ and $x_i' = x_i'(t-n,t'-n')$.

Because $E[\varepsilon(t,t')] = 0$, from (A.1) we get

$$R^{t,t'}_{n,n'} = (V(y,K)\otimes V(z,L))P^{t,t'}_{n,n'} \times (V(y,K)\otimes V(z,L))^{H} + Q_{n,n'}. \quad \text{(A.3)}$$

Let

$$E_p \triangleq \mathrm{diag}\,(\underset{\substack{\uparrow\\ \text{1st}}}{1}\ 0\ \cdots\ 0\ \underset{\substack{\uparrow\\ (l+2)\text{th}}}{1}\ 0\ \cdots\ 0\ \underset{\substack{\uparrow\\ l^2\text{th}}}{1}),$$

$$A_p \triangleq (V(y,K)\otimes V(z,L))E_p.$$

Since

$$Vec(\mathrm{diag}\,(x_1(t,t'),\dots,x_l(t,t'))) = E_p Vec(\mathrm{diag}\,(x_1(t,t'),\dots,x_l(t,t')))$$

and $P^{t,t'}_{n,n'} = E_pP^{t,t'}_{n,n'} = E_pP^{t,t'}_{n,n'}E_p$, equation (A.1) can be rewritten as

$$Vec(Z(t,t')) = A_p Vec(\mathrm{diag}\,(x_1(t,t'),\dots,x_l(t,t'))) + Vec(\varepsilon(t,t')). \quad \text{(A.4)}$$

It is easy to verify that $\mathrm{Range}\,(A_p) = \mathrm{Range}(A)$, because when the matrix $(V(y,K)\otimes V(z,L))$ is multiplied by the E_p

all columns become zero vectors except the 1st column, $(l+2)$th column, ..., and l^2th column. These non-zero columns form the matrix A. These facts and the expression (A.4) are particularly useful in simplifying some complicated expressions in this paper.

From (A.3), we have

$$R_{n,n'}^{l,l'} = A_v P_{n,n'}^{l,l'} A_v^{H} + Q_{n,n'}. \quad \text{(A.5)}$$

From the expression (A.2) for $P_{n,n'}^{l,l'}$ we know there exists a real unitary matrix $\bar{P}$ which is the product of several permutation matrices such that

$$\bar{P}^{T} P_{n,n'}^{l,l'} \bar{P} = \left(\begin{array}{c|c} B_{n,n'}^{l,l'} & 0 \\ \hline 0 & 0 \end{array}\right) \quad \text{(A.6)}$$

and $(V(y,K)\otimes V(z,L))E_p\bar{P} = (A \mid 0)$. Therefore

$$\begin{aligned} R_{n,n'}^{l,l'} &= (V(y,K)\otimes V(z,L))P_{n,n'}^{l,l'}(V(y,K)\otimes V(z,L))^{H} + Q_{n,n'} \\ &= (V(y,K)\otimes V(z,L))E_p\bar{P}\bar{P}^{T}P_{n,n'}^{l,l'}\bar{P}\bar{P}^{T}E_p \\ &\quad \times (V(y,K)\otimes V(z,L))^{H} + Q_{n,n'} \\ &= (A \mid 0)\bar{P}^{T}P_{n,n'}^{l,l'}\bar{P}\left(\frac{A^{H}}{0}\right) + Q_{n,n'} \\ &= AB_{n,n'}^{l,l'}A^{H} + Q_{n,n'}. \end{aligned}$$

This is (7) from which it is straightforward to get (8). □

APPENDIX B

Proof of Lemma 1.

For $|l| \ge K$ or $|l'| \ge L$, the identity (9) is true because both sides of (9) are zero.

Now we consider the case $|l| < K$ and $|l'| < L$. For $t = 0, 1, \ldots, M-K$, and $t' = 0, 1, \ldots, N-L$, define $\varepsilon_r(t,t') \triangleq (e(t,t'), e(t,t'+1), \ldots, e(t,t'+L-1))^{T}$. Then $Vec^{T}(\varepsilon(l,l')) = (\varepsilon_r^{T}(l,l'), \varepsilon_r^{T}(l+1,l'), \ldots, \varepsilon_r^{T}(l+K-1,l'))^{T}$.

