Matrix Formulas That Is Useful In Least Square Analysis

(You could always swap indexes properly to get similar results)

1 Difference of Quadratic Forms - The Regression Improvement by Introducing a Variate

For invertible $(n_1+n_2+n_3)\times(n_1+n_2+n_3)$ matrix R (no need symmetric), define Q

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

For any $\begin{bmatrix} \vec{r}^{(x|x)} & \vec{r}^{(x|y)} & \vec{r}^{(x|z)} \end{bmatrix}$ $(n_4\times(n_1+n_2+n_3)$ matrix),
define quadratic forms $\mathrm{SS}_{R[x,z]}$ and $\mathrm{SS}_{R[x,y,z]}$, solutions $\begin{bmatrix} \vec{a}^{(11)} & \vec{a}^{(12)} & \vec{a}^{(13)} \end{bmatrix}$ and $\begin{bmatrix} \vec{a}_T^{(11)} & \vec{a}_T^{(12)} & \vec{a}_T^{(13)} \end{bmatrix}$ as (replace transpose $\cdot^T$ to conjugate transpose $\cdot^H$ for matrix with complex entries)

$$\mathrm{SS}_{R[x,z]}=\begin{bmatrix} \vec{r}^{(x|x)} & \vec{r}^{(x|z)} \end{bmatrix}\begin{bmatrix} R^{(xx)} & R^{(xz)} \\ R^{(zx)} & R^{(zz)} \end{bmatrix}^{-1}\begin{bmatrix} \vec{r}^{(x|x)} & \vec{r}^{(x|z)} \end{bmatrix}^T,$$

$$\mathrm{SS}_{R[x,y,z]}=\begin{bmatrix} \vec{r}^{(x|x)} & \vec{r}^{(x|y)} & \vec{r}^{(x|z)} \end{bmatrix}\begin{bmatrix} R^{(xx)} & R^{(xy)} & R^{(xz)} \\ R^{(yx)} & R^{(yy)} & R^{(yz)} \\ R^{(zx)} & R^{(zy)} & R^{(zz)} \end{bmatrix}^{-1}\begin{bmatrix} \vec{r}^{(x|x)} & \vec{r}^{(x|y)} & \vec{r}^{(x|z)} \end{bmatrix}^T,$$

$$\begin{bmatrix} \vec{a}^{(11)} & \vec{a}^{(12)} & \vec{a}^{(13)} \end{bmatrix}\begin{bmatrix} R^{(xx)} & R^{(xy)} & R^{(xz)} \\ R^{(yx)} & R^{(yy)} & R^{(yz)} \\ R^{(zx)} & R^{(zy)} & R^{(zz)} \end{bmatrix}=\begin{bmatrix} \vec{r}^{(x|x)} & \vec{r}^{(x|y)} & \vec{r}^{(x|z)} \end{bmatrix},$$

$$\begin{bmatrix} \vec{a}_T^{(11)} & \vec{a}_T^{(12)} & \vec{a}_T^{(13)} \end{bmatrix}\begin{bmatrix} R^{(xx)} & R^{(xy)} & R^{(xz)} \\ R^{(yx)} & R^{(yy)} & R^{(yz)} \\ R^{(zx)} & R^{(zy)} & R^{(zz)} \end{bmatrix}^T=\begin{bmatrix} \vec{r}^{(x|x)} & \vec{r}^{(x|y)} & \vec{r}^{(x|z)} \end{bmatrix}.$$

Then the difference of the quadratic forms is
(Assume the inverse matries in the expressions are all valid. Pseudo-inverse might also work?)

$$\mathrm{SS}_{R[x,y,z]}-\mathrm{SS}_{R[x,z]}=\vec{a}^{(12)}\left(Q^{(yy)}\right)^{-1}\left(\vec{a}_T^{(12)}\right)^T. \tag{2}$$

The $\left(Q^{(yy)}\right)^{-1}$ might be expressed in R blocks as
(Swap index x and z will give another expression)

$$\begin{aligned}\left(Q^{(yy)}\right)^{-1} &= R^{(yy)}-R^{(yx)}\left(R^{(xx)}\right)^{-1}R^{(xy)} \\ &-\left(R^{(yz)}-R^{(yx)}\left(R^{(xx)}\right)^{-1}R^{(xz)}\right)\left(R^{(zz)}-R^{(zx)}\left(R^{(xx)}\right)^{-1}R^{(xz)}\right)^{-1}\left(R^{(zy)}-R^{(zx)}\left(R^{(xx)}\right)^{-1}R^{(xy)}\right)\end{aligned} \tag{3}$$

2 Difference of Solutions - The Correction to Coefficients When Introducing a Variate

If we solve

$$\begin{bmatrix} \vec{b}^{(11)} & \vec{b}^{(12)} \end{bmatrix} \begin{bmatrix} R^{(xx)} & R^{(xy)} \\ R^{(yx)} & R^{(yy)} \end{bmatrix} = \begin{bmatrix} \vec{r}^{(x|x)} & \vec{r}^{(x|y)} \end{bmatrix}$$

Then there is relation
(There are 6 of these relations)

$$b^{(12)} = a^{(12)} - a^{(13)} \left(Q^{(zz)}\right)^{-1} Q^{(zy)} \tag{4}$$

$$a^{(12)} \left(I - \left(Q^{(yy)}\right)^{-1} Q^{(yz)} \left(Q^{(zz)}\right)^{-1} Q^{(zy)}\right) = b^{(12)} + b^{(13)}_{[xz]} \left(Q^{(zz)}\right)^{-1} Q^{(zy)} \tag{5}$$

Define P

$$P = \begin{bmatrix} P^{(xx)} & P^{(xy)} \\ P^{(yx)} & P^{(yy)} \end{bmatrix} \triangleq \begin{bmatrix} R^{(xx)} & R^{(xy)} \\ R^{(yx)} & R^{(yy)} \end{bmatrix}^{-1}.$$

