Granger Causality Analysis Based on Quantized Minimum Error Entropy Criterion

Badong Chen, Senior Member, IEEE, Rongjin Ma, Member, IEEE, Siyu Yu, Shaoyi Du, and Jing Qin, Member, IEEE

Abstract—Linear regression model (LRM) based on mean square error (MSE) criterion is widely used in Granger causality analysis (GCA), which is the most commonly used method to detect the causality between a pair of time series. However, when signals are seriously contaminated by non-Gaussian noises, the LRM coefficients will be inaccurately identified. This may cause the GCA to detect a wrong causal relationship. Minimum error entropy (MEE) criterion can be used to replace the MSE criterion to deal with the non-Gaussian noises. But its calculation requires a double summation operation, which brings computational bottlenecks to GCA especially when sizes of the signals are large. To address the aforementioned problems, in this letter, we propose a new method called GCA based on the quantized MEE (QMEE) criterion (GCA-QMEE), in which the QMEE criterion is applied to identify the LRM coefficients and the quantized error entropy is used to calculate the causality indexes. Compared with the traditional GCA, the proposed GCA-QMEE not only makes the results more discriminative, but also more robust. Its computational complexity is also not high because of the quantization operation. Illustrative examples on synthetic and EEG datasets are provided to verify the desirable performance and the availability of the GCA-QMEE.

Index Terms—Granger causality analysis, mean square error criterion, minimum error entropy criterion, quantized minimum error entropy criterion, linear regression model.

I. INTRODUCTION

GRANGER causality analysis (GCA) is one of the most commonly used methods to detect the causality between a pair of time series, which finds successful applications in various areas, such as economics [1], [2], climate studies [3], [4], genetics [5], and neuroscience [6]–[9]. The original GCA is based on the linear regression model (LRM), so it is easy to understand and utilize [10], [11]. However, the mean square error (MSE) criterion used in LRM takes only the second order moments of the errors into account. When signals are seriously contaminated by non-Gaussian noises, it becomes hard to accurately identify the LRM coefficients [12] and this may cause the traditional GCA to wrongly detect the causal relationships. The minimum error entropy (MEE) criterion in information theoretic learning (ITL) [12]–[16], which takes higher order statistics of the errors into account, is generally superior to the MSE criterion in LRM identification especially when signals are contaminated by non-Gaussian noises, such as mixture Gaussian or Levy alpha-stable noises. But the calculation of the MEE based objective functions requires a double summation operation, which brings computational bottlenecks to GCA especially when the sizes of the signals are large. To address this issue, the quantized MEE (QMEE) criterion, a simplified version of the MEE criterion, was recently proposed in [17].

In this letter, we propose a new causality analysis method called GCA based on QMEE criterion (GCA-QMEE). The new method applies the QMEE to identify the LRM coefficients and uses the quantized error entropy to calculate the causality indexes. Compared with the traditional GCA, the proposed GCA-QMEE not only makes the results more discriminative, but also more robust. Because of the quantization operation, the computational complexity of GCA-QMEE is also not high.

The rest of the letter is organized as follows. Section II describes the original GCA method. Section III first introduces the QMEE criterion and its application to linear regression, then develops the GCA-QMEE method. In Section IV, experiments are conducted to demonstrate the desirable performance and the availability of the proposed method. Finally, conclusion is given in Section V.

II. GRANGER CAUSALITY ANALYSIS

In causality analysis, two vectors $X$ and $Y$ are considered as a pair of time series, which can be written as

$$
X = [x_1, x_2, \ldots, x_N]^T
$$

$$
Y = [y_1, y_2, \ldots, y_N]^T
$$

where $N$ is the size of the time series. Autoregressive (AR) models are used to fit the time series in Granger causality analysis (GCA) which are [18]

$$
x_t = \sum_{i=1}^{p_1} w_{11,i} x_{t-i} + e_{11,t}
$$

$$
y_t = \sum_{i=1}^{p_2} w_{21,i} y_{t-i} + e_{21,t}
$$

where $p_1$ and $p_2$ are the orders of the AR models, $e_{11}$ and $e_{21}$ are the errors, and $w_{11}$ and $w_{21}$ are the coefficients which can easily be solved by the least squares method based on the MSE criterion in linear regression. Under the MSE criterion, the
cost function for estimating the coefficients is \( J = \frac{1}{N} \sum_{i=1}^{N} e_i^2 \), where \( \{e_i\}_{i=1}^{N} \) are the error samples.