The matrix $Q_{l,l'} = E[Vec(\varepsilon(l,l'))Vec^{H}(\varepsilon(0,0))]$ is a $K\times K$ block matrix. The (u,v)-block in $Q_{l,l'}$ is

$$E[\varepsilon_r(l+u,l')\varepsilon_r^{H}(v,0)] = E\left[\begin{pmatrix} e(l+u,l') \\ e(l+u,l'+1) \\ \vdots \\ e(l+u,l'+L-1)\end{pmatrix} \times (e^{c}(v,0), e^{c}(v,1), \ldots, e^{c}(v,L-1))\right] = \begin{cases} \sigma^2 J_{l'}^{(L)}, & \text{if } v-u=1, \\ 0, & \text{elsewhere,}\end{cases}$$

i.e. $Q_{l,l'} = \sigma^2 J_l^{(K)}(J_{l'}^{(L)})$ where the matrix $J_l^{(K)}(J_{l'}^{(L)})$ is obtained from $J_l^{(K)}$ by substituting 1's in $J_l^{(K)}$ with $J_{l'}^{(L)}$ and 0's with $L\times L$ zero matrix. So $J_l^{(K)}(J_{l'}^{(L)}) = J_l^{(K)}\otimes J_{l'}^{(L)}$. Hence $Q_{l,l'} = \sigma^2 J_l^{(K)}\otimes J_{l'}^{(L)}$, and $J_{l,l'} = J_l^{(K)}\otimes J_{l'}^{(L)}$. □

APPENDIX C

Proof of Lemma 5.

Similar to the discussion in Appendix A in Stoica and Söderström (1991), we can get the approximation

$$U_s^{H}\hat{R}U_n \approx \hat{\Lambda}\hat{U}_s^{H}U_n \quad \text{or}$$

$$U_n^{H}\hat{U}_s \approx U_n^{H}\hat{R}U_s\hat{\Lambda}^{-1}. \quad \text{(C.1)}$$

So $U_n^{H}\hat{s}_i = U_n^{H}(\hat{s}_i - s_i) \approx U_n^{H}\hat{R}U_s\hat{\Lambda}^{-1}e_i = U_n^{H}\hat{R}U_s e_i \dfrac{1}{\lambda_i - \sigma^2}$, i.e. $U_n^{H}\hat{s}_i = U_n^{H}\hat{R}s_i\dfrac{1}{\lambda_i-\sigma^2}$. For u_1, $u_2 \in \mathrm{Range}(U_n)$, the column space of U_n, by the above equality,

$$(\lambda_i - \sigma^2)u_1^{H}(\hat{s}_i - s_i) \approx u_1^{H}\hat{R}s_i$$

and

$$(\lambda_j - \sigma^2)u_2^{H}(\hat{s}_j - s_j) \approx u_2^{H}\hat{R}s_j.$$

From above, we know

$$(\lambda_i - \sigma^2)(\lambda_j - \sigma^2)G_{ij} = E[u_1^{H}\hat{R}s_i(u_2^{H}\hat{R}s_j)^{H}], \quad \text{(C.2)}$$

where G_{ij} is the (i,j)th element of the matrix $G(u_1,u_2)$. To simplify notations, we assume that the data set $\{z(t,t'), 0 \le t \le M+K-1, 0 \le t' \le N+L-1\}$ is available. Then the sample estimation of R can be taken as

$$\hat{R} \triangleq \frac{1}{MN}\sum_{t=1}^{M}\sum_{t'=1}^{N} Vec(Z(t,t'))Vec^{H}(Z(t,t')).$$

Substituting the right-hand side of (A.4) into the above expression and then substituting $\hat{R}$ into (C.2), we can get the expression

$$\begin{aligned}(\lambda_i-\sigma^2)(\lambda_j-\sigma^2)G_{ij} &= \frac{1}{M^2N^2}E\Bigg[\sum_{t=1}^{M}\sum_{t'=1}^{N}\sum_{p=1}^{M}\sum_{p'=1}^{N} u_1^{H}Vec(Z(t,t'))Vec^{H} \\ &\quad \times (Z(t,t'))s_is_j^{H}Vec(Z(p,p'))Vec^{H}(Z(p,p'))u_2\Bigg]. \quad \text{(C.3)}\end{aligned}$$

The right-hand side of the above expressions can be further simplified by employing (A.4) and the fact that for $u\in U_n$, $A_v^{H}u = 0$ (because Range $(A_v) =$ Range (A)).