There are relations between P, Q and R

$$P = \begin{bmatrix} Q^{(xx)} & Q^{(xy)} \\ Q^{(yx)} & Q^{(yy)} \end{bmatrix} - \begin{bmatrix} Q^{(xz)} \\ Q^{(yz)} \end{bmatrix} \left(Q^{(zz)}\right)^{-1} \begin{bmatrix} Q^{(zx)} & Q^{(zy)} \end{bmatrix} \tag{6}$$

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

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

$$I - \left(Q^{(yy)}\right)^{-1} Q^{(yz)} \left(Q^{(zz)}\right)^{-1} Q^{(zy)} = \left(Q^{(yy)}\right)^{-1} P^{(yy)}$$

Sometimes $\vec{b}^{(11)}$ is also wanted

$$\vec{b}^{(11)} = \vec{a}^{(11)} - a^{(13)} \left(Q^{(zz)}\right)^{-1} Q^{(zx)}$$

In matrix form

$$\begin{bmatrix} \vec{b}^{(11)} & \vec{b}^{(12)} \\ \vec{b}^{(21)} & \vec{b}^{(22)} \end{bmatrix} = \begin{bmatrix} \vec{a}^{(11)} & \vec{a}^{(12)} \\ \vec{a}^{(21)} & \vec{a}^{(22)} \end{bmatrix} - \begin{bmatrix} a^{(13)} \\ a^{(23)} \end{bmatrix} \left(Q^{(zz)}\right)^{-1} \begin{bmatrix} Q^{(zx)} & Q^{(zy)} \end{bmatrix} \tag{7}$$

$$\begin{bmatrix} \vec{b}^{(11)} & \vec{b}^{(12)} \\ \vec{b}^{(21)} & \vec{b}^{(22)} \end{bmatrix} = \begin{bmatrix} \vec{a}^{(11)} & \vec{a}^{(12)} \\ \vec{a}^{(21)} & \vec{a}^{(22)} \end{bmatrix} + \begin{bmatrix} a^{(13)} \\ a^{(23)} \end{bmatrix} \begin{bmatrix} R^{(zx)} & R^{(zy)} \end{bmatrix} \begin{bmatrix} R^{(xx)} & R^{(xy)} \\ R^{(yx)} & R^{(yy)} \end{bmatrix}^{-1}$$

$$\begin{bmatrix} a^{(13)} \\ a^{(23)} \end{bmatrix} \begin{bmatrix} R^{(zx)} & R^{(zy)} \end{bmatrix} = \begin{bmatrix} \vec{r}^{(x|x)} & \vec{r}^{(x|y)} \\ \vec{r}^{(y|x)} & \vec{r}^{(y|y)} \end{bmatrix} - \begin{bmatrix} \vec{a}^{(11)} & \vec{a}^{(12)} \\ \vec{a}^{(21)} & \vec{a}^{(22)} \end{bmatrix} \begin{bmatrix} R^{(xx)} & R^{(xy)} \\ R^{(yx)} & R^{(yy)} \end{bmatrix}$$

$$\begin{bmatrix} a^{(13)} \\ a^{(23)} \end{bmatrix} \begin{bmatrix} R^{(zx)} & R^{(zy)} \end{bmatrix} \begin{bmatrix} R^{(xx)} & R^{(xy)} \\ R^{(yx)} & R^{(yy)} \end{bmatrix}^{-1} = \begin{bmatrix} \vec{r}^{(x|x)} & \vec{r}^{(x|y)} \\ \vec{r}^{(y|x)} & \vec{r}^{(y|y)} \end{bmatrix} \begin{bmatrix} R^{(xx)} & R^{(xy)} \\ R^{(yx)} & R^{(yy)} \end{bmatrix}^{-1} - \begin{bmatrix} \vec{a}^{(11)} & \vec{a}^{(12)} \\ \vec{a}^{(21)} & \vec{a}^{(22)} \end{bmatrix}$$

3 Relation to Linear Regression and Granger Causality For Time Series

Assume $\mathrm{E}(x_t) = \mathrm{E}(y_t) = \mathrm{E}(y_t) = O$, as usual, let's do a 3-var AR of order m:

$$\begin{cases} x_t = \sum_{j=1}^{m} a_j^{(11)} x_{t-j} + \sum_{j=1}^{m} a_j^{(12)} y_{t-j} + \sum_{j=1}^{m} a_j^{(13)} z_{t-j} + \epsilon_t^{(1)} \quad \text{(a)} \\ y_t = \sum_{j=1}^{m} a_j^{(21)} x_{t-j} + \sum_{j=1}^{m} a_j^{(22)} y_{t-j} + \sum_{j=1}^{m} a_j^{(23)} z_{t-j} + \epsilon_t^{(2)} \quad \text{(b)} \\ z_t = \sum_{j=1}^{m} a_j^{(31)} x_{t-j} + \sum_{j=1}^{m} a_j^{(32)} y_{t-j} + \sum_{j=1}^{m} a_j^{(33)} z_{t-j} + \epsilon_t^{(3)} \quad \text{(c)} \end{cases} \tag{8}$$

The least square solution for $\vec{a}^{(11)} \quad \vec{a}^{(12)} \quad \vec{a}^{(13)}$ in min var $\epsilon_t^{(1)}$ $(\mathrm{E}(\epsilon_t^{(u)} x_{t-k}^{(v)T}) = 0,\ \forall k = 1 \dots m)$ satisfies

$$\begin{bmatrix} \vec{a}^{(11)} & \vec{a}^{(12)} & \vec{a}^{(13)} \end{bmatrix} \begin{bmatrix} R^{(xx)} & R^{(xy)} & R^{(xz)} \\ R^{(yx)} & R^{(yy)} & R^{(yz)} \\ R^{(zx)} & R^{(zy)} & R^{(zz)} \end{bmatrix} = \begin{bmatrix} \vec{r}^{(x|x)} & \vec{r}^{(x|y)} & \vec{r}^{(x|z)} \end{bmatrix}, \tag{9}$$

where (here $\mathbb{E}$ can be expectation operator or $\frac{1}{L}\sum_{t=1}^{L}$)

$$\vec{a}^{(uv)} = \left(a_j^{(uv)}\right)_{j=1\dots m}^{\text{row}}, \ (u, v \in \{1, 2, 3\}),$$

