

Time Varying Autoregressive Moving Average Models for Covariance Estimation

Ami Wiesel, Ofir Bibi, and Amir Globerson

Abstract—We consider large scale covariance estimation using a small number of samples in applications where there is a natural ordering between the random variables. The two classical approaches to this problem rely on banded covariance and banded inverse covariance structures, corresponding to time varying moving average (MA) and autoregressive (AR) models, respectively. Motivated by this analogy to spectral estimation and the well known modeling power of autoregressive moving average (ARMA) processes, we propose a novel time varying ARMA covariance structure. Similarly to known results in the context of AR and MA, we address the completion of an ARMA covariance matrix from its main band, and its estimation based on random samples. Finally, we examine the advantages of our proposed methods using numerical experiments.

Index Terms—Autoregressive moving average, covariance estimation, matrix completion, instrumental variables.

I. INTRODUCTION

LARGE scale covariance estimation using a small number of samples is a fundamental problem in modern multivariate statistical analysis. In many applications, e.g., linear array processing, climatology, spectroscopy, and longitudinal data analysis, there is a natural ordering between the random variables and it is reasonable to assume that the statistical relation between two variables decays with their distance. Based on this assumption, low order parametric models may be imposed on the unknown covariance and allow for consistent estimation using a small number of samples n even when the dimension d is much greater.

A natural approach to covariance estimation is to formulate the decay in statistical relation using the notion of independence or correlation. Assuming that \mathbf{x}_i is uncorrelated with \mathbf{x}_j if $|i - j| > q$ leads an intuitive q -banded covariance structure [1]–[3]. This structure is a special case of the more general class of sparse covariance models [4]. An alternative approach uses the notion of conditional independence. In the Gaussian case, assuming that \mathbf{x}_i and \mathbf{x}_j with $|i - j| > p$ are conditionally independent given the rest of the elements in \mathbf{x} leads to a p -banded inverse covariance [1], [5]–[8]. Indeed, this structure is a special case of sparse inverse covariance models, also known as

Gaussian graphical models [9]. A remarkable result of [1] is that the condition $(\log d)/n \rightarrow 0$ is sufficient for consistent estimation in these models.

The starting point for this work is the analogy between covariance estimation and spectrum estimation in stationary random processes. The two classical models for spectrum estimation are moving average (MA) and autoregressive (AR) processes. Interestingly, the finite MA(q) process can be easily shown to be equivalent to a q -banded Toeplitz covariance model, whereas the finite AR(p) process corresponds to a p -banded Toeplitz inverse covariance model. In the non-stationary cases, the Toeplitz restriction is removed and the structures reduce to the banded and inverse banded structures discussed above. In stationary processes, it is well known that MA and AR are usually crude approximations of reality and that better modeling power may be obtained through autoregressive moving average (ARMA) processes [10]–[12]. Continuing this analogy, we propose a new non-stationary ARMA(p, q) covariance model.

In this paper, we introduce a novel time varying ARMA(p, q) covariance model. It reduces to sparse AR(p) structures when $q = 0$, and reduces to sparse MA(q) structures when $p = 0$. The general model leads to a dense covariance with a dense inverse, but is parameterized by a small number of unknowns. ARMA(p, q) are a special case of recursive linear models with correlated errors. A different special case, known as Bow-free acyclic path (BAP), has recently been studied in [13], [14].

The first contribution of this paper concerns ARMA(p, q) matrix completion. It is well known that AR(p) and MA(q) covariance matrices can be uniquely completed given their main p and q bands, respectively. Thus, it is natural to expect that ARMA(p, q) models can be reconstructed given their $p + q$ leading diagonals. Following this intuition, we derive an ARMA completion procedure denoted by \mathcal{AC} . The algorithm uses an instrumental variables (IVs) approach which is motivated by classical time series analysis techniques [10]–[12]. It does not require any ad hoc starting point, is not iterative, and is not prone to convergence to suboptimal local solutions. We provide a simple condition under which \mathcal{AC} is exact, as well as a counter example when the condition does not hold. Numerical results with random parameters suggest that, in practice, the condition almost always holds.

The second contribution of this paper addresses ARMA(p, q) covariance estimation with an emphasis on the high dimensional regime in which $p + q \ll n \ll d$. Following existing methods in AR(p) and MA(q), we propose to use \mathcal{AC} again but replace the unknown $p + q$ band with its sample version. This empirical band is much smaller than the full sample covariance matrix and is therefore more accurate. To formalize these statements,

Manuscript received March 20, 2012; revised July 31, 2012; accepted February 18, 2013. Date of publication April 04, 2013; date of current version May 08, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Shuguang (Robert) Cui. Preliminary results of this work have been presented in IEEE SAM-2012. This work was supported by the Israeli Smart Grid consortium, and also in part by the ISF Centers of Excellence Grant 1789/11 and ISF Grant 786/11.

The authors are with The Selim and Rachel Benin School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem 91905, Israel (e-mail: amiw@cs.huji.ac.il; ofirbb@cs.huji.ac.il; gamir@cs.huji.ac.il).

Digital Object Identifier 10.1109/TSP.2013.2256900

| | | |
|------------------|--------|-------|
| 1, 2, \dots, i-1 | | i |
| | p_i | i |
| \bar{r}_i | r_i | i |
| \bar{r}_i | iv_i | q_i |
| \bar{q}_i | q_i | i |

we analyze \mathcal{AC} under both deterministically and stochastically perturbed inputs. Similarly to [1], we show that, under technical conditions, it is consistent in the operator norm as long as $(\log d)/n \rightarrow 0$. In addition, we demonstrate the performance advantages of the \mathcal{AC} estimator using numerical experiments.

The outline of the paper is as follows. In Section II, we introduce the time-varying ARMA(p, q) covariance model. In Section III, we consider matrix completion and in Section IV we apply these results in the context of estimation. In Section V, we demonstrate the advantages of our proposed methods using numerical experiments. Finally, concluding remarks are given in Section VI.

The following notation is used. Boldface upper case letters denote matrices, boldface lower case letters denote column vectors, and standard lower case letters denote scalars. The superscripts $(\cdot)^T, (\cdot)^H, (\cdot)^{-1}$ and $(\cdot)^\dagger$ denote the transpose, the conjugate transpose, the inverse and the pseudoinverse, respectively. For sets a and b , the cardinality is denoted by $|a|$ and the set difference operator is denoted by $a \setminus b$. The sub-matrix of \mathbf{X} indexed by a and b is denoted by $\mathbf{X}_{a,b}$, and this operation precedes any other matrix operation, e.g., $\mathbf{X}_{a,b}^T = [\mathbf{X}_{a,b}]^T$. The operators $\|\mathbf{X}\|_1, \|\mathbf{X}\|_2, \|\mathbf{X}\|_\infty$ and $\|\mathbf{X}\|_{\max}$ denote the maximum absolute column sum norm, the operator norm, the maximum absolute row sum norm and the element-wise infinity norm, respectively. $\mathbf{X} \succ \mathbf{0}$ means that \mathbf{X} is positive definite. By $\mathbf{A} \in \mathcal{R}(\mathbf{B})$ we mean that the columns of \mathbf{A} lie within the range space of \mathbf{B} . The banding operator $\mathcal{B}_b(\mathbf{X})$ outputs a matrix with the principal $2b + 1$ diagonals of \mathbf{X} and zero-padding elsewhere. We use the following indices sets definitions. For $i = 1, \dots, d$ and given p, q and $r \geq p + q$, we define $q_i = \{\max\{1, i - q\}, \dots, i - 1\}$, $r_i = \{\max\{1, i - r\}, \dots, i - 1\}$, $iv_i = r_i \setminus q_i$, $\bar{r}_i = \{1, \dots, i - 1\} \setminus r_i$, $\bar{q}_i = \{1, \dots, i - 1\} \setminus q_i$ and $p_i = \{\max\{1, i - p, i - |r_i| + |q_i|\}, \dots, i - 1\}$. These sets satisfy:

$$\begin{aligned} |iv_i| &\geq |p_i| \\ iv_i \cap q_i &= \emptyset, \end{aligned} \quad (1)$$

and can be visualized as

Throughout the paper, we use a few technical lemmas provided in Appendix A.

II. MODELS

In this section, we review the two classical MA and AR covariance structures, and then generalize them to the more flexible class of ARMA models.

A natural model for the covariance of a random vector \mathbf{x} which satisfies a known ordering is a banded matrix, i.e., a matrix Σ such that [1]:

$$\Sigma = \mathcal{B}_q(\Sigma). \quad (2)$$

The rationale behind this structure is that, in the Gaussian case, a zero in the covariance matrix denotes statistical independence between variables:

$$[\Sigma]_{i,j} = 0 \Leftrightarrow p(\mathbf{x}_i, \mathbf{x}_j) = p(\mathbf{x}_i)p(\mathbf{x}_j), \quad (3)$$

and it is reasonable to assume that distant elements are statistically independent. It is well known that banded positive definite matrices have banded Cholesky factorization

$$\Sigma = \mathbf{B}\mathbf{B}^T, \quad (4)$$

where \mathbf{B} is a q -banded lower triangular matrix. This provides an elegant generative model for the random vector \mathbf{x} using an uncorrelated latent random vector \mathbf{w} . Namely,

$$\mathbf{x} = \mathbf{B}\mathbf{w}, \quad (5)$$

where \mathbf{w} is a driving vector which satisfies

$$\begin{aligned} E\{\mathbf{w}\} &= \mathbf{0} \\ E\{\mathbf{w}\mathbf{w}^T\} &= \mathbf{I}. \end{aligned} \quad (6)$$

Note that each variable in \mathbf{x} averages over adjacent elements in \mathbf{w} , and hence the name ‘‘moving average’’ (MA) process. More precisely, this is a time-varying or non-stationary MA model. The classical stationary MA process assumes that Σ is also a Toeplitz matrix. In what follows, we use MA(q) to denote covariance matrices which satisfy (4).

