Notes: dynamical differential covariance

Yusi Chen

October 8, 2021

1 Equivalence to Regularized AR(2) Process

Differential covariance method calculates the covariance between $\frac{dx(t)}{xt}$ and $x(t)$. More specifically $\frac{dx(t)}{dt} = \frac{x(t+dt)-x(t-dt)}{2dt}$ to mimic the maximum likelihood estimator (MLE) of a regularized second-order auto-regressive (AR(2)) model.
On one hand, the differential covariance matrix of d-dimensional time series with N time points could be calculated as Equation 1. $x_t \in \mathbb{R}^{(N-1)\times d}$, $x_{t+dt} \in \mathbb{R}^{(N-1)\times d}$ and $x_{t-dt} \in \mathbb{R}^{(N-1)\times d}$.

$$\begin{aligned} cov(\frac{x_{t+dt} - x_{t-dt}}{2dt}, x_t) &= \frac{1}{2dt}[cov(x_{t+dt}, x_t) - cov(x_{t-dt}, x_t)] \\ &= \frac{1}{2dt}\frac{1}{N}(x_{t+dt} - x_{t-dt})x_t^T \end{aligned} \tag{1}$$

On the other hand, let me first illustrate why differential covariance would look like an AR model. For a neural network with connection pattern as G and Gaussian noise $\mathcal{N}$, neuron states $\mathbf{x(t)}$ (e.x. firing rate) could be described as Equation 2, which is an AR(2) model with constrain that $a_{t-1} = 1$.

$$\begin{aligned} \frac{d\mathbf{x(t)}}{dt} &= G\mathbf{x(t)} + \mathcal{N} \\ \frac{x(t+1) - x(t-1)}{2dt} &= Gx(t) + \mathcal{N} \\ x(t+1) &= (2dtG)x(t) + x(t-1) + dt\mathcal{N} \end{aligned} \tag{2}$$

The G could be estimated using maximum likelihood. More spefically, for a linear regression model: $Y = \beta X + \mathcal{N}$. The MLE estimator of β is $\hat{\beta} = (X^T X)^{-1} X^T y$. Therefore G could be estimated from $x(t+1) - x(t-1) = 2dtGx(t) + dt\mathcal{N}$and $\hat{G} = (x_t^T x_t)^{-1} x_t^T (x_{t+1} - x_{t-1})$. This estimator is similar to the calculation of differential covariance equation 1.

2 Differential Precision Matrix

To control for the effect of a latent variable z on the correlation of x and y, partial covariance method is used to regress out the common factors. In the scenario of covariance-based

method, partial covariance matrix (i.e. precision matrix) is the inverse of a covariance matrix.
For the differential covariance method, the partial differential covariance Δ_p is no longer the inverse of the differential covariance matrix. Here I derived the differential precision matrix using linear regression.
First, we donote σ_{ij} as the covariance between i and j, $\Delta_{c,ij}$ as the differential covariance between i and j. To tease out the effect of a common factor z, we could first solve the coefficients of a linear regression problem.

$$\begin{aligned} w_x &= \arg\min_w E(dx - wz)^2 \\ w_y &= \arg\min_w E(y - wz)^2 \end{aligned} \tag{3}$$

By finding the minimal of a quadratic form of w , we have $w_x = \sigma_{zz}^{-1}\Delta_{c,xz}$ and $w_y = \sigma_{yz}\sigma_{zz}^{-1}$. Then we could find the residuals of dx and y as:

$$\begin{aligned} dr_x &= dx - w_x z \\ r_y &= y - w_y z \end{aligned} \tag{4}$$

Then the covariance between dr_x and r_y, i.e. the differential partial covariance between x and y, could be calculated as:

$$\begin{aligned} \Delta_{p,xy} &= cov(dr_x, r_y) = cov[(dx - w_x z), (y - w_y z)] \\ &= cov(dx, y) - w_x cox(z, y) - w_y cov(dx, z) + w_x w_y cov(z, z) \\ &= \Delta_{c,xy} - \Delta_{c,xz}\sigma_{zz}^{-1}\sigma_{yz} \end{aligned} \tag{5}$$

