Article

Guaranteed Stability of Autoregressive Models with Granger Causality Learned from Wald Tests

Nuntanut Raksasri[a] **and Jitkomut Songsiri**[b]

Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University
E-mail: [a]nuntanut.rak@gmail.com, [b]jitkomut.s@chula.ac.th (corresponding author)

Abstract. This paper aims to explain relationships between time series by using the Granger causality (GC) concept through autoregressive (AR) models and to assure the model stability. Examining such GC relationship is performed on the model parameters using the Wald test and the model stability is guaranteed by the infinity-norm constraint on the dynamic matrix of the AR process. The proposed formulation is a least-squares estimation with Granger causality and stability constraints which is a convex program with a quadratic objective subject to linear equality and inequality norm constraints. We show by simulations that various typical factors could lead to unstable estimated models when using an unconstrained method. Estimated models from our approach are guaranteed to be stable but the model fitting error could be conservatively increased due to the selected stability condition.

Keywords: Granger causality, Autoregressive model, Wald test, System stability

1 Introduction

A problem of understanding time series dynamic is to explain relationships between variables in multivariate time series by a dynamical model. One of typically used models is a vector autoregressive (AR) model with order p that is described by

$$y(t) = A_1 y(t-1) + \cdots + A_p y(t-p) + v(t) \tag{1}$$

where $y(t) = (y_1(t), \ldots, y_n(t)) \in \mathbf{R}^n$ is the time series of interest, $v(t)$ is a noise signal with covariance matrix Σ, and $A_k \in \mathbf{R}^{n \times n}$ for $k = 1, \ldots, p$ are the autoregressive coefficient matrices.

A relationship between variables in time series or a causal structure can be explained by applying the Granger causality concept through autoregressive (AR) model [1]. It is shown in [1] that y_j does not Granger-cause y_i if and only if for all $k = 1, 2, \ldots, p$

$$(A_k)_{ij} = 0 \tag{2}$$

when $(A_k)_{ij}$ is the (i, j) component of A_k, the parameter of the AR model. In other words, the Granger causality can be characterized as a linear parametric equation in AR matrix coefficients and can be included as a constraint in a model estimation problem.

Maximum likelihood (ML) estimation is chosen to estimate the model parameters. If an estimate of an (i, j) component of the model parameter A_k equals to zero then the ith and jth component of the time series are not related (in the sense of Granger causality). As an estimated A_k can have some components that are not equal to zero, or have significantly low values, a hypothesis test can be performed to determine whether (2) holds in a statistical sense. In particular, Wald test is considered to determine the zero structures of the model parameters. From [2], the Wald test considers a restriction function that explains our hypothesis and the Wald statistic is a quadratic form of the restriction function. If the hypothesis that the actual value of parameters are zero is true then an estimated value of those parameters should be significantly small. The Wald test can then provide an estimated zero structure in A, or the Granger causality revealed from the data. We then can consider a problem of estimating a model subject to the Granger causality constraint (2) learned from the Wald test. The formulation will be an optimization problem with zero constraints in A.

The random process (1) is stationary process if and only if the corresponding linear system is stable. Estimated parameters, determined from any methods, could cause an instability in the corresponding linear system. As a result, a stability condition should be taken into account in the parameter estimation in order for the estimated model to be meaningful and can be used in future purposes such as prediction or control design. From previous researches, there are several attempts to estimate a stable linear model. Jury's test [3, 4], for example, is used to guarantee the stability on a single input single output (SISO) linear system. By estimating the system parameters with conditions from the Jury's array which formulated as bilinear terms in [4], the system is guaranteed to have a bounded input bounded output (BIBO) stability; however, this approach cannot be readily applied to (1) as it is a multi-input multi-output (MIMO) system. From [5, 6], the Lyapunov's theory is applied to a vector autoregressive moving average (ARMA) model. The model parameters are estimated to be closest to the unconstrained estimate subject to the Lyapunov equation by using a change of variable and adjusting a weight matrix in the cost objective by P, the matrix variable in the Lyapunov equation. The change of variable simplifies the stability constraint which is a product between P and the dynamic matrix. A similar technique of using weighted least-squares objective and making a change of variable in the stability constraint but based on a subspace identification is also found in [7]. Its stability constraint is an LMI (linear matrix inequality) region that guarantees to cover all eigenvalues of the system dynamic matrix. Nevertheless, this technique cannot be applied to our work because the Granger causality conditions which are simple linear constraints becomes, on the other hand, more complicated after the change of variable. Problems of iteratively estimating a general dynamic matrix with convex stability constraints are explained in [8, 9, 10]. In [8], they propose the use of the Dikin ellipsoid that is proved to lie within the stable matrix set by using a property of log barrier function, as a feasible set. The formulation is to estimate a matrix closest to a given one subject to the ellipsoid constraint, which can be solved by computing projections onto the ellipsoid. In [9, 10], they use the least-squares objective from the subspace method and derive a hyperplane as a function of entries in the dynamic matrix implied from the condition that its maximum singular value exceeds one. The hyperplane is further served as a constraint in the estimation problem which is iteratively solved until a stability is reached. The authors show that in this iterative scheme, the stability is almost always obtained in the experiment but provide no mathematical proof. Many papers focus on the estimation of ARMAX model family directly. The authors in [11] consider only a special case, the first-order AR model and use an upper bound condition for the 2−norm of the dynamic matrix. Moreover, a Gershgorin circle's theory is used in estimating autoregressive moving average with exogenous terms (ARMAX) models in [12]. The model is arranged in a state equation form, and estimation formulation is to force the Gershgorin circle to lie within the unit circle. In this paper, we propose a sufficient stability condition in the form of a norm constraint on the dynamic matrix, derived from an upper bound of the spectral radius. Our approach is a general case of [11], and leads to similar solutions to [12]. Our stability feasible region may be more conservative than the ones proposed in [9, 10] but its convexity allow us to simply incorporate the additional Granger causality convex constraints in the problem.

In summary, our goal is to estimate multivariate autoregressive models with two significant characteristics. The first characteristic is the zero structure of the model parameters that presents the Granger causality, and the second one is the model stability. This paper first provides a short description of the maximum likelihood estimation method in section 2. An unconstrained estimate is performed through the Wald test with detailed explained in section 3. The main result of this paper which is a formulation of estimating models subject to Granger causality and stability constraints is proposed in section 4. Numerical results will be illustrated in section 5 with experiments tested on randomly generated data.

