Seminar for HIT

Investigation of long-term memory without Periodogram and Gaussianity

Presented by **Shao-Qun Zhang**

2021/11/17

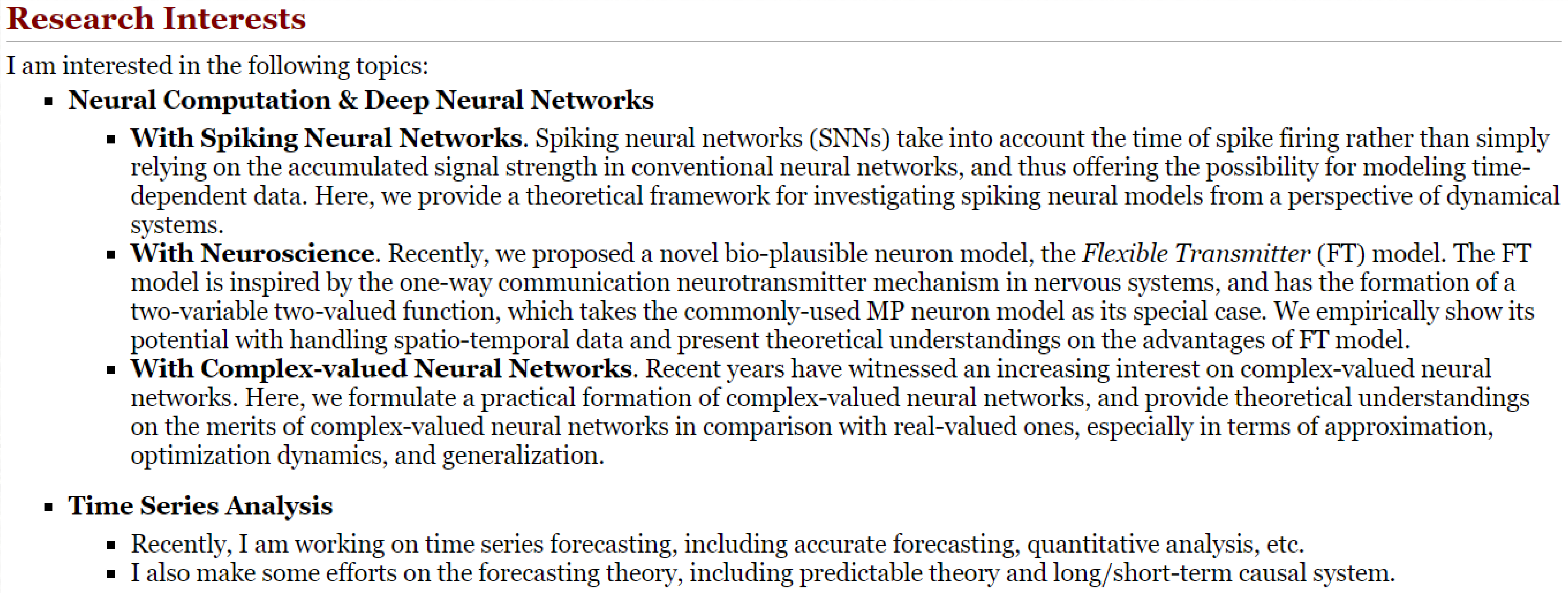
张绍群 (Shao-Qun Zhang)