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

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

$(a_j)_{j=1\dots m}^{\text{row}}$ means tiling a_j in a row: $(a_j)_{j=1\dots m}^{\text{row}} = \begin{bmatrix} a_1 & \cdots & a_m \end{bmatrix}$, and $(a_{jk})_{j,k=1\dots m}$ is a matrix which its j row k column is a_{jk}.

The residual variance of Eq. (8) can be obtained by

$$\mathrm{var}(x_t) = \vec{a}^{(11)} \vec{r}^{(x|x)T} + \vec{a}^{(12)} \vec{r}^{(x|y)T} + \vec{a}^{(13)} \vec{r}^{(x|z)T} + \mathrm{var}(\epsilon_t^{(1)})$$

or

$$\mathrm{var}(\epsilon_t^{(1)}) = \mathrm{var}(x_t) - \begin{bmatrix} \vec{r}^{(x|x)} & \vec{r}^{(x|y)} & \vec{r}^{(x|z)} \end{bmatrix} \begin{bmatrix} R^{(xx)} & R^{(xy)} & R^{(xz)} \\ R^{(yx)} & R^{(yy)} & R^{(yz)} \\ R^{(zx)} & R^{(zy)} & R^{(zz)} \end{bmatrix}^{-1} \begin{bmatrix} \vec{r}^{(x|x)} & \vec{r}^{(x|y)} & \vec{r}^{(x|z)} \end{bmatrix}^T. \tag{10}$$

Similarly for least square min var $\epsilon_t^{(1|3)}$ in (regression of x without y)

$$x_t = \sum_{j=1}^{m} c_j^{(11)} x_{t-j} + \sum_{j=1}^{m} c_j^{(13)} z_{t-j} + \epsilon_t^{(1|3)}, \tag{11}$$

that is

$$\mathrm{var}(\epsilon_t^{(1|3)}) = \mathrm{var}(x_t) - \begin{bmatrix} \vec{r}^{(x|x)} & \vec{r}^{(x|z)} \end{bmatrix} \begin{bmatrix} R^{(xx)} & R^{(xz)} \\ R^{(zx)} & R^{(zz)} \end{bmatrix}^{-1} \begin{bmatrix} \vec{r}^{(x|x)} & \vec{r}^{(x|z)} \end{bmatrix}^T. \tag{12}$$

Conditional Granger causality $F_{y \to x|z}$ is defined as

$$F_{y \to x|z} = \ln \frac{\mathrm{var}(\epsilon_t^{(1|3)})}{\mathrm{var}(\epsilon_t^{(1)})} = \ln \left(1 + \frac{\mathrm{var}(\epsilon_t^{(1|3)}) - \mathrm{var}(\epsilon_t^{(1)})}{\mathrm{var}(\epsilon_t^{(1)})} \right). \tag{13}$$

With help of Eq.(2), we get

$$F_{y \to x|z} = \ln \left(1 + \frac{\vec{a}^{(12)} \left(Q^{(yy)}\right)^{-1} \left(\vec{a}^{(12)}\right)^T}{\mathrm{var}(\epsilon_t^{(1)})} \right) \tag{14}$$

3.1 Conditional GC and pairwise GC

Due to expression Eq.(14), the pairwise GC can be expressed as

$$F_{y\to x|z} = \ln\left(1 + \frac{\vec{b}^{(12)}\left(P^{(yy)}\right)^{-1}\left(\vec{b}^{(12)}\right)^T}{\mathrm{var}(\epsilon_t^{(1|2)})}\right) \tag{15}$$

where (recall Eq.(4) (2) and (6))

$$b^{(12)} = a^{(12)} - a^{(13)}\left(Q^{(zz)}\right)^{-1}Q^{(zy)},$$

$$\mathrm{var}(\epsilon_t^{(1|2)}) = \mathrm{var}(\epsilon_t^{(1)}) + \vec{a}^{(13)}\left(Q^{(zz)}\right)^{-1}\left(\vec{a}^{(13)}\right)^T,$$

$$P^{(yy)} = Q^{(yy)} - Q^{(yz)}\left(Q^{(zz)}\right)^{-1}Q^{(zy)}.$$

The related regression is

$$\begin{cases} x_t = \sum_{j=1}^{m} b_j^{(11)} x_{t-j} + \sum_{j=1}^{m} b_j^{(12)} y_{t-j} + \epsilon_t^{(1|2)} & \text{(a)} \\ y_t = \sum_{j=1}^{m} b_j^{(21)} x_{t-j} + \sum_{j=1}^{m} b_j^{(22)} y_{t-j} + \epsilon_t^{(2|1)} & \text{(b)} \end{cases} \tag{16}$$

From Eq.(15), one can know that the pairwise GC can be expressed by joint regression coefficients and residuals, and Q matrix.

Later we will know, the Q matrix can be approximated by joint regression coefficients and residuals easily and accurately.