2 Autoregressive Model Estimation

The model (1) can be written in linear parametric equation form as

$$y(t) = AH(t) + v(t) \tag{3}$$

when $y(t) = (y_1(t), ..., y_n(t))$ is the output of the model, $v(t)$ is a zero-mean Gaussian noise with covariance Σ, $H(t) = \begin{bmatrix} y(t-1)^T & y(t-2)^T & \cdots & y(t-p)^T \end{bmatrix}^T$, and the model parameter to be estimated is

$$A = \begin{bmatrix} A_1 & A_2 & \cdots & A_p \end{bmatrix}.$$

In this work, the maximum likelihood (ML) is considered to estimate the parameter from the data $y(1), y(2), \ldots, y(N)$, and N is the number of time points. The ML estimation formulation is

$$\underset{A,\Sigma}{\text{maximize}} \quad \frac{N-p}{2} \log\det \Sigma^{-1} - \frac{1}{2}\|L(Y - AH)\|_F^2$$

when L is the function of Σ^{-1} as $L^T L = \Sigma^{-1}$, and the problem parameters are

$$Y = \begin{bmatrix} y(p+1) & y(p+2) & \cdots & y(N) \end{bmatrix}, \quad H = \begin{bmatrix} y(p) & y(p+1) & \cdots & y(N-1) \\ y(p-1) & y(p) & \cdots & y(N-2) \\ \vdots & \vdots & & \vdots \\ y(1) & y(2) & \cdots & y(N-p) \end{bmatrix}.$$

It is known that the closed-form ML solutions of A and Σ are given by

$$\hat{A} = YH^T(HH^T)^{-1}, \quad \hat{\Sigma} = \frac{1}{N-p}\sum_{t=p+1}^{N} (y(t) - \hat{A}H(t))(y(t) - \hat{A}H(t))^T,$$

where $\hat{\Sigma}$ can be calculated, when $\hat{A}$ is known. When the noise in a linear model is Gaussian, it is known that the ML formulation is equivalent to the least-squares (LS) estimation formulation

$$\underset{A}{\text{minimize}} \quad \|Y - AH\|_F^2. \tag{4}$$

In the next section, the LS estimation formulation will be considered to incorporate the Granger causality and stability conditions in the model.

3 Wald Test for Granger causality

To learn a Granger causality is to test whether the hypothesis that $(\hat{A}_k)_{ij} = 0$ is true. The test is done in order to determine the zero structures of the estimated model parameters. We can test each of the (i, j) entries of A by considering the restriction function

$$r(\hat{\theta}) = \Big((\hat{A}_1)_{ij}, (\hat{A}_2)_{ij}, \dots, (\hat{A}_p)_{ij} \Big) \tag{5}$$

when $\hat{\theta}$ is an estimation of θ, and $\theta = \mathrm{vec}(A)$ (vectorization of A). In Wald test, the null hypothesis is

$$H_0 : r(\hat{\theta}) = 0$$

Thus, the Wald statistic value (W_{ij}) for the test is of the form [2]

$$W_{ij} = r(\hat{\theta})^T \left[\widehat{\mathbf{Avar}}(\hat{\theta})_{ij} \right]^{-1} r(\hat{\theta}) \tag{6}$$

where $\widehat{\mathbf{Avar}}(\hat{\theta})_{ij}$ is $p \times p$ main diagonal block matrix of $\widehat{\mathbf{Avar}}(\hat{\theta})$ corresponding to $r(\hat{\theta})$ (see the details of evaluating W_{ij} in the Appendix.) Under the null hypothesis that $(A_k)_{ij} = 0$, the Wald statistic (W_{ij}) converges in distribution to a Chi-square distribution with p degrees of freedom. In Wald test, we denote $\mathcal{X}^2_{\alpha,p}$ the critical value obtaining from a significance value α defined as

$$\alpha = \mathbf{Prob}(W_{ij} > \mathcal{X}^2_{\alpha,p}). \tag{7}$$

We reject the null hypothesis H_0 if $W_{ij} > \mathcal{X}^2_{\alpha,p}$.

The test can be performed for each (i, j) entry of A_k's by repeating the test for all $j \neq i, i, j = 1, 2, \dots, n$; therefore, we can form the Wald statistic into a matrix $W = [W_{ij}]$ and referred to this as the Wald statistic matrix. Consequently, W_{ij} is compared with the corresponding critical value, and the result can be displayed in a binary matrix contains only ones and zeros. To clarify its meaning, if an (i, j) component of the binary matrix is one then y_j does not Granger-cause to y_i. On the other hand, if an (i, j) component is zero then y_j does not Granger-cause to y_i.

We illustrate performing the Wald test for learning a Granger causality in a model by a simulation. Time series $y(t)$ for $t = 1, 2, \dots, 1000$, are generated from an AR model with $n = 20, p = 4$, and the AR coefficients containing 50% of nonzero entries. The Wald test is done on the AR model parameters estimated by the LS method with the results shown in Fig. 1 and 2. We report the two types of errors:

$$\begin{aligned} \text{Type I Error} &= \mathbf{Prob}((\hat{A}_k)_{ij} \neq 0 \mid (A_k)_{ij} = 0), \\ \text{Type II Error} &= \mathbf{Prob}((\hat{A}_k)_{ij} = 0 \mid (A_k)_{ij} \neq 0). \end{aligned}$$

In other words, falsely identifying zero as nonzero is regarded as the type I error and falsely identifying nonzero as zero is called the type II error. From the definition of α in (7), as α increases, the chance of rejecting the null hypothesis also increases (we accept that there are more nonzero in $\hat{A}$, which can be seen from the decrease in $\mathcal{X}^2_{\alpha,p}$ as well.) As a result, Fig. 1 and Fig. 2 illustrate this conclusion that as α increases, we see an increase in the type I error and a decrease in type II error.