Vector autoregressive (VAR) models are also used to fit the time series in GCA, given by [18]

\[
x_t = \sum_{i=1}^{p_1} w_{12,i} x_{t-i} + \sum_{j=1}^{p_2} w_{12,j+p_1} y_{t-j} + e_{12,t} \\
y_t = \sum_{i=1}^{p_1} w_{22,i} y_{t-i} + \sum_{j=1}^{p_2} w_{22,j+p_1} x_{t-j} + e_{22,t}
\]

where \( p_1 \) and \( p_2 \) are the orders of the VAR models, \( w_{12} \) and \( w_{22} \) are the coefficients, and \( e_{12} \) and \( e_{22} \) are the error samples. If \( X \) and \( Y \) are independent, then \( \text{Var}(e_{11}) = \text{Var}(e_{12}) \) and \( \text{Var}(e_{21}) = \text{Var}(e_{22}) \), where \( \text{Var}(\epsilon) \) denotes the variance of the error \( \epsilon \). Otherwise, the two equations don’t hold. For example, if \( X \) is the cause of \( Y \), then \( \text{Var}(e_{21}) > \text{Var}(e_{22}) \). The Granger causality index from \( X \) to \( Y \) can be computed by [19]

\[
F_{X \rightarrow Y} = \log \frac{\text{Var}(e_{21})}{\text{Var}(e_{22})}
\]

Clearly, we have \( F_{X \rightarrow Y} \geq 0 \). With the above index, the causality can be analyzed. Specifically, if \( F_{X \rightarrow Y} > 0 \) in a significant way [10], then \( X \) is the cause of \( Y \); if \( F_{X \rightarrow Y} = 0 \), then there is no causality from \( X \) to \( Y \). Similarly, the Granger causality index from \( Y \) to \( X \) is defined by [19]

\[
F_{Y \rightarrow X} = \log \frac{\text{Var}(e_{11})}{\text{Var}(e_{12})}
\]

If \( F_{X \rightarrow Y} > F_{Y \rightarrow X} \) in a significant way, then the causality from \( X \) to \( Y \) is predominant to the causality from \( Y \) to \( X \).

### III. QMEE Based Granger Causality Analysis

#### A. LRM Identification Based on QMEE

In ITL, Renyi’s entropy of order \( \alpha (\alpha > 0, \alpha \neq 1) \) is widely used as a cost function, which is defined by [20]

\[
H_{\alpha} (\epsilon) = \frac{1}{1-\alpha} \log \left[ \int p^{\alpha} (\epsilon) d\epsilon \right]
\]

where \( p(\epsilon) \) denotes the probability density function (PDF) of the error variable \( \epsilon \), which is often estimated by the Parzen window approach [21]:

\[
p(\epsilon) \approx \frac{1}{N} \sum_{i=1}^{N} G_{\sigma} (\epsilon, e_i)
\]

where \( G_{\sigma} (\epsilon, e_i) \) is the Gaussian kernel function with bandwidth \( \sigma \) : \( G_{\sigma} (\epsilon, e_i) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{(\epsilon - e_i)^2}{2\sigma^2} \right] \).

In this work, without explicit mention we set \( \alpha = 2 \). In this case, we have [14]

\[
H_2 (\epsilon) \approx -\log \left[ \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} G_{\sqrt{2\sigma}} (e_i, e_j) \right]
\]

There is a double summation operator in (8). Thus the computational cost for evaluating the error entropy is \( \mathcal{O}(N^2) \), which is expensive especially for large scale datasets. To reduce the computational complexity of the MEE criterion, we proposed an efficient quantization approach in [17]. The details of this method is presented in Algorithm 1.