Ph.D. candidate, LAMDA Group

Supervisor: Professor Zhi-Hua Zhou

HomePage: http://www.lamda.nju.edu.cn/zhangsq/

Email: zhangsq@lamda.nju.edu.cn
zhangsqhndn@gmail.com



Research Interests

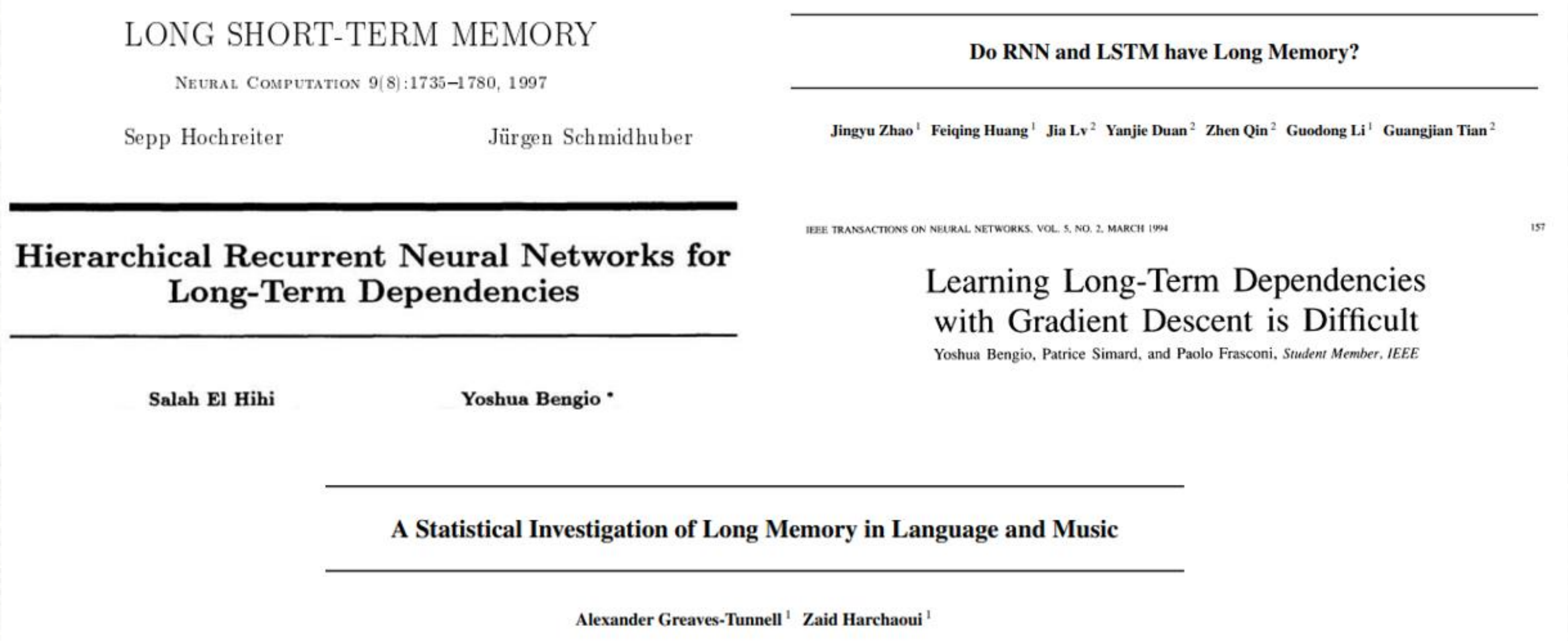
I am interested in the following topics:

- **Neural Computation & Deep Neural Networks**
  - **With Spiking Neural Networks**. Spiking neural networks (SNNs) take into account the time of spike firing rather than simply relying on the accumulated signal strength in conventional neural networks, and thus offering the possibility for modeling time-dependent data. Here, we provide a theoretical framework for investigating spiking neural models from a perspective of dynamical systems.
  - **With Neuroscience**. Recently, we proposed a novel bio-plausible neuron model, the *Flexible Transmitter* (FT) model. The FT model is inspired by the one-way communication neurotransmitter mechanism in nervous systems, and has the formation of a two-variable two-valued function, which takes the commonly-used MP neuron model as its special case. We empirically show its potential with handling spatio-temporal data and present theoretical understandings on the advantages of FT model.
  - **With Complex-valued Neural Networks**. Recent years have witnessed an increasing interest on complex-valued neural networks. Here, we formulate a practical formation of complex-valued neural networks, and provide theoretical understandings on the merits of complex-valued neural networks in comparison with real-valued ones, especially in terms of approximation, optimization dynamics, and generalization.