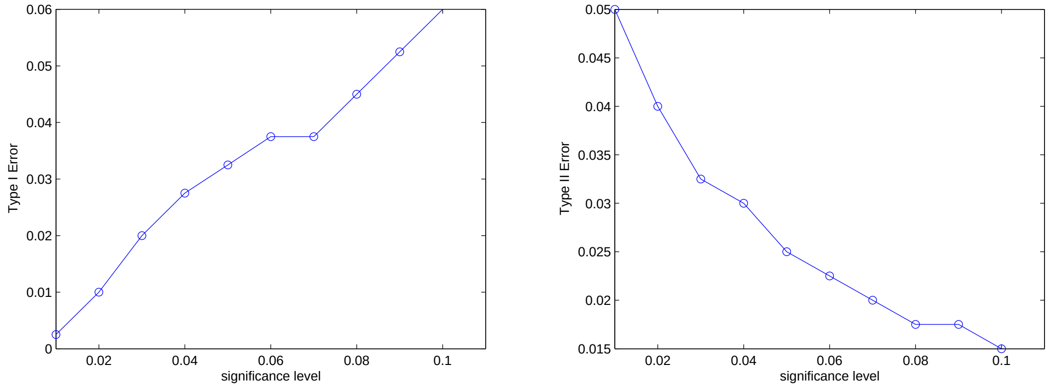
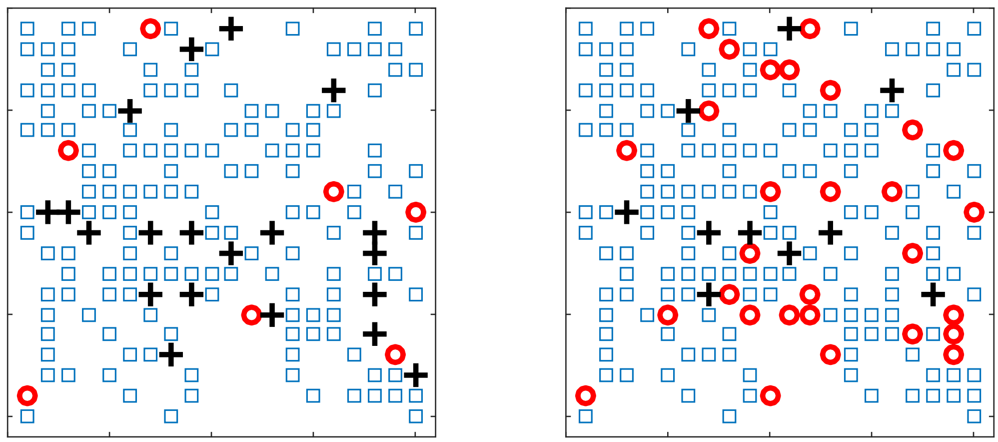


Fig. 1. Two type errors when varying the significance level (α)



(a) $\alpha = 0.01$

(b) $\alpha = 0.1$

Fig. 2. The zero structure of the estimated AR parameters compared with the true ones. The entries with $\square$ mark the common nonzero entries between the estimates and the true values; $\bigcirc$ represents the entries that the true value is zero while the estimated is nonzero; $+$ show the entries that the true value is nonzero while the estimated is zero, and the blank locations show the entries that both the true value and the estimates are zero.

The zero pattern in the estimated AR matrices which corresponds to Granger causality (GC) can be described through an index set I. For example, let $y(t) = (y_1(t), y_2(t), y_3(t))$ and the GC pattern we learned is that y_1 does not Granger-cause y_2 and y_3 does not Granger-cause y_1; thus, we have $I = \{(2,1),(1,3)\}$. The Granger causality concluded from the data allows us to reconsider an estimation problem that encodes the causality constraint in order to have a model that both explains the dynamics of the time series and reveal the relationships among the variables simultaneously. The estimation problem can be cast as the least-squares (LS) subject to Granger causality constraints:

$$\begin{array}{ll} \underset{A}{\text{minimize}} & \|Y - AH\|_F^2 \\ \text{subject to} & (A_k)_{ij} = 0, \; k = 1, \dots, p, \; (i,j) \in I. \end{array} \tag{8}$$

We refer to (8) as the *GC-constrained LS formulation* and it is regarded as a simple quadratic problem with linear constraints. Its closed-form solution can be efficiently computed and shown in [13]. We also note that the ML and LS estimation subject to GC constraint are no longer equivalent (shown in [14]) but we opt to use (8) for its simplicity while ML estimation with GC constraint leads to nonlinear optimality condition in A.

4 Model Estimation with Stability Constraints

Neither the unconstrained LS estimation in (4) nor the GC-constrained LS estimation in (8) guarantees that the estimated model is stable. It was addressed in [6] that obtaining unstable estimated models may occur if some of the eigenvalues of the true model are too close to the unit circle. If the unstable estimated model is further used in time series prediction, spectrum analysis, or other purposes that require the stationarity of the AR model, this would lead to a meaningless interpretation. As a result, stability conditions must be concerned in the parameter estimation in order for the estimated model to explain meaningful physical characteristics.

The model (1) can be written as a discrete-time linear dynamical system

$$x(t+1) = \mathcal{A}x(t) + \mathcal{B}u(t) \tag{9}$$

where x is a state variable and $\mathcal{A}$ is the dynamic matrix given by

$$\mathcal{A} = \begin{bmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I & 0 \end{bmatrix}_{np\times np}. \tag{10}$$

It is well-known that a discrete-time linear system is stable if and only if the magnitudes of all the eigenvalues of $\mathcal{A}$ are strictly less than one, *i.e.*, $|\lambda_i(\mathcal{A})| < 1$ for $i = 1, 2, \dots, np$. This condition leads to highly nonlinear constraints in the entries of $A_1, A_2, \dots, A_p$ and generally are difficult to be directly enforced in an estimation formulation.

Denote $\rho(\mathcal{A}) = \max_k |\lambda_k(\mathcal{A})|$ the spectral radius of $\mathcal{A}$. It then follows that the stability of $\mathcal{A}$ is equivalent to the condition $\rho(\mathcal{A}) < 1$. It can be shown that the spectral radius is related to an induced norm of $\mathcal{A}$, defined as $\|\mathcal{A}\| = \max_{\|x\|\leq 1} \|\mathcal{A}x\|$, by $\rho(A) \leq \|\mathcal{A}\|$. As a result,

$$\|\mathcal{A}\| < 1 \tag{11}$$

is a *sufficient* condition for the stability of $\mathcal{A}$. We can show that for some typical choices of norms in (11), it leads to a trivial model when being applied to $\mathcal{A}$ with the structure given in (10).