**Algorithm 1: Procedure of the Quantization Method [17].**

**Input:** error samples \( \{e_i\}_{i=1}^{N} \)

**Output:** quantized errors \( \{Q(e_i)\}_{i=1}^{N} \) and a codebook \( \{c_i\}_{i=1}^{M} \)

1: **Parameters setting:** quantization threshold \( \epsilon \)
2: **Initialization:** Set \( c_1 = \{e_1\} \), where \( c_i \) denotes the codebook at the iteration \( i \)
3: for \( i = 2, 3, \ldots, N \) do
4: compute the distance between \( e_i \) and \( c_{i-1} \):
5: \[ \text{dis}(e_i, c_{i-1}) = |e_i - c_{i-1}(j^*)| \]
6: if \( \text{dis}(e_i, c_{i-1}) \leq \epsilon \) then
7: keep the codebook unchanged: \( c_i = c_{i-1} \) and quantize \( e_i \) to the closest code word \( Q(e_i) = c_{i-1}(j^*) \)
8: else update the codebook: \( c_i = \{c_{i-1}, e_i\} \) and quantize \( e_i \) to itself: \( Q(e_i) = e_i \)
9: **end if**
10: **end for**

Given \( N \) error samples \( \{e_i\}_{i=1}^{N} \) and a quantization threshold \( \epsilon \), the outputs of this quantization method are \( \{Q(e_i)\}_{i=1}^{N} \) and a codebook \( \{c_i\}_{i=1}^{M} \) containing \( M \) real valued code words \( (M \ll N) \), where \( Q(e) \) denotes the quantization operator. The quantized error entropy is

\[
H_2^Q (\epsilon) = -\log \left[ \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} G_{\sqrt{2\sigma}} (e_i, Q(e_j)) \right]
\]

\[
= -\log \left[ \frac{1}{N^2} \sum_{i=1}^{N} \sum_{m=1}^{M} A_m G_{\sqrt{2\sigma}} (e_i, c_m) \right]
\]

\[
= -\log \left[ I_2^Q (\epsilon) \right]
\]

where \( A_m \) is the number of the error samples that are quantized to the code word \( c_m \), and \( I_2^Q (\epsilon) \) is called the quantized information potential [17]. Under the QMEE criterion, the optimal hypothesis can thus be solved by minimizing the quantized error entropy \( H_2^Q (\epsilon) \). Clearly, minimizing \( H_2^Q (\epsilon) \) is equivalent to maximizing \( I_2^Q (\epsilon) \).

Consider the LRM in which \( e_i = y_i - w^T x_i \), with \( w \) being the coefficient vector to be estimated. Taking the gradient of \( I_2^Q (\epsilon) \) with respect to \( w \), we have

\[
\frac{\partial I_2^Q (\epsilon)}{\partial w} = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{m=1}^{M} A_m \frac{\partial G_{\sqrt{2\sigma}} (e_i, c_m)}{\partial w}
\]

\[
= \tau \sum_{i=1}^{N} \sum_{m=1}^{M} A_m G_{\sqrt{2\sigma}} (e_i, c_m) (y_i - c_m) x_i
\]

\[
+ \tau \left[ \sum_{i=1}^{N} \sum_{m=1}^{M} A_m G_{\sqrt{2\sigma}} (e_i, c_m) x_i x_i^T \right] w
\]
where \( \tau = \frac{1}{N^2(2\pi)^m} \). Setting \( \frac{\partial I_2^Q(e)}{\partial w} = 0 \), a fixed-point equation of the LRM coefficients can be obtained as

\[
w = V^{-1} U \tag{11}\]

where \( U = \sum_{i=1}^{N} \sum_{m=1}^{M} A_m G_{\tau_{\pi}}(e_i, c_m) (y_i - c_m) x_i \) and \( V = \sum_{i=1}^{N} \sum_{m=1}^{M} A_m G_{\tau_{\pi}}(e_i, c_m) x_i x_i^T \). After reaching the steady state by multiple fixed-point iterations \( w_k = (V^{-1} U)_{|w_{k-1}}, k = 1, 2, \ldots, K \), the fixed-point solution of the coefficients can be obtained.