3.2 Application: Get Regression that is Used in Frequency Domain Conditional GC

The extra regression is

$$\begin{cases} x_t = \sum_{j=1}^{m} c_j^{(11)} x_{t-j} + \sum_{j=1}^{m} c_j^{(13)} z_{t-j} + \epsilon_t^{(1|3)} & \text{(a)} \\ z_t = \sum_{j=1}^{m} c_j^{(31)} x_{t-j} + \sum_{j=1}^{m} c_j^{(33)} z_{t-j} + \epsilon_t^{(3|1)} & \text{(b)} \end{cases} \tag{17}$$

Similarly to Eq. (7), we have

$$\begin{bmatrix} \vec{c}^{(11)} & \vec{c}^{(13)} \\ \vec{c}^{(31)} & \vec{c}^{(33)} \end{bmatrix} = \begin{bmatrix} \vec{a}^{(11)} & \vec{a}^{(13)} \\ \vec{a}^{(31)} & \vec{a}^{(33)} \end{bmatrix} - \begin{bmatrix} a^{(12)} \\ a^{(32)} \end{bmatrix}\left(Q^{(yy)}\right)^{-1}\begin{bmatrix} Q^{(yx)} & Q^{(yz)} \end{bmatrix}$$

And with help of Eq. (2)

$$\mathrm{var}(\epsilon_t^{(1,3)}) - \mathrm{var}(\epsilon_t^{(1,3|2)}) = \begin{bmatrix} \vec{a}^{(12)} \\ \vec{a}^{(32)} \end{bmatrix}\left(Q^{(yy)}\right)^{-1}\begin{bmatrix} \vec{a}^{(12)} \\ \vec{a}^{(32)} \end{bmatrix}^T$$

(not verified).

The frequency domain conditional GC then use $\epsilon_t^{(1|3)}$ and $\epsilon_t^{(3|1)}$ for further processing..

3.3 Approximation of GC

Due to Taylor expansion, we have approximation

$$\ln(1+x) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1} x^k}{k} \approx x, \quad \text{for } x \ll 1. \tag{18}$$

From Eq. (13) get

$$F_{y\to x|z} \approx \frac{\vec{a}^{(12)} \left(Q^{(yy)}\right)^{-1} \left(\vec{a}^{(12)}\right)^T}{\mathrm{var}(\epsilon_t^{(1)})}. \tag{19}$$

Or, one might choose a different definition of "GC":

$$F_{y\to x|z}^{(\text{linear})} = \exp(F_{y\to x|z}) - 1 = \frac{\vec{a}^{(12)} \left(Q^{(yy)}\right)^{-1} \left(\vec{a}^{(12)}\right)^T}{\mathrm{var}(\epsilon_t^{(1)})} \tag{20}$$

For whitened signal and a weak GC, we might even get a reasonable approximation

$$F_{y\to x|z} \approx \frac{\mathrm{var}\,(\epsilon_t^{(2)})}{\mathrm{var}\,(\epsilon_t^{(1)})} \vec{a}^{(12)} \vec{a}^{(12)T} \tag{21}$$

3.4 Partial correlation coefficient (seems useless)

$$\begin{bmatrix} \vec{a}^{(11)} & \vec{a}^{(12)} & \vec{a}^{(13)} \end{bmatrix} \begin{bmatrix} R^{(xx)} & R^{(xy)} & O \\ R^{(yx)} & R^{(yy)} & O \\ O & O & R^{(zz)} \end{bmatrix} = \begin{bmatrix} \vec{r}^{(x|x)} & \vec{r}^{(x|y)} & 0 \end{bmatrix}$$

$$\left(Q^{(yy)}\right)^{-1} = R^{(yy)} - R^{(yx)} \left(R^{(xx)}\right)^{-1} R^{(xy)}$$

$$x_t' = x_t - \alpha Y$$

What if z is totally whitened? i.e.

$$R^{(zz)} = I$$

not verified

$$c^{(33)} = a^{(33)} - \begin{bmatrix} a^{(31)} & a^{(32)} \end{bmatrix} \begin{bmatrix} Q^{(xx)} & Q^{(xy)} \\ Q^{(yx)} & Q^{(yy)} \end{bmatrix}^{-1} \begin{bmatrix} Q^{(xz)} \\ Q^{(yz)} \end{bmatrix}$$

4 Spectral Approximation

The $Q^{(uv)}$ and $R^{(uv)}$ can be approximated by circulant matrix. Hence, their related operations can be approxiamted by operations of spectrum.

4.1 Definitions

The Fourier transform used here is (Discrete-time Fourier transform)

$$\mathscr{F}_{\text{DTFT}}\left[\{x_j\}_{j\in\mathbb{Z}}\right] = \sum_{j=-\infty}^{\infty} x_j e^{-\mathrm{i}jw}, \tag{22}$$

$$\mathscr{F}_{\text{DTFT}}^{-1}\left[f(w)\right] = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(w) e^{\mathrm{i}jw} \mathrm{d}w. \tag{23}$$

Note in Eq. (22) and (23) both x_j and $f(w)$ can be matrix. And the convolution theorem still works.

The frequency domain quantities we will use include

$$\hat{a} = \mathscr{F}_{DTFT}\left[\{a_j\}_{j=0..m}\right], \quad \hat{b} = \mathscr{F}_{DTFT}\left[\{b_j\}_{j=0..m}\right], \quad \hat{c} = \mathscr{F}_{DTFT}\left[\{c_j\}_{j=0..m}\right],$$

$$\hat{Q} = \hat{a}^H \Sigma^{-1} \hat{a}$$

where $\Sigma = \operatorname{var}\left(\begin{bmatrix} \epsilon_t^{(1)T} & \epsilon_t^{(2)T} & \epsilon_t^{(3)T} \end{bmatrix}^T\right)$ is the residual variance.

We use $\hat{Q}_{uv}$ to denote the u-th row v-th column block of $\hat{Q}$, and similarly $-\hat{a}^{(uv)}$ to $\hat{a}$, $-a_j^{(uv)}$ to a_j, (and also b, c, see Eq. (16) and (17)) (also note the negative sign). Also note the time lag for coefficients are starting from 0, $a_0^{(uv)} = \delta_{uv} I$.