An alternative model for the covariance matrix of an ordered random vector is that its inverse is p -banded and satisfies [1], [15]

$$\Sigma^{-1} = \mathcal{B}_p(\Sigma^{-1}). \quad (7)$$

The rationale is that, in the Gaussian case, a zero in the inverse covariance matrix corresponds to conditional independence between variables:

$$[\Sigma^{-1}]_{i,j} = 0 \Leftrightarrow p(\mathbf{x}_i, \mathbf{x}_j | \mathbf{x}_r) = p(\mathbf{x}_i | \mathbf{x}_r)p(\mathbf{x}_j | \mathbf{x}_r), \quad (8)$$

where \mathbf{x}_r denotes a vector with all the elements in \mathbf{x} except \mathbf{x}_i and \mathbf{x}_j . Intuitively, this means that the dependency between distant variables disappears once we know the elements between them. Using the Cholesky factorization, any matrix whose inverse is p -banded can be written as

$$\Sigma = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{D}^{-1}(\mathbf{I} - \mathbf{A})^{-T}, \quad (9)$$

where $(\mathbf{I} - \mathbf{A})^{-1}\mathbf{D}^{-\frac{1}{2}}$ is the Cholesky factor, \mathbf{A} is a p -banded lower triangular matrix with zero valued diagonal elements and \mathbf{D} is a positive definite diagonal matrix. This leads to the dynamical ‘‘autoregressive’’ model

$$\begin{aligned} \mathbf{x} &= (\mathbf{I} - \mathbf{A})^{-1}\mathbf{v} \\ &= \mathbf{A}\mathbf{x} + \mathbf{v}, \end{aligned} \quad (10)$$

where \mathbf{v} is a vector of latent variables satisfying

$$\begin{aligned} E\{\mathbf{v}\} &= \mathbf{0} \\ E\{\mathbf{v}\mathbf{v}^T\} &= \mathbf{D}^{-1}. \end{aligned} \quad (11)$$

The name ‘‘autoregressive’’ expresses the causal dependency of \mathbf{x}_i on \mathbf{x}_{p_i} . The classical stationary AR model is a special case in which Σ has a Toeplitz structure. As before, we use $\text{AR}(p)$ to denote covariance matrices which satisfy (9).

Motivated by the MA and AR interpretation of the banded covariance and inverse covariance models, we now propose a new covariance structure based on the classical ARMA process. The scalar, infinite and stationary ARMA process $x[k]$ for $-\infty < k < \infty$ is defined as

$$x[k] = \sum_{i=1}^p a_i x[k-i] + \sum_{i=0}^q b_i w[k-i], \quad (13)$$

where $w[k]$ is a zero mean, white driving process. The natural vector, finite and non-stationary version of (13) is the length d random vector \mathbf{x} defined as

$$\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{w}, \quad (14)$$

where \mathbf{w} is a vector of latent variables which satisfies (6), \mathbf{A} is a p -banded, lower triangular matrix with zero diagonal elements and \mathbf{B} is a q -banded lower triangular matrix. Indeed, if we increase the dimension d without bound and constrain \mathbf{A} and \mathbf{B} to Toeplitz matrices, then model (13) can be interpreted as the limit to (14). The two extreme cases of the ARMA model are clearly the pure MA process in (5) where $p = 0$ and the pure AR process in (10) where $q = 0$, $\mathbf{v} = \mathbf{B}\mathbf{w}$ and $\mathbf{D} = \mathbf{B}\mathbf{B}^T$ is a diagonal matrix.

The ARMA model defines a low order parametric model for the covariance of \mathbf{x} . Rearranging (14), we obtain

$$\mathbf{x} = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{B}\mathbf{w}, \quad (15)$$

with a covariance matrix

$$\Sigma = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{B}\mathbf{B}^T (\mathbf{I} - \mathbf{A})^{-T}. \quad (16)$$

Hereinafter, we denote the class of matrices satisfying (16) as $\text{ARMA}(p, q)$ covariances. To simplify the notation, we will often use

$$\Sigma = \mathbf{U}^{-1} \Omega \mathbf{U}^{-T} \quad (17)$$

with

$$\begin{aligned} \mathbf{U} &= \mathbf{I} - \mathbf{A} \\ \Omega &= \mathbf{B}\mathbf{B}^T. \end{aligned} \quad (18)$$

Intuition on this structure can be obtained by examining its extreme cases. It is easy to see that this model has the pure MA model in (4) and the pure AR model in (9) as special cases. Unlike these special cases, both Σ and Σ^{-1} are non-banded and dense matrices in the general ARMA model. It is important to note that the decomposition of Σ in (16) is not unique and there may be different legitimate pairs of matrices $\{\mathbf{A}, \mathbf{B}\}$ which yield the same covariance. Thus, throughout this paper we only address the issue of recovering or estimating Σ rather than $\{\mathbf{A}, \mathbf{B}\}$.

For completeness, we note that the proposed ARMA model is a special case of recursive linear models with correlated errors.

Bow-free acyclic path (BAP) diagrams are another special case which has recently been studied in [13], [14]. The ARMA model is not necessarily a BAP, and is aimed at different applications where there is a natural ordering between the variables. Furthermore, the current results on BAPs assume that the number of samples is greater than the dimension and are therefore not suitable for large scale covariance estimation.

III. COVARIANCE COMPLETION

Matrix completion problems consider the recovery of a full matrix given the values of only a subset of its entries assuming some structure or objective function [16], [17]. We now review known results on these problems in MA and AR models and extend them to the ARMA setting.

Completion of an $\text{MA}(q)$ covariance matrix from its main q -band is trivial via zero padding [1]

$$\bar{\Sigma} = \mathcal{B}_q(\Sigma). \quad (19)$$

Completion of an $\text{AR}(p)$ covariance matrix from its main p -band is also well known. This procedure is known as positive definite completion and is equivalent to maximum likelihood estimation [9], [16]. In brief, we recover \mathbf{A}_{i,p_i} for $i = 1, \dots, d$ using simple linear regressions based on $\mathcal{B}_p(\Sigma)$. The i th row of (10) can be expressed as

$$\mathbf{x}_i = \mathbf{A}_{i,p_i} \mathbf{x}_{p_i} + \mathbf{v}_i. \quad (20)$$

We multiply this equation on the right by $\mathbf{x}_{p_i}^T$ and take the expectation. The residual \mathbf{v}_i is uncorrelated with $\mathbf{x}_{p_i}^T$ and therefore disappears so that

$$\Sigma_{i,p_i} = \mathbf{A}_{i,p_i} \Sigma_{p_i,p_i}. \quad (21)$$

Solving for \mathbf{A}_{i,p_i} yields the non-stationary Yule Walker formulas

$$\bar{\mathbf{A}}_{i,p_i} = \Sigma_{i,p_i} \Sigma_{p_i,p_i}^{-1} \quad (22)$$

from which the covariance can be easily completed.

We now extend these results to general ARMA structures. Similarly to the $\text{AR}(p)$ completion method, we use linear regressions to recover \mathbf{A} from $\mathcal{B}_r(\Sigma)$ and then reconstruct Σ . The main difference is that we cannot use simple regressions and require a more careful treatment based on IVs. As before, the i th row of (14) can be expressed as

$$\mathbf{x}_i = \mathbf{A}_{i,p_i} \mathbf{x}_{p_i} + [\mathbf{B}\mathbf{w}]_i \quad (23)$$

but now \mathbf{x}_{p_i} and $[\mathbf{B}\mathbf{w}]_i$ are correlated. IV methods are directly aimed at regressions when the explanatory variables (covariates) are correlated with the residual terms. Thus, instead of multiplying by \mathbf{x}_{p_i} , we multiply (23) by the IVs $\mathbf{x}_{iv_i}^T$ which, due to (1), are uncorrelated with $[\mathbf{B}\mathbf{w}]_i$

$$\mathbb{E} \{ [\mathbf{B}\mathbf{w}]_i \mathbf{x}_{iv_i}^T \} = \mathbf{0}^T. \quad (24)$$

Taking the expectation yields

$$\Sigma_{i,iv_i} = \mathbf{A}_{i,p_i} \Sigma_{p_i,iv_i} \quad (25)$$

for $i = 1, \dots, d$. Due to (1), the number of equations $|iv_i|$ will be greater than or equal to number of unknowns $|p_i|$. Thus, we can try to solve these linear systems as

$$\bar{\mathbf{A}}_{i,p_i} = \boldsymbol{\Sigma}_{i,iv_i} \boldsymbol{\Sigma}_{p_i,iv_i}^\dagger \quad (26)$$

and pad the rest of the matrix $\bar{\mathbf{A}}$ with zeros. Assuming that $\bar{\mathbf{A}}$ is identical to \mathbf{A} , we can easily recover $\boldsymbol{\Omega}$ as

$$\bar{\boldsymbol{\Omega}} = \mathcal{B}_q(\bar{\mathbf{U}}\mathcal{B}_r(\boldsymbol{\Sigma})\bar{\mathbf{U}}^T) \quad (27)$$

where

$$\bar{\mathbf{U}} = \mathbf{I} - \bar{\mathbf{A}} \quad (28)$$

and complete $\mathcal{B}_r(\boldsymbol{\Sigma})$ to

$$\bar{\boldsymbol{\Sigma}} = \bar{\mathbf{U}}^{-1}\bar{\boldsymbol{\Omega}}\bar{\mathbf{U}}^{-T}. \quad (29)$$

This ARMA completion procedure, denoted by \mathcal{AC} , is summarized below.

Procedure AC

Input: $\mathcal{B}_r(\boldsymbol{\Sigma}), p, q$

Output: $\bar{\boldsymbol{\Sigma}}$

$\bar{\mathbf{A}} = \mathbf{0};$

For $i = 1 : d$ **do**

$$[\bar{\mathbf{A}}_{i,p_i} = \boldsymbol{\Sigma}_{i,iv_i} \boldsymbol{\Sigma}_{p_i,iv_i}^\dagger$$

$\bar{\mathbf{U}} = \mathbf{I} - \bar{\mathbf{A}}$

$\bar{\boldsymbol{\Omega}} = \mathcal{B}_q(\bar{\mathbf{U}}\mathcal{B}_r(\boldsymbol{\Sigma})\bar{\mathbf{U}}^T)$

$\bar{\boldsymbol{\Sigma}} = \bar{\mathbf{U}}^{-1}\bar{\boldsymbol{\Omega}}\bar{\mathbf{U}}^{-T}$

Clearly, \mathcal{AC} is correct when $\bar{\mathbf{A}} = \mathbf{A}$. Due to the lack of identification in ARMA models, this may not be the situation in practice and it is very likely that $\bar{\mathbf{A}} \neq \mathbf{A}$. The next theorem characterizes the conditions for exact completion in this practical case.

