COMMUN. MATH. SCI.
Vol. 17, No. 5, pp. 1353–1386

REPRESENTING CONDITIONAL GRANGER CAUSALITY BY VECTOR AUTO-REGRESSIVE PARAMETERS*

YANYANG XIAO†, SONGTING LI‡, AND DOUGLAS ZHOU§

In memory of Professor David Shenou Cai

Abstract. Granger Causality (GC) has been widely applied to various scientific fields to reveal causal relationships between dynamical variables. The mathematical framework of GC is based on the vector auto-regression (VAR) model, and the GC value from one variable to the other is defined as the logarithmic ratio of the variance of two prediction errors obtained by excluding and including the second variable in the VAR model respectively. Besides its definition, GC shall also be reflected in the regression parameters of the VAR model, e.g., larger regression coefficients indicate stronger causal interactions in general. Yet an explicit description of how the GC value depends on the VAR parameters for a general multi-variable case remains lacking. In this work, we aim to bridge this gap by expressing conditional GC using the VAR parameters, which provides an alternative interpretation of GC with novel intuition. The analysis developed in this work may also benefit the study of the VAR model in the future.

Keywords. Granger Causality; VAR parameters; approximation.

AMS subject classifications. 62H20; 62F12; 62J05; 62M10.

1. Introduction

As a popular method to reveal causal relationship among multiple dynamical variables, Granger Causality (GC) [12, 14] has been widely applied in various fields, e.g., economics [4, 14], neuroscience [6, 8, 22], ecosystem [24] and climate science [20]. In general, GC is based on the vector auto-regression (VAR) modeling of observed time series (see an exception in [19] for non-linear GC). To identify GC, one compares the prediction of a variable x_t from the VAR model using all variables (known as joint-regression, full VAR, or simply VAR) and that using all but one variable y_t (known as "auto"-regression, or partial VAR). If the prediction is improved in the full VAR model, i.e., the variance of the prediction error of the full VAR model is smaller than that of the partial VAR model, then the variable y_t excluded in the partial VAR model is interpreted as having a "causality" to the variable x_t being predicted. To quantify the level of causality, the GC value from y_t to x_t is defined as the logarithmic ratio of the variance of the two prediction errors from the partial and full VAR models respectively.

GC can be applied to many systems of different types. The applicability of GC to a system relies on the validity of the VAR modeling of the system. In fact, a large number of systems can be reasonably described by the VAR model, as proved by using the multivariate Wold decomposition (Theorem 6.11 in [26]) which states that every wide-sense stationary multivariate time series can be decomposed into a one-sided moving average of white noise and a deterministic time series. By further assuming that the time series is purely nondeterministic with its power spectrum being bounded, the time series can be exactly represented by a VAR model.

*Received: July 01, 2019; Accepted (in revised form): September 10, 2019.

†Courant Institute of Mathematical Sciences, New York University, New York 10012, USA and NYUAD Institute, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates (yx742@nyu.edu).

‡School of Mathematical Sciences, MOE-LSC and Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai, P.R. China (songting@sjtu.edu.cn).

§School of Mathematical Sciences, MOE-LSC and Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai, P.R. China (zdz@sjtu.edu.cn).

The importance of GC partially attributes to its close relation with transfer entropy (TE) — an information-theoretic quantity being capable of detecting causality in general systems beyond the VAR model. GC has been proven to be equal to TE for Gaussian variables [3]. Although TE has a wider range of application in theory, GC can outperform TE in practice. For instance, for any system with a long memory, TE will suffer from the curse of dimensionality while GC can overcome the dimension issue.

Despite the simple framework of GC based on linear regressions of sample data, the interpretation of GC is often difficult to explicitly obtain in practice. The lack of an interpretation of the GC value largely lies in the fact that the relation between the GC value and the full VAR model is implicit. Although an approximate relation between the GC value and the VAR model has been derived in [11] and [29] for a two-variable system, it fails to be directly applied to multi-variable large systems. Therefore, in this work, we aim to establish a general relation between GC and the VAR model for high dimensional systems consisting of multiple variables.

The establishment of the relation between GC and the VAR model will make the GC value easy to understand. Note that the VAR model has a close relationship to the true dynamics of a system under certain conditions. In fact, the VAR model can be viewed as the (linearized) dynamics of a system when its underlying true dynamics is unknown or difficult to obtain. In both situations, an explicit relationship between GC and the VAR parameters could link the GC value to the system dynamics described by the VAR model, thus greatly simplifying the interpretation of the GC analysis associated with the dynamical system. In addition, it will provide an alternative way of investigating GC, different from the traditional way of computing GC from the second order statistics of the dynamics.

The establishment of the relation between GC and the VAR model will make GC easy to compare with other causality measures. In addition to GC, several other measures based on the VAR model have been proposed to detect the causal relations between dynamical variables. Among these quantities, partial directed coherence (PDC) [2] and its generalization gPDC [1] are defined explicitly using the VAR parameters. Interestingly, these methods behave similarly as GC in some cases but not in others. At present, the comparison between the performance of these extensively used methods and that of GC largely resorts to numerical computations. Therefore, the establishment of the relation between GC and the VAR parameters will allow one to to compare these methods analytically, i.e., the similarity and difference among these methods with respect to VAR parameters.

In this work, we derive both exact and approximated formulas of conditional GC using VAR parameters — the regression coefficients and the variance matrix of the prediction error in the full VAR model. Section 2 introduces the definition of conditional GC. Section 3 gives an exact formula of conditional GC semi-explicitly expressed by the VAR parameters. Section 4.1 gives an approximation of the GC value explicitly expressed by the VAR parameters in the frequency domain. Section 4.2 gives an alternative approximation of GC in the time domain which is more accurate than that in Section 4.1 but with a similar form, and its convergence to GC is proved. Section 4.3 numerically validates the approximations of GC derived from the theoretical analysis. Section 5 compares GC with other measures of "causality", notably gPDC. Section 6 discusses and summarizes these results. Appendix gives detailed proofs of lemmas and propositions used in the main text.

2. Definition of Granger causality

In this work, we focus on the general case when a system is composed of more than two dynamical variables. Here we first introduce the definition of conditional GC, i.e., the GC value from one variable to another variable conditioned on the remaining variables. Unless otherwise specified, in the following, GC is referred to as conditional GC.

Suppose we have a p-variate time series $X_t = \left[x_t^{(1)}\ x_t^{(2)}\ \cdots\ x_t^{(p)}\right]^T$. For the ease of illustration below, we set $x_t = x_t^{(1)}$, $y_t = x_t^{(2)}$ and $z_t = \left[x_t^{(3)}\ \cdots\ x_t^{(p)}\right]$, where x_t and y_t are scalars, z_t is a vector containing $p-2$ elements.[1] Assume $\mathrm{E}(x_t) = \mathrm{E}(y_t) = 0$, $\mathrm{E}(z_t) = O$, $O = [0\ 0\ \cdots\ 0]^T$. To define the GC value from y to x conditioned on z, we need to fit the following full and partial VAR models,

$$x_t = \sum_{j=1}^{m} a_j^{(xx)} x_{t-j} + \sum_{j=1}^{m} a_j^{(xy)} y_{t-j} + \sum_{j=1}^{m} a_j^{(xz)} z_{t-j} + \epsilon_t^{(x|y,z)}, \tag{2.1}$$

$$x_t = \sum_{j=1}^{m} c_j^{(xx)} x_{t-j} + \sum_{j=1}^{m} c_j^{(xz)} z_{t-j} + \epsilon_t^{(x|z)}, \tag{2.2}$$

where all variables are scalars except $a_j^{(xz)}$, $c_j^{(xz)} \in \mathbb{R}^{1\times(p-2)}$, $z_{t-j} \in \mathbb{R}^{(p-2)\times 1}$. The parameters $a_j^{(xx)}$, $a_j^{(xy)}$, $a_j^{(xz)}$, $c_j^{(xx)}$, and $c_j^{(xz)}$ for $j = 1,\ldots,m$ are solved by minimizing $\mathrm{var}(\epsilon_t^{(x|y,z)})$ and $\mathrm{var}(\epsilon_t^{(x|z)})$. In theory, the fitting order m is infinite, yet in practice with finite data length, a finite m is determined by certain criteria such as Akaike information criterion (AIC) [5] or Bayesian information criterion (BIC) [21].

The GC value from variable y to variable x in the time domain is defined as

$$F_{y\to x|z} = \ln \frac{\mathrm{var}(\epsilon_t^{(x|z)})}{\mathrm{var}(\epsilon_t^{(x|y,z)})}, \tag{2.3}$$

as x is assumed to be a scalar, or

$$F_{y\to x|z} = \ln \frac{\det\left(\mathrm{var}(\epsilon_t^{(x|z)})\right)}{\det\left(\mathrm{var}(\epsilon_t^{(x|y,z)})\right)}.$$

if x is multivariable (Ref. [12]).

By exchanging the variable indices, we can calculate the GC value for all pairs of variables. According to the definition of GC (Eq. (2.3)), it is evident that the relation between GC and the VAR model is implicit, which impedes one to understand the interpretation of the GC value in terms of the underlying dynamical system. This motivates us to express the GC value by the VAR parameters in this work.

It is worth mentioning that, in order to keep Equation (2.3) mathematically meaningful, in this work, we focus on the case that the time series is wide-sense stationary and purely non-deterministic [12, 26]. By *non-deterministic* we require X_t to lie out of

[1] In this paper, subscript denotes the time index of an element in a matrix or a time series; superscript denotes the index of variables; subscript with a square bracket denotes a block matrix with the number of blocks indexed by time.

the space spanned by all its history $\{X_{t-1}, X_{t-2}, \ldots\}$. This ensures $\mathrm{var}(\epsilon_t^{(x|y,z)}) > 0$ in Equation (2.3). By *purely* non-deterministic we additionally require that the projection of X_t to its history $\{X_{-k}, X_{-k-1}, \ldots\}$ converges to a zero vector as $k \to \infty$, i.e., the deterministic part of X_t in the Wold decomposition vanishes. This ensures that X_t can be represented by a one-sided moving average

$$X_t = \sum_{j=0}^{\infty} B_j \boldsymbol{\epsilon}_{t-j}, \tag{2.4}$$

where $B_j \in \mathbb{R}^{p\times p}$, $B_0 = I$, and $\boldsymbol{\epsilon}_t$ is white noise. Under the condition of X_t being purely non-deterministic with bounded power spectrum density (PSD), we have the VAR representation for the observed time series of the system

$$\sum_{j=0}^{\infty} A_j X_{t-j} = \boldsymbol{\epsilon}_t, \tag{2.5}$$

where $A_j \in \mathbb{R}^{p\times p}$, $A_0 = I$. Equation (2.5) serves as the mathematical foundation of Eqs. (2.1) and (2.2).

3. Exact expression of GC by VAR parameters

We first introduce an exact formula of GC semi-explicitly expressed by VAR parameters derived in the following theorem.

THEOREM 3.1. *Assume the power spectrum density $S(w)$ of X_t has an upper bound $S_{\max}$ and a lower bound $S_{\min}$, i.e.,*

$$\sigma_{\min}(S(w)) \geq S_{\min} > 0, \quad \sigma_{\max}(S(w)) \leq S_{\max} \quad \forall w \in [0, 2\pi].$$

And $\{X_t\}$ is fitted by an order m VAR model. Then the GC value defined in Equation (2.3) can be expressed in the quadratic form of the VAR parameters:

$$F_{y\to x|z} = \ln\left(1 + \frac{\boldsymbol{a}^{(xy)}\left(Q^{(yy)}\right)^{-1}\left(\boldsymbol{a}^{(xy)}\right)^T}{\mathrm{var}(\epsilon_t^{(x|y,z)})}\right), \tag{3.1}$$

where

$$\boldsymbol{a}^{(uv)} \triangleq \left(a_j^{(uv)}\right)_{j=1\ldots m}, \quad \textit{for} \quad u,v \in \{x,y,z\}, \tag{3.2}$$

and $Q^{(yy)}$ is a matrix coming from

$$Q = \begin{bmatrix} Q^{(xx)} & Q^{(xy)} & Q^{(xz)} \\ Q^{(yx)} & Q^{(yy)} & Q^{(yz)} \\ Q^{(zx)} & Q^{(zy)} & Q^{(zz)} \end{bmatrix} \triangleq R^{-1}, \tag{3.3}$$

with

$$R \triangleq \begin{bmatrix} R^{(xx)} & R^{(xy)} & R^{(xz)} \\ R^{(yx)} & R^{(yy)} & R^{(yz)} \\ R^{(zx)} & R^{(zy)} & R^{(zz)} \end{bmatrix} \tag{3.4}$$

and

$$R^{(uv)} \triangleq \left(\mathbb{E}(u_{t-j} v_{t-k}^T)\right)_{j,k=1\ldots m}, \quad \textit{for} \quad u,v \in \{x,y,z\}. \tag{3.5}$$

Here $(g_j)_{j=1\dots m}$ *is a row vector (a block matrix if* g_j *is a matrix) with its* j*-th element being* g_j*, and* $(g_{jk})_{j,k=1\dots m}$ *is a matrix with the element in its* j*-th row and* k*-th column being* g_{jk}*. And the superscript* $\cdot^T$ *means matrix transpose,* $\sigma_{\min}(S(w))$ *(*$\sigma_{\max}(S(w))$*) means the smallest (largest) singular value of* $S(w)$*.*[2]

Proof. By the definition of GC (Equation (2.3)), we have

$$F_{y\to x|z} = \ln\frac{\mathrm{var}(\epsilon_t^{(x|z)})}{\mathrm{var}(\epsilon_t^{(x|y,z)})} = \ln\left(1 + \frac{\mathrm{var}(\epsilon_t^{(x|z)}) - \mathrm{var}(\epsilon_t^{(x|y,z)})}{\mathrm{var}(\epsilon_t^{(x|y,z)})}\right).$$

Therefore, to prove the theorem, it is sufficient to prove that

$$\mathrm{var}(\epsilon_t^{(x|z)}) - \mathrm{var}(\epsilon_t^{(x|y,z)}) = \boldsymbol{a}^{(xy)}\left(Q^{(yy)}\right)^{-1}\left(\boldsymbol{a}^{(xy)}\right)^T. \tag{3.6}$$

The coefficient $\boldsymbol{a}^{(xy)}$ in Equation (3.6) comes from the following full VAR model,

$$x_t = \sum_{j=1}^m a_j^{(xx)} x_{t-j} + \sum_{j=1}^m a_j^{(xy)} y_{t-j} + \sum_{j=1}^m a_j^{(xz)} z_{t-j} + \epsilon_t^{(x|y,z)}, \tag{3.7}$$

which can be solved by the Yule–Walker equations (see Appendix A for the detailed derivation),

$$\begin{bmatrix} \boldsymbol{a}^{(xx)} & \boldsymbol{a}^{(xy)} & \boldsymbol{a}^{(xz)} \end{bmatrix} \begin{bmatrix} R^{(xx)} & R^{(xy)} & R^{(xz)} \\ R^{(yx)} & R^{(yy)} & R^{(yz)} \\ R^{(zx)} & R^{(zy)} & R^{(zz)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{r}^{(x|x)} & \boldsymbol{r}^{(x|y)} & \boldsymbol{r}^{(x|z)} \end{bmatrix}, \tag{3.8}$$

where

$$\boldsymbol{r}^{(u|v)} = \left(\mathbb{E}(u_t v_{t-k}^T)\right)_{k=1..m}, \quad u,v \in \{x,y,z\}. \tag{3.9}$$

Note that the matrix R is a symmetric matrix [3]. Therefore, the fitting coefficients $\boldsymbol{a}^{(xy)}$ can be solved in the following two ways derived from Equation (3.8) and its transposed version respectively,

$$\begin{aligned} \boldsymbol{a}^{(xy)} &= \begin{bmatrix} \boldsymbol{r}^{(x|x)} & \boldsymbol{r}^{(x|y)} & \boldsymbol{r}^{(x|z)} \end{bmatrix} \begin{bmatrix} Q^{(xy)} \\ Q^{(yy)} \\ Q^{(zy)} \end{bmatrix}, \\ \left(\boldsymbol{a}^{(xy)}\right)^T &= \begin{bmatrix} Q^{(yx)} & Q^{(yy)} & Q^{(yz)} \end{bmatrix} \begin{bmatrix} \boldsymbol{r}^{(x|x)} & \boldsymbol{r}^{(x|y)} & \boldsymbol{r}^{(x|z)} \end{bmatrix}^T. \end{aligned} \tag{3.10}$$

By right-multiplying x_t^T to both sides of Equation (3.7) and taking expectation, we get

$$\mathrm{var}(x_t) = \boldsymbol{a}^{(xx)}\left(\boldsymbol{r}^{(x|x)}\right)^T + \boldsymbol{a}^{(xy)}\left(\boldsymbol{r}^{(x|y)}\right)^T + \boldsymbol{a}^{(xz)}\left(\boldsymbol{r}^{(x|z)}\right)^T + \mathrm{var}(\epsilon_t^{(x|y,z)}).$$

Together with Equation (3.8), we get the prediction error

$$\mathrm{var}(\epsilon_t^{(x|y,z)}) = \mathrm{var}(x_t) - \mathrm{SS}^{(xyz)}, \tag{3.11}$$

[2]Theorem 3.1 also applies to complex-valued time series by replacing the transpose operator to the conjugate transpose operator.