Note: differential precision matrix is not symmetric (while the differential covariance matrix is anti-symmetric) because:

$$\begin{aligned} x = w_x z &+ r_x, \quad dy = w_y z + dr_y \\ \Delta_{p,yx} &= cov(r_x, dr_y) \\ &= cov(x - w_x z, dy - w_y z) \\ &= \Delta_{c,yx} - \Delta_{c,yz}\sigma_{zz}^{-1}\sigma_{xz} \end{aligned} \tag{6}$$

3 Sparse Latent Assumption/Regularization

4 Blind Deconvolution

See the lab meeting slides or the original paper (G.-R. Wu et al. / Medical Image Analysis 17 (2013) 365–374) for more information.

5 Significance test

It seems that differential covariance based methods works better if thresholding on the significance level. Then the question remains to be how to determine the significance level of each connection. Using the language of p value, the definition of significance would be the probability of this connectivity value occurring given null hypothesis (i.e. there is no correlation). There are two main approaches to get the significance level. I termed them as statistics-based and bootstrap-based method.

5.1 Correlation matrix

Here the correlation refers to Pearson's correlation, which is the covariance value divided by the standard deviation of two variables. This correlation is testing for the LINEAR dependence of two random variables. The classical way to test for the significance of Pearson's correlation is using t-distribution/normal distribution. Denote the true correlation between two random variables as ρ and the correlation calculated from N samples as r. When the null hypothesis is true, the sampling distribution of r is approximated normal given large N or t-distribution given smaller N with mean equals to zero. The probability of r appearing in this null distribution is the p value.

5.2 Partial covariance matrix / Precision matrix

5.3 Differential covariance matrix (bootstrap?)

6 Linear stochastic differential equations

Reference:

- Chapter 4.2, Applied Stochastic Differential Equations, Sarkka
- Chapter 4.1, Lyapunov Matrix Equation in System Stability and Control, Gajic and Qureshi
- Covariances of Linear Stochastic Differential Equations for Analyzing Computer Networks, Hua et al, Tsinghua Science and Technology

System setup:

$$\frac{d\vec{\mathbf{x}}}{dt} = \mathbf{A}\vec{\mathbf{x}} + \mathbf{D}\frac{d\vec{\beta}}{dt} \tag{7}$$

where random vector $\vec{\mathbf{x}} \in \mathbb{R}^{N\times 1}$ is the state variable and $d\vec{\beta}$ is Brownian motion: $\frac{d\vec{\beta}}{dt} \sim \mathcal{N}(0, \mathbf{Q})$. If we assume $\mathbf{D} = \mathbf{I}$ and $\mathbf{Q} = \alpha\mathbf{I}$ (α determines SNR), the system receives independent Gaussian white noise, which is the most commonly used case.

Mean and covariance of the random state variable:
Since Equation 7 involves a stochastic component, to solve for the state variable, we need to evaluate the Ito integral. Mean and covariance of the state variable could be calculated using Ito formula. Denote $\mathbb{E}(\vec{\mathbf{x}}) = \vec{\mathbf{m}}$ and $Cov(\vec{\mathbf{x}}) = \mathbb{E}[(\vec{\mathbf{x}} - \vec{\mathbf{m}})^2] = \mathbb{P}$, then:

$$\begin{aligned} \frac{d\vec{\mathbf{m}}}{dt} &= \mathbb{E}(\mathbf{A}\vec{\mathbf{x}}) = \mathbf{A}\vec{\mathbf{m}} \\ \frac{d\mathbf{P}}{dt} &= \mathbf{AP} + \mathbf{PA^T} + \mathbf{DQD^T} \end{aligned} \tag{8}$$