Define $\bar{A} = \begin{bmatrix} A_1 & A_2 & \dots & A_{p-1} \end{bmatrix}$ and consider $\mathcal{A}$ having the structure as (10). The condition $\|\mathcal{A}\|_2 < 1$ is equivalent to $\lambda_{\max}(\mathcal{A}^T\mathcal{A}) < 1$, which is also equivalent to the condition $\mathcal{A}^T\mathcal{A} \preceq I$. Therefore, $\mathcal{A}$ is stable if

$$I - \mathcal{A}^T\mathcal{A} = \begin{bmatrix} -\bar{A}^T\bar{A} & -\bar{A}^T A_p \\ -A_p^T\bar{A} & I - A_p^T A_p \end{bmatrix} \succeq 0.$$

The condition on the $(1,1)$ block of the above matrix suggests that $-\bar{A}^T\bar{A} \succeq 0$ and can only occur when $\bar{A} = 0$, or $A_1, \dots, A_{p-1}$ have to be zero. The results are trivial since all the AR coefficients are zero except A_p, while $A_1, \dots, A_p$ should have the same zero structure if we are concerned with the Granger causality. Moreover, it is too restrictive to have a model that $y(t)$ only depends on $y(t-p)$ (from the result that only A_p is nonzero). Likewise, if we consider the condition $\|\mathcal{A}\|_1 < 1$ and that $\|\mathcal{A}\|_1 = \max_j \sum_i |\mathcal{A}_{ij}|$ then it follows that each column block of $\mathcal{A}$ in (10) always contains 1 except the last block. Therefore, $\|\mathcal{A}\|_1 \leq 1$ implies that $A_1, \dots, A_{p-1}$ have to be zero which are also a trivial solution.

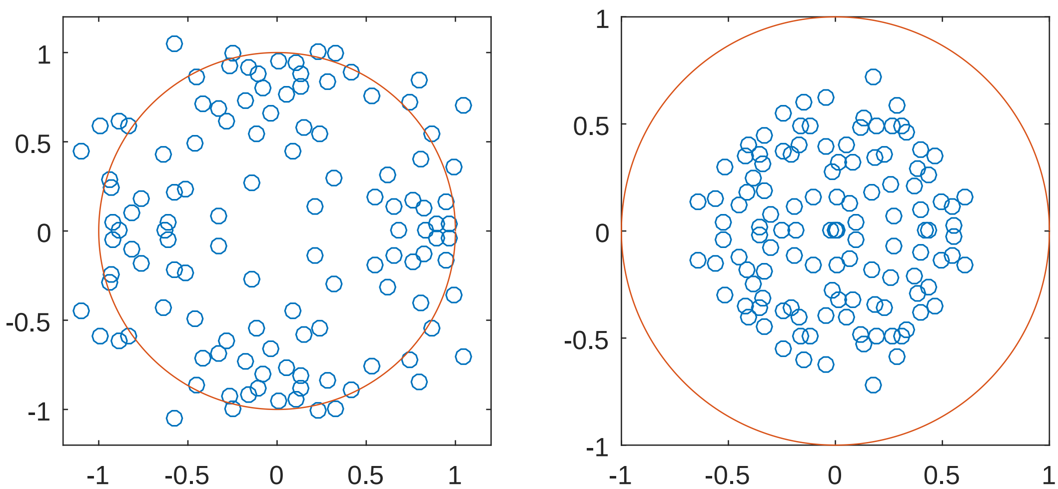
In this paper, we propose to use

$$\|\mathcal{A}\|_\infty < 1 \tag{12}$$

as a sufficient condition for the stability where $\|\mathcal{A}\|_1 = \max_i \sum_j |\mathcal{A}_{ij}|$. We see that the condition (12) with $\mathcal{A}$ in (10) enforces A_k's to have small magnitudes and some can be zero. Another benefit is the condition is also in a convex form and suitable for being included in the estimation formulation. Therefore, we present an AR estimation with stability and Granger causality constraints as

$$
\begin{aligned}
\underset{A,\mathcal{A}}{\text{minimize}} \quad & \|Y - AH\|_F^2 \\
\text{subject to} \quad & \mathcal{A} = \begin{bmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I & 0 \end{bmatrix}, \\
& A = \begin{bmatrix} A_1 & A_2 & \cdots & A_p \end{bmatrix}, \\
& \|\mathcal{A}\|_\infty \leq 1, \\
& (A_k)_{ij} = 0, \qquad (i,j) \in I
\end{aligned}
\tag{13}
$$

where $I \subset \{1,2,\dots,n\} \times \{1,2,\dots,n\}$ is the index set that contains the zero pattern in A_k. The problem (13) will be referred to as the *GC and stability constrained LS* formulation. The zero structure can be determined by the Wald test. The presented problem is convex which can be solved by many generic convex program solvers. CVX [15], for instance, is used for solving the problem (13) and we compare the eigenvalues of the estimated models with the model from the unconstrained LS estimation (4). The results are shown in Fig. 3 where all eigenvalues of the stable model lie inside the unit circle, while the unconstrained LS estimation does not provide a stable model.



(a) GC-constrained LS estimation (b) GC and stability constrained LS estimation

Fig. 3. The eigenvalues of a matrix $\mathcal{A}$ (blue) positions in complex plane and compare with a unit circle (red). Two estimated models are obtained from the GC-constrained LS and the GC and stability constrained LS formulations.

Conservatism on the proposed stability constraint. With the additional ∞-norm constraint in the problem (13), we note that the minimized objective function will increase relatively to the objective of the unconstrained problem (4) or to the GC-constrained problem (8). In other words, to guarantee the stability and reveal a Granger causality, we trade off with the model accuracy. To illustrate the conservatism of the stability constraints in problem (13), we design an experiment by generating time series data and the model parameter A with $n = 40, p = 3$. Using the collected data $N = 200$, we estimate the parameter by LS estimation formulation (4), the GC-constrained LS formulation (8), and the GC and stability constrained LS formulation (13), respectively. The minimized objective function values representing the goodness of the model are computed versus N, while A is fixed. Fig 4 shows that adding stability constraints introduces much higher gaps in the cost objective than adding GC constraints alone. In particular, if N is large and therefore the unconstrained LS estimates should be close to the true values, then the GC-constrained LS estimates provide stable models and there is no need to add the stability constraints.