[3]R is a Hermitian matrix in the case of complex-valued time series.

where

$$\mathrm{SS}^{(xyz)} \triangleq \left[\boldsymbol{r}^{(x|x)} \ \boldsymbol{r}^{(x|y)} \ \boldsymbol{r}^{(x|z)} \right] \begin{bmatrix} R^{(xx)} & R^{(xy)} & R^{(xz)} \\ R^{(yx)} & R^{(yy)} & R^{(yz)} \\ R^{(zx)} & R^{(zy)} & R^{(zz)} \end{bmatrix}^{-1} \left[\boldsymbol{r}^{(x|x)} \ \boldsymbol{r}^{(x|y)} \ \boldsymbol{r}^{(x|z)} \right]^T. \tag{3.12}$$

The inverse of R in Equation (3.12) exists because $\sigma_{\min}(R) \geq S_{\min} > 0$ (see Proposition E.1 in Appendix E).

Similarly, by solving the partial VAR model in the absence of the variable y (Equation (2.2)), we get the prediction error

$$\mathrm{var}(\epsilon_t^{(x|z)}) = \mathrm{var}(x_t) - \mathrm{SS}^{(xz)},$$

where

$$\mathrm{SS}^{(xz)} \triangleq \left[\boldsymbol{r}^{(x|x)} \ \boldsymbol{r}^{(x|z)} \right] \begin{bmatrix} R^{(xx)} & R^{(xz)} \\ R^{(zx)} & R^{(zz)} \end{bmatrix}^{-1} \left[\boldsymbol{r}^{(x|x)} \ \boldsymbol{r}^{(x|z)} \right]^T. \tag{3.13}$$

The inverse matrix in Equation (3.13) exists because it is a principal submatrix of R. Now the left-hand side of Equation (3.6) becomes

$$\mathrm{var}(\epsilon_t^{(x|z)}) - \mathrm{var}(\epsilon_t^{(x|y,z)}) = \mathrm{SS}^{(xyz)} - \mathrm{SS}^{(xz)}.$$

Before calculating $\mathrm{SS}^{(xyz)} - \mathrm{SS}^{(xz)}$, we first calculate the difference between the following two terms, and express it in terms of the VAR coefficients $\boldsymbol{a}^{(xy)}$,

$$\begin{aligned} I_1 &\triangleq \left[\boldsymbol{r}^{(x|x)} \ \boldsymbol{r}^{(x|y)} \ \boldsymbol{r}^{(x|z)} \right] \begin{bmatrix} Q^{(xx)} & Q^{(xy)} & Q^{(xz)} \\ Q^{(yx)} & Q^{(yy)} & Q^{(yz)} \\ Q^{(zx)} & Q^{(zy)} & Q^{(zz)} \end{bmatrix} \left[\boldsymbol{r}^{(x|x)} \ \boldsymbol{r}^{(x|y)} \ \boldsymbol{r}^{(x|z)} \right]^T \\ &\quad - \left[\boldsymbol{r}^{(x|x)} \ \boldsymbol{r}^{(x|z)} \right] \begin{bmatrix} Q^{(xx)} & Q^{(xz)} \\ Q^{(zx)} & Q^{(zz)} \end{bmatrix} \left[\boldsymbol{r}^{(x|x)} \ \boldsymbol{r}^{(x|z)} \right]^T \\ &= \left[\boldsymbol{r}^{(x|x)} \ \boldsymbol{r}^{(x|y)} \ \boldsymbol{r}^{(x|z)} \right] \begin{bmatrix} O & Q^{(xy)} & O \\ Q^{(yx)} & Q^{(yy)} & Q^{(yz)} \\ O & Q^{(zy)} & O \end{bmatrix} \left[\boldsymbol{r}^{(x|x)} \ \boldsymbol{r}^{(x|y)} \ \boldsymbol{r}^{(x|z)} \right]^T \\ &= \boldsymbol{a}^{(xy)} \left(\boldsymbol{r}^{(x|y)} \right)^T + \boldsymbol{r}^{(x|y)} \left(\boldsymbol{a}^{(xy)} \right)^T - \boldsymbol{r}^{(x|y)} Q^{(yy)} \left(\boldsymbol{r}^{(x|y)} \right)^T. \end{aligned} \tag{3.14}$$

The last equality in Equation (3.14) holds by applying the decomposition

$$\begin{bmatrix} O & Q^{(xy)} & O \\ Q^{(yx)} & Q^{(yy)} & Q^{(yz)} \\ O & Q^{(zy)} & O \end{bmatrix} = \begin{bmatrix} O & Q^{(xy)} & O \\ O & Q^{(yy)} & O \\ O & Q^{(zy)} & O \end{bmatrix} + \begin{bmatrix} O & O & O \\ Q^{(yx)} & Q^{(yy)} & Q^{(yz)} \\ O & O & O \end{bmatrix} - \begin{bmatrix} O & O & O \\ O & Q^{(yy)} & O \\ O & O & O \end{bmatrix}$$

and Equation (3.10).

We next calculate the difference between the following two terms denoted by I_2, and express it in terms of the VAR coefficients $\boldsymbol{a}^{(xy)}$,

$$\begin{aligned} I_2 &\triangleq \left[\boldsymbol{r}^{(x|x)} \ \boldsymbol{r}^{(x|z)} \right] \left(\begin{bmatrix} Q^{(xx)} & Q^{(xz)} \\ Q^{(zx)} & Q^{(zz)} \end{bmatrix} - \begin{bmatrix} R^{(xx)} & R^{(xz)} \\ R^{(zx)} & R^{(zz)} \end{bmatrix}^{-1} \right) \left[\boldsymbol{r}^{(x|x)} \ \boldsymbol{r}^{(x|z)} \right]^T \\ &= \left[\boldsymbol{r}^{(x|x)} \ \boldsymbol{r}^{(x|z)} \right] \begin{bmatrix} Q^{(xy)} \\ Q^{(zy)} \end{bmatrix} \left(Q^{(yy)} \right)^{-1} \left[Q^{(yx)} \ Q^{(yz)} \right] \left[\boldsymbol{r}^{(x|x)} \ \boldsymbol{r}^{(x|z)} \right]^T \end{aligned}$$

$$=\left(\boldsymbol{a}^{(xy)}-\boldsymbol{r}^{(x|y)}Q^{(yy)}\right)\left(Q^{(yy)}\right)^{-1}\left(\left(\boldsymbol{a}^{(xy)}\right)^{T}-Q^{(yy)}\left(\boldsymbol{r}^{(x|y)}\right)^{T}\right). \qquad (3.15)$$

Here $Q^{(yy)}$ is invertible because it is a principal submatrix of $Q=R^{-1}$, $\sigma_{\min}(Q^{(yy)})\geq\sigma_{\min}(Q)=\sigma_{\max}^{-1}(R)\geq S_{\max}^{-1}$. In Equation (3.15), the second equality holds because

$$\begin{bmatrix} R^{(xx)} & R^{(xz)} \\ R^{(zx)} & R^{(zz)} \end{bmatrix}^{-1}=\begin{bmatrix} Q^{(xx)} & Q^{(xz)} \\ Q^{(zx)} & Q^{(zz)} \end{bmatrix}-\begin{bmatrix} Q^{(xy)} \\ Q^{(zy)} \end{bmatrix}\left(Q^{(yy)}\right)^{-1}\begin{bmatrix} Q^{(yx)} & Q^{(yz)} \end{bmatrix}$$

obtained by applying Theorem B.1 in Appendix B, and the last equality holds due to Equation (3.10).

Combining Equations (3.14) and (3.15), now we can express the difference $\mathrm{SS}^{(xyz)}-\mathrm{SS}^{(xz)}$ in terms of the VAR coefficients $\boldsymbol{a}^{(xy)}$,

$$\begin{aligned}\mathrm{SS}^{(xyz)}-\mathrm{SS}^{(xz)}&=I_1+I_2\\ &=\boldsymbol{a}^{(xy)}\left(\boldsymbol{r}^{(x|y)}\right)^{T}+\boldsymbol{r}^{(x|y)}\left(\boldsymbol{a}^{(xy)}\right)^{T}-\boldsymbol{r}^{(x|y)}Q^{(yy)}\left(\boldsymbol{r}^{(x|y)}\right)^{T}\\ &\quad+\left(\boldsymbol{a}^{(xy)}-\boldsymbol{r}^{(x|y)}Q^{(yy)}\right)\left(Q^{(yy)}\right)^{-1}\left(\left(\boldsymbol{a}^{(xy)}\right)^{T}-Q^{(yy)}\left(\boldsymbol{r}^{(x|y)}\right)^{T}\right)\\ &=\boldsymbol{a}^{(xy)}\left(Q^{(yy)}\right)^{-1}\left(\boldsymbol{a}^{(xy)}\right)^{T}.\end{aligned}$$

This completes the proof. ∎

Theorem 3.1 can be extended to the general case of multivariable x_t and y_t as follows.

PROPOSITION 3.1. *Under the same condition in Theorem 3.1, for multivariable x_t and y_t, we have*

$$F_{y\to x|z}=\ln\det\left(I+\left(\mathrm{var}(\epsilon_t^{(x|y,z)})\right)^{-\frac{1}{2}}\boldsymbol{a}^{(xy)}\left(Q^{(yy)}\right)^{-1}\left(\left(\mathrm{var}(\epsilon_t^{(x|y,z)})\right)^{-\frac{1}{2}}\boldsymbol{a}^{(xy)}\right)^{T}\right)$$

where $M^{\frac{1}{2}}$ means the decomposition of the positive definite matrix M such that $M^{\frac{1}{2}}\left(M^{\frac{1}{2}}\right)^{T}=M$.

Proof. By the definition of the GC value, and noting that $\mathrm{var}(\epsilon_t^{(x|y,z)})$ is positive definite, we have

$$\begin{aligned}&F_{y\to x|z}\\ &=\ln\frac{\det\left(\mathrm{var}(\epsilon_t^{(x|z)})\right)}{\det\left(\mathrm{var}(\epsilon_t^{(x|y,z)})\right)}=\ln\frac{\det\left(\mathrm{var}(\epsilon_t^{(x|y,z)})+\left(\mathrm{var}(\epsilon_t^{(x|z)})-\mathrm{var}(\epsilon_t^{(x|y,z)})\right)\right)}{\det\left(\mathrm{var}(\epsilon_t^{(x|y,z)})\right)}\\ &=\ln\det\left(I+\left(\mathrm{var}(\epsilon_t^{(x|y,z)})\right)^{-\frac{1}{2}}\left(\mathrm{var}(\epsilon_t^{(x|z)})-\mathrm{var}(\epsilon_t^{(x|y,z)})\right)\left(\left(\mathrm{var}(\epsilon_t^{(x|y,z)})\right)^{-\frac{1}{2}}\right)^{T}\right).\end{aligned}$$

Then by following Equation (3.6) and thereafter, it is straightforward to complete the proof of the proposition. ∎

Theorem 3.1 gives a semi-explicit formula of GC, and the non-explicit part lies in $Q^{(yy)}$ which, in principle, can be determined by the VAR parameters also. Yet it provides several additional information of GC that cannot be gained from the original GC definition (Equation (2.3)):

- $Q^{(yy)}$ is a positive definite matrix. This implies that $\boldsymbol{a}^{(xy)}\left(Q^{(yy)}\right)^{-1}\left(\boldsymbol{a}^{(xy)}\right)^T$ is approximately quadratic (skewed by $\left(Q^{(yy)}\right)^{-1}$) when $\left\|\boldsymbol{a}^{(xy)}\right\|_2$ is small, thus $F_{y\to x|z}$ is also a quadratic function of $\boldsymbol{a}^{(xy)}$ approximately.
- Assume the variance of prediction error Σ is a diagonal matrix. Then after whitening X_t by causal filters on each variable and normalizing X_t so that $\Sigma = I$ — a process that does not change GC [10, 12], if the interaction among x, y and z is weak ($\left\|\boldsymbol{a}^{(uv)}\right\|_2 \ll 1$, $\forall u,v \in \{x,y,z\}$), then we have the approximation $R\approx I$, $Q\approx I$, thus $Q^{(yy)}\approx I$. In this case, GC is approximately $F_{y\to x|z}\approx \boldsymbol{a}^{(xy)}\left(\boldsymbol{a}^{(xy)}\right)^T = \left\|\boldsymbol{a}^{(xy)}\right\|_2^2$, a clean quadratic function of $\boldsymbol{a}^{(xy)}$.
- Equation (3.1) provides a more accurate way to compute GC numerically than using its original definition by Equation (2.3). In particular, when the interaction is weak, $\mathrm{var}(\epsilon_t^{(x|z)})$ and $\mathrm{var}(\epsilon_t^{(x|y,z)})$ would be close, thus the ratio between them in $F_{y\to x|z}$ will lose significant digits. Equation (3.1) avoids this problem. For example, in some extreme cases, the GC value computed by Equation (3.1) can achieve the order of accuracy of 10^{-30} when using double precision arithmetic, in contrast to 10^{-16} when computed by Equation (2.3).
- Equation (3.1) substantially saves computation time by avoiding computing the partial VAR model Equation (2.2). Note that, for a large dynamical system, a substantial amount of partial VAR models need to be calculated according to GC definition (Equation (2.3)) in order to identify the causal relation between all pairs of variables in the system. Therefore, Equation (3.1) can lead to fast algorithms of conditional GC.
- The finite regression order requirement in Theorem 3.1 can be dropped. In fact, as the regression order $m\to\infty$, we can show that the computed $F_{y\to x|z}$ will converge to its theoretical limit due to the convergence of the monotone sequences $\mathrm{var}(\epsilon_t^{(x|z)})$ and $\mathrm{var}(\epsilon_t^{(x|y,z)})$. We can modify the proof in Theorem 3.1 for the case of $m=\infty$ by replacing matrices with linear operators since these matrices are all norm bounded regardless of m.
- R is a block Toeplitz matrix, correspondingly Q and $Q^{(yy)}$ can be proved to be approximately Toeplitz matrices. Therefore, $Q^{(yy)}$ can be approximated by its power spectrum [16], which inspires us to approximate GC in the Fourier domain. We will discuss this issue in Section 4.1.

4. Approximation of GC by VAR parameters

Equation (3.1) links the GC value $F_{y\to x|z}$ with the VAR coefficient $\boldsymbol{a}^{(xy)}$, which provides an intuitive interpretation for $F_{y\to x|z}$. For instance, $F_{y\to x|z}$ takes a large value when $\boldsymbol{a}^{(xy)}$ is large. However, there is still an implicit part $Q^{(yy)}$ in the expression of the GC value in Equation (3.1). This section aims to develop expressions that approximate $Q^{(yy)}$ and further derive the GC value using VAR coefficients and prediction variances explicitly.

4.1. Asymptotically equivalent approximation of GC. We note that $Q^{(yy)}$ is approximately a Toeplitz matrix, which has a tight relationship to the Fourier transform [15, 16]. In fact, the Fourier transform has become an effective tool to solve time series problems involving a Toeplitz matrix [9]. This subsection utilizes this idea to link $Q^{(yy)}$ with the VAR parameters and eventually derive an approximation of GC expressed by VAR parameters.