Here we present an illustrative example to compare the performance of the MSE, MEE and QMEE criterions. Consider the following linear system:

\[
y_i = w^T x_i + \varphi_i \tag{12}\]

where \( w^* = [2, 1]^T \), \( x_i \) is assumed to be uniformly distributed over \([-2, 2] \times [-2, 2] \), and \( \varphi_i \) is a noise drawn from:

- Case 1) Gaussian density with zero mean and unit variance:
  \[ f(x) = \frac{1}{\sqrt{2\pi}x_1} \exp\left[-\frac{(x_1-0)^2}{2}\right]; \]
- Case 2) symmetric Gaussian mixture density:
  \[ f(x) = 0.5 \times \frac{1}{\sqrt{2\pi}x_1} \exp\left[-\frac{(x_1-0)^2}{2}\right] + 0.5 \times \frac{1}{\sqrt{2\pi}x_1} \exp\left[-\frac{(x_1+1.5)^2}{2}\right]; \]
- Case 3) asymmetric Gaussian mixture density:
  \[ f(x) = 0.6 \times \frac{1}{\sqrt{2\pi}x_1} \exp\left[-\frac{(x_1-0.5)^2}{2}\right] + 0.4 \times \frac{1}{\sqrt{2\pi}x_1} \exp\left[-\frac{(x_1+1.5)^2}{2}\right]; \]
- Case 4) Levy alpha-stable distribution with characteristic exponent \((0 < \alpha \leq 2)\), skewness \((-1 \leq \beta \leq 1)\), scale parameter \((0 < \gamma < \infty)\) and location parameter \((-\infty < \delta < \infty)\), where \( [\alpha, \beta, \gamma, \delta] = [1.3, 0, 0.4, 0] \).

The root mean squared error (RMSE) is employed to measure the performance, computed by \( RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (w^* - \bar{w})^2} \), where \( w^* \) and \( \bar{w} \) denote the target and the estimated weight vectors respectively.

For the MSE criterion, there is a closed-form solution, so no iteration is needed. For other two criterions, the fixed-point iteration is used to solve the model. In the simulations, the size of samples is \( N = 500 \), the iteration number is \( K = 100 \), the Gaussian kernel bandwidth is \( \sigma = 0.5 \) and the quantization threshold is set at \( \epsilon = 0.4 \). The mean \( \pm \) deviation results of the RMSE over 100 Monte Carlo runs are presented in Table I. From Table I, we observe i) the MEE and QMEE criterions can significantly outperform the traditional MSE criterion in the cases of non-Gaussian noise; ii) the QMEE criterion can achieve almost the same (or even better) performance as the original MEE criterion.

In the Case 2), the training times of the QMEE and MEE are 12.1469 \( \pm \) 0.8445 s and 98.2965 \( \pm \) 2.8661 s respectively. Obviously, compared with the MEE criterion, the computational complexity of the QMEE criterion is much lower.

### B. GCA Based on QMEE

When signals are seriously contaminated by non-Gaussian noises, the traditional GCA may not work well. We propose a new causality analysis method called GCA-QMEE to solve this problem, in which the QMEE criterion is applied to identify (by fixed-point iterations) the LRM coefficients, namely, the coefficients of the AR and VAR models in (2) and (3). The quantized error entropy is used to calculate the causality indexes:

\[
F_{X \rightarrow Y} = H_Q^2(e_{21}) - H_Q^2(e_{22}) = \log \left[ \frac{I_Q^2(e_{22})}{I_Q^2(e_{21})} \right] \tag{13}
\]

\[
F_{Y \rightarrow X} = H_Q^2(e_{11}) - H_Q^2(e_{12}) = \log \left[ \frac{I_Q^2(e_{12})}{I_Q^2(e_{11})} \right]
\]

where \( H_Q^2(e_{11}), H_Q^2(e_{21}), H_Q^2(e_{12}) \) and \( H_Q^2(e_{22}) \) are the trained quantized error entropies of the LRMs, and \( I_Q^2(e_{11}), I_Q^2(e_{21}), I_Q^2(e_{12}) \) and \( I_Q^2(e_{22}) \) are the trained quantized information potentials (after convergence). The orders (or the embedding dimensions) of the LRMs in our approach are determined by the following modified Bayesian Information Criterion (BIC):

\[
p = \min_{1 \leq i \leq p_{\text{max}}} \left\{ N \log \left( H_Q^2(e) \right) + i \log (N) \right\} \tag{14}
\]

where \( p_{\text{max}} \) is the largest embedding dimension we set.