- **Time Series Analysis**
  - Recently, I am working on time series forecasting, including accurate forecasting, quantitative analysis, etc.
  - I also make some efforts on the forecasting theory, including predictable theory and long/short-term causal system.

Topic

Representation and learning of long-term memory is a fundamental problem confronted in machine learning to sequential data.



LONG SHORT-TERM MEMORY

Neural Computation 9(8):1735–1780, 1997

Sepp Hochreiter Jürgen Schmidhuber

Do RNN and LSTM have Long Memory?

Jingyu Zhao[1] Feiqing Huang[1] Jia Lv[2] Yanjie Duan[2] Zhen Qin[2] Guodong Li[1] Guangjian Tian[2]

Hierarchical Recurrent Neural Networks for Long-Term Dependencies

Salah El Hihi Yoshua Bengio *

IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 5, NO. 2, MARCH 1994 157

Learning Long-Term Dependencies with Gradient Descent is Difficult

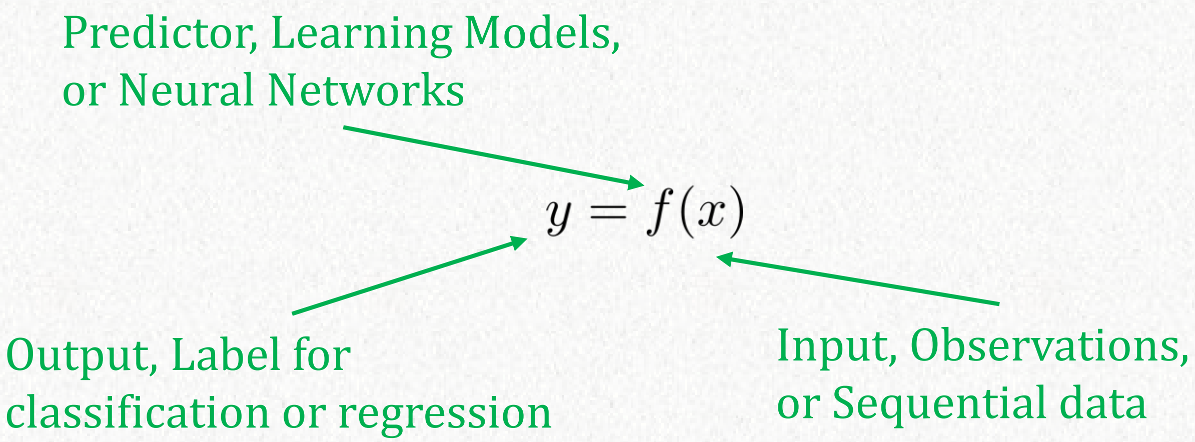
Yoshua Bengio, Patrice Simard, and Paolo Frasconi, *Student Member, IEEE*

A Statistical Investigation of Long Memory in Language and Music

Alexander Greaves-Tunnell[1] Zaid Harchaoui[1]

largely limited to heuristic tools

Long-term Memory & Long-range Dependency



Predictor, Learning Models, or Neural Networks

$$y = f(x)$$

Output, Label for classification or regression

Input, Observations, or Sequential data

Predictor $\boldsymbol{f}$ learns the “rules” or patterns from observations $(\boldsymbol{x}, \boldsymbol{y})$.

Linear, Kernel, Long-term Memory, etc.

The long-term memory of the concerned model is closely related to the long-range dependency of the data.