The block parts of $\hat{Q}$, explicitly write down is:

$$\hat{Q}_{uv} = \sum_{j,k=1}^{3} \left(\hat{a}^{(ju)}\right)^H \left(\Sigma^{-1}\right)^{(jk)} \hat{a}^{(kv)}$$

where $\left(\Sigma^{-1}\right)^{(jk)}$ means the j-th row k-th column block of Σ^{-1}.

Sometimes we need the operator $[\cdot]_+$ to extract the positive time part (work in frequency domain):

$$[f(w)]_+ = \mathscr{F}_{\text{DTFT}}\left[u(t)\mathscr{F}_{\text{DTFT}}^{-1}\left[f(w), t\right](t), w\right]$$

$$u(t) = \begin{cases} 1 & t > 0 \\ 0 & t \le 0 \end{cases}.$$

4.2 The Approximations

The conditions of these approximations to work good will be discussed later. Roughly speaking, that condition is weak conditional GC.

The $x - y$ regression coefficient formula Eq. (4) can be approximated by

$$\hat{b}^{(12)} \approx \hat{a}^{(12)} - \left[\hat{a}^{(13)} \hat{Q}_{zz}^{-1} \hat{Q}_{zy}\right]_+. \tag{24}$$

GC is

$$F_{y\to x|z} \approx F_{y\to x|z}^{(\text{linear})} \approx \frac{1}{2\pi} \int_{[-\pi,\pi]} \frac{\hat{a}^{(12)} \hat{Q}_{yy}^{-1} \left(\hat{a}^{(12)}\right)^H}{\operatorname{var}(\epsilon_t^{(1)})} \mathrm{d}w. \tag{25}$$

See Eq.(20) for definition of $F_{y\to x|z}^{(\text{linear})}$.

Actually the frequency domain GC can also be approximated

$$f_{y\to x|z} \approx f_{y\to x|z}^{(\text{linear})} \approx f_{y\to x|z}^{(\text{app freq})} = \frac{\hat{a}^{(12)} \hat{Q}_{yy}^{-1} \left(\hat{a}^{(12)}\right)^H}{\operatorname{var}(\epsilon_t^{(1)})}. \tag{26}$$

where $f_{y\to x|z}^{(\text{linear})} = \exp(f_{y\to x|z}) - 1$.

One beauty thing is that the frequency domain GC approximation Eq. (26) does not involve complicated "renormalization" operations discussed in Geweke [2] when the residual variance Σ is non-diagonal.

See sec 4.3.2 for numerical example.

4.3 Approximations Under Certain (Restricted) Conditions

4.3.1 Assumptions and approximation to GC

(Assumption #1) Under the assumption that the variance of residual is blocked diagonal matrix

$$\Sigma = \operatorname{diag}\left(\begin{bmatrix} \Sigma_{xx} & \Sigma_{yy} & \Sigma_{zz} \end{bmatrix}\right)$$

We have

$$\hat{Q}_{yy} = \left(\hat{a}^{(12)}\right)^H \left(\Sigma^{-1}\right)^{(11)} \hat{a}^{(12)} + \left(\hat{a}^{(22)}\right)^H \left(\Sigma^{-1}\right)^{(22)} \hat{a}^{(22)} + \left(\hat{a}^{(32)}\right)^H \left(\Sigma^{-1}\right)^{(33)} \hat{a}^{(32)}$$

(Assumption #2) To further simplify, the next assumption is: Σ is diagonal matrix (i.e. each block is also diagonal). With this assumption, and the invariant property of GC under scalar transform to variates, we can assume $\Sigma = I$. This leads to

$$\hat{Q}_{yy} = \left(\hat{a}^{(12)}\right)^H \hat{a}^{(12)} + \left(\hat{a}^{(22)}\right)^H \hat{a}^{(22)} + \left(\hat{a}^{(32)}\right)^H \hat{a}^{(32)}, \tag{27}$$

Note the coefficients here (Eq. (27), and later this sub-section) are properly scaled to reflect the normalization of $\Sigma = I$:

$$a^{(uv)(\text{normalized})} = \frac{\Sigma^{(vv)}}{\Sigma^{(uu)}} a^{(uv)(\text{original})}. \tag{28}$$

Under these assumptions, the expressions Eq. (26) now become more clean (When we calculate GC, x and y are scalars, so the matrix inversion become division)

$$f_{y\to x|z}^{(\text{app freq})} = \frac{\hat{a}^{(12)}\left(\hat{a}^{(12)}\right)^H}{\left(\hat{a}^{(12)}\right)^H \hat{a}^{(12)} + \left(\hat{a}^{(22)}\right)^H \hat{a}^{(22)} + \left(\hat{a}^{(32)}\right)^H \hat{a}^{(32)}}. \tag{29}$$

(Assumption #3) After that, an aggressive approximation would be to drop high order terms in Eq. (29), usually (for small system and weak GC) that is $\left(\hat{a}^{(12)}\right)^H \hat{a}^{(12)}$ and $\left(\hat{a}^{(32)}\right)^H \hat{a}^{(32)}$:

$$f_{y\to x|z}^{(\text{app freq})} \approx \frac{\hat{a}^{(12)}\left(\hat{a}^{(12)}\right)^H}{\left(\hat{a}^{(22)}\right)^H \hat{a}^{(22)}}. \tag{30}$$

Note, for system involves hundreds of neurons, $\left(\hat{a}^{(32)}\right)^H \hat{a}^{(32)}$ might not be drop.