$$\begin{aligned}&(\lambda_i-\sigma^2)(\lambda_j-\sigma^2)G_{ij}M^2N^2 \\ &= E\Bigg[\sum_{t,t',p,p'} u_1^{H}Vec(\varepsilon(t,t'))[Vec^{H}(X^{d}(t,t'))A_v^{H} \\ &\quad + Vec^{H}(\varepsilon(t,t'))]s_is_j^{H}[A_vVec(X^{d}(p,p')) \\ &\quad + Vec(\varepsilon(p,p'))]Vec^{H}(\varepsilon(p,p'))u_2\Bigg] \\ &= E\Bigg[\sum_{t,t',p,p'} [u_1^{H}Vec(\varepsilon(t,t'))Vec^{H}(X^{d}(t,t'))A_v^{H}s_is_j^{H}A_vVec \\ &\quad \times (X^{d}(p,p'))Vec^{H}(\varepsilon(p,p'))u_2 \\ &\quad + u_1^{H}Vec(\varepsilon(t,t'))Vec^{H}(X^{d}(t,t'))A_v^{H}s_is_j^{H}Vec \\ &\quad \times (\varepsilon(p,p'))Vec^{H}(\varepsilon(p,p'))u_2 \\ &\quad + u_1^{H}Vec(\varepsilon(t,t'))Vec^{H}(\varepsilon(t,t'))s_is_j^{H}A_vVec \\ &\quad \times (X^{d}(p,p'))Vec^{H}(\varepsilon(p,p'))u_2 \\ &\quad + u_1^{H}Vec(\varepsilon(t,t'))Vec^{H}(\varepsilon(t,t'))s_is_j^{H}Vec \\ &\quad \times (\varepsilon(p,p'))Vec^{H}(\varepsilon(p,p'))u_2]\Bigg] \\ &= \sum_{t,t',p,p'} E[\text{Term 1} + \text{Term 2} + \text{Term 3} + \text{Term 4}]. \quad \text{(C.4)}\end{aligned}$$

Term 2 and 3 are zero since for three independent and zero mean Gaussian random variables X_1, X_2 and X_3, $E[X_i^3] = 0$, $i = 1, 2, 3$, $E[X_2X_1^2] = E[X_3X_1^2] = E[X_2X_3^2] = 0$, and $E[X_1X_2X_3] = 0$. So the right-hand side of (C.4) is equal to

$$\sum_{t,t',p,p'}\{u_1^{H}Q_{t-p,t'-p'}u_2s_j^{H}A_vVec(X^{d}(p,p'))Vec^{H} \times (X^{d}(t,t'))A_v^{H}s_i + E[\text{Term 4}]\} \quad \text{(C.5)}$$

in which the expectation of Term 4 (the product of the four complex Gaussian random variables) can be simplified by using the formula

$$\begin{aligned}E[x_1x_2x_3x_4] &= E[x_1x_2]E[x_3x_4] \\ &\quad + E[x_1x_3]E[x_2x_4] + E[x_1x_4]E[x_2x_3]\end{aligned}$$

(see Stoica and Söderström 1991). So (C.5) is equal to

$$\sum_{t,t',p,p'} \{u_1^H Q_{t-p,t'-p'} u_2 s_j^H A_v Vec(X^d(p,p')) Vec^H(X^d(t,t')) A_v^H s_i$$
$$+ E[u_1^H Vec(\varepsilon(t,t')) Vec^H(\varepsilon(t,t')) s_i]$$
$$\times E[s_j^H Vec(\varepsilon(p,p')) Vec^H(\varepsilon(p,p')) u_2]$$
$$+ E[u_1^H Vec(\varepsilon(t,t')) s_j^H Vec(\varepsilon(p,p'))]$$
$$\times E[Vec^H(\varepsilon(t,t')) s_i Vec^H(\varepsilon(p,p')) u_2]$$
$$+ E[u_1^H Vec(\varepsilon(t,t')) Vec^H(\varepsilon(p,p')) u_2]$$
$$\times E[Vec^H(\varepsilon(t,t')) s_i s_j^H Vec(\varepsilon(p,p'))]\}$$
$$= \sum_{t,t',p,p'} u_1^H Q_{t-p,t'-p'} u_2 s_j^H A_v P_{p-t,p'-t'}^{p,p'} A_v^H s_i$$
$$+ \sum_{t,t',p,p'} u_1^H Q_{t-p,t'-p'} u_2 s_j^H Q_{p-t,p'-t'} s_i. \tag{C.6}$$

The second term in (C.6) dividing by M^2N^2 is

$$\frac{1}{M^2N^2} \sum_{t=1}^{M} \sum_{t'=1}^{N} \sum_{p=1}^{M} \sum_{p'=1}^{N} u_1^H Q_{t-p,t'-p'} u_2 s_j^H Q_{p-t,p'-t'} s_i$$
$$= \frac{1}{M^2N^2} \sum_{|k|<K} \sum_{|k'|<L} (M-|k|)(N-|k'|) u_1^H Q_{k,k'} u_2 s_j^H Q_{-k,-k'} s_i$$
$$= \frac{1}{MN} \sum_{|k|<K} \sum_{|k'|<L} u_1^H Q_{k,k'} u_2 s_j^H Q_{-k,-k'} s_i, \tag{C.7}$$

for the large data sizes M and N.

Comparing (A.5) and (7), we know that $A_v P_{p-t,p'-t'}^{p,p'} A_v^H = A B_{p-t,p'-t'}^{p,p'} A^H$. Therefore the first term in (C.6) is equal to

$$\frac{1}{M^2N^2} \sum_{t=1}^{M} \sum_{t'=1}^{N} \sum_{p=1}^{M} \sum_{p'=1}^{N} u_1^H Q_{t-p,t'-p'} u_2 s_j^H A B_{p-t,p-t'}^{p,p'} A^H s_i. \tag{C.8}$$

The method used in deriving (C.7) is not applicable to simplify the sum (C.8) because the matrix $B_{n,n'}^{p,p'}$ depends on p and p'. This is a major difficulty in deriving the large sample covariance due to the deterministic phases.