4.1.1. Notation of the power spectrum of time series. Let $S(w)$ denote the power spectrum density of X_t. By Wiener–Khinchin theorem [26], under the condition of X_t being purely non-deterministic, we have

$$S(w)=\mathscr{F}_{\mathrm{DTFT}}\left[\{R_j\}_{j\in\mathbb{Z}}\right](w),$$

where

$$R_j \triangleq \mathbb{E}\left(X_t X_{t-j}^T\right), \tag{4.1}$$

and the discrete-time Fourier transform (DTFT) is defined as

$$\mathscr{F}_{\mathrm{DTFT}}\left[\{R_j\}_{j\in\mathbb{Z}}\right]=\sum_{j\in\mathbb{Z}} R_j e^{-\mathrm{i}jw}.$$

More details of DTFT on a matrix function can be found in Appendix C. In the following, to simplify the notation, we use $\tilde{A}$ to denote the DTFT of a (matrix) time series $\{A_j\}$.

According to Section 2, a time series X_t with bounded PSD has a VAR representation

$$\sum_{j=0}^{m} A_j X_{t-j}=\boldsymbol{\epsilon}_t, \tag{4.2}$$

where

$$A_j=-\begin{bmatrix} a_j^{(xx)} & a_j^{(xy)} & a_j^{(xz)} \\ a_j^{(yx)} & a_j^{(yy)} & a_j^{(yz)} \\ a_j^{(zx)} & a_j^{(zy)} & a_j^{(zz)} \end{bmatrix}, \quad A_0=\begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}, \quad A_{-j}=\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \forall j>0,$$

and m can be ∞. According to Lemma 3.8 in Ref. [27], we have

$$S(w)=\tilde{A}^{-1}(w)\Sigma\left(\tilde{A}^{-1}(w)\right)^H, \tag{4.3}$$

where $\cdot^H$ means conjugate transpose, and $\Sigma=\mathrm{var}(\boldsymbol{\epsilon}_t)$ is the variance of the prediction error. We denote the inverse of $S(w)$ as

$$\tilde{Q}(w)\triangleq S^{-1}(w)=\tilde{A}^H(w)\Sigma^{-1}\tilde{A}(w). \tag{4.4}$$

4.1.2. Derivation of asymptotically equivalent approximation of GC. In this subsection, we will derive an asymptotically equivalent approximation of GC in the following theorem.

THEOREM 4.1. *Under the same condition of Theorem 3.1, for small GC value $F_{y\to x|z}$, we have the approximation relation*

$$F_{y\to x|z}\approx\frac{1}{2\pi}\int_{-\pi}^{\pi} f_{y\to x|z}^{(app)}(w)\mathrm{d}w,$$

where

$$f_{y\to x|z}^{(app)}(w)\triangleq\tilde{A}^{(xy)}(w)\frac{\left(\tilde{Q}^{(yy)}(w)\right)^{-1}}{\Sigma^{(xx)}}\left(\tilde{A}^{(xy)}(w)\right)^H. \tag{4.5}$$

Here $\Sigma^{(xx)}=\mathrm{var}(\epsilon_t^{(x|y,z)})$, $\tilde{Q}^{(yy)}$ is the second-row-second-column element of $\tilde{Q}$ (Equation (4.4)), similarly $\tilde{A}^{(xy)}$ is the first-row-second-column element of $\tilde{A}$.[4]

[4] $\cdot^{(uv)}$ means extracting the submatrix corresponding to a pair of variables u and v.

Proof. To define the mathematical meaning of the approximation in the theorem, we first introduce the concept of asymptotic equivalence [16] as follows.

Suppose we have two sequences of matrices B_m, $C_m \in \mathbb{C}^{m\times m}$, if $\|B_m\|_2$ and $\|C_m\|_2$ are bounded [5] (irrelevant to m), and if

$$\frac{\|B_m - C_m\|_F}{\sqrt{m}} \to 0 \quad \text{as} \quad m \to \infty,$$

then we define the sequences $\{B_m\}$ and $\{C_m\}$ as *asymptotically equivalent*, denoted by $\{B_m\} \sim \{C_m\}$.

Our aim is to approximate $\left(Q^{(yy)}\right)^{-1}$ in Equation (3.1) by an asymptotically equivalent matrix, such that the matrix can be explicitly expressed by the AR parameters. We start from a block Toeplitz matrix $R_{[m]}$ by permuting the entries in R to group variables of the same time index together

$$R_{[m]} \triangleq \begin{bmatrix} R_0 & R_1 & \cdots & R_{m-1} \\ R_{-1} & R_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & R_1 \\ R_{1-m} & \cdots & R_{-1} & R_0 \end{bmatrix}, \tag{4.6}$$

where R_j is defined in Equation (4.1). Then $\left\{R_{[m]}\right\} \sim \{\mathcal{C}_m(S)\}$ according to Lemma 6.1 in Ref. [16], where $\mathcal{C}_m$ is an operator that constructs a block circulant matrix using elements from the inverse discrete Fourier transform,

$$\mathcal{C}_m(S) \triangleq \begin{bmatrix} c_0 & c_1 & \cdots & c_{m-1} \\ c_{m-1} & c_0 & \cdots & c_{m-2} \\ \vdots & \ddots & \ddots & \vdots \\ c_1 & c_2 & \cdots & c_0 \end{bmatrix}$$

and

$$c_k = \sum_{j=0}^{m-1} S(2\pi\frac{j}{m})e^{2\pi \mathrm{i} jk/m}. \tag{4.7}$$

By applying Theorem 6.4 in Ref. [16], we also have

$$\left\{R_{[m]}^{-1}\right\} \sim \{\mathcal{C}_m(S^{-1})\} = \left\{\mathcal{C}_m(\tilde{Q})\right\}. \tag{4.8}$$

The matrix $Q^{(yy)}$ in Equation (3.1) is a permuted submatrix of $R_{[m]}^{-1}$, let's denote its size in subscript $Q_{[m]}^{(yy)}$. Similarly, $\mathcal{C}_m(\tilde{Q}^{(yy)})$ is a permuted submatrix of $\mathcal{C}_m(\tilde{Q})$, we thus have

$$\left\{Q_{[m]}^{(yy)}\right\} \sim \left\{\mathcal{C}_m(\tilde{Q}^{(yy)})\right\}.$$

Note that $\tilde{Q}^{(yy)}(w)$ is continuous and invertible, and according to Proposition E.1 in Appendix E we have

$$\sigma_{\min}\left(Q_{[m]}^{(yy)}\right) \geq \sigma_{\min}\left(R_{[m]}^{-1}\right) = \frac{1}{\sigma_{\max}\left(R_{[m]}\right)} \geq \frac{1}{S_{\max}},$$

[5] $\|\cdot\|_2$ means the matrix 2-norm, $\|\cdot\|_F$ means the Frobenius norm.

i.e., $\left\|\left(Q_{[m]}^{(yy)}\right)^{-1}\right\|_2$ is bounded. Therefore, by using Theorem 6.4 in Ref. [16], we have

$$\left\{\left(Q_{[m]}^{(yy)}\right)^{-1}\right\} \sim \left\{\mathcal{C}_m\left(\left(\tilde{Q}^{(yy)}\right)^{-1}\right)\right\},$$

which is the desired asymptotically equivalent matrix of $\left(Q^{(yy)}\right)^{-1}$, since $\tilde{Q}^{(yy)}$ can be explicitly expressed by AR parameters as shown in Equation (4.4).

With the help of the approximated $\left(Q^{(yy)}\right)^{-1}$, we derive the approximation of GC in the following. Note that the product of a circulant matrix and a vector can be computed from the circular convolution of the matrix element and the vector. For our case, we have

$$\mathcal{C}_m\left(\left(\tilde{Q}^{(yy)}\right)^{-1}\right)\left(\boldsymbol{a}^{(xy)}\right)^T = \mathcal{C}_{1,m}\left(\left(\tilde{Q}^{(yy)}\right)^{-1}\left(\tilde{A}^{(xy)}e^{\mathrm{i}w}\right)^H\right), \tag{4.9}$$

where $\mathcal{C}_{1,m}$ is defined as

$$\mathcal{C}_{1,m}(S) \triangleq \left[c_0\ c_1\ \cdots\ c_{m-1}\right],$$

with c_i defined in Equation (4.7). The factor $e^{\mathrm{i}w}$ appears due to the fact that the index for elements in $\boldsymbol{a}^{(xy)}$ starts from zero instead of one when performing DFT.

By applying Parseval's theorem of DFT, from Equation (4.9) we obtain

$$\boldsymbol{a}^{(xy)}\mathcal{C}_m\left(\left(\tilde{Q}^{(yy)}\right)^{-1}\right)\left(\boldsymbol{a}^{(xy)}\right)^T = \sum_{j=0}^{m-1}\tilde{A}^{(xy)}(w_j)\left(\tilde{Q}^{(yy)}(w_j)\right)^{-1}\left(\tilde{A}^{(xy)}(w_j)\right)^H,$$

where $w_j = \frac{2\pi j}{m}$. And in the limiting case

$$\boldsymbol{a}^{(xy)}\mathcal{C}_m\left(\left(\tilde{Q}^{(yy)}\right)^{-1}\right)\left(\boldsymbol{a}^{(xy)}\right)^T \to \frac{1}{2\pi}\int_{-\pi}^{\pi}\tilde{A}^{(xy)}\left(\tilde{Q}^{(yy)}\right)^{-1}\left(\tilde{A}^{(xy)}\right)^H \mathrm{d}w \quad \text{as } m\to\infty. \tag{4.10}$$

Based on the right-hand side of Equation (4.10), we define $f_{y\to x|z}^{(\mathrm{app})}(w)$ as Equation (4.5). For small $F_{y\to x|z}$ such that

$$F_{y\to x|z} = \ln\left(1+\frac{\boldsymbol{a}^{(xy)}\left(Q^{(yy)}\right)^{-1}\left(\boldsymbol{a}^{(xy)}\right)^T}{\mathrm{var}(\epsilon_t^{(x|y,z)})}\right) \approx \frac{\boldsymbol{a}^{(xy)}\left(Q^{(yy)}\right)^{-1}\left(\boldsymbol{a}^{(xy)}\right)^T}{\mathrm{var}(\epsilon_t^{(x|y,z)})},$$

by using Equations (4.5) and (4.10), we can approximate GC as

$$F_{y\to x|z} \approx \frac{\boldsymbol{a}^{(xy)}\left(Q^{(yy)}\right)^{-1}\left(\boldsymbol{a}^{(xy)}\right)^T}{\mathrm{var}(\epsilon_t^{(x|y,z)})} \approx \frac{1}{2\pi}\int_{-\pi}^{\pi} f_{y\to x|z}^{(\mathrm{app})}(w)\mathrm{d}w, \tag{4.11}$$

which ends the proof. ∎

Theorem 4.1 shows that $f_{y\to x|z}^{(\mathrm{app})}$ is in fact an approximate frequency-domain decomposition of time-domain GC $F_{y\to x|z}$. It is now straightforward to write down $f_{y\to x|z}^{(\mathrm{app})}$ expressed fully by parameters in the full VAR model, by combining Equation (4.4) and

Equation (4.5)

$$f^{(\text{app})}_{y\to x|z}(w) = \left(\Sigma^{(xx)}\right)^{-\frac{1}{2}} \tilde{A}^{(xy)} \left(\begin{bmatrix} \tilde{A}^{(xy)} \\ \tilde{A}^{(yy)} \\ \tilde{A}^{(zy)} \end{bmatrix}^H \Sigma^{-1} \begin{bmatrix} \tilde{A}^{(xy)} \\ \tilde{A}^{(yy)} \\ \tilde{A}^{(zy)} \end{bmatrix} \right)^{-1} \left(\left(\Sigma^{(xx)}\right)^{-\frac{1}{2}} \tilde{A}^{(xy)} \right)^H. \tag{4.12}$$

The approximation of GC can be further simplified by assuming Σ is diagonal as often observed in practice. In such a case, we have

$$\tilde{Q}^{(yy)} = \sum_{u\in\{x,y,z\}} \left(\tilde{A}^{(uy)}\right)^H \left(\Sigma^{(uu)}\right)^{-1} \tilde{A}^{(uy)}. \tag{4.13}$$

Therefore, we can obtain a second approximation of GC in the frequency domain as

$$f^{(\text{app2})}_{y\to x|z} \triangleq \frac{\tilde{A}^{(xy)} \left(\Sigma^{(xx)}\right)^{-1} \left(\tilde{A}^{(xy)}\right)^H}{\sum_{u\in\{x,y,z\}} \left(\tilde{A}^{(uy)}\right)^H \left(\Sigma^{(uu)}\right)^{-1} \tilde{A}^{(uy)}}. \tag{4.14}$$

With Equation (4.14), it becomes evident how GC depends on the VAR parameters. For instance, some coefficients will not affect GC, e.g., $F_{y\to x|z}$ is only affected by $\tilde{A}^{(xy)}$, $\tilde{A}^{(yy)}$, $\tilde{A}^{(zy)}$, and Σ, but not, say $\tilde{A}^{(yx)}$ and $\tilde{A}^{(yz)}$.

4.2. Tight convergent approximation of GC. It is mathematically challenging to estimate the accuracy of the approximation of GC given by Equations (4.11), (4.12) and (4.14), which could limit its application. In this subsection, we aim to derive an alternative GC approximation using VAR parameters in a similar form. Importantly, we prove that our derived GC approximation becomes exact in the limit of the regression order m being infinity.

The key to derive the GC approximation lies in representing the inverse covariance matrix $R^{-1}_{[m]}$ by VAR parameters. Ref. [28] and Ref. [13] provide ways to calculate $R^{-1}_{[m]}$ using VAR parameters for a simple univariate AR process. We next generalize the result to the multivariable case, and finally derive a convergent approximation for GC. In this subsection, the main theorem of GC approximation is Theorem 4.3, and its proof requires proving Theorem 4.2 and Corollary 4.1 as follows.

THEOREM 4.2 (Representing the inverse of covariance matrix by VAR parameters). *Under the same condition of Theorem 3.1, in addition, assume the time series X_t matches an order m_{true} VAR model exactly, we then have the following representation of $R^{-1}_{[m]}$ when choosing the regression order $m \geq m_{\text{true}}$ in the full VAR model (Equation (2.1)),*

$$R^{-1}_{[m]} = V^{-1}_{[m]} - L_{[m]} C^{-1}_{[m]} L^H_{[m]}, \tag{4.15}$$

where $R_{[m]}, V_{[m]}, L_{[m]}, C_{[m]} \in \mathbb{R}^{mp\times mp}$, $R_{[m]}$ is defined in Equation (4.6), and $V_{[m]}$ is defined as

$$V_{[m]} \triangleq \grave{A}^{-1}_{[m]} \Gamma_{[m]} \left(\grave{A}^{-1}_{[m]}\right)^H,$$

and $\grave{A}_{[m]}$ and $\Gamma_{[m]}$ are defined by VAR parameters

$$\grave{A}_{[m]} \triangleq \begin{bmatrix} A_0 & A_1 & \cdots & A_{m-1} \\ O & A_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & A_1 \\ O & \cdots & O & A_0 \end{bmatrix}, \quad \Gamma_{[m]} \triangleq \begin{bmatrix} \Sigma & O & \cdots & O \\ O & \Sigma & \ddots & \vdots \\ \vdots & \ddots & \ddots & O \\ O & \cdots & O & \Sigma \end{bmatrix} \text{(same size as } \grave{A}_{[m]}\text{)},$$

A_j is defined in Equation (4.2) and $A_j = O$ for $j > m_{\text{true}}$, O means a zero matrix, and $R_{[m]}$ is defined in Equation (4.6).

In addition, $L_{[m]}$ has the structure

$$L_{[m]} = \begin{bmatrix} O_{(m-m_{true})p \times m_{true}p} & O_{(m-m_{true})p\times(m-m_{true})p} \\ L_{[m_{true}]} & O_{m_{true}p\times(m-m_{true})p} \end{bmatrix}, \tag{4.16}$$

where $O_{m\times n}$ is an $m \times n$ matrix of zeros and $\left\| C_{[m]}^{-1} \right\|_2$ is bounded regardless of m.