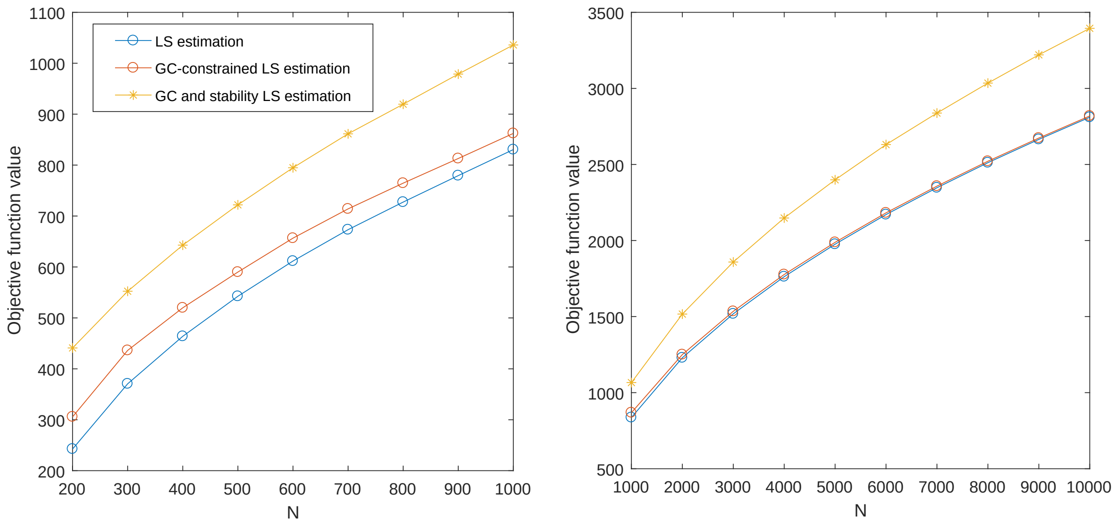


Fig. 4. Increases in the minimized cost objectives (model accuracy) when more constraints are added to the estimation. *Left:* small to moderate sample sizes. *Right:* large sample sizes.

To reduce the conservatism of the stability constraints introduced in the problem (13), we propose a scheme, shown in Fig 5, of estimating stable AR models that reveal Granger causality from time series based on the three estimation formulations: (4), (8) and (13). Firstly, the Wald test is applied on the unconstrained LS estimate of AR Model and reveals a Granger causality (GC) from the time series. This GC relationship is encoded as constraints and an AR model is re-estimated by the GC constrained LS formulation. Secondly, the stability of the resulting model is checked; if it is stable, the estimated model can be readily used for other purposes. However, if the model is not stable, we re-estimate the model again by using the GC and stability constrained LS formulation. Performing the estimation this way provides us an AR model with both guaranteed stability and GC pattern underlying the time series.

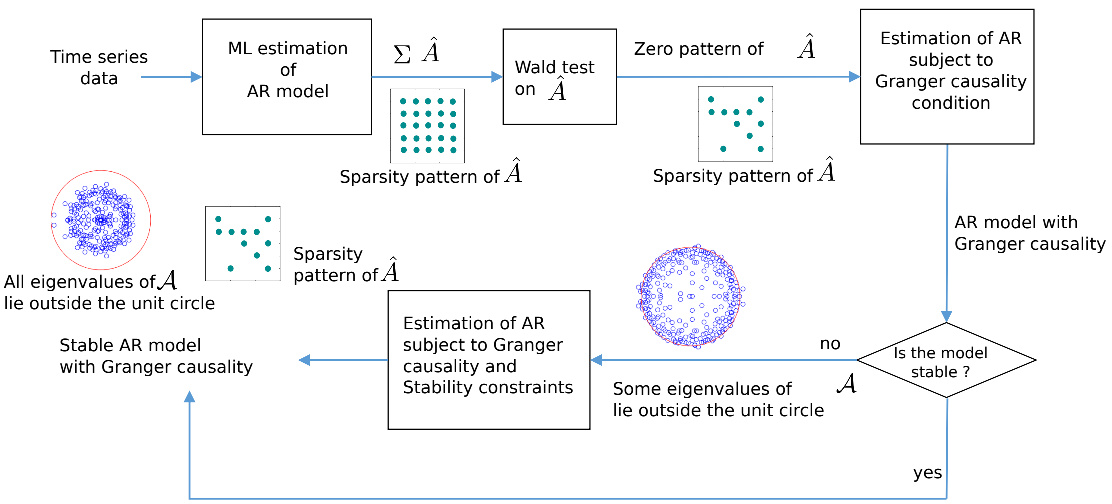


Fig. 5. Diagram of stable AR model estimation with Granger causality.

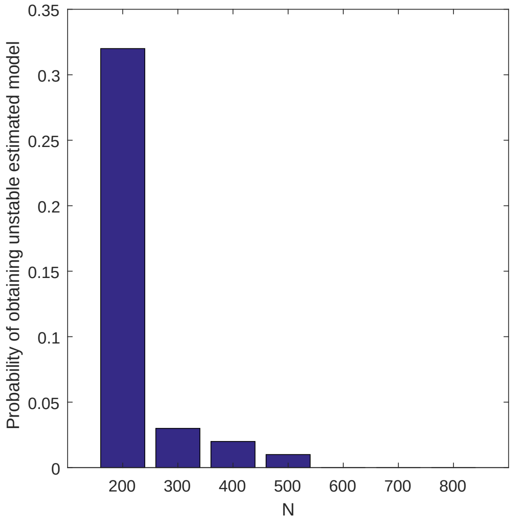
5 Experiments

In this section, we perform two experiments in order to illustrate our approach. The first experiment is to explore possible factors that lead to unstable estimated model using the unconstrained LS and GC-constrained LS methods. Those factors include i) the number of data samples (N) ii) the AR model order p and iii) the pattern of GC constraints used in the estimation. In this experiment, we generate a sparse AR model with $n = 20, p = 4$ and the nonzero AR coefficient density of 0.5 where some of the system eigenvalues are chosen to lie closely to the unit circle.

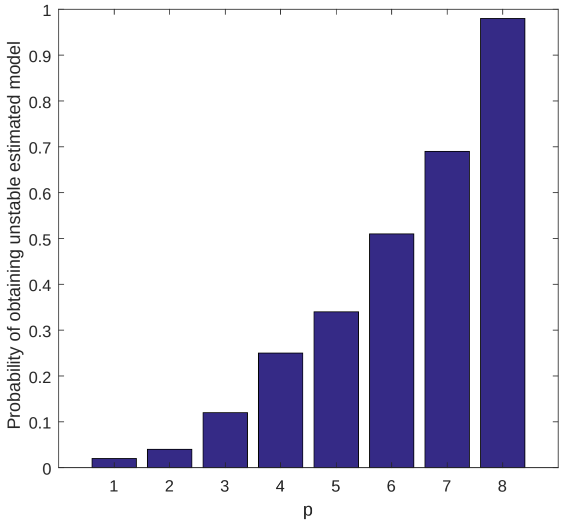
- Effect of N: We vary N from small to large and for each N, 200 time series are generated and we estimate the models using the unconstrained LS method. The probability of obtaining unstable estimated model calculated on the 200 trials is presented in Figure 6a. As expected, when N is small, the LS estimates are not accurate and the estimated models are more likely to be unstable.