Theorem 1: Let $\boldsymbol{\Sigma}$ follow an ARMA(p, q) model. Define a bandwidth $r \geq p + q$ and assume that

$$\boldsymbol{\Sigma}_{p_i, \bar{q}_i} \in \mathcal{R}(\boldsymbol{\Sigma}_{p_i, iv_i}), \quad i = 1, \dots, d. \quad (30)$$

Then, \mathcal{AC} reconstructs the unknown covariance exactly so that $\bar{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}$. In particular, if $r = d - 1$ then (30) always hold, and the reconstruction is exact (yet trivial).

Proof: The core of the proof lies in the identity

$$\begin{aligned} \boldsymbol{\Sigma}_{\bar{q}_i, p_i} \bar{\mathbf{A}}_{i, p_i}^T &= \boldsymbol{\Sigma}_{\bar{q}_i, p_i} \boldsymbol{\Sigma}_{iv_i, p_i}^\dagger \boldsymbol{\Sigma}_{iv_i, i} \\ &\stackrel{(25)}{=} \boldsymbol{\Sigma}_{\bar{q}_i, p_i} \boldsymbol{\Sigma}_{iv_i, p_i}^\dagger \boldsymbol{\Sigma}_{iv_i, p_i} \mathbf{A}_{i, p_i}^T \\ &\stackrel{(30)}{=} \boldsymbol{\Sigma}_{\bar{q}_i, p_i} \mathbf{A}_{i, p_i}^T \\ &= \boldsymbol{\Sigma}_{\bar{q}_i, i}, \quad i = 1, \dots, d. \end{aligned} \quad (31)$$

The last equality is similar to (24)–(25) and is based on the fact that, like $\mathbf{x}_{iv_i}, \mathbf{x}_{\bar{q}_i}$ is also uncorrelated with $[\mathbf{B}\mathbf{w}]_i$. The identity suggests a simple recursive completion method. Assume that we have already reconstructed $\boldsymbol{\Sigma}_{1:i-1, 1:i-1}$, then using $\bar{\mathbf{A}}, \mathcal{B}_r(\boldsymbol{\Sigma})$ and (31), we can recover $\boldsymbol{\Sigma}_{1:i, 1:i}$. In what follows, we show that

the last two lines of \mathcal{AC} are simply a non-recursive implementation of this idea.

Using (28), identity (31) can be compactly expressed as

$$\boldsymbol{\Sigma}_{\bar{q}_i, 1:i} \bar{\mathbf{U}}_{i, 1:i}^T = \mathbf{0}. \quad (32)$$

For any matrix \mathbf{Z} , we have the following recursive properties

$$\begin{aligned} [\bar{\mathbf{U}}^{-1}]_{1:i, 1:i} &= \bar{\mathbf{U}}_{1:i, 1:i}^{-1} \\ \bar{\mathbf{U}}_{1:i, 1:i} \mathbf{Z}_{1:i, 1:i} &= [\bar{\mathbf{U}}\mathbf{Z}]_{1:i, 1:i} \\ \bar{\mathbf{U}}_{1:i, 1:i}^{-1} \mathbf{Z}_{1:i, 1:i} &= [\bar{\mathbf{U}}^{-1}\mathbf{Z}]_{1:i, 1:i} \\ [\mathcal{B}_q(\mathbf{Z})]_{1:i, 1:i} &= \mathcal{B}_q([\mathbf{Z}]_{1:i, 1:i}) \end{aligned} \quad (33)$$

for $i = 1, \dots, d$, where we have used the triangular structure of $\bar{\mathbf{U}}$. Thus, the last two lines of \mathcal{AC} can be expressed as

$$\bar{\boldsymbol{\Omega}}_{1:i, 1:i} = \mathcal{B}_q(\bar{\mathbf{U}}_{1:i, 1:i} \mathcal{B}_r(\boldsymbol{\Sigma}_{1:i, 1:i}) \bar{\mathbf{U}}_{1:i, 1:i}^T) \quad (34)$$

$$\bar{\boldsymbol{\Sigma}}_{1:i, 1:i} = \bar{\mathbf{U}}_{1:i, 1:i}^{-1} \bar{\boldsymbol{\Omega}}_{1:i, 1:i} \bar{\mathbf{U}}_{1:i, 1:i}^{-T}. \quad (35)$$

Based on this observation, the proof that $\bar{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}$ proceeds by induction on the correctness of $\bar{\boldsymbol{\Sigma}}_{1:i, 1:i}$ starting with $i = 1$ and up to $i = d$. The basis of the induction

$$\bar{\boldsymbol{\Sigma}}_{1,1} = \boldsymbol{\Sigma}_{1,1} \quad (36)$$

clearly holds and we now show that the hypothesis

$$\bar{\boldsymbol{\Sigma}}_{1:i-1, 1:i-1} = \boldsymbol{\Sigma}_{1:i-1, 1:i-1} \quad (37)$$

leads to

$$\bar{\boldsymbol{\Sigma}}_{1:i, 1:i} = \boldsymbol{\Sigma}_{1:i, 1:i}. \quad (38)$$

For this purpose, we define the following partitioning

$$\bar{\boldsymbol{\Omega}}_{1:i, 1:i} = \begin{bmatrix} \bar{\boldsymbol{\Omega}}_{1:i-1, 1:i-1} & \boldsymbol{\omega}_i \\ \boldsymbol{\omega}_i^T & \boldsymbol{\omega}_i \end{bmatrix} \quad (39)$$

$$\bar{\boldsymbol{\Sigma}}_{1:i, 1:i} = \begin{bmatrix} \bar{\boldsymbol{\Sigma}}_{1:i-1, 1:i-1} & \boldsymbol{\sigma}_i \\ \boldsymbol{\sigma}_i^T & \boldsymbol{\sigma}_i \end{bmatrix}. \quad (40)$$

We begin by analyzing $\boldsymbol{\omega}_i$ and $\boldsymbol{\omega}_i$. The matrix $\bar{\mathbf{U}}$ is p -banded and where $r \geq p$. Thus, the inner banding in (34) can be omitted

$$\bar{\boldsymbol{\Omega}}_{1:i, 1:i} = \mathcal{B}_q(\bar{\mathbf{U}}_{1:i, 1:i} \boldsymbol{\Sigma}_{1:i, 1:i} \bar{\mathbf{U}}_{1:i, 1:i}^T) \quad (41)$$

$$\boldsymbol{\omega}_i = \begin{bmatrix} \mathbf{0} \\ \bar{\mathbf{U}}_{q_i, 1:i} \boldsymbol{\Sigma}_{1:i, 1:i} \bar{\mathbf{U}}_{i, 1:i}^T \end{bmatrix} \quad (42)$$

$$\stackrel{(32)}{=} \begin{bmatrix} \mathbf{0} \\ \bar{\mathbf{U}}_{q_i, q_i} \boldsymbol{\Sigma}_{q_i, 1:i} \bar{\mathbf{U}}_{i, 1:i}^T \end{bmatrix} \quad (43)$$

$$\boldsymbol{\omega}_i = \bar{\mathbf{U}}_{i, 1:i} \boldsymbol{\Sigma}_{1:i, 1:i} \bar{\mathbf{U}}_{i, 1:i}^T. \quad (44)$$

Due to the triangular structure, we have

$$\begin{aligned} [\bar{\mathbf{U}}^{-1}]_{1:i-1, 1:i-1} \boldsymbol{\omega}_i &= \begin{bmatrix} \mathbf{0} \text{ (triangular)} \\ \overbrace{[\bar{\mathbf{U}}^{-1}]_{\bar{q}_i, q_i} \bar{\mathbf{U}}_{q_i, q_i} \boldsymbol{\Sigma}_{q_i, 1:i} \bar{\mathbf{U}}_{i, 1:i}^T} \\ \mathbf{I} \text{ (triangular)} \\ \overbrace{[\bar{\mathbf{U}}^{-1}]_{q_i, q_i} \bar{\mathbf{U}}_{q_i, q_i} \boldsymbol{\Sigma}_{q_i, 1:i} \bar{\mathbf{U}}_{i, 1:i}^T} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\Sigma}_{q_i, 1:i} \bar{\mathbf{U}}_{i, 1:i}^T \end{bmatrix}. \end{aligned} \quad (45)$$

$$[\bar{\mathbf{U}}^{-1}]_{1:i, 1:i} = \begin{bmatrix} \bar{\mathbf{U}}_{1:i-1, 1:i-1}^{-1} & \mathbf{0} \\ -\bar{\mathbf{U}}_{i, 1:i-1} \bar{\mathbf{U}}_{1:i-1, 1:i-1}^{-1} & \mathbf{1} \end{bmatrix}. \quad (46)$$

Finally, we complete the induction by deriving σ_i and σ_i

$$\begin{aligned}
\sigma_i &= [\bar{\mathbf{U}}^{-1}]_{1:i-1,1:i} \bar{\boldsymbol{\Omega}}_{1:i,1:i} [\bar{\mathbf{U}}^{-T}]_{1:i,i} \\
&\stackrel{(46)}{=} -\bar{\mathbf{U}}_{1:i-1,1:i-1}^{-1} \bar{\boldsymbol{\Omega}}_{1:i-1,1:i-1} \\
&\quad \cdot \bar{\mathbf{U}}_{1:i-1,1:i-1}^{-T} \bar{\mathbf{U}}_{i,1:i-1}^T + \bar{\mathbf{U}}_{1:i-1,1:i-1}^{-1} \boldsymbol{\omega}_i \\
&\stackrel{(35)}{=} -\bar{\boldsymbol{\Sigma}}_{1:i-1,1:i-1} \bar{\mathbf{U}}_{i,1:i-1}^T + \bar{\mathbf{U}}_{1:i-1,1:i-1}^{-1} \boldsymbol{\omega}_i \\
&\stackrel{(37)}{=} -\boldsymbol{\Sigma}_{1:i-1,1:i-1} \bar{\mathbf{U}}_{i,1:i-1}^T + \bar{\mathbf{U}}_{1:i-1,1:i-1}^{-1} \boldsymbol{\omega}_i \\
&\stackrel{(45)}{=} \begin{bmatrix} -\boldsymbol{\Sigma}_{q_i,1:i-1} \bar{\mathbf{U}}_{i,1:i-1}^T \\ -\boldsymbol{\Sigma}_{q_i,1:i-1} \bar{\mathbf{U}}_{i,1:i-1}^T + \boldsymbol{\Sigma}_{q_i,1:i} \bar{\mathbf{U}}_{i,1:i}^T \end{bmatrix} \\
&\stackrel{(32)}{=} \bar{\mathbf{U}}_{i,i=1}^{-1} \boldsymbol{\Sigma}_{1:i-1,i}
\end{aligned} \tag{47}$$