4.3.2 Comparison to exact frequency domain GC

The exact bivariate frequency domain GC is

$$\exp(f_{y\to x}) - 1 = \frac{1-\Upsilon^2}{\left|\frac{1}{r} - \Upsilon\right|^2} = \frac{1-\Upsilon^2}{\frac{1}{rr^H} - \frac{\Upsilon}{r^H} - \frac{\Upsilon}{r} + \Upsilon^2} \tag{31}$$

where $\Upsilon = \Sigma^{(12)}$, $r = \hat{a}^{(12)}/\hat{a}^{(22)}$. In Eq. (31), normalization to residual variance Eq. (28) is applied, but none of the assumptions #1, #2 and #3 is used.

Case 1: $\Upsilon = 0$

In this case,

$$\exp(f_{y\to x}) - 1 = rr^H$$

which is exactly the Eq.(30). One might also start from Eq.(29), and apply the function

$$g(f) = \frac{1}{1/f - 1}$$

to get a exact "approximation" $\exp(f_{y\to x}) - 1 = g(f_{y\to x|z}^{(\text{app freq})})$.

Case 2: $\Upsilon \neq 0$ (no instantaneous correlation)

Still, the $f_{y\to x|z}^{(\text{app freq})}$ without assumptions (Eq. (26)) (but with the residual variance normalization) is a good approximation

$$\begin{aligned} f_{y\to x|z}^{(\text{app freq})} &= \hat{a}^{(12)}\left(\hat{a}^{(12)H}\hat{a}^{(12)} + \hat{a}^{(22)H}\hat{a}^{(22)} - (\hat{a}^{(22)H}\hat{a}^{(12)} + \hat{a}^{(12)H}\hat{a}^{(22)})\frac{\Upsilon}{1-\Upsilon^2}\right)^{-1}\left(\hat{a}^{(12)}\right)^H \\ &= \frac{1}{\frac{1}{rr^H} - \frac{\Upsilon/(1-\Upsilon^2)}{r^H} - \frac{\Upsilon/(1-\Upsilon^2)}{r} + 1} = \frac{1}{\left|\frac{1}{r} - \frac{\Upsilon}{1-\Upsilon^2}\right|^2 + 1 - \left(\frac{\Upsilon}{1-\Upsilon^2}\right)^2} \end{aligned}$$

If apply the function

$$g_\Upsilon(f) = \frac{1-\Upsilon^2}{1/f - \frac{1}{1-\Upsilon^2}} \tag{32}$$

The approximation can be more accurate: $\exp(f_{y\to x}) - 1 \approx g_\Upsilon(f^{(\text{app freq})}_{y\to x|z})$. See Fig. 1, for an example model

$$a = \begin{bmatrix} -0.8 & -0.16k_a & 0.5 & 0.2k_a \\ 0 & -0.9 & 0 & 0.5 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0.2k_a \\ 0.2k_a & 1 \end{bmatrix}, \quad k_a = \sqrt{1/0.7} \tag{33}$$

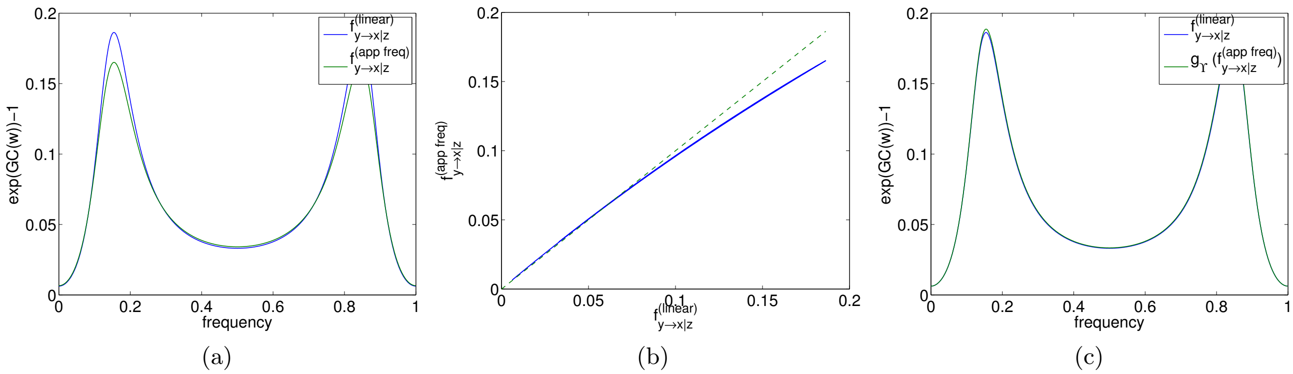

Figure 1: Two variable frequency domain GC example Eq.(33). Decrease Υ will lead to better approximation. (a) Comparison of exact result and approximation. (b) The relation (curve) between true result and approximation, w is the varying parameter. A thin curve here means that the difference can be compromised by a static transform. (c) The approximation after apply correction Eq.(32).

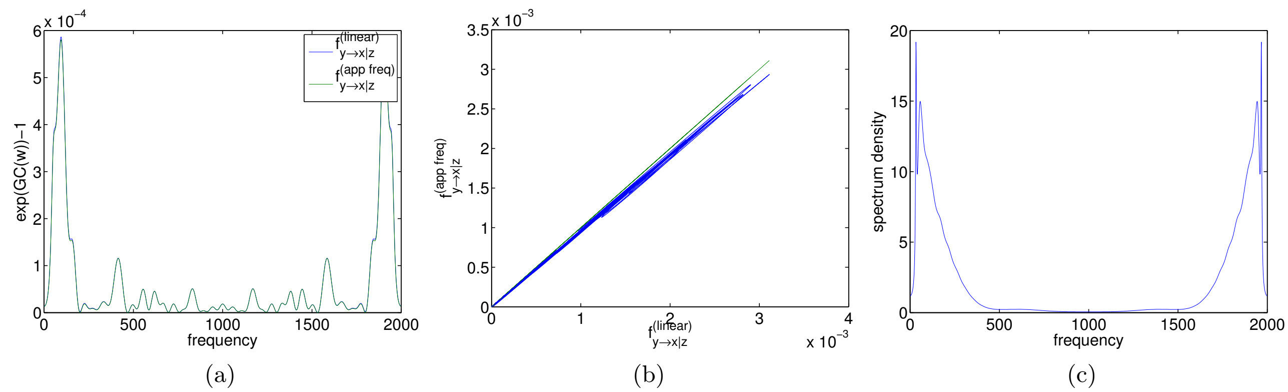

Figure 2: 100 neurons case. (a) the frequency domain GC. (b) the relation (curve) between true result and approximation. 20 pairs of frequency GC been ploted. (c) The spectral density of a neuron.