- Effect of p used in the estimation: Supposedly the true AR model order is not known. We vary p from 1 to 8 and for each p, 200 time series are generated; each of which contains 200 time points. Models are estimated by the unconstrained LS method and the probability of obtaining unstable estimated model calculated on the 200 data sets is presented in Figure 6b. Using p that is higher than the true order yields more chances of obtaining unstable models since there are more parameters to be estimated while N is fixed. Therefore, the estimates which are less closer to the true values, could be highly deviated from the stable region.

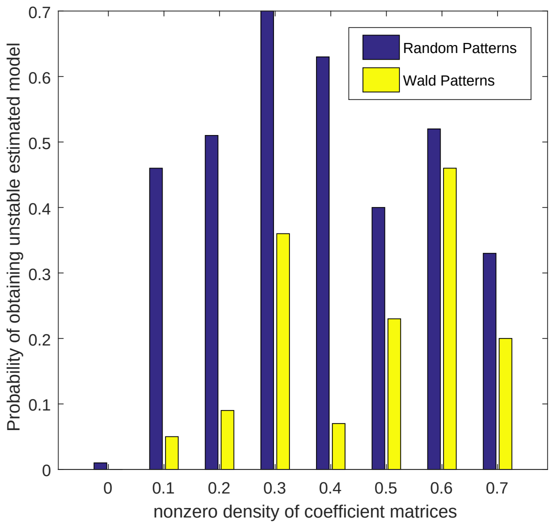
- Effect of GC-constraint pattern used in the estimation: In this experiment, we generate stable AR models with various densities of AR matrices. For each density value, 200 trials of time series are generated and we randomize 200 different zero patterns of GC constraints corresponding to the same density value of nonzero entries in AR matrices. Each of these 200 patterns may or may not be the same as the true GC pattern embedded in the true model. We also consider the zero pattern of GC constraint as a result from the Wald test performed on each of 200 trials. Models are estimated by the GC-constrained LS method and the probability of obtaining unstable estimated models calculated on the 200 trials is presented in Figure 6c. The sparsest models (corresponding to uncoupled AR models) turn to be most likely to be stable but do not reflect intrinsic GC relationships underlying in the data. The models estimated by using GC constraints explained from the Wald test give a smaller chance of being unstable compared to models estimated using other random GC constraints.



(a) The effect of the number of samples (N)



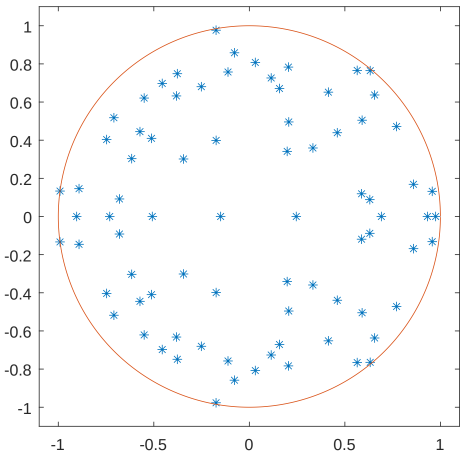
(b) The effect of the AR model order (p)



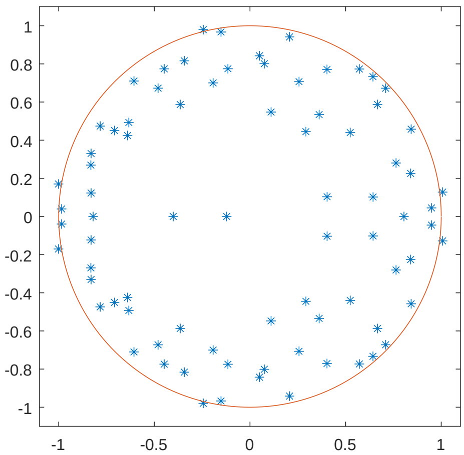
(c) The effect of GC-constraint patterns

Fig. 6. The probability of obtaining unstable estimated models affected by the number of samples N, the AR model order p, and GC-constraint patterns used in the estimation process.

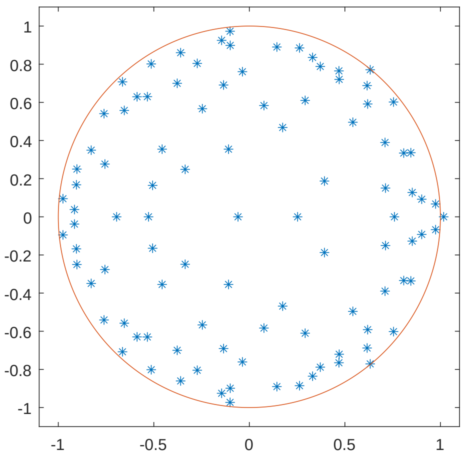
In the second experiment, we compare the estimation results from the three methods: i) unconstrained LS (4), ii) GC-constrained LS (8) and iii) GC and stability constrained LS (13). A true sparse AR model with $n = 20, p = 4$ and the nonzero AR coefficient density of 0.5. This model is particularly chosen such that some of its eigenvalue lie closely to the boundary of the unit circle. In other words, the true model is almost unstable. Then this model is used to generate a time series of 200 time points for estimation. We compare the eigenvalue locations of the models estimated from the three approaches plotted on the complex plane. The results are illustrated in Fig 7 and show that although the true AR model is stable, the unconstrained LS and GC-constrained LS estimation method can lead to unstable models, as well as using a different AR model order from the true order. Some of the eigenvalues are still close to the boundary but no longer inside the unit circle. Figure 7f shows that GC and stability constrained LS estimation can guarantee the model stability where the region of eigenvalues are shrunk inside the unit circle.