and

$$\begin{aligned}
\sigma_i &= [\bar{\mathbf{U}}^{-1}]_{i,1:i} \bar{\boldsymbol{\Omega}}_{1:i,1:i} [\bar{\mathbf{U}}^{-T}]_{1:i,i} \\
&\stackrel{(46)}{=} \bar{\mathbf{U}}_{i,1:i-1} \bar{\mathbf{U}}_{1:i-1,1:i-1}^{-1} \bar{\boldsymbol{\Omega}}_{1:i-1,1:i-1} \bar{\mathbf{U}}_{1:i-1,1:i-1}^{-T} \\
&\quad \cdot \bar{\mathbf{U}}_{i,1:i-1}^T - 2\bar{\mathbf{U}}_{i,1:i-1} \bar{\mathbf{U}}_{1:i-1,1:i-1}^{-1} \boldsymbol{\omega}_i + \boldsymbol{\omega}_i \\
&\stackrel{(37),(45)}{=} \bar{\mathbf{U}}_{i,1:i-1} \boldsymbol{\Sigma}_{1:i-1,1:i-1} \bar{\mathbf{U}}_{i,1:i-1}^T \\
&\quad - 2\bar{\mathbf{U}}_{i,q_i} \boldsymbol{\Sigma}_{q_i,1:i} \bar{\mathbf{U}}_{i,1:i}^T + \bar{\mathbf{U}}_{i,1:i} \boldsymbol{\Sigma}_{1:i,1:i} \bar{\mathbf{U}}_{i,1:i}^T \\
&\stackrel{(32)}{=} \bar{\mathbf{U}}_{i,i=1}^{-1} \boldsymbol{\Sigma}_{i,i}
\end{aligned} \tag{48}$$

Theorem 1 extends classical results on MA(q) and AR(p) to the case of ARMA(p, q), and shows that, under assumption (30), the covariance matrix can be completed given its $r = p + q$ leading diagonals. The assumption is similar to classical invertibility assumptions in stationary ARMA process identification via IVs [10]–[12]. Numerical results in Section V with randomly generated parameters suggest that it usually holds in practice. However, it is possible to construct specific counter examples as follows.

Theorem 2: There exist ARMA(p, q) covariances which cannot be completed from their main $r = p + q$ band.

Proof: Consider the following counter example. First, define the ARMA(1, 1) matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 1 & .5 & .25 \\ 1 & 2 & 0 & 0 \\ .5 & 0 & 1.5 & 1.75 \\ .25 & 0 & 1.75 & 3.375 \end{bmatrix} \tag{49}$$

whose parameters are

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ .5 & 0 & 0 & 0 \\ 0 & .5 & 0 & 0 \\ 0 & 0 & .5 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ .5 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}. \tag{50}$$

Next, define the ARMA(1, 1) matrix

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 1 & .5 & 0 \\ 1 & 2 & 0 & 0 \\ .5 & 0 & 1.5 & 1.75 \\ 0 & 0 & 1.75 & 3.375 \end{bmatrix} \tag{51}$$

whose parameters are

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ .5 & 0 & 0 & 0 \\ 0 & .5 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ .5 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 1.75 & \sqrt{.3125} \end{bmatrix} \tag{52}$$

Both are legitimate ARMA(1, 1) matrices. Their main r -bands are identical but each has a different value in the (4, 1)'th entry which needs to be completed. Thus, it is impossible to uniquely decide between them given the main band. ■

IV. COVARIANCE ESTIMATION

In this section, we consider the estimation of the covariance matrix of an observed sample of random vectors. Specifically, let \mathbf{x} be a length d , zero mean normal vector with covariance $\boldsymbol{\Sigma}$. Our goal is to estimate $\boldsymbol{\Sigma}$ based on n independent and identically distributed realizations of \mathbf{x} denoted by $\mathbf{x}[i]$ for $i = 1, \dots, n$.

The simplest estimator is the well known sample covariance

$$\mathbf{S} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}[i] \mathbf{x}^T[i]. \tag{53}$$

In the Gaussian case, it coincides with the unstructured Maximum Likelihood (ML) estimate as long as $n \geq d$ and is therefore consistent and efficient when $n \gg d$. Unfortunately, the number of samples in many practical applications is not sufficient, and better performance may be obtained through low order parametric models.

The standard methods for estimation in pure MA and AR structures use only the main band of the sample covariance [1]. In what follows, we extend this approach to ARMA models. Our estimator is obtained by taking the \mathbf{S} matrix and applying the \mathcal{AC} procedure (see Section III) to it, with $r = p + q$. In other words, it takes the main r band of the \mathbf{S} matrix and uses \mathcal{AC} to obtain a new matrix $\hat{\boldsymbol{\Sigma}}$. The matrix $\hat{\boldsymbol{\Sigma}}$ will serve as our estimator of the true covariance $\boldsymbol{\Sigma}$. We now analyze the number of samples needed for this scheme to achieve a desired accuracy.

As in similar analyses, we require certain conditions on the $\boldsymbol{\Sigma}$ that generate the data. This is a common approach in analyzing the covariance estimation and its dependence on dimension (e.g., see [1]). Specifically, we shall assume that these belong to a subclass of ARMA models of arbitrary dimension d . Define the following set of $\boldsymbol{\Sigma}$ matrices:

$$\mathcal{U}(\sigma, \zeta, p, q) = \left\{ \begin{array}{l} \boldsymbol{\Sigma} \in \text{ARMA}(p, q) : \\ \max\{\|\boldsymbol{\Sigma}\|_2, \|\boldsymbol{\Sigma}^{-1}\|_2, \|\bar{\mathbf{U}}\|_2, \|\bar{\mathbf{U}}^{-1}\|_2, \|\bar{\boldsymbol{\Omega}}\|_2\} \leq \sigma \\ \max\{\|\boldsymbol{\Sigma}_{p_i, iv_i}\|_2, \|\boldsymbol{\Sigma}_{p_i, iv_i}^\dagger\|_2\} \leq \zeta, \quad i = 1, \dots, d. \\ \boldsymbol{\Sigma}_{p_i, \bar{q}_i} \in \mathcal{R}(\boldsymbol{\Sigma}_{p_i, iv_i}), \quad i = 1, \dots, d. \end{array} \right\} \tag{54}$$

It is instructive to explain the meaning of the above conditions on $\boldsymbol{\Sigma} \in \mathcal{U}(\sigma, \zeta, p, q)$. The first condition is similar to standard

assumptions in high dimensional analysis of covariance estimation methods. It basically states that the elements and the parameters of the unknown covariance and its inverse are bounded by constants which do not depend on the dimension. Here we also require similar conditions on the matrices $\bar{\mathbf{U}}, \bar{\mathbf{\Omega}}$ which are obtained as intermediate steps of \mathcal{AC} applied to $\mathbf{\Sigma}$ (we assume that \mathcal{AC} is applied with $r = p + q$). The second condition addresses the continuity of the pseudo-inversions in (26), and guarantees stability under perturbation.¹ The third condition is simply (30), which is necessary for \mathcal{AC} to work.

The following theorem states that when \mathbf{S} is close to $\mathbf{\Sigma}$, then our estimator $\hat{\mathbf{\Sigma}}$ will also be close to $\mathbf{\Sigma}$. We shall later use it in Theorem 4 to obtain convergence rates for our estimator.

Theorem 3: Let $\mathbf{\Sigma}$ be a matrix in $\mathcal{U}(\sigma, \zeta, p, q)$. Assume that the matrix \mathbf{S} satisfies

$$\|\mathbf{S} - \mathbf{\Sigma}\|_{\max} \leq \epsilon_S \quad (55)$$

where

$$\epsilon_S \leq \min \left\{ \frac{1}{2\zeta r}, \frac{1}{2\sigma C_U} \right\} \quad (56)$$

Then the matrix $\hat{\mathbf{\Sigma}}$, which results from applying \mathcal{AC} to $\mathcal{B}_r(\mathbf{S})$, satisfies

$$\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_2 \leq C_T \epsilon_S \quad (57)$$

where

$$C_T = 7 \max\{C_{U-1}, C_{\Omega}\} \max^2\{C_{U-1}\epsilon_S, C_{\Omega}\epsilon_S, \sigma\} \quad (58)$$

The constants C_U, C_{Ω}, C_{U-1} are functions of σ and ζ and do not depend on the dimension d . They are defined in (65), (71), and (74) below.

Proof: Recall that \mathcal{AC} calculates the matrices $\bar{\mathbf{A}}, \bar{\mathbf{U}}, \bar{\mathbf{\Omega}}$ which are then used to obtain the completed matrix $\bar{\mathbf{\Sigma}}$. We will be interested in two such sets of matrices. Those obtained when applying \mathcal{AC} to the true covariance $\mathbf{\Sigma}$, and those obtained when applying \mathcal{AC} to the sample covariance \mathbf{S} . The former will be denoted by $\bar{\mathbf{A}}, \bar{\mathbf{U}}, \bar{\mathbf{\Omega}}, \bar{\mathbf{\Sigma}}$ and the latter by $\hat{\mathbf{A}}, \hat{\mathbf{U}}, \hat{\mathbf{\Omega}}, \hat{\mathbf{\Sigma}}$. Since under assumption (30), applying \mathcal{AC} to $\mathbf{\Sigma}$ results in $\mathbf{\Sigma}$ itself, we have that:

$$\bar{\mathbf{\Sigma}} = \bar{\mathbf{U}}^{-1} \bar{\mathbf{\Omega}} \bar{\mathbf{U}} = \mathbf{\Sigma} \quad (59)$$

Recall also that $\hat{\mathbf{A}}, \hat{\mathbf{U}}, \hat{\mathbf{\Omega}}$ are given by:

$$\begin{aligned} \hat{\mathbf{A}}_{i,p_i} &= \mathbf{S}_{i,iv_i} \mathbf{S}_{p_i,iv_i}^\dagger \\ \hat{\mathbf{U}} &= \mathbf{I} - \hat{\mathbf{A}} \\ \hat{\mathbf{\Omega}} &= \mathcal{B}_q(\hat{\mathbf{U}} \mathcal{B}_r(\mathbf{S}) \hat{\mathbf{U}}^T) \\ \hat{\mathbf{\Sigma}} &= \hat{\mathbf{U}}^{-1} \hat{\mathbf{\Omega}} \hat{\mathbf{U}}^{-T} \end{aligned} \quad (60)$$

In what follows, we will analyze how $\hat{\mathbf{A}}, \hat{\mathbf{U}}, \hat{\mathbf{\Omega}}$ differ from $\bar{\mathbf{A}}, \bar{\mathbf{U}}, \bar{\mathbf{\Omega}}$ and use this to analyze how $\hat{\mathbf{\Sigma}}$ differs from $\bar{\mathbf{\Sigma}}$.