Proof. For any VAR process with a finite regression order $m \geq m_{\text{true}}$, the Z-transform representation of Equation (4.3), or the so-called spectral polynomial satisfies

$$\sum_{k=-\infty}^{\infty} R_k z^k = \left(\sum_{k=0}^{m} A_k z^k \right)^{-1} \Sigma \left(\sum_{k=0}^{m} A_k^H z^{-k} \right)^{-1},$$

where R_j, defined in Equation (4.1), are the covariance matrices of time series generated by $\{A_j\}$ and Σ through Equation (4.2). Rearranging terms

$$\left(\sum_{k=-\infty}^{\infty} R_k z^k \right) \left(\sum_{k=0}^{m} A_k^H z^{-k} \right) = \left(\sum_{k=0}^{m} A_k z^k \right)^{-1} \Sigma$$

and matching the coefficients of z^j in both sides gives

$$\sum_{k=0}^{m} R_{k+j} A_k^H = \begin{cases} B_j \Sigma & j \geq 0 \\ O & j < 0 \end{cases}, \tag{4.17}$$

where B_j is defined as

$$\sum_{k=0}^{\infty} B_k z^k = \left(\sum_{k=0}^{m} A_k z^k \right)^{-1}. \tag{4.18}$$

Note that B_j is well defined because $\sum_{k=0}^{m} A_k z^k$ does not have any zero roots within the unit disk on the z-complex plane, i.e., it is minimum-phase [25]. By defining

$$R_{[m]} \triangleq \begin{bmatrix} R_0 & R_1 & \cdots & R_{m-1} \\ R_{-1} & R_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & R_1 \\ R_{1-m} & \cdots & R_{-1} & R_0 \end{bmatrix}, \quad R_{\star[m]} \triangleq \begin{bmatrix} R_{-m} & \cdots & R_{-2} & R_{-1} \\ R_{-m-1} & R_{-m} & & R_{-2} \\ \vdots & \ddots & \ddots & \vdots \\ R_{1-2m} & \cdots & R_{-m-1} & R_{-m} \end{bmatrix}, \tag{4.19}$$

$$\grave{A}^H_{[m]} \triangleq \begin{bmatrix} A_0^H & O & \cdots & O \\ A_1^H & A_0^H & \ddots & \vdots \\ \vdots & \ddots & \ddots & O \\ A_{m-1}^H & \cdots & A_1^H & A_0^H \end{bmatrix}, \quad \grave{A}^H_{\star[m]} \triangleq \begin{bmatrix} A_m^H & \cdots & A_2^H & A_1^H \\ O & A_m^H & & A_2^H \\ \vdots & & \ddots & \vdots \\ O & \cdots & O & A_m^H \end{bmatrix},$$

$$\grave{B}_{[m]} \triangleq \begin{bmatrix} B_0 & B_1 & \cdots & B_{m-1} \\ O & B_0 & \ddots & \vdots \\ \vdots & & \ddots & B_1 \\ O & \cdots & O & B_0 \end{bmatrix}, \quad \Gamma_{[m]} \triangleq \begin{bmatrix} \Sigma & O & \cdots & O \\ O & \Sigma & \ddots & \vdots \\ \vdots & \ddots & \ddots & O \\ O & \cdots & O & \Sigma \end{bmatrix} \ (m \times m \text{ blocks}),$$

Equation (4.17) can be cast into block matrix form

$$\begin{bmatrix} R_{[m]} & R^H_{\star[m]} \end{bmatrix} \begin{bmatrix} \grave{A}^H_{[m]} \\ \grave{A}^H_{\star[m]} \end{bmatrix} = \grave{B}_{[m]} \Gamma_{[m]}, \tag{4.20}$$

$$\begin{bmatrix} R_{\star[m]} & R_{[m]} \end{bmatrix} \begin{bmatrix} \grave{A}^H_{[m]} \\ \grave{A}^H_{\star[m]} \end{bmatrix} = O. \tag{4.21}$$

Here we introduce the notation of " ` " such that $\grave{A}_{[m]}$ and $\grave{B}_{[m]}$ are special forms of block Toeplitz matrix[6], and the notation of "$\star$" such that $\begin{bmatrix} R_{\star[m]} & R_{[m]} \end{bmatrix}$ and $\begin{bmatrix} \grave{A}^H_{[m]} \\ \grave{A}^H_{\star[m]} \end{bmatrix}$ are Toeplitz. Note that in Equation (4.18), only the first m terms of $\{B_j\}$ instead of infinite terms are needed to form a complete set of linear equations with $\{A_j\}$, i.e., $\grave{B}_{[m]} \grave{A}_{[m]} = I$. Therefore

$$\grave{B}_{[m]} = \grave{A}^{-1}_{[m]}, \tag{4.22}$$

where the inverse of $\grave{A}_{[m]}$ exists because $S(w)$ is bounded. For simplicity, now we omit the subscript $[m]$ in Equations (4.20)-(4.22). From Equations (4.20)-(4.22) we get

$$R\grave{A}^H + R^H_\star \grave{A}^H_\star = \grave{A}^{-1}\Gamma, \tag{4.23}$$

$$R_\star \grave{A}^H + R\grave{A}^H_\star = O. \tag{4.24}$$

In Equation (4.23), by left-multiplying $\grave{A}$, and replacing $\grave{A}R^H_\star$ by taking the conjugate transpose of Equation (4.24), we arrive at

$$\grave{A}R\grave{A}^H - \grave{A}_\star R \grave{A}^H_\star = \Gamma. \tag{4.25}$$

By defining

$$G \triangleq \grave{A}^{-1}\grave{A}_\star, \tag{4.26}$$

Equation (4.25) becomes

$$GRG^H - R + V = 0, \tag{4.27}$$

[6]There is one exception of $\grave{Q}$ defined in Equation (4.32).

which is known as the discrete Lyapunov equation.

Applying binomial inverse theorem [7] to Equation (4.27) to solve R^{-1}, we get

$$\begin{aligned} R^{-1} &= \left(V + GRG^H\right)^{-1} \\ &= V^{-1} - V^{-1}G(R^{-1} + G^H V^{-1} G)^{-1} G^H V^{-1} \\ &= V^{-1} - \grave{A}^H \Gamma^{-1} \grave{A}_\star (R^{-1} + \grave{A}_\star^H \Gamma^{-1} \grave{A}_\star)^{-1} \grave{A}_\star^H \Gamma^{-1} \grave{A}. \end{aligned} \tag{4.28}$$

The proof of Equation (4.15) completes by defining

$$L_{[m]} \triangleq \grave{A}_{[m]}^H \Gamma_{[m]}^{-1} \grave{A}_{\star[m]}, \tag{4.29}$$

$$C_{[m]} \triangleq R_{[m]}^{-1} + \grave{A}_{\star[m]}^H \Gamma_{[m]}^{-1} \grave{A}_{\star[m]}, \tag{4.30}$$

and substituting $L_{[m]}$ and $C_{[m]}$ into Equation (4.28).

To examine the structure of $L_{[m]}$ described by Equation (4.16), let's first focus on Equation (4.29). For any $m \geq m_{\text{true}}$, the non-zero blocks of $\grave{A}_\star$ will be located at its lower left corner, and the non-zero blocks of $\grave{A}$ will be located around its upper diagonal. Hence the non-zero terms of $\grave{A}^H \Gamma^{-1} \grave{A}_\star$ form a lower triangular block matrix of size $m_{\text{true}} \times m_{\text{true}}$ blocks located in its lower left corner. These nonzero elements do not change as m changes. This proves the structure of $L_{[m]}$ (Equation (4.16)).

For the bound on $\left\|C_{[m]}^{-1}\right\|_2$, we have

$$\sigma_{\min}\left(C_{[m]}\right) \geq \sigma_{\min}\left(R_{[m]}^{-1}\right) = \sigma_{\max}^{-1}\left(R_{[m]}\right), \tag{4.31}$$

$$\left\|C_{[m]}^{-1}\right\|_2 = \sigma_{\min}^{-1}\left(C_{[m]}\right) \leq \sigma_{\max}\left(R_{[m]}\right).$$

Due to Proposition E.1 in Appendix E, $\sigma_{\max}\left(R_{[m]}\right)$ is bounded regardless of m, so is $\left\|C_{[m]}^{-1}\right\|_2$.

□

Based on Theorem 4.2, we have the following remarks:

- The error of estimating $R_{[m]}^{-1}$ by $V_{[m]}^{-1}$ is a low-rank and sparse matrix due to the sparse structure of $L_{[m]}$, which will be further seen in Proposition 4.1. This provides us a way to approximate Q by only VAR parameters through $V_{[m]}^{-1}$.
- It happens that only the upper left of Q (including $Q^{(yy)}$) is important for GC computation, because GC can be computed from vector-matrix product $\boldsymbol{a}^{(xy)}\left(Q^{(yy)}\right)^{-1}$ and $a_j \to 0$ as $j \to \infty$ under the condition of Theorem 3.1. The convergence of $V_{[m]}^{-1}$ under this type of vector-matrix product will be discussed in Corollary 4.1.
- According to Equation (4.28), we can design an iterative scheme to compute the full matrix $R_{[m]}^{-1}$ accurately by choosing $m \geq 2m_{\text{true}}$.
- Another version of Theorem 4.2 can be proved by reordering the elements according to variable index instead of time index in $R_{[m]}^{-1}$, $\grave{A}_{[m]}$, and $\Gamma_{[m]}$, which is concluded in Proposition 4.2.

PROPOSITION 4.1 (Element-wise approximation to $R_{[m]}^{-1}$). *Under the condition of Theorem 4.2, $R_{[m]}^{-1}$ can be approximated by $V_{[m]}^{-1}$ according to*

$$R_{[m]}^{-1} \circ \hat{\mathbf{1}}_{[m]} = V_{[m]}^{-1} \circ \hat{\mathbf{1}}_{[m]},$$

where $\circ$ means Hadamard product (element-wise product), and $\hat{\mathbf{1}}_{[m]}$ is defined as

$$\hat{\mathbf{1}}_{[m]} \triangleq \begin{bmatrix} \mathbf{1}_{(m-m_{\text{true}})p\times(m-m_{\text{true}})p} & \mathbf{1}_{(m-m_{\text{true}})p\times m_{\text{true}}p} \\ \mathbf{1}_{m_{\text{true}}p\times(m-m_{\text{true}})p} & O_{m_{\text{true}}p\times m_{\text{true}}p} \end{bmatrix},$$

in which $\mathbf{1}_{m\times n}$ is an $m\times n$ matrix of ones.

Proof. By Theorem 4.2, we have

$$R_{[m]}^{-1} = V_{[m]}^{-1} - L_{[m]} C_{[m]}^{-1} L_{[m]}^{H}.$$

The non-zero elements of the $L_{[m]} C_{[m]}^{-1} L_{[m]}^{H}$ is a block matrix of size $m_{\text{true}} \times m_{\text{true}}$ located at its lower right corner, independent of the value of R^{-1}. Therefore, R^{-1} equals $V_{[m]}^{-1}$ except its lower right $m_{\text{true}} \times m_{\text{true}}$ blocks. The Hadamard product $\circ\hat{\mathbf{1}}_{[m]}$ picks out all the identical elements in the two matrices. We point out that $m = m_{\text{true}}$ is a trivial case. ∎

PROPOSITION 4.2 (Element-wise approximation to Q). *Under the condition of Theorem 4.2, the inverse covariance matrix Q can be approximated by $\grave{Q}$ according to* [7]

$$Q \circ \hat{\mathbf{1}}^{[xyz]} = \grave{Q} \circ \hat{\mathbf{1}}^{[xyz]}, \tag{4.32}$$

where we define

$$\grave{Q} \triangleq \left(\grave{A}^{[xyz]}\right)^T \left(\Gamma^{[xyz]}\right)^{-1} \grave{A}^{[xyz]},$$

$$\grave{A}^{[xyz]} \triangleq \begin{bmatrix} \grave{A}^{(xx)} & \grave{A}^{(xy)} & \grave{A}^{(xz)} \\ \grave{A}^{(yx)} & \grave{A}^{(yy)} & \grave{A}^{(yz)} \\ \grave{A}^{(zx)} & \grave{A}^{(zy)} & \grave{A}^{(zz)} \end{bmatrix}, \quad \Gamma^{[xyz]} \triangleq \begin{bmatrix} \Gamma^{(xx)} & \Gamma^{(xy)} & \Gamma^{(xz)} \\ \Gamma^{(yx)} & \Gamma^{(yy)} & \Gamma^{(yz)} \\ \Gamma^{(zx)} & \Gamma^{(zy)} & \Gamma^{(zz)} \end{bmatrix},$$

$$\grave{A}^{(uv)} \triangleq \begin{bmatrix} a_0^{(uv)} & a_1^{(uv)} & \cdots & a_{m-1}^{(uv)} \\ 0 & a_0^{(uv)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_1^{(uv)} \\ 0 & \cdots & 0 & a_0^{(uv)} \end{bmatrix}, \quad \Gamma^{(uv)} \triangleq \begin{bmatrix} \mathbb{E}(\epsilon_t^{(u)}\epsilon_t^{(v)T}) & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbb{E}(\epsilon_t^{(u)}\epsilon_t^{(v)T}) \end{bmatrix},$$

and here $\Gamma^{(uv)}$ is an $m\times m$ block matrix, $a_0^{(uv)} = \delta_{uv} I$, $u,v \in \{x,y,z\}$, and $\hat{\mathbf{1}}^{[xyz]}$ is a permuted $\hat{\mathbf{1}}_{[m]}$ following the same way as permuting $\grave{A}^{[xyz]}$ to $\grave{A}_{[m]}$. (The proof is omitted here.)

[7] The superscript $\cdot^{[xyz]}$ means rearranging the matrix in the order of variables x, y and z (variable-major order), instead of time-major order ($\cdot_{[m]}$). Note that this subscript is omitted if self-evident, for example, Q is equivalent to $Q^{[xyz]}$.

We comment that although Q contains additional information of $a_m^{(uv)}$ for $u,v \in \{x,y,z\}$ compared with $\grave{Q}$ by their definitions, both of them contain identical VAR information when $m \geq m_{\text{true}}+1$. In addition, $\grave{Q}$ consists of blocks

$$\grave{Q}^{(uv)} = \sum_{j\in\{x,y,z\}} \sum_{k\in\{x,y,z\}} \left(\grave{A}^{(ju)}\right)^H \left(\left(\Gamma^{[xyz]}\right)^{-1}\right)^{(jk)} \grave{A}^{(kv)}. \tag{4.33}$$

Here $\left(\left(\Gamma^{[xyz]}\right)^{-1}\right)^{(jk)}$ means the block submatrix in the j-th row k-th column of $\left(\Gamma^{[xyz]}\right)^{-1}$ following the same partition order as $\Gamma^{[xyz]}$. We can prove the convergence relation between $Q^{(uv)}$ and $\grave{Q}^{(uv)}$ as well as the inverse case in Corollary 4.1 below.

Before the corollary, we first introduce the permutation matrices $P_{[m]}^{[xyz]}$ and $P_{[m]}^{(u)}$ to reorder matrices such that

$$R = P_{[m]}^{[xyz]} R_{[m]} P_{[m]}^{[xyz]T},$$

and

$$R^{(uv)} = P_{[m]}^{(u)} R_{[m]} P_{[m]}^{(v)T}, \tag{4.34}$$

where R and $R_{[m]}$ are defined in Equation (3.4) and Equation (4.19) respectively. We will use the permutation matrices to prove Corollary 4.1 in the following.

Corollary 4.1 (Approximation to Q under vector multiplication). *Under the same condition as Theorem 4.2, for any vector $\boldsymbol{b} = \left[b_0 \cdots b_{m-1}\right]$, $\sum_{j\geq 0} \|b_j\|_2^2 < \infty$, we have the following element-wise convergence relation*

$$\boldsymbol{b}\grave{Q}^{(uv)} \to \boldsymbol{b}Q^{(uv)}, \tag{4.35}$$

$$\boldsymbol{b}\left(\grave{Q}^{(uu)}\right)^{-1} \to \boldsymbol{b}\left(Q^{(uu)}\right)^{-1}, \tag{4.36}$$

for $m \to \infty$, $u,v \in \{x,y,z\}$. Here b_j has the same matrix size as $a_j^{(uu)}$ (see also Equation (3.2)).