Using linear system dynamics, the mean vector $\vec{\mathbf{m}} = \exp(\mathbf{A}t)\vec{\mathbf{m}}_0$. The equation of covariance matrix $\mathbf{P}$ is a standard Lyapunov matrix equation. Its solution is given by:

$$\mathbf{P} = \exp(\mathbf{A}t)\mathbf{P}_0 \exp(\mathbf{A^T}t) + \int_0^t \exp(\mathbf{A}(t-s))\mathbf{DQD^T}\exp(\mathbf{A^T}(t-s))ds \tag{9}$$

On the other hand, the Lyapunov matrix equation could also be solved using vectorization and Kronecker product. Note that $(\mathbf{AXB})_v = (\mathbf{B^T} \otimes \mathbf{A})\mathbf{X}_v$ where subscript v stands for column vectorization and $\otimes$ stands for Kronecker product. Then the Lyapunov matrix equation could be rewritten into a linear differential equation:

$$\frac{d\mathbf{P}_v}{dt} = (\mathbf{I} \otimes \mathbf{A} + \mathbf{A} \otimes \mathbf{I})\mathbf{P}_v + (\mathbf{D} \otimes \mathbf{D})\mathbf{Q}_v \tag{10}$$

Denote $(\mathbf{I} \otimes \mathbf{A} + \mathbf{A} \otimes \mathbf{I}) = \tilde{\mathbf{A}}$, $(\mathbf{D} \otimes \mathbf{D})\mathbf{Q}_v = \vec{\mathbf{c}}$ and assume that $\tilde{\mathbf{A}}$ is invertible, the vectorized solution is:

$$\mathbf{P}_v = \exp(\tilde{\mathbf{A}}t)(\mathbf{P}_{v0} + \tilde{\mathbf{A}}^{-1}\vec{\mathbf{c}}) - \tilde{\mathbf{A}}^{-1}\vec{\mathbf{c}} \tag{11}$$

If $\mathbf{I} \otimes \mathbf{A} + \mathbf{A} \otimes \mathbf{I}$ is stable and in the limit of t goes to infinity:

$$\mathbf{P}_v = -(\mathbf{I} \otimes \mathbf{A} + \mathbf{A} \otimes \mathbf{I})^{-1}(\mathbf{D} \otimes \mathbf{D})\mathbf{Q}_v \tag{12}$$

When we have independent white noise in the system:

$$\mathbf{P}_v = -(\mathbf{I} \otimes \mathbf{A} + \mathbf{A} \otimes \mathbf{I})^{-1}\mathbf{I}_v \tag{13}$$

where $\mathbf{A}$ is the real connection pattern in the original SDE system. This equation serves as an explicit solution of the covariance matrix in a linear SDE system.

Differential covariance matrix:

For differential covariance matrix defined as $\Delta c_{ij} = cov(\frac{d\vec{\mathbf{x}}_i}{dt}, \vec{\mathbf{x}}_j)$, using the SDE system, it can be solved as:

$$\Delta c = cov(\frac{d\vec{x}}{dt}, \vec{x}) = \mathbb{E}_t(\frac{d\vec{x}}{dt}\vec{x}^T) - \mathbb{E}_t(\frac{d\vec{x}}{dt})\mathbb{E}_t(\vec{x})^T \tag{14}$$

Plug in the SDE system equation: $\frac{d\vec{x}}{dt} = \mathbf{A}\vec{x} + \mathbf{D}\frac{d\vec{\beta}}{dt}$ and as before denote $\mathbb{E}_t(\vec{x}) = \vec{m}$:

$$\begin{aligned} \mathbb{E}_t(\frac{d\vec{x}}{dt}) &= \mathbb{E}_t(\mathbf{A}\vec{x} + \mathbf{D}\frac{d\vec{\beta}}{dt}) = \mathbf{A}\vec{m} + \mathbf{D}\mathbb{E}_t(\frac{d\vec{\beta}}{dt}) \\ \mathbb{E}_t(\frac{d\vec{x}}{dt}\vec{x}^T) &= \mathbb{E}_t[(\mathbf{A}\vec{x} + \mathbf{D}\frac{d\vec{\beta}}{dt})\vec{x}^T] = \mathbf{A}\mathbb{E}_t(\vec{x}\vec{x}^T) + \mathbf{D}\mathbb{E}_t(\frac{d\beta}{dt}\vec{x}^T) \end{aligned} \tag{15}$$