We begin by analyzing $\hat{\mathbf{A}}$. From (56) we have that

$$r\epsilon_S \leq \frac{1}{2\zeta}. \quad (61)$$

¹We believe this condition may be dropped via more careful analysis.

Thus,

$$\begin{aligned} \|\mathbf{\Sigma}_{p_i,iv_i} - \mathbf{S}_{p_i,iv_i}\|_2 &\leq r \|\mathbf{\Sigma}_{p_i,iv_i} - \mathbf{S}_{p_i,iv_i}\|_{\max} \\ &\leq r\epsilon_S \leq \frac{1}{2\zeta} \end{aligned}$$

and due to Lemma 2

$$\|\mathbf{\Sigma}_{p_i,iv_i}^\dagger - \mathbf{S}_{p_i,iv_i}^\dagger\|_2 \leq 2 \max \left\{ \zeta^2, \frac{1}{4\zeta^2} \right\} r\epsilon_S. \quad (62)$$

Using the sub-multiplicative and triangle inequalities we get

$$\begin{aligned} \|\hat{\mathbf{A}}_{i,p_i} - \bar{\mathbf{A}}_{i,p_i}\|_{\infty} &\leq \|\hat{\mathbf{A}}_{i,p_i} - \bar{\mathbf{A}}_{i,p_i}\|_2 \\ &\leq \|\mathbf{\Sigma}_{i,iv_i} - \mathbf{S}_{i,iv_i}\|_2 \|\mathbf{\Sigma}_{p_i,iv_i}^\dagger\|_2 \\ &\quad + \|\mathbf{S}_{i,iv_i}\|_2 \|\mathbf{\Sigma}_{p_i,iv_i}^\dagger - \mathbf{S}_{p_i,iv_i}^\dagger\|_2 \\ &\leq \|\mathbf{\Sigma}_{i,iv_i} - \mathbf{S}_{i,iv_i}\|_2 \|\mathbf{\Sigma}_{p_i,iv_i}^\dagger\|_2 \\ &\quad + (\|\mathbf{\Sigma}_{i,iv_i} - \mathbf{S}_{i,iv_i}\|_2 + \|\mathbf{\Sigma}_{i,iv_i}\|_2) \|\mathbf{\Sigma}_{p_i,iv_i}^\dagger - \mathbf{S}_{p_i,iv_i}^\dagger\|_2 \\ &\leq r\epsilon_S \zeta + 2(r\epsilon_S + \sigma) \max \left\{ \zeta^2, \frac{1}{4\zeta^2} \right\} r\epsilon_S. \end{aligned} \quad (63)$$

The matrices $\hat{\mathbf{A}}$ and $\bar{\mathbf{A}}$ are p -banded and lower triangular. Thus,

$$\begin{aligned} \|\hat{\mathbf{U}} - \bar{\mathbf{U}}\|_2 &= \|\hat{\mathbf{A}} - \bar{\mathbf{A}}\|_2 \\ &\leq \sqrt{\|\hat{\mathbf{A}} - \bar{\mathbf{A}}\|_1 \|\hat{\mathbf{A}} - \bar{\mathbf{A}}\|_{\infty}} \\ &\leq \sqrt{p \|\hat{\mathbf{A}} - \bar{\mathbf{A}}\|_{\max} p \|\hat{\mathbf{A}} - \bar{\mathbf{A}}\|_{\max}} \\ &= p \|\hat{\mathbf{A}} - \bar{\mathbf{A}}\|_{\max} \\ &\leq p \left[r\zeta + 2(r\epsilon_S + \sigma) \max \left\{ \zeta^2, \frac{1}{4\zeta^2} \right\} r \right] \epsilon_S \\ &= C_U \epsilon_S \end{aligned} \quad (64)$$

where

$$C_U = p \left[r\zeta + 2(r\epsilon_S + \sigma) \max \left\{ \zeta^2, \frac{1}{4\zeta^2} \right\} r \right]. \quad (65)$$

Next, we analyze the difference between $\bar{\mathbf{\Omega}}$ and $\hat{\mathbf{\Omega}}$. We define

$$\begin{aligned} \bar{\mathbf{Q}} &= \bar{\mathbf{U}} \mathcal{B}_r(\mathbf{\Sigma}) \bar{\mathbf{U}}^T \\ \hat{\mathbf{Q}} &= \hat{\mathbf{U}} \mathcal{B}_r(\mathbf{S}) \hat{\mathbf{U}}^T \end{aligned}$$

so that

$$\begin{aligned} \bar{\mathbf{\Omega}} &= \mathcal{B}_q(\bar{\mathbf{Q}}) \\ \hat{\mathbf{\Omega}} &= \mathcal{B}_q(\hat{\mathbf{Q}}). \end{aligned}$$

We first note two facts regarding the banded matrices $\mathcal{B}_r(\mathbf{\Sigma})$ and $\mathcal{B}_r(\mathbf{S})$. First, due to Lemma 1 the matrix $\mathcal{B}_r(\mathbf{\Sigma})$ satisfies

$$\|\mathcal{B}_r(\mathbf{\Sigma})\|_2 \leq (2r + 1)\sigma. \quad (66)$$

Second, the two banded matrices satisfy

$$\begin{aligned} \|\mathcal{B}_r(\mathbf{\Sigma}) - \mathcal{B}_r(\mathbf{S})\|_2 &\leq \|\mathcal{B}_r(\mathbf{\Sigma}) - \mathcal{B}_r(\mathbf{S})\|_1 \\ &\leq (2r + 1) \|\mathbf{S} - \mathbf{\Sigma}\|_{\max} \\ &\leq (2r + 1) \epsilon_S, \end{aligned} \quad (67)$$

where we have used basic inequalities between norms of symmetric matrices, and (55). This inequality is actually the core of the pure MA(q) analysis. It states that the operator norm of the banded matrix error is bounded by the element wise error and is independent of the dimension which may be much larger. We next combine inequalities (66) and (67) with Lemma 3 to obtain

$$\begin{aligned} \|\hat{\mathbf{Q}} - \bar{\mathbf{Q}}\|_2 &\leq 7 \max\{\|\hat{\mathbf{U}} - \bar{\mathbf{U}}\|_2, \|\mathcal{B}_r(\mathbf{\Sigma} - \mathbf{S})\|_2\} \\ &\quad \cdot \max^2\{\|\hat{\mathbf{U}} - \bar{\mathbf{U}}\|_2, \|\mathcal{B}_r(\mathbf{\Sigma} - \mathbf{S})\|_2, \|\bar{\mathbf{U}}\|_2, \|\mathcal{B}_r(\mathbf{\Sigma})\|_2\} \\ &\leq 7 \max\{C_U \epsilon_S, (2r+1)\epsilon_S\} \\ &\quad \cdot \max^2\{C_U \epsilon_S, (2r+1)\epsilon_S, \sigma, (2r+1)\sigma\} \\ &= C_Q \epsilon_S \end{aligned} \quad (68)$$

where

$$\begin{aligned} C_Q &= 7 \max\{C_U, (2r+1)\} \\ &\quad \cdot \max^2\{C_U \epsilon_S, (2r+1)\epsilon_S, \sigma, (2r+1)\sigma\}. \end{aligned} \quad (69)$$

Using Lemma 1 again yields

$$\begin{aligned} \|\hat{\mathbf{\Omega}} - \bar{\mathbf{\Omega}}\|_2 &= \|\mathcal{B}_q(\hat{\mathbf{Q}}) - \mathcal{B}_q(\bar{\mathbf{Q}})\|_2 \\ &\leq (2q+1)C_Q \epsilon_S \\ &\leq C_\Omega \epsilon_S \end{aligned} \quad (70)$$

where we define

$$C_\Omega = (2q+1)C_Q. \quad (71)$$

Finally, we turn to the overall error in $\mathbf{\Sigma}$. If

$$C_U \epsilon_S \leq \frac{1}{2\sigma} \quad (72)$$

then due to Lemma 2

$$\begin{aligned} \|\hat{\mathbf{U}}^{-1} - \bar{\mathbf{U}}^{-1}\|_2 &\leq 2 \max\left\{\sigma^2, \frac{1}{4\sigma^2}\right\} C_U \epsilon_S \\ &= C_{U^{-1}} \epsilon_S \end{aligned} \quad (73)$$

where

$$C_{U^{-1}} = 2 \max\left\{\sigma^2, \frac{1}{4\sigma^2}\right\} C_U. \quad (74)$$

Recall that $\hat{\mathbf{\Sigma}} = \hat{\mathbf{U}}^{-1} \hat{\mathbf{\Omega}} \hat{\mathbf{U}}^{-T}$, so that application of Lemma 3 yields

$$\begin{aligned} \|\hat{\mathbf{\Sigma}} - \bar{\mathbf{\Sigma}}\|_2 &\leq 7 \max\{C_{U^{-1}}, C_\Omega\} \\ &\quad \cdot \max^2\{C_{U^{-1}} \epsilon_S, C_\Omega \epsilon_S, \sigma\} \epsilon_S \end{aligned} \quad (75)$$

as long as (61) and (72) are satisfied. Rearranging these conditions results in (57) as required. ■

Theorem 3 can now be used to derive the non-asymptotic behavior of our estimator. The theorem below states that when estimating a $\mathbf{\Sigma}$ of dimension d with n samples, the estimation error is of order $\sqrt{\frac{\log d}{n}}$, implying only a weak dependence on d . This is similar to results obtained for MA and AR estimators in [1].