Proof. According to Theorem 4.2, we have

$$R_{[m]}^{-1} = V_{[m]}^{-1} - L_{[m]} C_{[m]}^{-1} L_{[m]}^H. \tag{4.37}$$

In addition, the matrices $Q^{(uv)}$ and $\grave{Q}^{(uv)}$ can be expressed through a reordering of the elements in $R_{[m]}^{-1}$ and $V_{[m]}^{-1}$ according to variable index

$$Q^{(uv)} = P_{[m]}^{(u)} R_{[m]}^{-1} P_{[m]}^{(v)T}, \tag{4.38}$$

$$\grave{Q}^{(uv)} = P_{[m]}^{(u)} V_{[m]}^{-1} P_{[m]}^{(v)T}, \tag{4.39}$$

where P is the permutation matrix that has been used previously in Equation (4.34).

Now we are ready to prove Equation (4.35). From Equations (4.37)-(4.39), we have

$$Q^{(uv)} - \grave{Q}^{(uv)} = -P^{(u)} L C^{-1} L^H P^{(v)T}, \tag{4.40}$$

where the subscript $[m]$ in all matrices are omitted in the proof to further simplify the notation.

From Theorem 4.2 we know $P^{(u)}L$ and $L^H P^{(v)T}$ are low rank (at most $m_{\text{true}} \times m_{\text{true}}$ nonzero blocks) matrices. Actually,

$$P^{(u)} L C^{-1} L^H P^{(v)T} = \begin{bmatrix} O_{[(m-m_{\text{true}})\times(m-m_{\text{true}})]} & O_{[(m-m_{\text{true}})\times m_{\text{true}}]} \\ O_{[m_{\text{true}}\times(m-m_{\text{true}})]} & M \end{bmatrix},$$

where M is a submatrix of the same size as $O_{[m_{\text{true}}\times m_{\text{true}}]}$, $O_{[m\times n]}$ means a block matrix of zeros with $m\times n$ blocks, and each block has the same size as $a_j^{(uv)}$ (size can be different depending on u and v). Therefore, we have

$$\boldsymbol{b}\left(Q^{(uv)} - \grave{Q}^{(uv)}\right) = \left(\boldsymbol{b} \circ \left[O_{[1\times(m-m_{\text{true}})]} \ \mathbf{1}_{[1\times m_{\text{true}}]}\right]\right)\left(Q^{(uv)} - \grave{Q}^{(uv)}\right).$$

Correspondingly, we can estimate the bound of $\boldsymbol{b}\left(Q^{(uv)} - \grave{Q}^{(uv)}\right)$ as

$$\left\|\boldsymbol{b}\left(Q^{(uv)} - \grave{Q}^{(uv)}\right)\right\|_2 \leq \left\|\boldsymbol{b} \circ \left[O_{[1\times(m-m_{\text{true}})]} \ \mathbf{1}_{[1\times m_{\text{true}}]}\right]\right\|_2 \left\|P^{(u)} L C^{-1} L^H P^{(v)T}\right\|_2,$$

where $\left\|P^{(u)} L C^{-1} L^H P^{(v)T}\right\|_2 \leq \left\|C^{-1}\right\|_2 \left\|P^{(u)} L\right\|_2^2$.

Clearly the nonzero elements in $L_{[m]}$ are always located in a block $L_{[m_{\text{true}}]}$ which does not change as m increases. Therefore $\left\|L_{[m_{\text{true}}]}\right\|_2$ and $\left\|L_{[m]}\right\|_2$ is upper bounded, regardless of m. Because the operation of permutation $P^{(u)}$ does not increase matrix 2-norm of $L_{[m]}$, and the bound on $\left\|C^{-1}\right\|_2$ has been given in Theorem 4.2, $\left\|P^{(u)} L C^{-1} L^H P^{(v)T}\right\|_2$ is bounded.

Note that we also have $\left\|\boldsymbol{b} \circ \left[O_{[1\times(m-m_{\text{true}})]} \ \mathbf{1}_{[1\times m_{\text{true}}]}\right]\right\|_2 \to 0$ as $m\to\infty$, because $\sum_{j\geq 0}\|b_j\|_2^2 < \infty$. Therefore,

$$\frac{1}{\sqrt{p}}\left\|\boldsymbol{b}\left(Q^{(uv)} - \grave{Q}^{(uv)}\right)\right\|_F \leq \left\|\boldsymbol{b}\left(Q^{(uv)} - \grave{Q}^{(uv)}\right)\right\|_2 \to 0.$$

This proves Equation (4.35).

We next prove Equation (4.36). From Equation (4.40), we have

$$\grave{Q}^{(uu)} = Q^{(uu)} + P^{(u)} L C^{-1} L^H P^{(u)T}. \tag{4.41}$$

By applying binomial inverse theorem to Equation (4.41), we have

$$\begin{aligned}\left(\grave{Q}^{(uu)}\right)^{-1} = \left(Q^{(uu)}\right)^{-1} - \left(Q^{(uu)}\right)^{-1} P^{(u)} L \left(C + L^H P^{(u)T}\left(Q^{(uu)}\right)^{-1} P^{(u)} L\right)^{-1} \\ \cdot L^H P^{(u)T}\left(Q^{(uu)}\right)^{-1}. \end{aligned}\tag{4.42}$$

We point out that $Q^{(uu)}$ is invertible because $Q = R^{-1}$ is strictly positive definite, $Q^{(uu)}$ is a principal submatrix of Q, from Proposition E.1 in Appendix E, we have $\sigma_{\min}\left(Q^{(uu)}\right) \geq \sigma_{\min}(Q) > 0$. Similarly, $\grave{Q}^{(uu)}$ is invertible according to Proposition E.2 in Appendix E.

Correspondingly, we have

$$\left\|\boldsymbol{b}\left(\left(\grave{Q}^{(uu)}\right)^{-1} - \left(Q^{(uu)}\right)^{-1}\right)\right\|_2 \leq \left\|\boldsymbol{b}\left(Q^{(uu)}\right)^{-1} P^{(u)} L\right\|_2$$

$$\cdot \left\| \left(C + L^H P^{(u)T} \left(Q^{(uu)} \right)^{-1} P^{(u)} L \right)^{-1} \right\|_2$$
$$\cdot \left\| L^H P^{(u)T} \left(Q^{(uu)} \right)^{-1} \right\|_2 .$$

It is straightforward to bound the norms

$$\begin{aligned} \left\| \left(C + L^H P^{(u)T} \left(Q^{(uu)} \right)^{-1} P^{(u)} L \right)^{-1} \right\|_2 &= 1/\sigma_{\min} \left(C + L^H P^{(u)T} \left(Q^{(uu)} \right)^{-1} P^{(u)} L \right) \\ &\leq \frac{1}{\sigma_{\min}(C)}, \end{aligned} \tag{4.43}$$

$$\left\| \left(Q^{(uu)} \right)^{-1} P^{(u)} L \right\|_2 \leq \left\| \left(Q^{(uu)} \right)^{-1} \right\|_2 \|L\|_2 = \frac{\|L\|_2}{\sigma_{\min}\left(Q^{(uu)} \right)}. \tag{4.44}$$

Note that

$$\sigma_{\min}^{-1} \left(Q^{(uu)} \right) \leq \sigma_{\min}^{-1}(Q) = \sigma_{\min}^{-1}\left(R^{-1} \right) = \sigma_{\max}^{-1}(R), \tag{4.45}$$

together with Equations (4.16) and (4.31), the right-hand side of Equation (4.43) and Equation (4.44) are upper bounded regardless of m.

In order to prove Equation (4.36), now it becomes sufficient to prove that

$$\left\| \boldsymbol{b} \left(Q^{(uu)} \right)^{-1} P^{(u)} L \right\|_2 \to 0 \quad (m \to \infty). \tag{4.46}$$

This is true because $P^{(u)}L$ picks only the last m_{true} of m elements in $\boldsymbol{d} = \boldsymbol{b}\left(Q^{(uu)}\right)^{-1}$, and the elements in $\boldsymbol{d} = \left[d_0 \cdots d_{m-1} \right]$ have the property of $d_m \to 0$ as $m \to \infty$ due to

$$\|\boldsymbol{d}\|_2 \leq \|\boldsymbol{b}\|_2 \left\| \left(Q^{(uu)} \right)^{-1} \right\|_2 = \|\boldsymbol{b}\|_2 \sigma_{\min}^{-1} \left(Q^{(uu)} \right) < \infty \quad (\forall m \in \mathbb{N}). \tag{4.47}$$

A uniformly bounded $\|\boldsymbol{d}\|_2$ regardless of m implies that all rows of $\boldsymbol{d}$ are 2-norm bounded also, thus all rows of d_m converge to 0 as $m \to \infty$.

Now let's examine the bound on Equation (4.47). We have seen the bound on $\sigma_{\min}^{-1}\left(Q^{(uu)}\right)$ in Equation (4.45). To estimate the bound on $\|\boldsymbol{b}\|_2$ we need the property of matrix 2-norm: for any fixed m, there exists a column vector $\boldsymbol{c}$ such that $\|\boldsymbol{b}\|_2 = \|\boldsymbol{bc}\|_2$ and $\|\boldsymbol{c}\|_2 = 1$. By splitting the vector $\boldsymbol{c}$ to match the partition of $\boldsymbol{b}$, we have

$$\|\boldsymbol{bc}\|_2 = \left\| \sum_{j=0}^{m-1} b_j c_j \right\|_2 \leq \sum_{j=0}^{m-1} \|b_j c_j\|_2 \leq \sum_{j=0}^{m-1} \|b_j\|_2 \|c_j\|_2 \leq \sum_{j=0}^{\infty} \|b_j\|_2 < \infty,$$

i.e., $\|\boldsymbol{b}\|_2 < \infty$ regardless of m.

Therefore, from Equations (4.43)-(4.46), we have

$$\left\| \boldsymbol{b} \left(\left(\grave{Q}^{(uu)} \right)^{-1} - \left(Q^{(uu)} \right)^{-1} \right) \right\|_2 \to 0$$

as $m \to \infty$, which proves Equation (4.36). $\square$

We next introduce the main theorem in this subsection, which gives a convergent approximation of GC explicitly expressed by VAR parameters.

THEOREM 4.3 (Convergent explicit expression of GC). *Under the same condition as Theorem 4.2, we have the following element-wise convergence relations*

$$\boldsymbol{a}^{(xy)}\left(\grave{Q}^{(yy)}\right)^{-1}\left(\boldsymbol{a}^{(xy)}\right)^{T} \to \boldsymbol{a}^{(xy)}\left(Q^{(yy)}\right)^{-1}\left(\boldsymbol{a}^{(xy)}\right)^{T}, \tag{4.48}$$

$$\grave{F}_{y\to x|z} \triangleq \ln\left(1+\frac{\boldsymbol{a}^{(xy)}\left(\grave{Q}^{(yy)}\right)^{-1}\left(\boldsymbol{a}^{(xy)}\right)^{T}}{\Sigma^{(xx)}}\right) \to F_{y\to x|z}, \tag{4.49}$$

as $m\to\infty$. For definitions of symbols in the theorem, see Eq. (3.2) for $\boldsymbol{a}^{(xy)}$, Equation (3.3) for $Q^{(yy)}$ and Equation (4.33) for $\grave{Q}^{(yy)}$, $\Sigma^{(xx)} = \mathrm{var}(\epsilon_t^{(x|y,z)})$.

Proof. To prove the theorem, we only need to verify that the VAR coefficients $\boldsymbol{a}^{(xy)}$ satisfy the assumption in Corollary 4.1 i.e., $\sum_{j=0}^{\infty}\left\|a_j^{(xy)}\right\|_2^2 < \infty$.

Because $S(w)$ is lower bounded, we have

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left\|S^{-1}(w)\right\|_2 \mathrm{d}w \le \frac{1}{2\pi}\int_{-\pi}^{\pi}\max_w\left(\left\|S^{-1}(w)\right\|_2\right)\mathrm{d}w < \infty.$$

From Equation (4.4) and the relation between norms $\|\cdot\|_2$ and $\|\cdot\|_F$, we have

$$\frac{1}{\sqrt{p}}\left\|\tilde{A}(w)\right\|_F^2 = \frac{1}{\sqrt{p}}\left\|\tilde{A}^H(w)\tilde{A}(w)\right\|_F \le \left\|\tilde{A}^H(w)\tilde{A}(w)\right\|_2 \le \frac{1}{\sigma_{\min}\left(\Sigma^{-1}\right)}\left\|S^{-1}(w)\right\|_2.$$

Therefore, we have

$$\sum_{j=0}^{\infty}\|A_j\|_2^2 \le \sum_{j=0}^{\infty}\|A_j\|_F^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left\|\tilde{A}(w)\right\|_F^2 \mathrm{d}w < \infty,$$

and

$$\sum_{j=0}^{\infty}\left\|a_j^{(xy)}\right\|_2^2 \le \sum_{j=0}^{\infty}\|A_j\|_2^2 < \infty.$$

By Corollary 4.1,

$$\boldsymbol{a}^{(xy)}\left(\grave{Q}^{(yy)}\right)^{-1} \to \boldsymbol{a}^{(xy)}\left(Q^{(yy)}\right)^{-1}, \quad (m\to\infty)$$

together with the arithmetic rules of limit, we complete the proof of Equations (4.48) and (4.49). ∎

We should point out that the finite m_{true} condition in Theorem 4.3 can also be dropped by the following observation: the only difference between the condition of Theorem 3.1 and Theorem 4.3 is that the latter one requires the time series $\{X_t\}$ to match a finite order (m_{true}) VAR model. However, under the condition in Theorem 3.1, for any infinite order VAR model there is a sequence of finite order VAR models converging to it, in the sense that their second-order moments (i.e. $\{R_j\}$) are eventually all matched. Then the limit in Theorem 4.3 is valid for every VAR model in this converging sequence because they are all finite-ordered. Thus by combining these two embedded limiting processes, we have the convergence of Equation (4.49) for infinite m_{true}.

By definition, $Q^{(yy)}$can be explicitly expressed by the parameters of the full VAR model,

$$\grave{Q}^{(yy)} = \left[\left(\grave{A}^{(xy)}\right)^H \left(\grave{A}^{(yy)}\right)^H \left(\grave{A}^{(zy)}\right)^H\right]\left(\Gamma^{[\mathrm{xyz}]}\right)^{-1}\begin{bmatrix}\grave{A}^{(xy)}\\ \grave{A}^{(yy)}\\ \grave{A}^{(zy)}\end{bmatrix}.$$

If we further assume that $\Gamma^{[\mathrm{xyz}]}$ is diagonal as often observed in practice, $\grave{Q}^{(yy)}$ can be simplified as

$$\grave{Q}^{(yy)} = \sum_{u\in\{x,y,z\}}\left(\grave{A}^{(uy)}\right)^H\left(\Gamma^{(uu)}\right)^{-1}\grave{A}^{(uy)},$$

which takes a similar form as Equation (4.13).

Based on Theorem 4.3, we have the following remarks,

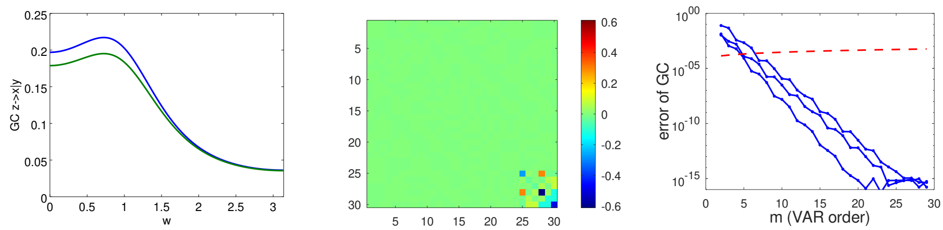
- In practice, with finite data length, we need to determine a finite value of regression order for Equation (2.1) by AIC or BIC. Then to accurately approximate GC using Equation (4.49), we use the VAR parameters in Equation (2.1) with a sufficiently large m, and set $A_m = 0$ for m being greater than the regression order determined by AIC or BIC.
- The increase of m does not increase the difficulty of interpreting the GC value. Note that quadratic part in Equation (4.49) is the same as that in Equation (4.5). Therefore, all the interpretation or intuition originated from Equation (4.5) can be applied to Equation (4.49), giving the interpretation of GC as a function of VAR parameters a solid (convergent) foundation.