In Fig. 2, we see that under such double peak spectral density (see (c)), the approximation is still good.

4.3.3 Approximation to "auto"-regression coefficients

Recall Eq.(24)

$$\hat{b}^{(12)} \approx \hat{a}^{(12)} - \left[\hat{a}^{(13)}\hat{Q}_{zz}^{-1}\hat{Q}_{zy}\right]_+ .$$

Follow the same assumptions in Sec 4.3.1, we get:

- Apply assumptions #1 and #2:

$$\hat{Q}_{zz} = \left(\hat{a}^{(13)}\right)^H \hat{a}^{(13)} + \left(\hat{a}^{(23)}\right)^H \hat{a}^{(23)} + \left(\hat{a}^{(33)}\right)^H \hat{a}^{(33)}, \tag{34}$$

$$\hat{Q}_{zy} = \left(\hat{a}^{(13)}\right)^H \hat{a}^{(12)} + \left(\hat{a}^{(23)}\right)^H \hat{a}^{(22)} + \left(\hat{a}^{(33)}\right)^H \hat{a}^{(32)}. \tag{35}$$

- Apply also assumption #3:

$$\begin{aligned} \hat{b}^{(12)} &\approx \hat{a}^{(12)} - \left[\hat{a}^{(13)} \left(\left(\hat{a}^{(33)}\right)^H \hat{a}^{(33)}\right)^{-1} \left(\left(\hat{a}^{(23)}\right)^H \hat{a}^{(22)} + \left(\hat{a}^{(33)}\right)^H \hat{a}^{(32)}\right)\right]_+ . \\ &= \hat{a}^{(12)} - \left[\hat{a}^{(13)} \left(\hat{a}^{(33)}\right)^{-1} \left(\left(\hat{a}^{(33)H}\right)^{-1} \left(\hat{a}^{(23)}\right)^H \hat{a}^{(22)} + \hat{a}^{(32)}\right)\right]_+ . \end{aligned}$$

(assumption #4) If further, y and z been jointly whiten before, i.e. $\hat{a}^{(33)} \approx I$ and $\hat{a}^{(22)} \approx I$. then

$$\hat{b}^{(12)} \approx \hat{a}^{(12)} - \left[\hat{a}^{(13)} \left(\hat{a}^{(23)H} + \hat{a}^{(32)}\right)\right]_+ .$$

Full matrix form

$$\begin{bmatrix} O & A_{xy} & A_{xz} \\ & & \end{bmatrix} \left(\begin{bmatrix} A_{xy} \\ O \\ A_{zy} \end{bmatrix} + \begin{bmatrix} A_{yz} & O & A_{yz} \end{bmatrix}^{*T} \right) = A(A + A^*)$$

5 Approximation of Toeplitz Matrix Operations (not useful)

Mul of band toeplitz matrix

$$\begin{bmatrix} a_0 & a_1 & \\ a_{-1} & a_0 & a_1 \\ & a_{-1} & a_0 \end{bmatrix} \begin{bmatrix} b_0 & b_1 & \\ b_{-1} & b_0 & b_1 \\ & b_{-1} & b_0 \end{bmatrix} = \begin{bmatrix} c_0 - a_{-1}b_1 & c_1 & c_2 \\ c_{-1} & c_0 & c_1 \\ c_{-2} & c_{-1} & c_0 - a_1 b_{-1} \end{bmatrix} \tag{36}$$

where $c_0 = a_{-1}b_1 + a_0 b_0 + a_1 b_{-1}$, $c_{-1} = a_{-1}b_0 + a_0 b_{-1}$, $c_1 = a_0 b_1 + a_1 b_0$.

Inverse of band toeplitz ??????????

$$R(Q + G) = I$$

$$GR = I - QR$$

$$GRQ = Q - QRQ$$

since

$$RQ \approx I$$

so

$$G \approx Q - QRQ$$

$$R^{-1} \approx Q + Q - QRQ$$

In summary, that is

$$K = I - QR$$

$$R^{-1} = (I - K)^{-1} Q$$

Bad......since it possible that $\|K\| > 1$. Indeed, there is a 10 neuron case that $\|K\| = 8.34$ (for non-whitenned covz). But seems work for whitened signal and $p < 40$.

6 Appendix: Formulas Used to Derive Above Results

6.1 Binomial inverse theorem

$$(A+UBV)^{-1} = A^{-1} - A^{-1}UB\left(B+BVA^{-1}UB\right)^{-1}BVA^{-1}$$

Specially

$$\left(A-BD^{-1}C\right)^{-1} = A^{-1} + A^{-1}B\left(D-CA^{-1}B\right)^{-1}CA^{-1}$$

6.2 Inverse of $(n_1+n_2)\times(n_1+n_2)$ matrix

(Assume the inverse matries in the expressions are all valid)

$$\left(R^{(zy)} - R^{(zx)}\left(R^{(xx)}\right)^{-1}R^{(xy)}\right)\left(R^{(yy)} - R^{(yx)}\left(R^{(xx)}\right)^{-1}R^{(xy)}\right)^{-1}$$