Theorem 4: Let $\mathbf{\Sigma}$ be a matrix in $\mathcal{U}(\sigma, \zeta, p, q)$. Assume that \mathbf{S} in (53) is a sample covariance constructed using n independent and identically distributed realizations of a multivariate Gaussian distribution with zero mean and covariance $\mathbf{\Sigma}$. Then, there exist constants C_1, C_2 and $\delta > 0$ which depend only on σ , such that for

$$n \geq \frac{2M^2 \log d}{C_2 \min^2\left\{\delta, \frac{1}{2\zeta r}, \frac{1}{\alpha C_U}\right\}} \quad (76)$$

the inequality

$$\|\hat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_2 \leq \frac{C_T M}{\sqrt{0.5C_2}} \sqrt{\frac{\log d}{n}} \quad (77)$$

holds with probability greater than $1 - \frac{C_1}{d^{2M^2-2}}$, where the constants are defined in Theorem 3, and M is arbitrary.

Proof: Lemma A.3 and the union bound in page 221 of [1] show that

$$\Pr\left(\|\mathbf{S} - \mathbf{\Sigma}\|_{\max} \leq \sqrt{\frac{\log d^2}{nC_2}} M\right) \geq 1 - \frac{C_1}{d^{2M^2-2}} \quad (78)$$

for

$$\sqrt{\frac{\log d^2}{nC_2}} M \leq \delta. \quad (79)$$

Combining this with the deterministic inequalities in Theorem 3 yields the required result. ■

Finally, we now briefly discuss the issue of order selection. In order to enjoy the advantages of low order parametric models it is vital to choose the orders p and q in an efficient and accurate manner. In some applications, these bandwidths are a priori known from previous experiments or the physical characteristics of the system. In other scenarios, these parameters must be inferred from the available data. There are different approaches to this task. Hypothesis testing and expected likelihood based criteria have been recently proposed in the context of time varying AR covariance models [5], [15], [18]. Alternatively, cross validation approaches can be used as in [1], [2]. We follow these last works and propose a simple K -fold cross validation procedure. We divide the indices set $S = \{1, \dots, n\}$ into $K = 10$ non-overlapping groups S_k such that $S = \cup_{k=1}^K S_k$, and consider a set of candidate pairs (p_l, q_l) for $l = 1, \dots, L$. We define $\hat{\mathbf{\Sigma}}_l[k]$ as the output of \mathcal{A} applied to

$$\mathbf{S}[k] = \frac{1}{\frac{K-1}{K}n} \sum_{i \in S \setminus S_k} \mathbf{x}_i \mathbf{x}_i^T \quad (80)$$

with parameters (p_l, q_l) . Our criterion for choosing l is then mean squared Frobenius error:

$$\hat{l} = \arg \min_{l \in [1, \dots, L]} \sum_{k=1}^K \left\| \hat{\mathbf{\Sigma}}_l[k] - \frac{1}{K} \sum_{i \in S_k} \mathbf{x}_i \mathbf{x}_i^T \right\|_F^2 \quad (81)$$

After finding this optimal \hat{l} , we re-estimate the covariance using the bands $(p_{\hat{l}}, q_{\hat{l}})$ using the full set of samples.

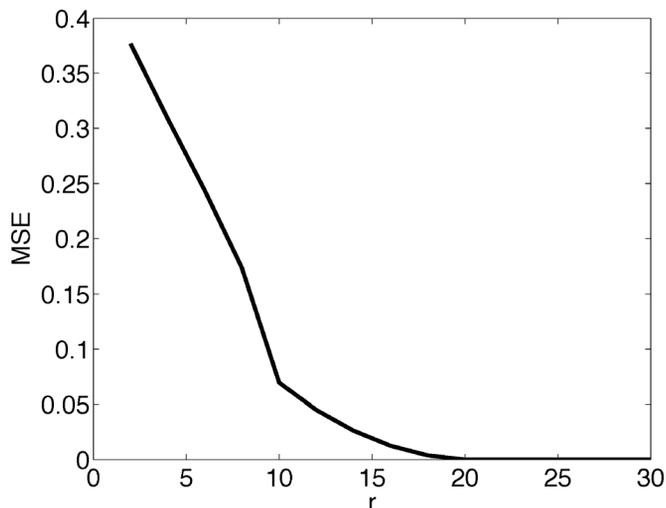


Fig. 1. Error in reconstruction of ARMA(10, 10) covariances of dimension $d = 100$ from their main r band, as a function of r .

V. EXPERIMENTS

In what follows we experiment with the \mathcal{AC} procedure on two tasks. The first is the completion of ARMA covariance matrices from their main r band (see Section III). The second is estimation of ARMA models from finite samples (see Section IV).

A. Evaluating the Completion Algorithm

In Section III we show that the \mathcal{AC} procedure can be used to complete an ARMA covariance from its main r band. The completion is guaranteed to work under the condition in (30), which appears to hold quite generally. To show this, we draw well conditioned ARMA models with $p = 10, q = 10$ and $d = 100$ from a random ensemble. At each experiment, we randomly generate two 10-banded positive definite matrices as the Hadamard product between Wishart matrices with d degrees of freedom and banded Toeplitz masks with (i, j) elements equal to $0.95^{|i-j|}$ within the band. We use one of these matrices to compute the AR component \mathbf{A} via (9), and the other to compute the MA component \mathbf{B} via (4). Plugging these \mathbf{A} and \mathbf{B} into (16) yields a well conditioned ARMA covariance. We then use \mathcal{AC} to complete Σ from its main r band, for different values of r . Fig. 1 shows the Frobenius norm error of this procedure for different r values averaged over 100 independent experiments. It can be seen that we achieve perfect reconstruction for $r \geq p + q$.

B. Evaluating the Estimation Algorithm

In what follows we evaluate our ARMA estimation method (Section IV) on various simulated covariance matrices. We compare four different estimators: the sample covariance, an autoregressive model $\text{AR}(p + q)$, a moving average model $\text{MA}(p + q)$, and our autoregressive moving average model $\text{ARMA}(p, q)$. Note that to ensure fairness, all three parametric models have the same number of degrees of freedom. The ARMA model is estimated using the matrix completion approach corresponding to \mathcal{AC} with $r = p + q$. The AR models

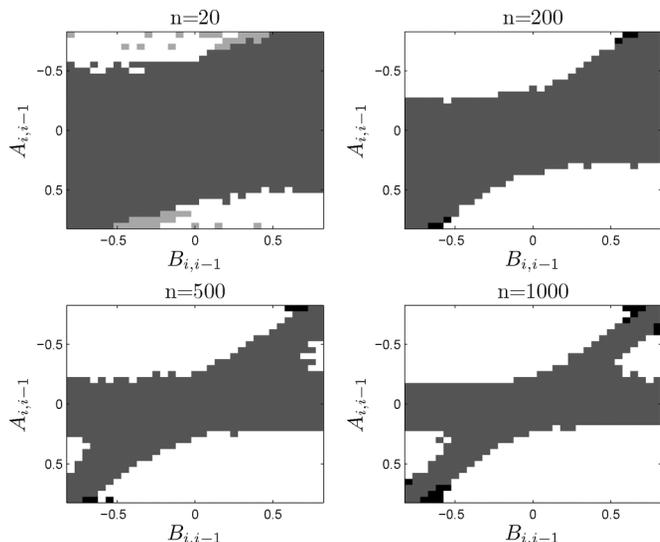


Fig. 2. Best estimators as a function of the ARMA model: ARMA is white, AR, MA and sample covariance are grays.

are estimated using standard Yule Walker estimation and the MA models are estimated using the procedure in [2].²

In our first experiment, we examine the performance as a function of the ARMA(1, 1) parameters. Specifically, we let $d = 10$ and define \mathbf{A} and \mathbf{B} as Toeplitz matrices with $\mathbf{A}_{i,i-1}$ and $\mathbf{B}_{i,i-1}$ evaluated on the grid $[-0.8, -0.7, \dots, 0.7, 0.8]$. For each point, we simulate all four estimators and compute the squared Frobenius norm error in the covariance. In Fig. 2 we plot the labels of the best estimators for each pole and zero values and for different numbers of samples. As expected, the advantage of the ARMA estimator is emphasized when the AR and MA components are of the same order with different signs. Also, the region in which ARMA dominates becomes larger with the number of samples.

In our second experiment, we consider the stationary narrow band ARMA(2, 2) example in [19]:

$$x(t) = 1.4x(t-1) - 0.95x(t-2) + w(t) - 0.86w(t-1) + 0.431w(t-2) \quad (82)$$

We construct the time varying ARMA covariance model as follows. First we define a size $d = 500$ Toeplitz matrix $\Sigma_{\text{stationary}}$ with the corresponding parameters $\mathbf{A}_{i,i} = 0, \mathbf{A}_{i,i-1} = 1.4, \mathbf{A}_{i,i-2} = -0.95, \mathbf{B}_{i,i} = 1, \mathbf{B}_{i,i-1} = -0.86$ and $\mathbf{B}_{i,i-2} = 0.431$. Next, following [5] in the context of time varying AR covariance estimation, we add non-stationarity using phase path variations along propagation channels. We define the Doppler matrix as:

$$\mathbf{D} = \text{diag} \left[\exp \left(\frac{i2\pi k}{d} \left[1 - \cos \frac{2\pi t}{d} \right] \right) \right] \quad (83)$$

with $k = 40$, and the overall time varying covariance matrix as

$$\Sigma_{\text{time varying}} = \mathbf{D} \Sigma_{\text{stationary}} \mathbf{D}^H. \quad (84)$$

²The MA estimation method in [2] typically outperforms simple banding of Σ .