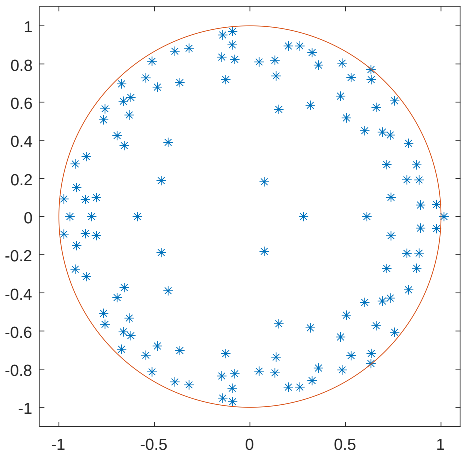
(a) True AR model ($p = 4$)



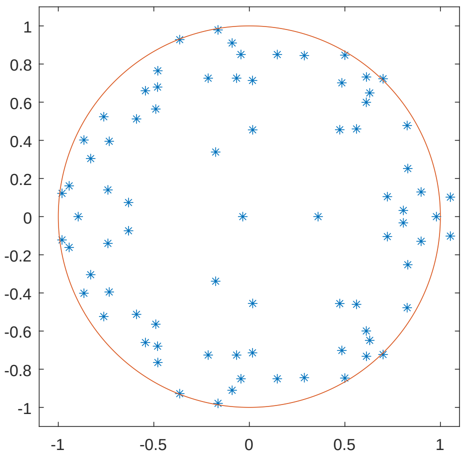
(b) Unconstrained LS using $p = 4$



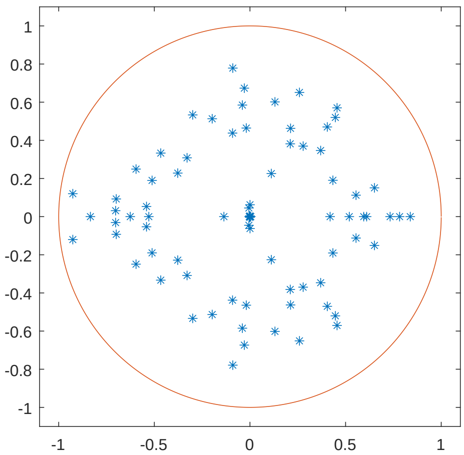
(c) Unconstrained LS using $p = 5$



(d) Unconstrained LS using $p = 6$



(e) GC-constrained LS using $p = 4$



(f) GC and stability constrained LS using $p = 4$

Fig. 7. The eigenvalues of a matrix $\mathcal{A}$ estimated by different methods. The eigenvalues are marked by the blue stars which are located on the complex plane compared to the unit circle (red).

6 Conclusion

This paper aims to provide an estimation formulation of AR models that explains a Granger causality (GC) through the zero structure of estimated AR coefficients, and to guarantee the estimated model stability. The scheme in Fig 5 concludes our framework. Firstly, we estimate an AR model by ML (or LS) method which typically reveals no information about causality as the AR coefficient matrices are usually dense. Thus, the Wald test is performed on the significance of the LS estimate values in order to test the hypothesis for Granger causality. Then the obtained causal structure will be used as the GC constraints in the LS estimation. Subsequently, the estimated model is checked the stability. If the model is not stable, we can re-estimate the model using the formulation with stability constraints, proposed as the unit-bounded infinity-norm of the dynamic matrix. The resulting model provides not only the Granger causality explaining the relationships of the variables in time series, but also is assured to be stable. Our LS estimation formulations are proposed in a convex framework and therefore can be efficiently solved by existing convex program generic solvers.

References

[1] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*. Springer, 2005.

[2] W. H. Greene, "Econometric analysis," *Upper Saddle River, NJ: Prentice Hall*, vol. 24, pp. 25–26, 2008.

[3] V. Cerone, D. Piga, and D. Regruto, "Bounding the parameters of linear systems with stability constraints," in *American Control Conference (ACC), 2010*. IEEE, 2010, pp. 2152–2157.

[4] U. Nallasivam, B. Srinivasan, V. Kuppuraj, M. N. Karim, and R. Rengaswamy, "Computationally efficient identification of global ARX parameters with guaranteed stability," *IEEE Transactions on Automatic Control*, vol. 56, no. 6, pp. 1406–1411, 2011.

[5] S. L. Lacy and D. S. Bernstein, "Subspace identification with guaranteed stability using constrained optimization," *Automatic Control, IEEE Transactions on*, vol. 48, no. 7, pp. 1259–1263, 2003.

[6] J. Mari, P. Stoica, and T. McKelvey, "Vector ARMA estimation: A reliable subspace approach," *Signal Processing, IEEE Transactions on*, vol. 48, no. 7, pp. 2092–2104, 2000.

[7] D. N. Miller and R. A. De Callafon, "Subspace identification with eigenvalue constraints," *Automatica*, vol. 49, no. 8, pp. 2468–2473, 2013.

[8] F.-X. Orbandexivry, Y. Nesterov, and P. Van Dooren, "Nearest stable system using successive convex approximations," *Automatica*, vol. 49, no. 5, pp. 1195–1203, 2013.

[9] B. Boots, "Learning stable linear dynamical systems," Ph.D. dissertation, MS Thesis in Machine Learning, Carnegie Mellon University, 2009.

[10] G. Mallet, G. Gasso, and S. Canu, "New methods for the identification of a stable subspace model for dynamical systems," in *2008 IEEE Workshop on Machine Learning for Signal Processing*. IEEE, 2008, pp. 432–437.

[11] L. Buesing, J. H. Macke, and M. Sahani, "Learning stable, regularised latent models of neural population dynamics," *Network: Computation in Neural Systems*, vol. 23, no. 1-2, pp. 24–47, 2012.

[12] K. Turksoy, E. S. Bayrak, L. Quinn, E. Littlejohn, and A. Cinar, "Guaranteed stability of recursive multi-input-single-output time series models," in *American Control Conference (ACC), 2013*. IEEE, 2013, pp. 77–82.

[13] A. Pongrattarakul, P. Lerdkultanon, and J. Songsiri, "Sparse system identification for discovering brain connectivity from fMRI time series," in *Proceedings of SICE Annual Conference*, 2013, pp. 949–954.

[14] N. Raksasri and J. Songsiri, "Exploring Granger causality for time series via statistical tests on estimated models," Chulalongkorn University, Tech. Rep., 2016, available in Thai at http://jitkomut.eng.chula.ac.th/group.html.

[15] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[16] A. C. Cameron and P. K. Trivedi, *Microeconometrics: methods and applications*. Cambridge university press, 2005.