4.3. Numerical results. Here we demonstrate the accuracy of our derived GC approximations Equation (4.12) and Equation (4.49) using the following numerical example (a model from Ref. [8]),

$$\begin{cases} x_t = 0.8x_{t-1} - 0.5x_{t-1} + 0.4z_{t-1} + 0.2y_{t-2} + \epsilon_t \\ y_t = 0.9y_{t-1} - 0.8y_{t-2} + \xi_t \\ z_t = 0.5z_{t-1} - 0.2z_{t-2} + 0.5y_{t-1} + \eta_t \end{cases} \tag{4.50}$$

with $\operatorname{var}(\epsilon_t) = 0.3$, $\operatorname{var}(\xi_t) = 1.0$, and $\operatorname{var}(\eta_t) = 0.2$. The theoretical conditional GC value between x, y, and z are $F_{y\to x|z} \approx 0.06742$, $F_{z\to x|y} \approx 0.12360$, $F_{y\to z|z} \approx 1.06839$ respectively and the GC values for the remaining pairs of variables are zeros. These theoretical values are calculated by definition using Equations (2.1)-(2.3) with m being sufficiently large, and with the theoretical covariances defined by Equations (3.5) and (3.9) that are numerically computed by Equation (4.3) with sufficiently large discrete Fourier transform (DFT) length (1024) and inverse DFT. Then VAR parameters are solved through Yule–Walker equations. The approximations Equations (4.12), (4.32), and (4.49) are computed through these VAR parameters.

Figure 4.1a demonstrates the accuracy of the GC approximation in the frequency domain (Equation (4.12)), which well approximates the ground truth but with visible error. To verify Equation (4.32), Figure 4.1b shows that Q and $\grave{Q}$ are approximately the same. The difference between them only appears at the lower right block of the two matrices, and the size of that block is exactly $m_{\text{true}}p \times m_{\text{true}}p = 6 \times 6$, as stated in Proposition 4.2. To verify Equation (4.49), Figure 4.1c shows the errors of the GC approximation by Equation (4.49) rapidly converge to zero as the regression order m increases. In addition, zero GC values have zero error.



(a) (b) (c)

Figure 4.1: *(a) Comparison between the exact frequency domain GC (blue) and its approximation by Equation (4.12) (green). (b) Difference between matrices Q and $\grave{Q}$, with $m = 10$ (thus the block matrix has a size of 30×30). (c) Comparison between the theoretical GC value and its tight approximation given by Equation (4.49). The abscissa is the fitting order m. The red dotted line indicates the significance level $\alpha = 10^{-3}$ for data length 10^5. There are six GC value curves in total, but the errors for GC values $F_{x\to y}$, $F_{x\to z}$, and $F_{z\to y}$ are below 10^{-16} thus not visible.*

5. Comparison between GC and other causality measures

In addition to GC, several other measurements of causality have been proposed and extensively applied in practice, for instance, partial directed coherence (PDC) [2], generalized partial directed coherence (gPDC) [1] and directed transfer function (DTF) [18]. In this subsection, we introduce the definitions of the three causality measures and identify the relation between them and GC using our derived GC expressions.

The PDC from variable y to variable x is defined as

$$\pi_{xy}(w) \triangleq \frac{\tilde{A}^{(xy)}(w)}{\sqrt{\sum_{u\in\{x,y,z\}} \left(\tilde{A}^{(uy)}(w)\right)^H \tilde{A}^{(uy)}(w)}}. \tag{5.1}$$

The gPDC from variable y to variable x is defined as

$$\pi^{(g)}_{xy}(w) \triangleq \frac{\tilde{A}^{(xy)}(w)\left(\Sigma^{(xx)}\right)^{-1/2}}{\sqrt{\sum_{u\in\{x,y,z\}} \left(\tilde{A}^{(uy)}(w)\right)^H \left(\Sigma^{(uu)}\right)^{-1} \tilde{A}^{(uy)}(w)}}. \tag{5.2}$$

The DTF from variable y to variable x is defined as

$$\gamma_{xy}(w) = \frac{\tilde{B}^{(xy)}(w)}{\sqrt{\sum_{u\in\{x,y,z\}} \tilde{B}^{(xu)}(w)\left(\tilde{B}^{(xu)}(w)\right)^H}},$$

where $\tilde{B}^{(xu)}(w)$ is the submatrix of $\tilde{B}(w)$ corresponding to variables x and $u \in \{x,y,z\}$, and $\tilde{B}(w)$ is the DTFT of B_j defined in Equation (2.4). $\tilde{B}(w)$ is also called the transfer function.

By comparing Equation (5.1) and Equation (5.2) with the GC approximation Equation (4.14), we immediately see that PDC and gPDC are special forms of frequency domain GC approximation. Specifically, we have

$$\pi_{xy}^{(g)}(w)\left(\pi_{xy}^{(g)}(w)\right)^* = f_{y\to x|z}^{(\mathrm{app2})}(w)$$

under the assumptions of Σ being diagonal, where $\cdot^*$ means complex conjugate. And PDC is identical to $f_{y\to x|z}^{(\mathrm{app2})}$ under an additional assumption that Σ has identical entries along its diagonal. If further assuming $F_{y\to x|z}$ being sufficiently small so that $\ln(1+f_{y\to x|z}^{(\mathrm{app2})}) \approx f_{y\to x|z}^{(\mathrm{app2})}$, we can reveal the relation between time domain GC and gPDC by

$$F_{y\to x|z} \approx \frac{1}{2\pi}\int_{-\pi}^{\pi} \pi_{xy}^{(g)}(w)\left(\pi_{xy}^{(g)}(w)\right)^* \mathrm{d}w.$$

The relationship between DTF and GC can be revealed by assuming that the interaction between variables are sufficiently small but not vanishing (i.e., $\|A_j\|_2 \ll 1$, $\forall j \geq 1$) so that the Taylor expansion

$$\grave{B}_{[m]} = \left(I - \left(I - \grave{A}_{[m]}\right)\right)^{-1} \approx I + \left(I - \grave{A}_{[m]}\right) \tag{5.3}$$

is relatively accurate. Here $\grave{A}_{[m]}$ and $\grave{B}_{[m]}$ are defined in Equation (4.22). Note that the diagonal of $I - \grave{A}_{[m]}$ are zeros, and the off-diagonal elements of RHS are $-A_1, -A_2, \cdots$, i.e., $B_j \approx -A_j$ and $\tilde{B}(w) \approx -\tilde{A}(w)$ due to the approximation Equation (5.3). Thus

$$\gamma_{xy}(w)(\gamma_{xy}(w))^* = \frac{\tilde{B}^{(xy)}(w)\left(\tilde{B}^{(xy)}(w)\right)^H}{\sum_{u\in\{x,y,z\}} \tilde{B}^{(xu)}(w)\left(\tilde{B}^{(xu)}(w)\right)^H} \tag{5.4}$$

$$\approx \frac{\tilde{A}^{(xy)}(w)\left(\tilde{A}^{(xy)}(w)\right)^H}{\sum_{u\in\{x,y,z\}} \tilde{A}^{(xu)}(w)\left(\tilde{A}^{(xu)}(w)\right)^H}. \tag{5.5}$$

Comparing Equation (5.5) with Equation (4.14), we see that if further assuming $\Sigma = I$, $\sum_{u\in\{y,z\}} \tilde{A}^{(xu)}(w)\left(\tilde{A}^{(xu)}(w)\right)^H \ll 1$, and $\sum_{u\in\{y,z\}} \left(\tilde{A}^{(uy)}(w)\right)^H \tilde{A}^{(uy)}(w) \ll 1$, we can have $\gamma_{ij}(w)(\gamma_{ij}(w))^* \approx f_{y\to x|z}^{(\mathrm{app2})}$.

Here we have shown that PDC, gPDC, and DTF are closely related to the frequency domain GC approximation Equation (4.14) under some conditions. By examining those conditions we can obtain qualitative and sometimes quantitative differences among these alternative measurements of causality. For example, the approximation of B_j by A_j in Equation (5.3) may not be accurate enough and needs to include higher order terms of A_j if the number of variables is large. Consequently, the DTF value is likely to deviate from GC when applied to a large system. In contrast, the relationship between PDC and $f_{y\to x|z}^{(\mathrm{app2})}$ does not depend on the number of variables, and is actually exact.

6. Summary and discussion

In this work, we have derived a new formula of GC (Equation (3.1)), which reveals a semi-explicit relation between the GC value and the full VAR model. This derived formula greatly helps interpret the GC value in terms of the underlying dynamics. In order to write down a fully explicit relationship between the GC value and the VAR parameters, we have developed two types of approximations described by Equation (4.11) and Equation (4.49) respectively, which are summarized in Table 6.1. Equation (4.11) is easy to interpret and compute, generally a non-convergent approximation. Equation (4.49) provides an alternative convergent GC approximation with a similar form as Equation (4.11), thus provides one with novel interpretation and intuition of GC.

Methodology	$F_{y\to x\mid z}$	"core part"
Original form	$\ln\frac{\mathrm{var}(\epsilon_t^{(x\mid z)})}{\mathrm{var}(\epsilon_t^{(x\mid y,z)})}$	—
Variance reduction form	$\ln\left(1+F_{y\to x\mid z}^{(\mathrm{linear})}\right)$	$F_{y\to x\mid z}^{(\mathrm{linear})} \triangleq \frac{\boldsymbol{a}^{(xy)}\left(Q^{(yy)}\right)^{-1}\left(\boldsymbol{a}^{(xy)}\right)^T}{\Sigma^{(xx)}}$
$Q^{(yy)}$ by circulant approximation	$\approx \frac{1}{2\pi}\int_{-\pi}^{\pi} f_{y\to x\mid z}^{(\mathrm{app})}(w)\mathrm{d}w$	$f_{y\to x\mid z}^{(\mathrm{app})} \triangleq \tilde{A}^{(xy)}\frac{\left(\tilde{Q}^{(yy)}\right)^{-1}}{\Sigma^{(xx)}}\left(\tilde{A}^{(xy)}\right)^H$
$Q^{(yy)}$ by Toeplitz approximation	$\approx \ln\left(1+\grave{F}_{y\to x\mid z}^{(\mathrm{linear})}\right)$	$\grave{F}_{y\to x\mid z}^{(\mathrm{linear})} \triangleq \frac{\boldsymbol{a}^{(xy)}\left(\grave{Q}^{(yy)}\right)^{-1}\left(\boldsymbol{a}^{(xy)}\right)^T}{\Sigma^{(xx)}}$

Table 6.1: *Different GC expressions. Note that the last two "core part" can be explicitly expressed by VAR parameters: for circulant approximation, the expression utilizes frequency domain VAR coefficients; for Toeplitz approximation, the expression involves "Toeplitz-rized" form of coefficients, i.e.,* $\grave{A}$ *and* Γ.

The GC approximation Equation (4.49) can be exact for finite m under a proper assumption and a specific computational procedure, i.e., the time series is zero before time zero and we do the regression by using its first m data points only, the covariances needed in Yule–Walker equations are computed by taking average over multiple realizations. From there we get an intuitive understanding of Theorem 4.3. As m goes to infinity, data at the time zero point becomes a far elder history and tends to be uncorrelated to most of the m samples. In other words, the difference between the "positive time" signal and the full time signal is vanished, thus the GC of both time series coincide.

Noting that both Equation (4.11) and Equation (4.49) can be expressed by similar operators involving Toeplitz matrices. It is thus possible to further investigate the difference between Equation (4.11) and Equation (4.49), which may help to derive an error bound. The methodology used to derive Equation (4.11) is also seen on various applications in the field of signal processing and time series analysis [9], thus the theorems in Subsection 4.2 could be helpful for investigating errors in these applications also.

Furthermore, the Toeplitz structure in Equation (3.1) also allows us to obtain an approximation of the GC value in the frequency domain given by Equation (4.5), which is derived by applying the Fourier analysis. As shown in Figure 4.1c, Equation (4.5) well approximates the conditional frequency domain GC qualitatively. However, several questions regarding the relation between the frequency domain GC and its approximation (Equation (4.5)) require further investigations, including (1) analyzing the error bound of the approximation; and (2) deriving an exact or convergence expression of frequency domain GC using VAR parameters on the basis of Equation (3.1). To analyze the error and convergence of the approximation, the difficulty largely lies in the fact that the inversion of a Toeplitz matrix is no longer Toeplitz, although they are asymptotically equivalent. To circumvent this difficulty, here we have derived an alternative approximation Equation (4.49) to approximate GC in the time domain rather than the frequency domain. We also comment that, with the development of recent advances in spectral matrix factorization techniques [17], there is a hope to derive an exact expression of frequency domain GC using VAR parameters.

As a byproduct, the approximation to the inverse covariance matrix derived in Theorem 4.2 is also useful in tackling many other problems involving covariance matrix and linear regression, such as investigating the sparsity of the covariance matrix and the VAR model.

The theorems in this paper might be generalized to nonlinear GC problems. e.g., when the nonlinear regression for the GC can be expressed as a linear combination of certain nonlinear basis, or when the regression can be linearized after embedding data points into certain space, i.e., the kernel trick [19].

Acknowledgments. This work was supported by Student Innovation Center at Shanghai Jiao Tong University, New York University Abu Dhabi Institute Grant G1301 (to Y.X., S.L., and D.Z.); National Natural Science Foundation of China Grants 11901388, Shanghai Sailing Program 19YF1421400, Shanghai Chengguang program (S.L.); Shanghai Rising-Star Program 15QA1402600; and National Natural Science Foundation of China Grants 11671259, 11722107, 91630208 (D.Z.).

Appendix A. Yule–Walker equation. The ordinary least squares problem

$$x_t = \sum_{j=1}^{m} a_j^{(xx)} x_{t-j} + \sum_{j=1}^{m} a_j^{(xy)} y_{t-j} + \sum_{j=1}^{m} a_j^{(xz)} z_{t-j} + \epsilon_t^{(x|y,z)},$$

can be solved by Yule–Walker equations

$$\begin{bmatrix} \boldsymbol{a}^{(xx)} & \boldsymbol{a}^{(xy)} & \boldsymbol{a}^{(xz)} \end{bmatrix} \begin{bmatrix} R^{(xx)} & R^{(xy)} & R^{(xz)} \\ R^{(yx)} & R^{(yy)} & R^{(yz)} \\ R^{(zx)} & R^{(zy)} & R^{(zz)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{r}^{(x|x)} & \boldsymbol{r}^{(x|y)} & \boldsymbol{r}^{(x|z)} \end{bmatrix}, \tag{A.1}$$

which is obtained by right-multiplying x_{t-k}^T and taking expectation on both sides,

$$\begin{aligned} \mathbb{E}\left(x_t x_{t-k}^T\right) = &\sum_{j=1}^{m} a_j^{(xx)} \mathbb{E}\left(x_{t-j} x_{t-k}^T\right) + \sum_{j=1}^{m} a_j^{(xy)} \mathbb{E}\left(y_{t-j} x_{t-k}^T\right) \\ &+ \sum_{j=1}^{m} a_j^{(xz)} \mathbb{E}\left(z_{t-j} x_{t-k}^T\right) + \mathbb{E}\left(\epsilon_t^{(x|y,z)} x_{t-k}^T\right). \end{aligned}$$

The least square minimization will lead to $\mathbb{E}\left(\epsilon_t^{(x|y,z)} x_{t-k}^T\right) = 0$ for $k = 1,\dots,m$, this leads to Equation (A.2).

$$\mathbb{E}\left(x_t x_{t-k}^T\right) = \sum_{j=1}^{m} a_j^{(xx)} \mathbb{E}\left(x_{t-j} x_{t-k}^T\right) + \sum_{j=1}^{m} a_j^{(xy)} \mathbb{E}\left(y_{t-j} x_{t-k}^T\right) + \sum_{j=1}^{m} a_j^{(xz)} \mathbb{E}\left(z_{t-j} x_{t-k}^T\right). \tag{A.2}$$

By using symbols defined in Equation (3.5) and Equation (3.9), i.e.

$$R^{(uv)} = \left(\mathbb{E}(x_{t-j}^{(u)} \left(x_{t-k}^{(v)}\right)^T) \right)_{j,k=1\dots m}, \quad (u,v \in \{x,y,z\}), \tag{A.3}$$

$$\boldsymbol{r}^{(u|v)} = \left(\mathbb{E}(x_t^{(u)} x_{t-k}^{(v)T})\right)_{k=1..m}^{\text{row}} \qquad (u,v \in \{x,y,z\}), \tag{A.4}$$

we reach Equation (A.1).