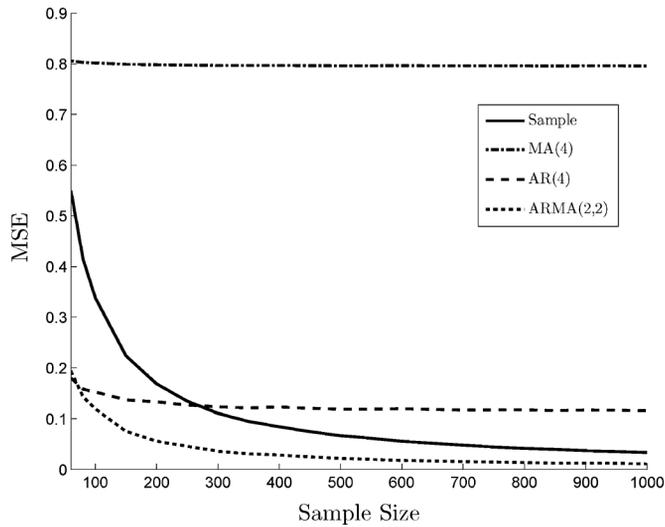


Fig. 3. Error in Gaussian ARMA(2, 2) as a function of the number of samples.

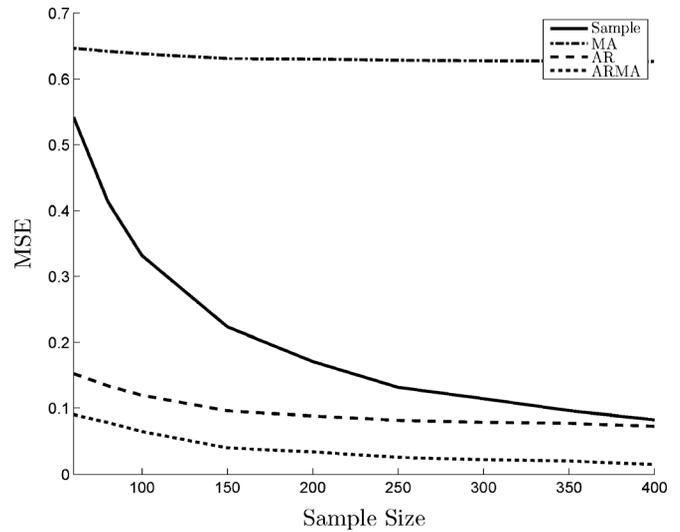


Fig. 5. Error in Gaussian ARMA(2, 2) with model selection as a function of the number of samples.

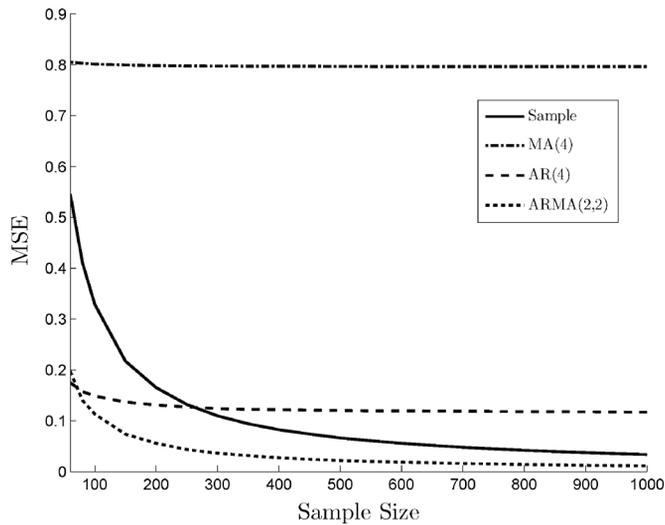


Fig. 4. Error in Uniform ARMA(2, 2) as a function of the number of samples.

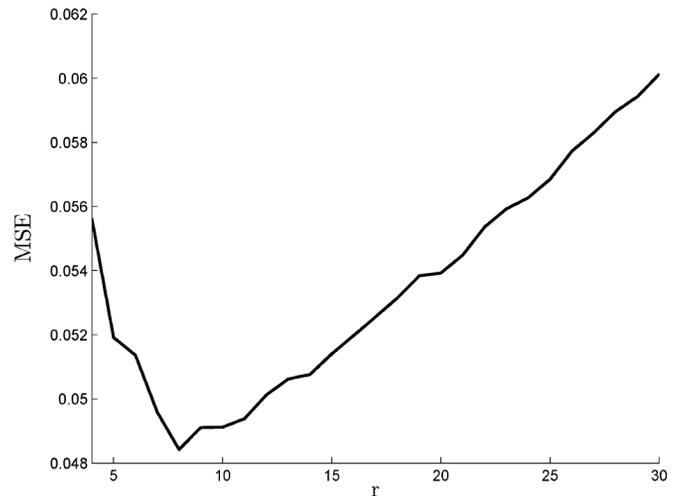


Fig. 6. Error in Gaussian ARMA(2, 2) as a function of the bandwidth r .

Note that this construction involves the use of complex variables. For this purpose, we replace the transpose and pseudo-inverse operators in \mathcal{A} with their complex counterparts, i.e., conjugate transpose and complex valued pseudo-inverse, respectively.

We report the results of four experiments performed on this model. In Fig. 3 we provide the normalized Frobenius norm errors of the AR(4), MA(4) and ARMA(2, 2) with estimators as a function of the number of samples. The samples are drawn from a multivariate normal distribution. In order to test the robustness to non Gaussian samples, we also repeat the experiment using uniform distribution for the driving noise \mathbf{w} as described in [19]. The results are provided in Fig. 4. Next, in Fig. 5 we examine the effect of model selection and run the simulation again without providing the estimators their band parameter, but choosing it using the proposed cross validation technique. In all three cases, it is easy to see the advantage of the ARMA estimator over its competitors.

Finally, we consider the performance as a function of the bandwidth parameter r of the \mathcal{A} algorithm. The previous results are all based on the choice $r = p + q = 4$. To illustrate the effect of r , we evaluate performance for different values of r . This is done for the ARMA(2, 2) model above with normal noise, and with a sample size of $n = 200$. Results are shown in Fig. 6. The curve suggests that there is an optimal bandwidth around $r \approx 8$ which can provide even better performance. Clearly, this comes with the additional cost of optimizing this additional hyper parameter.

In our third experiment, we consider the ARMA(4, 2) in Exercise C2.22 in the spectral analysis book [20]. As before, we construct a $d = 500$ Toeplitz covariance with the parameters $\mathbf{A}_{i,i} = 0, \mathbf{A}_{i,i-1} = -1.6408, \mathbf{A}_{i,i-2} = 2.2044, \mathbf{A}_{i,i-3} = -1.4808, \mathbf{A}_{i,i-4} = 0.8145, \mathbf{B}_{i,i} = 1, \mathbf{B}_{i,i-1} = 1.5857, \mathbf{B}_{i,i-2} = 0.9604$ and zero otherwise. We then add non stationarity as formulated in (84). We report the error as a function of the number of samples with and without cross validation in Figs. 7 and 8, respectively. Both figures

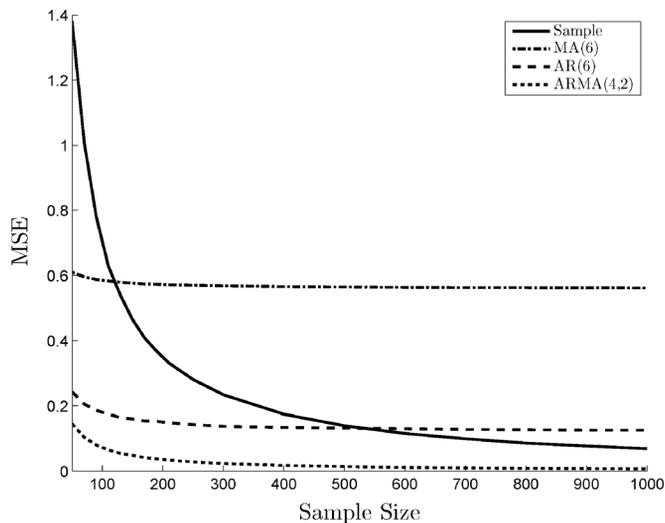


Fig. 7. Error in Gaussian ARMA(4, 2) as a function of the number of samples.

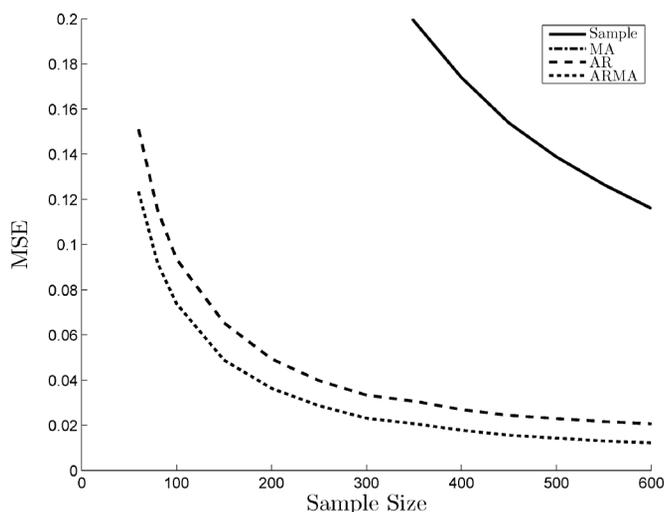


Fig. 8. Error in Gaussian ARMA(4, 2) with cross validation as a function of the number of samples (MA error is larger than 0.2).

demonstrate the advantages of ARMA approach both when the band parameters are a priori known, and when these must be estimated from the available data.

VI. DISCUSSION

In this paper, we introduced a low order parametric model for covariances based on a time varying ARMA structure. We considered the completion of such covariances from their main band, and their estimation from random samples. Our contributions generalize existing results on the special cases of pure MA and pure AR structures. In particular, we provide conditions under which our proposed ARMA estimates are consistent in the operator norm as long as $(\log d)/n \rightarrow 0$.

The ARMA(p, q) model assumes that the locations of the non-zero valued parameters are known in advance based on the ordering of the variables. Recently, there has been a growing interest in learning sparse models where the sparsity pattern has to be estimated from the data as well. This has been done in the context of sparse covariance matrices as well as sparse in-

verse covariance matrices [21]–[25]. Future work should address these approaches in the context of ARMA models. By interpreting the MA part as latent variables, such extensions will be related to the recent works on learning covariance matrices with hidden variables [26].

APPENDIX A TECHNICAL LEMMAS

Lemma 1: Let \mathbf{X} be an arbitrary matrix, then $\|\mathcal{B}_b(\mathbf{X})\|_2 \leq (2b + 1)\|\mathbf{X}\|_2$.