Use row elimination, get

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \left(a_{11}-a_{12}a_{22}^{-1}a_{21}\right)^{-1} & -\left(a_{11}-a_{12}a_{22}^{-1}a_{21}\right)^{-1}a_{12}a_{22}^{-1} \\ -\left(a_{22}-a_{21}a_{11}^{-1}a_{12}\right)^{-1}a_{21}a_{11}^{-1} & \left(a_{22}-a_{21}a_{11}^{-1}a_{12}\right)^{-1} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}^{-1} + a_{11}^{-1}a_{12}\left(a_{22}-a_{21}a_{11}^{-1}a_{12}\right)^{-1}a_{21}a_{11}^{-1} & -a_{11}^{-1}a_{12}\left(a_{22}-a_{21}a_{11}^{-1}a_{12}\right)^{-1} \\ -a_{22}^{-1}a_{21}\left(a_{11}-a_{12}a_{22}^{-1}a_{21}\right)^{-1} & a_{22}^{-1} + a_{22}^{-1}a_{21}\left(a_{11}-a_{12}a_{22}^{-1}a_{21}\right)^{-1}a_{12}a_{22}^{-1} \end{bmatrix}$$

6.2.1 Known inverse matrix, get inverse of sub matrix

Known

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

Then a_{22}^{-1} can get from

$$a_{22}^{-1} = b_{22} - b_{21}\left(b_{11}\right)^{-1}b_{12}$$

6.3 Inverse of $(n_1+n_2+n_3)\times(n_1+n_2+n_3)$ matrix

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}^{-1} = UD^{-1}L$$

$$U = \begin{bmatrix} I & -a_{11}^{-1}a_{12} & a_{11}^{-1}a_{12}\left(a_{22}-a_{21}a_{11}^{-1}a_{12}\right)^{-1}\left(a_{23}-a_{21}a_{11}^{-1}a_{13}\right) - a_{11}^{-1}a_{13} \\ O & I & -\left(a_{22}-a_{21}a_{11}^{-1}a_{12}\right)^{-1}\left(a_{23}-a_{21}a_{11}^{-1}a_{13}\right) \\ O & O & I \end{bmatrix}$$

$$L = \begin{bmatrix} I & O & O \\ -a_{21}a_{11}^{-1} & I & O \\ \left(a_{32}-a_{31}a_{11}^{-1}a_{12}\right)\left(a_{22}-a_{21}a_{11}^{-1}a_{12}\right)^{-1}a_{21}a_{11}^{-1} - a_{31}a_{11}^{-1} & -\left(a_{32}-a_{31}a_{11}^{-1}a_{12}\right)\left(a_{22}-a_{21}a_{11}^{-1}a_{12}\right)^{-1} & I \end{bmatrix}$$

$$D = \begin{bmatrix} a_{11} & O & O \\ O & a_{22}-a_{21}a_{11}^{-1}a_{12} & O \\ O & O & a_{33}-a_{31}a_{11}^{-1}a_{13} - \left(a_{32}-a_{31}a_{11}^{-1}a_{12}\right)\left(a_{22}-a_{21}a_{11}^{-1}a_{12}\right)^{-1}\left(a_{23}-a_{21}a_{11}^{-1}a_{13}\right) \end{bmatrix}$$

Note:

$$\begin{bmatrix} 1 & a_{12} & a_{13} \\ 0 & 1 & a_{23} \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} b_1 & 0 & 0 \\ 0 & b_2 & 0 \\ 0 & 0 & b_3 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ c_{21} & 1 & 0 \\ c_{31} & c_{32} & 1 \end{bmatrix} = \begin{bmatrix} b_1 + a_{12}b_2c_{21} + a_{13}b_3c_{31} & a_{12}b_2 + a_{13}b_3c_{32} & a_{13}b_3 \\ b_2c_{21} + a_{23}b_3c_{31} & b_2 + a_{23}b_3c_{32} & a_{23}b_3 \\ b_3c_{31} & b_3c_{32} & b_3 \end{bmatrix}$$

7 Useful Theorems

7.1 Spectral Representation of Multivariate Stationary Time Series [1]

For $X(h)$ widesense stationary, for its covariance $R(h)$, there exist $F(w)$ such that

$$R(h) = \int_{-\pi}^{\pi} e^{ihw} \mathrm{d}F(w), \quad h = 0, \pm 1, \dots$$

$F(-\pi) = 0$, F is right-continuous, $F(\mu) - F(\lambda)$ non-negative define for all $\lambda < \mu$. And satisfies

- $E\left(Z(\lambda)Z^*(\lambda)\right)$ finite
- $E\left(Z(\lambda)\right) = 0$, for $-\pi < \lambda < \pi$
- $E\left(Z(\lambda_4) - Z(\lambda_3)\right)\left(Z(\lambda_2) - Z(\lambda_1)\right)^* = 0$, for $(\lambda_1, \lambda_2] \cap (\lambda_3, \lambda_4] = \phi$
- $E\left(Z(\lambda + \delta) - Z(\lambda)\right)\left(Z(\lambda + \delta) - Z(\lambda)\right)^* \to 0$, as $\delta \downarrow 0$.

unique matrix distribution F, $F(\mu) - F(\lambda) = E\left(\left(Z(\mu) - Z(\lambda)\right)\left(Z(\mu) - Z(\lambda)\right)^*\right)$, $\lambda \leq \mu$. o

$$E\left(\mathrm{d}Z(\lambda)\mathrm{d}Z^*(\lambda)\right) = \begin{cases} \mathrm{d}F(\lambda) & \mu = \lambda \\ 0 & \text{otherwise} \end{cases}$$

$$X_t = \int_{-\pi}^{\pi} e^{itv} \mathrm{d}Z(v), \quad t = 0, \pm 1, \dots \quad \text{with probability } 1$$

References

[1] Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods*. Springer Science & Business Media, November 2013.

[2] John F. Geweke. Measures of Conditional Linear Dependence and Feedback between Time Series. *Journal of the American Statistical Association*, 79(388):907–915, 1984.