Appendix B. Inverse of submatrix.

THEOREM B.1. *For a block matrix with known inversion (assume all occurred inversion of matrices are valid)*

$$A^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix},$$

the sub-block inverse A_{11}^{-1} can be expressed in terms of submatrix of B:

$$A_{11}^{-1} = B_{11} - B_{12} B_{22}^{-1} B_{21}. \tag{B.1}$$

Proof. A column transform of matrix B to eliminate its lower-left block reads:

$$\begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} I & O \\ -B_{22}^{-1} B_{21} & I \end{bmatrix} = \begin{bmatrix} B_{11} - B_{12} B_{22}^{-1} B_{21} & B_{12} \\ O & B_{22} \end{bmatrix}.$$

Then left-multiplying matrix A to both sides, we get

$$\begin{bmatrix} I & O \\ -B_{22}^{-1} B_{21} & I \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} - B_{12} B_{22}^{-1} B_{21} & B_{12} \\ O & B_{22} \end{bmatrix}.$$

Now its upper-left submatrix reveals the desired inverse A_{11}^{-1}:

$$I = A_{11} \left(B_{11} - B_{12} B_{22}^{-1} B_{21}\right).$$

∎

In the main context, one more trick is used: a block submatrix follow the same rule as "single patched" submatrix when performing inversion. e.g. submatrix

$$A_{11} = \begin{bmatrix} R^{(xx)} & R^{(xz)} \\ R^{(zx)} & R^{(zz)} \end{bmatrix} \tag{B.2}$$

of the whole matrix

$$R = \begin{bmatrix} R^{(xx)} & R^{(xy)} & R^{(xz)} \\ R^{(yx)} & R^{(yy)} & R^{(yz)} \\ R^{(zx)} & R^{(zy)} & R^{(zz)} \end{bmatrix}$$

is related through column and row permutation

$$\begin{bmatrix} I & O & O \\ O & O & I \\ O & I & O \end{bmatrix} \begin{bmatrix} [R^{(xx)}] & R^{(xy)} & [R^{(xz)}] \\ R^{(yx)} & R^{(yy)} & R^{(yz)} \\ [R^{(zx)}] & R^{(zy)} & [R^{(zz)}] \end{bmatrix} \begin{bmatrix} I & O & O \\ O & O & I \\ O & I & O \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} R^{(xx)} & R^{(xz)} \\ R^{(zx)} & R^{(zz)} \end{bmatrix} & \begin{matrix} R^{(xy)} \\ R^{(zy)} \end{matrix} \\ \begin{matrix} R^{(yx)} & R^{(yz)} \end{matrix} & R^{(yy)} \end{bmatrix}.$$

Note the inversion of the permutation here is itself:

$$\begin{bmatrix} I & O & O \\ O & O & I \\ O & I & O \end{bmatrix} \begin{bmatrix} I & O & O \\ O & O & I \\ O & I & O \end{bmatrix} = \begin{bmatrix} I & O & O \\ O & I & O \\ O & O & I \end{bmatrix}.$$

Thus the inversion $Q = R^{-1}$ relates to the inversion of permuted matrix in the same way:

$$B = \begin{bmatrix} I & O & O \\ O & O & I \\ O & I & O \end{bmatrix} \begin{bmatrix} R^{(xx)} & R^{(xy)} & R^{(xz)} \\ R^{(yx)} & R^{(yy)} & R^{(yz)} \\ R^{(zx)} & R^{(zy)} & R^{(zz)} \end{bmatrix}^{-1} \begin{bmatrix} I & O & O \\ O & O & I \\ O & I & O \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} R^{(xx)} & R^{(xz)} \\ R^{(zx)} & R^{(zz)} \end{bmatrix} & \begin{matrix} R^{(xy)} \\ R^{(zy)} \end{matrix} \\ \begin{matrix} R^{(yx)} & R^{(yz)} \end{matrix} & R^{(yy)} \end{bmatrix}^{-1}.$$

Due to Equation (B.1) we know the inversion of the submatrix A_{11} from Equation (B.2) can be computed from submatrices of B and thus from the submatrices of Q, i.e. by defining B by

$$B_{11} = \begin{bmatrix} Q^{(xx)} & Q^{(xz)} \\ Q^{(zx)} & Q^{(zz)} \end{bmatrix}, \quad B_{12} = \begin{bmatrix} Q^{(xy)} \\ Q^{(zy)} \end{bmatrix},$$

$$B_{21} = \begin{bmatrix} Q^{(yx)} & Q^{(yz)} \end{bmatrix}, \quad B_{22} = Q^{(yy)}.$$

Appendix C. Discrete-time Fourier transform on matrix function.

For any scalar or matrix valued time series $A_j \in \mathbb{R}^{m\times n}$, we define its discrete-time Fourier transform (DTFT) $\tilde{A}(w)$ as:

$$\tilde{A}(w) = \mathscr{F}_{\mathrm{DTFT}}\left[\{A_j\}_{j\in\mathbb{Z}}\right](w) \triangleq \sum_{j\in\mathbb{Z}} A_j e^{-\mathrm{i}jw}.$$

And its inverse discrete-time Fourier transform:

$$A_j = \mathscr{F}^{-1}_{\mathrm{DTFT}}\left[\tilde{A}\right](j) \triangleq \frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{A}(w) e^{\mathrm{i}jw}\,\mathrm{d}w.$$

For simplicity, we sometimes use the symbol with "tilde" to denote its DTFT, like the A_j and $\tilde{A}(w)$ above.

For DTFT and inverse DTFT to be meaningful, A_j or $\tilde{A}(w)$ need to satisfy some mild condition, such as $\{A_j\}$ and $\tilde{A}(w)$ both being absolutely summable, or $\sum_{j\in\mathbb{Z}} \|A_j\|_2^2 < \infty$ ($\|\cdot\|_2$ is the matrix spectral norm). Refer to standard text book (e.g. [23]) for that topic.

Due to Wold's theorem, DTFT is always valid in our case.

In case of matrix A_j and $\tilde{A}(w)$, DTFT and inverse DTFT is defined entry-wise. But still the convolution theorem applies

$$\mathscr{F}_{\mathrm{DTFT}}\left[\left\{\sum_{j\in\mathbb{Z}} A_j B_{k-j}\right\}_{k\in\mathbb{Z}}\right] = \mathscr{F}_{\mathrm{DTFT}}\left[\{A_j\}_{j\in\mathbb{Z}}\right]\mathscr{F}_{\mathrm{DTFT}}\left[\{B_j\}_{j\in\mathbb{Z}}\right] \tag{C.1}$$

here $A_j \in \mathbb{R}^{m\times n}$, $B_j \in \mathbb{R}^{n\times l}$ (and suitable condition for DTFT to exist), the usual matrix multiplication is used. As a corollary, the Parseval's theorem in matrix form is

$$\sum_{j\in\mathbb{Z}} A_j A_j^T = \frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{A}(w)\tilde{A}^H(w)\,\mathrm{d}w. \tag{C.2}$$

which can be seen by letting $B_j = A_{-j}^T$ in Equation (C.1).

LEMMA C.1. *Given power spectrum $S(w)$, for any block vector $\vec{B} = [B_0\ B_1\ \cdots]$ $\left\|\vec{B}\vec{B}^T\right\|_2$ finite and its DTFT $\tilde{B}(w)$, we have bound*

$$\left\|\vec{B}\vec{B}^T\right\|_2 \min_w \sigma_{\min}(S(w)) \leq \left\|\frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{B}(w)S(w)\tilde{B}^H(w)\mathrm{d}w\right\|_2 \leq \left\|\vec{B}\vec{B}^T\right\|_2 \max_w \|S(w)\|_2 \tag{C.3}$$

Proof. By definition of norm $\|\cdot\|_2$,

$$\begin{aligned}\left\|\frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{B}(w)S(w)\tilde{B}^H(w)\mathrm{d}w\right\|_2 &= \max_{\|\vec{c}\|=1} \frac{1}{2\pi}\int_{-\pi}^{\pi} \vec{c}\tilde{B}(w)S(w)\tilde{B}^H(w)\vec{c}^H\mathrm{d}w \\ &\leq \max_{\|\vec{c}\|=1} \frac{1}{2\pi}\int_{-\pi}^{\pi} \vec{c}\tilde{B}(w)\|S(w)\|_2\tilde{B}^H(w)\vec{c}^H\mathrm{d}w \\ &= \max_w \|S(w)\|_2 \left\|\frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{B}(w)\tilde{B}^H(w)\mathrm{d}w\right\|_2.\end{aligned}$$

Due to Equation (C.2)

$$\left\|\frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{B}(w)\tilde{B}^H(w)\mathrm{d}w\right\|_2 = \left\|\vec{B}\vec{B}^T\right\|_2,$$

thus we get the right-hand side inequality. Similarly for the left-hand side. ∎

Appendix D. Block circulant matrix and DFT. It is well known that circulant matrix is diagonalizable by discrete Fourier transform (DFT) matrix. Here we present a block circulant matrix version of this fact.

For circulant matrix $C = (C_{jk})_{j,k=0}^{m-1}$, $C_{jk} = C_{k-j}$, $C_{j+m} = C_j$

$$C = \begin{bmatrix} C_0 & C_1 & \cdots & C_{m-1} \\ C_{m-1} & C_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & C_1 \\ C_1 & \cdots & C_{m-1} & C_0 \end{bmatrix}$$

We have diagonalization

$$\frac{1}{m}F^H CF = \begin{bmatrix} \tilde{C}_0 & O & \cdots & O \\ O & \tilde{C}_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & O \\ O & \cdots & O & \tilde{C}_{m-1} \end{bmatrix} \tag{D.1}$$

where F is the block DFT matrix

$$F = \left(z^{jk}I\right)_{j,k=0}^{m-1} = \begin{bmatrix} z^0I & z^0I & \cdots & z^0I \\ z^0I & z^1I & \cdots & z^{m-1}I \\ \vdots & \vdots & & \vdots \\ z^0I & z^{m-1}I & \cdots & z^{(m-1)(m-1)}I \end{bmatrix}$$

$$z = e^{-2\pi\mathrm{i}/m} \tag{D.2}$$

$$\tilde{C}_j = \sum_{k=0}^{m-1} C_k e^{-2\pi \mathrm{i} jk/m}.$$

This can be proved by straightforward verification.

The unit root z in Equation (D.2) has period m, i.e. $z^{j+m} = z^j$. For any integers l, n, we have

$$\sum_{j=0}^{m-1} z^{-lj} z^{n(q+j)} = z^{nq} \sum_{j=0}^{m-1} z^{(n-l)j},$$

$$\sum_{j=0}^{m-1} z^{(n-l)j} = \begin{cases} \frac{1-z^{(n-l)m}}{1-z^{(n-l)}} = 0 & n-l \nmid m \\ m & n-l \mid m \end{cases}.$$

Let $B_{l,n}$ be the block entry of $F^H C F = (B_{l,n})_{l,n=0}^{m-1}$

$$\begin{aligned} B_{l,n} &= \sum_{j=0}^{m-1} \sum_{k=0}^{m-1} F_{lj}^H C_{jk} F_{nk} \\ &= \sum_{q=0}^{m-1} \sum_{j=0}^{m-1} F_{lj}^H C_q F_{n,q+j} \quad (q = k-j) \\ &= \begin{cases} O & l \neq n \\ m \sum_{q=0}^{m-1} C_q z^{nq} = m\tilde{C}_n & l = n \end{cases}. \end{aligned}$$

This proves Equation (D.1).

As a consequence, $F/\sqrt{m}$ is a unitary matrix (proved by letting $C_j = \delta_j I$).

Another useful fact is, any general block matrix can be thought of as a part of block circulant matrix.

For example

$$\grave{A}_{[m]} = \begin{bmatrix} A_0 & A_1 & \cdots & A_{m-1} \\ O & A_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & A_1 \\ O & \cdots & O & A_0 \end{bmatrix}$$

is part of block circulant matrix (size $2m \times 2m$)

$$A_{C[m]} = \begin{bmatrix} \grave{A}_{[m]} & \grave{A}_{\star[m]} \\ \grave{A}_{\star[m]} & \grave{A}_{[m]} \end{bmatrix}, \tag{D.3}$$

where

$$\grave{A}_{\star[m]} = \begin{bmatrix} A_m & O & \cdots & O \\ A_{m-1} & A_m & \ddots & \vdots \\ \vdots & \ddots & \ddots & O \\ A_1 & \cdots & A_{m-1} & A_m \end{bmatrix}.$$

As in Equation (D.1), $A_{C[m]}$ is block diagonalizable, where the block diagonal is DTFT of $\{A_j\}_{j=0}^{m}$.

Appendix E. Relation between power spectrum and the covariance matrix.

PROPOSITION E.1 (Norm of covariance matrix). *For any wide-sense stationary multi-variable time series X_t, we have bounds on singular value of its covariance matrix*

$$\sigma_{\max}\left(R_{[m]}\right) = \left\|R_{[m]}\right\|_2 \leq \max_w \|S(w)\|_2, \tag{E.1}$$

$$\sigma_{\min}\left(R_{[m]}\right) \geq \min_w \sigma_{\min}\left(S(w)\right), \tag{E.2}$$

where $\sigma_{\min}(C)$ and $\sigma_{\max}(C)$ means minimum and maximum singular values of matrix C. These are also the bounds for eigenvalues ($\lambda_{\max}\left(R_{[m]}\right)$, $\lambda_{\min}\left(R_{[m]}\right)$) due to $R_{[m]} = R_{[m]}^T$ and $R_{[m]}$ positive semidefinite. (This is a multi-variable extension to Equation (2.38) in Chap. 2.4.1 in [25].)

Proof. Consider a time series Y_t as a filter result of X_t by $\vec{B} = \left[B_0\ B_1\ \cdots\ B_{m-1}\right]$,

$$Y_t = \sum_{j=0}^{m-1} B_j X_{t-j}.$$

Then the "variance" of Y_t can be expressed in time domain,

$$\mathbb{E}\left(Y_t Y_t^T\right) = \sum_{j=0}^{m-1}\sum_{k=0}^{m-1} B_j \mathbb{E}\left(X_{t-j}X_{t-k}^T\right) B_k^T = \vec{B} R_{[m]} \vec{B}^T, \tag{E.3}$$

and in frequency domain (note m finite, so $\tilde{B}(w)$ is always well-defined)

$$\mathbb{E}\left(Y_t Y_t^T\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{B}(w) S(w) \tilde{B}^H(w) \mathrm{d}w \tag{E.4}$$

By definition of norm $\|\cdot\|_2$, for $c \in \mathbb{R}^{1\times mp}$,

$$\begin{aligned}
\left\|R_{[m]}\right\|_2 &= \max_{\|\vec{c}\|_2 = 1} \vec{c} R_{[m]} \vec{c}^T \\
&\leq \max_{\left\|\vec{B}\vec{B}^T\right\|_2 = 1} \left\|\mathbb{E}\left(Y_t Y_t^T\right)\right\|_2 \quad \text{(by Equation (E.3))} \\
&\leq \max_{\left\|\vec{B}\vec{B}^T\right\|_2 = 1} \left\|\frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{B}(w) S(w) \tilde{B}^H(w) \mathrm{d}w\right\|_2 \quad \text{(by Equation (E.4))} \\
&\leq \max_w \|S(w)\|_2 \quad \text{(by Equation (C.3))}.
\end{aligned}$$

Similarly for Equation (E.2). ∎

LEMMA E.1 (Bounds on $\grave{A}_{[m]}$ by DTFT). *Given white noise $W_j \in \mathbb{R}^{p\times 1}$, $\mathbb{E}\left(W_j W_k^T\right) = \delta_{jk}\Sigma$ (δ_{jk} is Kronecker delta), and filter coefficients $C_j \in \mathbb{R}^{p\times p}$, in matrix*

$$W_{[m]} = \begin{bmatrix} W_0 \\ W_1 \\ \vdots \\ W_{m-1} \end{bmatrix}$$

$$C_{[m]} = \begin{bmatrix} C_0 & O & \cdots & O \\ C_1 & C_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & O \\ C_{m-1} & \cdots & C_1 & C_0 \end{bmatrix},$$

$$Z = C_{[m]} W$$

we have bound

$$\left\| \mathbb{E}\left(ZZ^T\right) \right\|_2 \leq m \left\| \Sigma \right\|_2 \left\| \vec{C}\vec{C}^T \right\|_2,$$

where

$$\vec{C} = \left[C_0 \ C_1 \ \cdots \ C_{m-1}\right].$$

Proof. We want to prove $\mathbb{E}\left(ZZ^T\right) = C_{[m]}\Gamma_{[m]}C_{[m]}^T$ is bounded, where

$$\Gamma_{[m]} = \begin{bmatrix} \Sigma & O & \cdots & O \\ O & \Sigma & \ddots & \vdots \\ \vdots & \ddots & \ddots & O \\ O & \cdots & O & \Sigma \end{bmatrix} \quad (m \times m \text{ blocks}).$$

Zero extend $C_{-j} = C_{m-1+j} = O$, for all integers $j \geq 1$, and define

$$C_{\star[m]} = \begin{bmatrix} C_m & C_{m-1} & \cdots & C_1 \\ O & C_m & \ddots & \vdots \\ \vdots & \ddots & \ddots & C_{m-1} \\ O & \cdots & O & C_m \end{bmatrix}.$$

Also zero extend the white noise so that $W_{-j} = W_{m-1+j} = \vec{0}$, $\forall j \geq 1$.