Let $C_t \triangleq \{(p,t): |p-t|<K,\ 1\le t\le M,\ 1\le p\le M\}$ and $C_t' \triangleq \{(p',t'): |p'-t'|<L,\ 1\le t'\le N,\ 1\le p'\le N\}$ be two index sets.

Since $Q_{t-p,t'-p'}=0$ for $|t-p|\ge K$ or $|t'-p'|\ge L$, the sum (C.8) only includes those terms with $(p,t)\in C_t$ and $(p',t')\in C_t'$. C_t can be partitioned into three index sets

$C_1 = \{(p,t): |p-t|<K,\ 1\le t\le M,\ 1\le p\le K\}$,

$C_2 = \{(p,t): |p-t|<K,\ 1\le t\le M,\ K+1\le p\le M-K\}$, and

$C_3 = \{(p,t): |p-t|<K,\ 1\le t\le M,\ M-K+1\le p\le M\}$.

The number of points in C_2 is $(2K-1)(M-2K)$ which depends on the data size M. But the number of points in $C_1 \cup C_3$ is $K(3K-1)$ which does not depend on the data size M. Similarly, we can define C_1', C_2' and C_3' which are partitions of C_t'. The number of points in $C_1' \cup C_3'$ does not depend on the data size N. So the sum (C.8) is equal to

$$\frac{1}{M^2N^2} \sum_{(p,t)\in C_2} \sum_{(p',t')\in C_2'} u_1^H Q_{t-p,t'-p'} u_2 s_j^H A B_{p-t,p'-t'}^{p,p'} A^H s_i + \frac{1}{MN}\left(o\left(\frac{1}{M}\right) + o\left(\frac{1}{N}\right)\right)$$

in which the first term is equal to

$$= \frac{1}{M^2N^2} \sum_{p=K+1}^{M-K} \sum_{t=p-K+1}^{p+K-1} \sum_{p'=L+1}^{N-L} \sum_{t'=p'-L+1}^{p'+L-1}$$
$$\times u_1^H Q_{t-p,t'-p'} u_2 s_j^H A B_{p-t,p'-t'}^{p,p'} A^H s_i$$
$$= \frac{1}{M^2N^2} \sum_{p=K+1}^{M-K} \sum_{p'=L+1}^{N-L} \sum_{|k|<K} \sum_{|k'|<L} u_1^H Q_{k,k'} u_2 s_j^H A B_{-k,-k'}^{p,p'} A^H s_i$$
$$= \frac{1}{MN} \frac{(M-2K)(N-2L)}{MN} \sum_{|k|<K} \sum_{|k'|<L} u_1^H Q_{k,k'} u_2 s_j^H$$
$$A\left(\frac{1}{(M-2K)(N-2L)} \sum_{p=K+1}^{M-K} \sum_{p'=L+1}^{N-L} B_{-k,-k'}^{p,p'}\right) A^H s_i$$
$$\approx \frac{1}{MN} \sum_{|k|<K} \sum_{|k'|<L} u_1^H Q_{k,k'} u_2 s_j^H A B_{-k,-k'} A^H s_i.$$

Finally, from (C.6) and (C.7) we have

$$(\lambda_i - \sigma^2) G_{ij} (\lambda_j - \sigma^2) = \frac{1}{MN} \sum_{|k|<K} \sum_{|k'|<L} u_1^H Q_{k,k'} u_2 s_j^H$$
$$\times (A B_{-k,-k'} A^H + Q_{-k,-k'}) s_i$$
$$= \frac{1}{MN} \sum_{|k|<K} \sum_{|k'|<L} u_1^H Q_{k,k'} u_2 s_j^H R_{-k,-k'} s_i,$$

for large M and N.

Since $R_{-k,-k'} = R_{k,k'}^H$, from the above expression we have

$$G_{ij} = \frac{1}{MN} \sum_{|k|<K} \sum_{|k'|<L} u_1^H Q_{k,k'} u_2 s_j^H R_{k,k'}^H s_i \frac{1}{(\lambda_i - \sigma^2)(\lambda_j - \sigma^2)}$$
$$= \frac{1}{MN} \sum_{|k|<K} \sum_{|k'|<L} u_1^H Q_{k,k'} u_2 e_i^T \tilde{\Lambda}^{-1} U_s^T R_{k,k'}^c U_s^c \tilde{\Lambda}^{-1} e_j.$$

So we have the equality (17). In the same way as above, we can prove (18). □