7 Appendix

This section provides the details of deriving the Wald statistic in (6). In estimation, suppose there are N measurement samples of random variable y and the goal is to estimate a model parameter θ. The estimate of θ is typically a function of y and its asymptotic covariance matrix is given by

$$\mathbf{Avar}(\theta) = \frac{1}{N}\mathcal{I}(\theta)^{-1} \tag{14}$$

when $\mathcal{I}(\theta)$ is Fisher Information of θ calculated from 1 sample of y [2, §14]. The definition of $\mathcal{I}(\theta)$ is

$$\mathcal{I}(\theta) = -\mathbf{E}\left[\nabla_\theta^2 \log f(y|\theta)\right],$$

where $\log f(y|\theta)$ is the loglikelihood function of y paramitrized by model parameter θ. A Wald hypothesis test begins with assuming a null hypothesis

$$H_0 : r(\theta) = 0$$

where $r : \mathbf{R}^n \to \mathbf{R}^m, m \leq n$ is a restriction function on the parameter θ. To test whether H_0 is true, the Wald statistic is given by [16, §5]

$$W = r(\hat{\theta})^T \left[D_r(\hat{\theta})\widehat{\mathbf{Avar}}(\hat{\theta})D_r(\hat{\theta})^T\right]^{-1} r(\hat{\theta}) \tag{15}$$

when $D_r(\theta)$ is Jacobian of r with respect to θ and $\widehat{\mathbf{Avar}}(\hat{\theta})$ is an estimate of $\mathbf{Avar}(\hat{\theta})$ from (14). For ML estimation, $\widehat{\mathbf{Avar}}(\hat{\theta})$ can be calculated from

$$\widehat{\mathbf{Avar}}(\hat{\theta}) = \left[-\sum_{i=1}^{N} \nabla^2 \log f(y_i|\hat{\theta})\right]^{-1}. \tag{16}$$

In what follows, we will characterize $\widehat{\mathbf{Avar}}(\hat{\theta})$ and W by arranging (3) in a vector form:

$$y(t) = \bar{H}(t)\theta + v(t), \quad v(t) \sim \mathcal{N}(0, \Sigma)$$

where the model parameters are vectorized to

$$\begin{aligned}
\theta &= \begin{bmatrix} B_{11}^T & \cdots & B_{1n}^T & B_{21}^T & \cdots & B_{2n}^T & \cdots & B_{nn}^T \end{bmatrix}^T \in \mathbf{R}^{n^2p}, \\
B_{ij} &= ((A_1)_{ij}, (A_2)_{ij}, ..., (A_p)_{ij}) \in \mathbf{R}^p \\
\bar{y}_i &= \begin{bmatrix} y_i(t-1) \\ y_i(t-2) \\ \vdots \\ y_i(t-p) \end{bmatrix}_{p\times 1}, \quad \bar{y} = \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_n \end{bmatrix}_{np\times 1}, \quad \bar{H}(t) = \begin{bmatrix} \bar{y}^T & 0 & 0 & \cdots & 0 \\ 0 & \bar{y}^T & 0 & \cdots & 0 \\ 0 & 0 & \bar{y}^T & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \bar{y}^T \end{bmatrix}_{n\times n^2p}.
\end{aligned}$$

With Gaussian assumption on the additive noise, the ML estimation is equivalent to the LS problem (4) whose form is equivalent to

$$\underset{\theta}{\text{minimize}} \quad \|F\theta - z\|_2^2$$

for

$$\begin{aligned} F &= \begin{bmatrix} \bar{H}^T(p+1) & \bar{H}^T(p+2) & \cdots & \bar{H}^T(N) \end{bmatrix}^T_{n(N-p)\times n^2 p}, \\ z &= \begin{bmatrix} y^T(p+1) & y^T(p+2) & \cdots & y^T(N) \end{bmatrix}^T_{n(N-p)\times 1} \end{aligned}$$

The Wald test of Granger causality (whether y_j Granger causes y_i) therefore corresponds to considering the restriction function

$$r(\hat{\theta}) = \begin{bmatrix} (\hat{A}_1)_{ij} \\ (\hat{A}_2)_{ij} \\ \vdots \\ (\hat{A}_p)_{ij} \end{bmatrix} = \hat{B}_{ij}$$

The estimated asymptotic covariance matrix of $\hat{\theta}$, computed from (16) will be

$$\widehat{\mathbf{Avar}}(\hat{\theta}) = \left[\sum_{t=p+1}^{N} \bar{H}(t)^T \hat{\Sigma}^{-1} \bar{H}(t) \right]^{-1},$$

with size of $n^2 p \times n^2 p$. Since the parameter θ is a vector of n^2 blocks; each of which has p components, we can regard $\widehat{\mathbf{Avar}}(\hat{\theta})$ as a matrix of $n \times n$ blocks and each block has size $p \times p$.

Next, we derive the Wald statistic (W_{ij}) for $r(\theta)$ from (15). Note that while testing Granger causality for each (i, j) pair, the Jacobian matrix of $r(\hat{\theta})$ has size $p \times n^2 p$ or can be viewed as a fat matrix of n^2 blocks; each block has size $p \times p$. Since $r(\hat{\theta})$ is linear in θ and only a function of B_{ij}, then the column blocks of $D_r(\hat{\theta})$ are all zero, except the one corresponds to the (i, j) entries in $\hat{\theta}$. The term $D_r(\hat{\theta})\widehat{\mathbf{Avar}}(\hat{\theta})D_r(\hat{\theta})^T$ is simplified to the main diagonal block of $\widehat{\mathbf{Avar}}(\hat{\theta})$ corresponding to B_{ij} in θ, which will be denoted by $\widehat{\mathbf{Avar}}(\hat{\theta})_{ij}$. For example, $n = 3, p = 2$ or $y = (y_1, y_2, y_3)$. Testing y_2 causes y_1 gives

$$D_r(\hat{\theta}) = \begin{bmatrix} 0 & I_p & 0 & \cdots & 0 \end{bmatrix}_{p\times n^2 p}.$$

As a result, the Wald statistic of testing if y_j Granger causes y_i in (15) is

$$W_{ij} = \hat{B}_{ij}^T \left[\widehat{\mathbf{Avar}}(\hat{\theta})_{ij} \right]^{-1} \hat{B}_{ij}$$

where $\widehat{\mathbf{Avar}}(\hat{\theta})_{ij}$ is $p \times p$ main diagonal block matrix of $\widehat{\mathbf{Avar}}(\hat{\theta})$ corresponding to $r(\hat{\theta})$.