### IV. EXPERIMENTAL RESULTS

#### A. Synthetic Dataset

Suppose the time series \( X \) and \( Y \) are generated by

\[
x_t = \varphi_t \Rightarrow y_t = x_{t-1} + \psi_t \tag{15}
\]

where \( \varphi_t \) denotes a noise following the uniform distribution over \([-2, 2] \), and \( \psi_t \) is a noise drawn from the four PDFs in the illustrative example of the previous section. In this example, obviously, \( X \) is the cause of \( Y \) but not vice versa.

We compare the performance of several GCA methods: GCA-MEE, GCA-QMEE and GCA-MEE, where GCA-MEE represents the traditional GCA, and GCA-MEE corresponds to the GCA-QMEE with quantization threshold \( \epsilon = 0 \). Since \( X \) is the cause of \( Y \), we define the following discrimination index to measure the performance of causality detection:

\[
\rho = \frac{F_{X \rightarrow Y} - F_{Y \rightarrow X}}{F_{X \rightarrow Y}} \tag{16}
\]

Clearly, if \( \rho > 0 \), then \( F_{X \rightarrow Y} > F_{Y \rightarrow X} \), and thus the causal relationship is correctly detected; if \( \rho < 0 \), then the causal relationship is wrongly detected. The bigger the \( \rho \) is, the more discriminative the causality analysis result is (i.e. the difference between \( F_{X \rightarrow Y} \) and \( F_{Y \rightarrow X} \) is more significant). The parameter settings are \( N = 500, K = 100, \sigma = 0.5, \epsilon = 0.4 \) and \( p_{\text{max}} = 10 \). The causality indexes over 100 Monte Carlo experiments are presented in Table II. From Table II, we observe: i) all methods can correctly detect that \( X \) is the cause of \( Y \); ii) the GCA-QMEE and the GCA-MEE can significantly outperform the GCA-MEE in terms of the discrimination index except the case of Gaussian background noise.

To further show the robustness of our proposed method, we evaluate the relative variation ratio (RVR) of the causality index
 TABLE II
CAUSALITY INDEXES OF DIFFERENT METHODS

| Method          | C3→C3 | C3→C4 | C4→C3 | C4→C4 |
|-----------------|-------|-------|-------|-------|
| GCA-MSE         | 0.009 | 0.005 | 0.009 | 0.005 |
| Case 1          | 0.002 | 0.0035| 0.009 | 0.0013|
| Case 2          | 0.005 | 0.0095| 0.009 | 0.0011|
| Case 3          | 0.055 | 0.071 | 0.0016| 0.004 |
| Case 4          | 0.027 | 0.010 | 0.004| 0.0038|
| max              |       |       | 0.0064| 0.004 |

 Fig. 1. Relative variation ratio of the causality from X to Y versus α.

TABLE III
CAUSALITY INDEXES AMONG C3, Cz AND C4 FOR LEFT-HAND MI AND RIGHT-HAND MI AND THEIR STATISTICAL SIGNIFICANCE

| Method          | C3→C3 | C3→C4 | C4→C3 | C4→C4 |
|-----------------|-------|-------|-------|-------|
| GCA-QMEE        | 0.007 | 0.006 | 0.0013| 0.1381|
| GCA-MEE         | 0.036 | 0.057 | 0.0036| 0.0193|
| GCA-MSE         | 0.0013| 0.0919| 0.0028| 0.1163|
| GCA-QMEE        | 0.0924| 0.0800| 0.0018| 0.0481|
| 0.0008 | 0.2137| 0.0111| 0.1711|
| 0.0036 | 0.0422| 0.0049| 0.0719|
| max             | 0.0505| 0.0505|       |       |