The filtering result Y_j of W_j by C_j is

$$Y_j = \sum_{k \in \mathbb{Z}} C_k W_{j-k},$$

in matrix form

$$Y_{L[m]} = \left[Y_0^T \ Y_1^T \ \cdots \ Y_{m-1}^T\right]^T, \quad Y_{U[m]} = \left[Y_m^T \ Y_{m+1}^T \ \cdots \ Y_{2m-1}^T\right]^T$$

$$\vec{Y} = \begin{bmatrix} Y_{L[m]} \\ Y_{U[m]} \end{bmatrix} = \begin{bmatrix} C_{[m]} \\ C_{\star[m]} \end{bmatrix} W_{[m]}.$$

It is the matrix $C_{\star[m]}$ that helped us to express the problem in DTFT. Also note that we have only finite terms of nonzero $\{W_j\}$, so its DTFT always exists:

$$\tilde{C} = \mathscr{F}_{\mathrm{DTFT}}\left[\{C_j\}_{j \in \mathbb{Z}}\right],$$

$$\tilde{W} = \mathscr{F}_{\mathrm{DTFT}}\left[\{W_j\}_{j\in\mathbb{Z}}\right],$$

$$\tilde{Y} = \mathscr{F}_{\mathrm{DTFT}}\left[\{Y_j\}_{j\in\mathbb{Z}}\right],$$

$$\tilde{Y}(w) = \tilde{C}(w)\tilde{W}(w).$$

Denote $S(w) = \mathbb{E}\left(\tilde{W}(w)\tilde{W}^H(w)\right)$, so $\mathbb{E}\left(\tilde{Y}(w)\tilde{Y}^H(w)\right) = \tilde{C}(w)S(w)\tilde{C}^H(w)$. Due to Lemma C.1:

$$\left\|\vec{C}\vec{C}^T\right\|_2 \min_w \sigma_{\min}(S(w)) \le \left\|\frac{1}{2\pi}\int_{-\pi}^{\pi}\tilde{C}(w)S(w)\tilde{C}^H(w)\mathrm{d}w\right\|_2 \le \left\|\vec{C}\vec{C}^T\right\|_2 \max_w \|S(w)\|_2.$$

More explicitly, by convolution theorem

$$\mathbb{E}\left(\tilde{W}(w)\tilde{W}^H(w)\right) = \mathscr{F}_{\mathrm{DTFT}}\left[\left\{\mathbb{E}\left(\sum_{k\in\mathbb{Z}} W_k W_{k-j}^T\right)\right\}_{j\in\mathbb{Z}}\right] = \mathscr{F}_{\mathrm{DTFT}}\left[\{\delta_j m\Sigma\}_{j\in\mathbb{Z}}\right] = m\Sigma.$$

So we get upper bound

$$\begin{aligned}\left\|\mathbb{E}\left(Y_{L[m]}Y_{L[m]}^T\right)\right\|_2 &\le \left\|\mathbb{E}\left(\vec{Y}\vec{Y}^T\right)\right\|_2 \\ &= \left\|\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathbb{E}\left(\tilde{Y}(w)\tilde{Y}^H(w)\right)\mathrm{d}w\right\|_2 \\ &\le m\left\|\vec{C}\vec{C}^T\right\|_2 \|\Sigma\|_2,\end{aligned}$$

here $Z = Y_{L[m]}$. □

PROPOSITION E.2 (bounds of coefficient matrix). *For any wide-sense stationary multi-variable time series X_t that has auto-regression representation*

$$\sum_{j=0}^{m} A_j X_{t-j} = E_t \tag{E.5}$$

where $A_j = O\ \forall j > m_{true}$. And has bounded power spectrum

$$0 < S_{\min} \le \sigma_{\min}(S(w)),\ \ \|S(w)\|_2 \le S_{\max} < \infty.$$

Then we have bounds

$$\left\|\grave{A}_{[m]}^T\Gamma_{[m]}^{-1}\grave{A}_{[m]}\right\|_2 \le \max_{w\in[0,2\pi]}\left\|S^{-1}(w)\right\|_2 + \left\|\grave{A}_{\star[m]}^T\Gamma_{[m]}^{-1}\grave{A}_{\star[m]}\right\|_2, \tag{E.6}$$

$$\sigma_{\min}\left\|\grave{A}_{[m]}^T\Gamma_{[m]}^{-1}\grave{A}_{[m]}\right\|_2 \ge \min_{w\in[0,2\pi]}\sigma_{\min}\left(S^{-1}(w)\right). \tag{E.7}$$

where

$$\grave{A}_{[m]} = \begin{bmatrix} A_0 & A_1 & \cdots & A_{m-1} \\ O & A_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & A_1 \\ O & \cdots & O & A_0 \end{bmatrix}, \quad \Gamma_{[m]} = \begin{bmatrix} \Sigma & O & \cdots & O \\ O & \Sigma & \ddots & \vdots \\ \vdots & \ddots & \ddots & O \\ O & \cdots & O & \Sigma \end{bmatrix} \text{(same size as } \grave{A}_{[m]}\text{)},$$

and $\Sigma = \mathbb{E}\left(E_t E_t^T\right)$.

In particular, Equation (E.6) is uniformly bounded with respect to m.

Proof. For convenience, redefine $A_j \leftarrow \Sigma^{-1/2} A_j$, where $\Sigma^{-1/2}$ is square root of positive semidefinite matrix Σ. i.e. we only need to prove the case of $\Sigma = I$, $\left\|\grave{A}_{[m]}^H \grave{A}_{[m]}\right\|_2$ is uniformly bounded for $\forall m \geq m_{\text{true}}$.

Follow the construction of Equation (D.3) in Subsection D. Consider the big upper left block of $A_{C[m]}^T A_{C[m]}$, that is $\grave{A}_{[m]}^T \grave{A}_{[m]} + \grave{A}_{\star[m]}^T \grave{A}_{\star[m]}$. Spectral norm of upper left submatrix and the full matrix $A_{C[m]}$ has relation

$$\left\|\grave{A}_{[m]}^T \grave{A}_{[m]} + \grave{A}_{\star[m]}^T \grave{A}_{\star[m]}\right\|_2 \leq \left\|A_{C[m]}^T A_{C[m]}\right\|_2.$$

Spectral norm of sum of matrices obey triangle inequality

$$\left\|\grave{A}_{[m]}^T \grave{A}_{[m]}\right\|_2 \leq \left\|\grave{A}_{[m]}^T \grave{A}_{[m]} + \grave{A}_{\star[m]}^T \grave{A}_{\star[m]}\right\|_2 + \left\|\grave{A}_{\star[m]}^T \grave{A}_{\star[m]}\right\|_2.$$

With Equation (D.1), we have

$$\left\|A_{C[m]}^T A_{C[m]}\right\|_2 = \left\|\frac{1}{2m} F A_{C[m]}^T F^H F A_{C[m]} F^H \frac{1}{2m}\right\|_2 = \left\|D^H D\right\|_2$$

where

$$D = \begin{bmatrix} \tilde{A}_0 & O & \cdots & O \\ O & \tilde{A}_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & O \\ O & \cdots & O & \tilde{A}_{2m-1} \end{bmatrix},$$

and

$$\left\|D^H D\right\|_2 = \max_{j \in \{0,\ldots,2m-1\}} \left\|\tilde{A}_j^H \tilde{A}_j\right\|_2 \leq \max_{w \in [0,2\pi]} \left\|\tilde{A}(w)^H \tilde{A}(w)\right\|_2.$$

Here $\tilde{A}(w)^H \tilde{A}(w) = S^{-1}(w)$. Together we get

$$\left\|\grave{A}_{[m]}^T \grave{A}_{[m]}\right\|_2 \leq \max_{w \in [0,2\pi]} \left\|S^{-1}(w)\right\|_2 + \left\|\grave{A}_{\star[m]}^T \grave{A}_{\star[m]}\right\|_2.$$

Note that $\left\|\grave{A}_{\star[m]}^T \grave{A}_{\star[m]}\right\|_2$ will not grow with m when $m \geq m_{\text{true}}$ because its nonzero submatrix is the same for different m. Therefore the spectral norm $\left\|\grave{A}_{[m]}^T \grave{A}_{[m]}\right\|_2$ is also uniformly bounded with regard to m. This give us the conclusion Equation (E.6).

For the lower bound Equation (E.7), due to Equation (4.28), i.e. (note $\Sigma = I$, thus $\Gamma = I$)

$$\grave{A}_{[m]}^T \grave{A}_{[m]} = R^{-1} + \grave{A}_{[m]}^T \grave{A}_{\star[m]} (R^{-1} + \grave{A}_{\star[m]}^T \grave{A}_{\star[m]})^{-1} \grave{A}_{\star[m]}^T \grave{A}_{[m]}$$

where the second term of right-hand side is positive semidefinite matrix. Hence

$$\sigma_{\min}\left(\grave{A}_{[m]}^T \grave{A}_{[m]}\right) \geq \sigma_{\min}\left(R^{-1}\right) = 1/\sigma_{\max}(R).$$

With help of Equation (E.1), we have

$$1/\sigma_{\max}(R) \geq 1/\max_w \left\|S(w)\right\|_2 = \min_{w \in [0,2\pi]} \sigma_{\min}\left(S^{-1}(w)\right),$$

which gives Equation (E.7). ∎

REFERENCES

[1] L.A. Baccala, K. Sameshima, and D.Y. Takahashi, *Generalized partial directed coherence*, in 15th International Conference on Digital Signal Processing, 163–166, 2007. 1, 5

[2] L.A. Baccalá and K. Sameshima, *Partial directed coherence: a new concept in neural structure determination*, Biol. Cybern., 84(6):463–474, 2001. 1, 5

[3] L. Barnett, A.B. Barrett, and A.K. Seth, *Granger causality and transfer entropy are equivalent for Gaussian variables*, Phys. Rev. Lett., 103(23):238701, 2009. 1

[4] F. Benhmad, *Modeling nonlinear Granger causality between the oil price and U.S. dollar: A wavelet based approach*, Econ. Model., 29(4):1505–1514, 2012. 1

[5] H. Bozdogan, *Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions*, Psychometrika, 52(3):345–370, 1987. 2

[6] Y. Chen, G. Rangarajan, J. Feng, and M. Ding, *Analyzing multiple nonlinear time series with extended Granger causality*, Phys. Lett. A, 324(1):26–35, 2004. 1

[7] Wikipedia contributors, *Woodbury matrix identity*, March 2019, Version ID: 890122248. 4.2

[8] M. Ding, Y. Chen, and S.L. Bressler, *Granger causality: Basic theory and application to neuroscience*, in B. Schelter, M. Winterhalder and J. Timmer (eds.), Handbook of Time Series Analysis, John Wiley & Sons, Ltd., 437–460, 2006. 1, 4.3

[9] F.-W. Sun, Y. Jiang, and J.S. Baras, *On the convergence of the inverses of Toeplitz matrices and its applications*, IEEE Trans. Inf. Theory, 49(1):180–190, 2003. 4.1, 6

[10] E. Florin, J. Gross, J. Pfeifer, G.R. Fink, and L. Timmermann, *The effect of filtering on Granger causality based multivariate causality measures*, NeuroImage, 50(2):577–588, 2010. 3

[11] K.J. Friston, A.M. Bastos, A. Oswal, B. van Wijk, C. Richter, and V. Litvak, *Granger causality revisited*, NeuroImage, 101:796–808, 2014. 1

[12] J. Geweke, *Measurement of linear dependence and feedback between multiple time series*, J. Amer. Stat. Associ., 77(378):304–313, 1982. 1, 2, 3

[13] E.J. Godolphin and J.M. Unwin, *Evaluation of the covariance matrix for the maximum likelihood estimator of a Gaussian autoregressive-moving average process*, Biometrika, 70(1):279–284, 1983. 4.2

[14] C.W.J. Granger, *Investigating causal relations by econometric models and cross-spectral methods*, Econometrica, 37(3):424–438, 1969. 1

[15] R.M. Gray, *Toeplitz and Circulant Matrices: A Review*, Now Publishers Inc, 2006. 4.1

[16] J. Gutiérrez-Gutiérrez and P.M. Crespo, *Block Toeplitz matrices: asymptotic results and applications*, Found. Trends Commun. Inf. Theory, 8(3):179–257, 2012. 3, 4.1, 4.1.2, 4.1.2, 4.1.2, 4.1.2

[17] G. Janashia, E. Lagvilava, and L. Ephremidze, *A new method of matrix spectral factorization*, IEEE Trans. Inf. Theory, 57(4):2318–2326, 2011. 6

[18] M.J. Kaminski and K.J. Blinowska, *A new method of the description of the information flow in the brain structures*, Biol. Cybern., 65(3):203–210, 1991. 5

[19] D. Marinazzo, M. Pellicoro, and S. Stramaglia, *Kernel method for nonlinear Granger causality*, Phys. Rev. Lett., 100(14):144103, 2008. 1, 6

[20] T.J. Mosedale, D.B. Stephenson, M. Collins, and T.C. Mills, *Granger causality of coupled climate processes: ocean feedback on the north Atlantic oscillation*, J. Clim., 19(7):1182–1194, 2006. 1

[21] A.A. Neath and J.E. Cavanaugh, *The Bayesian information criterion: background, derivation, and applications*, Wiley Interdiscip. Rev. Comput. Stat., 4(2):199–203, 2012. 2

[22] A.K. Seth, A.B. Barrett, and L. Barnett, *Granger causality analysis in neuroscience and neuroimaging*, J. Neurosci, 35(8):3293–3297, 2015. 1

[23] E.M. Stein and R. Shakarchi, *Fourier Analysis: An Introduction*, Princeton University Press, 2011. C

[24] G. Sugihara, R. May, H. Ye, C.-H. Hsieh, E. Deyle, M. Fogarty, and S. Munch, *Detecting causality in complex ecosystems*, Science, 338(6106):496–500, 2012. 1

[25] P.P. Vaidyanathan, *The theory of linear prediction*, in J. Moura (ed.), Synthesis Lectures on Signal Processing, 2(1):1–184, 2007. 4.2, E.1

[26] N. Wiener and P. Masani, *The prediction theory of multivariate stochastic processes*, Acta Math., 98(1):111–150, 1957. 1, 2, 4.1.1

[27] N. Wiener and P. Masani, *The prediction theory of multivariate stochastic processes, II*, Acta Math., 99(1):93–137, 1958. 4.1.1

[28] J. Wise, *The autocorrelation function and the spectral density function*, Biometrika, 42(1/2):151–159, 1955. 4.2

[29] D. Zhou, Y. Xiao, Y. Zhang, Z. Xu, and D. Cai, *Granger causality network reconstruction of conductance-based integrate-and-fire neuronal systems*, PLOS ONE, 9(2):e87636, 2014. 1