Taken together:

$$\begin{aligned} \Delta c &= \mathbf{A}(\mathbb{E}_t(\vec{x}\vec{x}^T) - \vec{m}\vec{m}^T) + \mathbf{D}[\mathbb{E}_t(\frac{d\vec{\beta}}{dt}\vec{x}^T) - \mathbb{E}_t(\frac{d\vec{\beta}}{dt})\vec{m}^T] \\ &= \mathbf{AP} + \mathbf{D}[\mathbb{E}_t(\frac{d\vec{\beta}}{dt}\vec{x}^T) - \mathbb{E}_t(\frac{d\vec{\beta}}{dt})\vec{m}^T] \\ &= \mathbf{AP} + \mathbf{D}\mathbb{E}_t[\frac{d\vec{\beta}}{dt}(\vec{x}^T - \vec{m}^T)] \\ &= \mathbf{A}cov(\vec{x}, \vec{x}^T) + \mathbf{D}cov(\frac{d\vec{\vec{\beta}}}{dt}, \vec{x}) \end{aligned} \tag{16}$$

Under the assumption that $\frac{d\vec{\beta}}{dt} \sim \mathcal{N}(0, \mathbf{Q})$, i.e. $\vec{\beta}$ is Brownian motion and independent of $\vec{x}$, the second terms become zero. **Actually, as long as the noise satisfies independence assumption, the following expression will hold**. Then:

$$\Delta c = \mathbf{AP} \tag{17}$$

Note, if $\mathbf{A}$ is a lower triangle matrix, Δc could help resolve the confounder problem, at least proved in a three-neuron case.

Extended differential covariance: Extended differential covariance is defined as $\Delta \mathrm{L} = \Delta c\mathbf{P}^{-1}$. From the above calculation, under independent Brownian noise assumption:

$$\Delta \mathrm{L} = \mathbf{APP}^{-1} = \mathbf{A} \tag{18}$$

Forced system: Similarly in a forced stochastic system governed by:

$$\frac{d\vec{x}}{dt} = \mathbf{A}\vec{x} + \mathbf{B}\vec{u} + \mathbf{D}\frac{d\vec{\beta}}{dt} \tag{19}$$

The expression of differential covariance is:

$$\Delta c = \mathbf{A}cov(\vec{x}, \vec{x}) + \mathbf{B}cov(\vec{u}, \vec{x}) + \mathbf{D}cov(\frac{d\vec{\beta}}{dt}, \vec{x}) \tag{20}$$

The issue is that the independence between $\vec{u}$ and $\vec{x}$ cannot be achieved because this is a forced system.

7 Numerical methods for derivative computation

It turns out that the specific numerical methods chosen to compute the derivative will affect the estimation process a lot. Let's consider a noise-free linear system and its discretization through Euler integration:
System:

$$\frac{dx}{dt} = Ax; \quad x_{t+1} = x_t + Ax_t * dt \tag{21}$$

First order derivative: P is the sample covariance matrix of the state variable x

$$\begin{aligned}(\frac{dx}{dt})_t &= \frac{1}{dt}(x_{t+1} - x_t) = Ax_t \\ \Delta c &= \frac{1}{N-1}\sum_{t=1}^{N-1}((\frac{dx}{dt})_t x_t^T) = \frac{1}{N-1}\sum_{t=1}^{N-1}(Ax_t x_t^T) = AP \\ \hat{A} &= \Delta c P^{-1}\end{aligned} \tag{22}$$

Second order derivative: Similarly, we could iterate the computation of x_{t+1} and x_t to get the following results. Worth notice is that certain symmetric choice of A (when $AP = PA$) may lead to cancellation of the term $AP - PA^T$.