B. EEG Dataset

Now we apply the GCA-QMEE to analyze the dataset IIb of BCI competition IV [22]. The dataset are recorded from nine subjects with three electrodes (C3, Cz and C4). The subjects are right-handed, and have normal or corrected-to-normal vision. There are two different motor imagery (MI) tasks, namely, left-hand MI and right-hand MI. The sample frequency is 250 Hz. The EEG signals are already band-pass filtered between 0.5 Hz and 100 Hz with a 50 Hz notch filter enabled. For each subject, we only use the first three sessions. Hence, there are 27 sessions for our analysis. Each session contains some trials of C3, Cz and C4 under left-hand MI and right-hand MI. For each electrode of each session under each task, we extract all the valid time series, then average the multiple time series into one time series. We analyze the causalities among C3, Cz and C4 by the GCA-QMEE. The parameter settings are K = 100, σ = 0.5, ε = 0.05 and pmax = 20. After calculation, we get 27 causality indexes among C3, Cz and C4. The results are averaged in Table III. Additionally, we perform the significance test for all the causality indexes. The process of a commonly used significance test is as follows [23], [24]: for a pair of time series U and V, randomly sample U by 100 times; each time analyze the causality from univariate phase-randomized Fourier-transform surrogate data U to V, and obtain 100 causality indexes F(U→V); count the number of times F(U→V) ≥ F(\text{causality})/n; calculate the statistical significance t = \frac{\text{causality}}{\text{max}}. Here, t is an indicator that reflects the statistical significance of the causality analysis. The smaller the t is, the more significant the causality analysis is. All the results of statistical significance are averaged in Table III.

From Table III, we observe: a) causality indexes from Cz to C3 are greater than those from C3 to Cz during MI; b) causality indexes from Cz to C4 are greater than those from C4 to Cz during MI; c) causality index from C4 to C3 is larger than that from C3 to C4 during left-hand MI, and causality index from C3 to C4 is larger than that from C4 to C3 during right-hand MI; d) the causality index from Cz to C4 during right-hand MI is larger than that from Cz to C3 during left-hand MI, and the causality index from C3 to C4 during right-hand MI is larger than that from C4 to C3 during left-hand MI; which may demonstrate the influence of the brain asymmetry of right-handedness on effective connectivity networks. These results also validate the previous findings [25]–[28].

V. CONCLUSION

In this work, we proposed a new causality analysis method called Granger causality analysis (GCA) based on the quantized minimum error entropy (QMEE) criterion (GCA-QMEE), in which the QMEE criterion is applied to identify the LRM coefficients and the quantized error entropy is used to calculate the causality indexes. Compared with the traditional GCA, the GCA-QMEE not only makes the results more discriminative, but also more robust. Its computational complexity is also not high because of the quantization operation. Experimental results with synthetic and EEG datasets have been provided to confirm the desirable performance of the GCA-QMEE. For the future study, it is interesting to explore the nonlinear causality or the causal relationships among multivariates based on the QMEE criterion.
REFERENCES

[1] S. Paul and R. N. Bhattacharya, “Causality between energy consumption and economic growth in India: A note on conflicting results,” Energy Econ., vol. 26, no. 6, pp. 977–983, 2004.

[2] M. Kar, S. Nazlioglu, and H. Agür, “Financial development and economic growth nexus in the mena countries: Bootstrap panel granger causality analysis,” Econ. Model., vol. 28, no. 1–2, pp. 685–693, 2011.

[3] 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. Climate, vol. 19, no. 7, pp. 1182–1194, 2006.

[4] L. Faes, G. Nollo, S. Stramaglia, and D. Marinazzo, “Multiscale Ganger causality,” Phys. Rev. E, vol. 96, no. 4, 2017. Art. no. 042150.

[5] J. Zhu et al., “Characterizing dynamic changes in the human blood transcriptional network,” PLoS Comput. Biol., vol. 6, no. 2, p. e1000671, 2010.

[6] A. Roebroeck, E. Formisano, and R. Goebel, “Mapping directed influence over the brain using Granger causality and FMRI,” Neuroimage, vol. 25, no. 1, pp. 230–242, 2005.

[7] R. Goebel, A. Roebroeck, D.-S. Kim, and E. Formisano, “Investigating directed cortical interactions in time-resolved FMRI data using vector autoregressive modeling and Granger causality mapping,” Magn. Reson. Imaging., vol. 21, no. 10, pp. 1251–1261, 2003.

[8] D. Marinazzo, M. Pelliccoro, and S. Stramaglia, “Kernel method for nonlinear Granger causality,” Phys. Rev. Lett., vol. 100, no. 14, p. 144103, 2008.