Proof: Let $\mathbf{B} = \mathcal{B}_b(\mathbf{O})$ where \mathbf{O} is the all ones matrix, then $\mathcal{B}_b(\mathbf{X}) = \mathbf{B} \circ \mathbf{X}$ where \circ is the Hadamard elementwise product. The spectral norm is Hadamard submultiplicative so that $\|\mathbf{B} \circ \mathbf{X}\|_2 \leq \|\mathbf{B}\|_2 \|\mathbf{X}\|_2$ [27]. Finally, $\|\mathbf{B}\|_2 \leq \|\mathbf{B}\|_1$ for any symmetric matrix and $\|\mathbf{B}\|_1 = (2b + 1)$.

Lemma 2 (Full Rank Pseudoinverse Perturbation): If \mathbf{A} is a tall full rank with $\|\mathbf{A}\|_2 \leq \alpha$, $\|\mathbf{A}^\dagger\|_2 \leq \alpha$, $\mathbf{B} = \mathbf{A} + \mathbf{E}$ and $\|\mathbf{E}\|_2 \leq \frac{1}{2\alpha}$, then \mathbf{B} is full rank with

$$\|\mathbf{B}^\dagger - \mathbf{A}^\dagger\|_2 \leq 2 \max \left\{ \alpha^2, \frac{1}{4\alpha^2} \right\} \|\mathbf{E}\|_2 \quad (85)$$

Proof: Corollary 22 and Lemma 23 in [28] with $\epsilon = 1/2$ and $\frac{1+\sqrt{5}}{2} \leq 2$.

Lemma 3: Let \mathbf{X}_i for $i = 1, 2, 3$ satisfy $\|\mathbf{X}_i\|_2 \leq \nu$, and $\hat{\mathbf{X}}_i = \mathbf{X}_i + \mathbf{E}_i$ with $\|\mathbf{E}_i\|_2 \leq \delta$. Then,

$$\|\hat{\mathbf{X}}_1 \hat{\mathbf{X}}_2 \hat{\mathbf{X}}_3 - \mathbf{X}_1 \mathbf{X}_2 \mathbf{X}_3\|_2 \leq 7\delta \max^2 \{\delta, \nu\} \quad (86)$$

Proof: Apply the submultiplicative and triangle inequalities to

$$\begin{aligned} \hat{\mathbf{X}}_1 \hat{\mathbf{X}}_2 \hat{\mathbf{X}}_3 - \mathbf{X}_1 \mathbf{X}_2 \mathbf{X}_3 &= \mathbf{E}_1 \mathbf{E}_2 \mathbf{E}_3 + \mathbf{E}_1 \mathbf{E}_2 \mathbf{X}_3 + \mathbf{E}_1 \mathbf{X}_2 \mathbf{E}_3 \\ &+ \mathbf{E}_1 \mathbf{X}_2 \mathbf{X}_3 + \mathbf{X}_1 \mathbf{E}_2 \mathbf{E}_3 + \mathbf{X}_1 \mathbf{E}_2 \mathbf{X}_3 + \mathbf{X}_1 \mathbf{X}_2 \mathbf{E}_3 \end{aligned} \quad (87)$$

REFERENCES

- [1] P. J. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *Annal. Statist.*, vol. 36, no. 1, pp. 199–227, 2008.
- [2] A. J. Rothman, E. Levina, and J. Zhu, "A new approach to cholesky-based covariance regularization in high dimensions," *Biometrika*, vol. 97, no. 3, p. 539, 2010.
- [3] T. T. Cai, C. H. Zhang, and H. H. Zhou, "Optimal rates of convergence for covariance matrix estimation," *Annal. Statist.*, vol. 38, no. 4, pp. 2118–2144, 2010.
- [4] S. Chaudhuri, M. Drton, and T. S. Richardson, "Estimation of a covariance matrix with zeros," *Biometrika*, 2007.
- [5] Y. I. Abramovich, N. K. Spencer, and M. D. E. Turley, "Order estimation and discrimination between stationary and time-varying (TVAR) autoregressive models," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2861–2876, Jun. 2007.
- [6] Y. I. Abramovich, N. K. Spencer, and M. D. E. Turley, "Time-varying autoregressive (TVAR) models for multiple radar observations," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1298–1311, Apr. 2007.
- [7] A. Kavcic and J. M. F. Moura, "Matrices with banded inverses: Inversion algorithms and factorization of Gauss-Markov processes," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1495–1509, Apr. 2000.
- [8] A. Asif and J. M. F. Moura, "Block matrices with l-block-banded inverse: Inversion algorithms," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 630–642, Feb. 2005.
- [9] S. L. Lauritzen, *Graphical Models*. New York, ser. Oxford Statistical Science Series, 1996, vol. 17.
- [10] P. Stoica and M. Jansson, "MIMO system identification: State-space and subspace approximations versus transfer function and instrumental variables," *IEEE Tran. Signal Process.*, vol. 48, no. 11, pp. 3087–3099, Nov. 2000.

- [11] P. Stoica, T. Soderstrom, and B. Friedlander, "Optimal instrumental variable estimates of the AR parameters of an ARMA process," *IEEE Trans. Automat. Control*, vol. 30, no. 11, pp. 1066–1074, Nov. 1985.
- [12] M. Viberg, P. Stoica, and B. Ottersten, "Array processing in correlated noise fields based on instrumental variables and subspace fitting," *IEEE Trans. Signal Process.*, vol. 43, no. 5, pp. 1187–1199, May 1995.
- [13] C. Brito and J. Pearl, "A new identification condition for recursive models with correlated errors," *Structural Equation Model.*, vol. 9, no. 4, pp. 459–474, 2002.
- [14] M. Drton, M. Eichler, and T. S. Richardson, "Computing maximum likelihood estimates in recursive linear models with correlated errors," *J. Mach. Learn. Res.*, vol. 10, pp. 2329–2348, 2009.
- [15] Y. I. Abramovich, N. K. Spencer, and M. D. E. Turley, "Time-varying autoregressive (TVAR) models for multiple radar observations," *IEEE Trans. on Signal Process.*, vol. 55, no. 4, pp. 1298–1311, Apr. 2007.
- [16] B. Grone, C. R. Johnson, E. Marques de Sa, and H. Wolkowicz, "Positive definite completions of partial Hermitian matrices," *Linear Algebra Appl.*, vol. 58, pp. 109–124, 1985.
- [17] E. Ben David and B. Rajaratnam, "Positive definite completion problems for directed acyclic graphs," Arxiv Preprint arXiv:1201.0310 2011.
- [18] Y. Abramovich and B. A. Johnson, "Expected likelihood approach for covariance matrix estimation: Complex angular central Gaussian Case," in *Proc. IEEE SAM-2012*, June 2012, 2012.
- [19] A. K. Rao, Y. F. Huang, and S. Dasgupta, "ARMA parameter estimation using a novel recursive estimation algorithm with selective updating," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 3, pp. 447–457, 1990.
- [20] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ, USA: Pearson/Prentice Hall, 2005.
- [21] P. J. Bickel and E. Levina, "Covariance regularization by thresholding," *The Annal. Statist.*, vol. 36, no. 6, pp. 2577–2604, 2008.
- [22] N. Meinshausen and P. Buhlmann, "High-dimensional graphs and variable selection with the LASSO," *Annal. Statist.*, pp. 1436–1462, 2006.
- [23] A. J. Rothman, P. J. Bickel, E. Levina, and J. Zhu, "Sparse permutation invariant covariance estimation," *Electron. J. Statist.*, vol. 2, pp. 494–515, 2008.
- [24] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the LASSO," *Biostat.*, vol. 9, no. 3, pp. 432–441, July 2008.
- [25] G. Cao, L. R. Bacheaga, and C. A. Bouman, "The sparse matrix transform for covariance estimation and analysis of high dimensional signals," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 625–640, Mar., 2011.
- [26] V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky, "Latent variable graphical model selection via convex optimization," Arxiv Preprint arXiv:1008.1290 2010.
- [27] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, MA, USA: Cambridge University Press, 1991.
- [28] D. Hsu, S. M. Kakade, and T. Zhang, "A spectral algorithm for learning hidden Markov models," *J. Comput. Syst. Sci.*, 2012.



Ami Wiesel received the B.Sc. and M.Sc. degrees in electrical engineering from Tel-Aviv University, Tel-Aviv, Israel, in 2000 and 2002, respectively, and the Ph.D. degree in electrical engineering from the Technion—Israel Institute of Technology, Haifa, Israel, in 2007.

He was a postdoctoral fellow with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, in 2007–2009. Since Jan. 2010, he has been a Faculty Member at the Rachel and Selim Benin School of Computer Science and Engineering at the Hebrew University of Jerusalem, Jerusalem Israel.

Dr. Wiesel was a recipient of the Young Author Best Paper Award for a 2006 paper in the IEEE TRANSACTIONS ON SIGNAL PROCESSING and a Student Paper Award for a 2005 Workshop on Signal Processing Advances in Wireless Communications (SPAWC) paper. He was awarded the Weinstein Study Prize in 2002, the Intel Award in 2005, the Viterbi Fellowship in 2005 and 2007, and the Marie Curie Fellowship in 2008.



Ofir Bibi received the B.Sc. degree in physics and computer science from The Hebrew University, Jerusalem, Israel, in 2009. He is currently working toward the Ph.D. degree in The Alice and Jack Ormut International Ph.D. Program for outstanding students in Brain Research: Computation and Information Processing at The Hebrew University.

Mr. Bibi was awarded the Intel Award in 2005.



Amir Globerson (M'07) received the B.Sc. degree in computer science and physics from the Hebrew University, Jerusalem, Israel, in 1995. He received the Ph.D. degree in neural computation in 2007, also from the Hebrew University.

He was a postdoctoral fellow at the University of Toronto and at MIT from 2007 to 2009. Since September 2009 he has been a Faculty Member at the Rachel and Selim Benin School of Computer Science and Engineering at the Hebrew University of Jerusalem, Jerusalem, Israel.

Dr. Globerson was a recipient of two best paper awards at the Uncertainty in Artificial Intelligence (UAI) conference (2007 and 2008), and an outstanding student paper award at the Advances in Neural Information Processing Systems (NIPS) conference (2004). He was awarded the Rothschild Yad Hanadiv postdoctoral fellowship in 2008 and the HP labs innovation research award in 2010.