$$\begin{aligned}(\frac{dx}{dt})_t &= \frac{1}{2dt}(x_{t+1} - x_{t-1}) \\ \Delta c &= 2(AP - PA^T) \\ \Delta c P^{-1} &= 2(A - PA^T P^{-1})\end{aligned} \tag{23}$$

8 Differential covariance in the Fourier domain

Still consider the linear dynamic system and given that Fourier transform is a linear operation:

$$\begin{aligned} \frac{dx}{dt} &= \mathbf{A}x \\ \widehat{\tfrac{dx}{dt}} &= \mathbf{A}\hat{x} \\ i\omega\hat{x} &= \mathbf{A}\hat{x} \end{aligned} \tag{24}$$

Now let's consider the hemodynamic effects as convolution by the kernel $h(t)$:

$$\begin{aligned} \widehat{i\omega x \otimes h} &= \mathbf{A}\widehat{x \otimes h} \\ i\omega\hat{x}\hat{h} &= \mathbf{A}\hat{x}\hat{h} \end{aligned} \tag{25}$$

Define $\hat{x}\hat{h} = Y(w)$:

$$\begin{aligned} i\omega Y(\omega) &= \mathbf{A}Y(\omega) \\ i\omega Y(\omega)Y(\omega)^H &= \mathbf{A}Y(\omega)Y(\omega)^H \end{aligned} \tag{26}$$

9 Time decaying system equation

For neural dynamics, there's usually a self-decaying factor in the system equation. For example, the leaky dynamics in LIF. In this case, we could modify the system equation to be:

$$\begin{aligned}\tau\frac{d\mathbf{x}}{dt} &= -\mathbf{x} + \mathbf{W}\mathbf{x} \\ \tau\frac{d\mathbf{x}}{dt} &= -\mathbf{x} + \mathbf{W}R(\mathbf{x})\end{aligned} \tag{27}$$

For the linear case, compared to the original derivation, DDC estimation only differs up to a scaling factor and the diagonal terms:

$$\Delta\mathrm{L} = \langle\frac{d\mathbf{x}}{dt}, \mathbf{x}\rangle\langle\mathbf{x}, \mathbf{x}\rangle^{-1} = \frac{1}{\tau}(\mathbf{W} - \mathbf{I}) \tag{28}$$

But for the nonlinear system, there's a bigger influence on the orginal DDCs.

$$\begin{aligned}\tau\langle\frac{d\mathbf{x}}{dt}, \mathbf{x}\rangle &= -\langle\mathbf{x}, \mathbf{x}\rangle + \mathbf{W}\langle R(\mathbf{x}), \mathbf{x}\rangle \\ \Delta\mathrm{L} = \langle\frac{d\mathbf{x}}{dt}, \mathbf{x}\rangle\langle\mathbf{x}, \mathbf{x}\rangle^{-1} &= -\mathbf{I} + \mathbf{W}\langle R(\mathbf{x}), \mathbf{x}\rangle\langle\mathbf{x}, \mathbf{x}\rangle^{-1} \\ \Delta\mathrm{R} = \langle\frac{d\mathbf{x}}{dt}, \mathbf{x}\rangle\langle\mathrm{R'}(\mathbf{x}), \mathbf{x}\rangle^{-1} &= -\langle\mathbf{x}, \mathbf{x}\rangle\langle\mathrm{R'}(\mathbf{x}), \mathbf{x}\rangle^{-1} + \mathbf{W}\langle R(\mathbf{x}), \mathbf{x}\rangle\langle\mathrm{R'}(\mathbf{x}), \mathbf{x}\rangle^{-1}\end{aligned} \tag{29}$$

A direct modification is to compute a decaying version:

$$\Delta\mathrm{D} = (\tau\langle\frac{d\mathbf{x}}{dt}, \mathbf{x}\rangle + \langle\mathbf{x}, \mathbf{x}\rangle)\langle\mathrm{R'}(\mathbf{x}), \mathbf{x}\rangle^{-1} \tag{30}$$

In this case, we have to guess/estimate the self-decaying constant a prior or optimize for this constant.