[9] S. Stramaglia, L. Angelini, G. Wu, J. M. Cortes, L. Faes, and D. Marinazzo, “Synergistic and redundant information flow detected by unnormalized Granger causality: Application to resting state FMRI,” IEEE Trans. Biomed. Eng., vol. 63, no. 12, pp. 2518–2524, Dec. 2016.

[10] M. Kamifuruki, M. Ding, W. A. Truccolo, and S. L. Bressler, “Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance,” Biol. Cybern., vol. 85, no. 2, pp. 145–157, 2001.

[11] L. Barnett and A. K. Seth, “Granger causality for state-space models,” Phys. Rev. E Stat. Nonlinear Soft Matter Phys., vol. 91, no. 4, p. 040101, 2015.

[12] B. Chen, L. Xing, B. Xu, H. Zhao, and J. C. Principe, “Insights into the robustness of minimum error entropy estimation,” IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 3, pp. 731–737, Mar. 2018.

[13] D. Erdogmus and J. C. Principe, “From linear adaptive filtering to nonlinear information processing—the design and analysis of information processing systems,” IEEE Signal Process. Mag., vol. 23, no. 6, pp. 14–33, Nov. 2006.

[14] J. C. Principe, Information Theoretic Learning: Renyi’s Entropy and Kernel Perspectives. Berlin, Germany: Springer-Verlag, 2010.

[15] B. Chen, Y. Zhu, J. Hu, and J. C. Principe, System Parameter Identification: Information Criteria and Algorithms. Oxford, U.K.: Newnes, 2013.

[16] D. Erdogmus and J. C. Principe, “Convergence properties and data efficiency of the minimum error entropy criterion in adaline training,” IEEE Trans. Signal Process., vol. 51, no. 7, pp. 1966–1978, Jul. 2003.

[17] B. Chen, L. Xing, N. Zheng, and J. C. Principe, “Quantized minimum error criterion,” 2017, arXiv:1710.04089.

[18] W. Hesse, E. Müller, M. Arnold, and B. Schack, “The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies,” J. Neurosci. Methods, vol. 124, no. 1, pp. 27–44, 2003.

[19] S. L. Bressler and A. K. Seth, “Wiener–Granger causality: A well established methodology,” Neuroimage, vol. 58, no. 2, pp. 323–329, 2011.

[20] A. Renyi, “On measures of information and entropy,” Maximum-Entropy Bayesian Methods Sci. Eng., vol. 1, no. 2, pp. 547–561, 1961.

[21] B. W. Silverman, Density Estimation for Statistics and Data Analysis. London, England: Routledge, 2018.

[22] R. Leeb, C. Brunner, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, “Bci competition 2008–graz data set b,” Graz University Technol., Austria, 2008.

[23] D. Prichard and J. Theiler, “Generating surrogate data for time series with several simultaneously measured variables,” Phys. Rev. Lett., vol. 73, no. 7, p. 951, 1994.

[24] F. Luca, P. Alberto, and N. Giandomenico, “Testing frequency-domain causality in multivariate time series,” IEEE Trans. Biomed. Eng., vol. 57, no. 8, pp. 1897–1906, Aug. 2010.

[25] H. Chen, Q. Yang, W. Liao, Q. Gong, and S. Shen, “Evaluation of the effective connectivity of supplementary motor areas during motor imagery using Granger causality mapping,” Neuroimage, vol. 47, no. 4, pp. 1844–1853, 2009.

[26] Q. Gao, X. Duan, and H. Chen, “Evaluation of effective connectivity of motor areas during motor imagery and execution using conditional Granger causality,” Neuroimage, vol. 54, no. 2, pp. 1280–1288, 2011.

[27] S. Hu, H. Wang, J. Zhang, W. Kong, and Y. Cao, “Causality from c2 to c3/c4 or between c3 and c4 revealed by Granger causality and new causality during motor imagery,” in Proc. Int. Joint Conf. Neural Netw., 2014, pp. 3178–3185.

[28] S. Hu, H. Wang, J. Zhang, W. Kong, Y. Cao, and R. Koizma, “Comparison analysis: Granger causality and new causality and their applications to motor imagery,” IEEE Trans. Neural Netw. Learn. Syst., vol. 27, no. 7, pp. 1429–1444, Jul. 2016.