Definition 1 (Dependency in Statistic) *Let $\{X_t, t \in \mathbb{Z}\}$ be a second-order stationary univariate process with auto-covariance function $\gamma_X(k)$ for all $k \in \mathbb{Z}$. Then the concerned process $\{X_t, t \in \mathbb{Z}\}$ has (1) long-term dependency, or (2) short-term dependency if*

$$(1) \sum_{k=-\infty}^{\infty} \gamma_X(k) = \infty, \quad or \quad (2)\ 0 < \sum_{k=-\infty}^{\infty} \gamma_X(k) < \infty, \tag{1}$$

under a well-defined Fourier spectral density $f_X(\omega) = (2\pi)^{-1} \sum_{k=-\infty}^{\infty} \gamma_X(k) exp(-ik\omega)$ for $\omega \in [-\pi, \pi]$.

Auto-Regressive Moving Average (ARMA) model

$$\text{AR(p)}: \quad \sum_{i=1}^{p} \alpha_i x_{t-i} = \epsilon_t.$$

$$\gamma_k = \sum_{i=1}^{p} c_i \lambda_i^k \to 0 \quad \text{with} \quad |\lambda_i| < 1, \quad \text{as} \quad k \to \infty.$$

Long-range Dependency

The auto-correlation of AR(p):

$$\gamma_k = \sum_{i=1}^{p} c_i \lambda_i^k \to 0 \quad \text{with} \quad |\lambda_i| < 1, \quad \text{as} \quad k \to \infty.$$

Auto-correlation exponential decays, not at polynomials.

Intuitively,

First, strengthen the model.

- Complex enough.
- Clear formulation or specification.

RNN, LSTM, and their variants X

Second, develop the measure.

- Polynomial correlation.

Besides, for reality.

- Long-range dependency.
- Non-stationarity.
- Aperiodic spectrum.

ARISE: ApeRIodic SEmi-parametric Process

Part I: Parametric Integrated Process

$$\begin{pmatrix} (1-\mathfrak{B})^{d_1} & & 0 \\ & \ddots & \\ 0 & & (1-\mathfrak{B})^{d_l} \end{pmatrix} \begin{pmatrix} \mathbf{X}_{1t} \\ \vdots \\ \mathbf{X}_{lt} \end{pmatrix} = \begin{pmatrix} \epsilon_{1t} \\ \vdots \\ \epsilon_{lt} \end{pmatrix}$$

where

l-dimensional generation process $\{\mathbf{X}_t\}_{t=0}^{\infty}$ with $\mathbb{E}(\mathbf{X}_{it}) = 0$ for $i \in [q]$

source process $\{\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \dots, \epsilon_{lt})^\top\}_{t=0}^{\infty}$ is weakly stationary whose spectral density $f_\epsilon(\lambda)$ is bounded and bounded away from zero when frequency λ tends to zero

$\boldsymbol{d} = (d_1, \dots, d_l)^\top \in (-1/2, 1/2)^l$

$\mathfrak{B}$ is the backward-shift operator, satisfying that $\mathfrak{B}^k \mathbf{X}_{it} = \mathbf{X}_{i(t-k)}$ for $k \in \mathbb{N}^+$ and $i \in [l]$

Part II: Aperiodic Spectrum Estimation

$$\mathbf{J}_T(\lambda_j) = \sum_{(i,\kappa)\in\mathfrak{J}_T} \tau(\cdot; \alpha_{i,\kappa}, \rho_{i,\kappa}) \varphi_{i,\kappa}(\lambda_j).$$

Parametric Integrated Process

$$\begin{pmatrix} (1-\mathfrak{B})^{d_1} & & 0 \\ & \ddots & \\ 0 & & (1-\mathfrak{B})^{d_l} \end{pmatrix} \begin{pmatrix} \mathbf{X}_{1t} \\ \vdots \\ \mathbf{X}_{lt} \end{pmatrix} = \begin{pmatrix} \epsilon_{1t} \\ \vdots \\ \epsilon_{lt} \end{pmatrix}$$

Inverse

Provided $d_i > 0$ for $i \in [l]$, one has

$$\mathbf{X}_t = \mathrm{diag}_{i\in\{1,\cdots,l\}}\left\{(1-\mathfrak{B})^{-d_i}\right\}\boldsymbol{\epsilon}_t = \mathrm{diag}_{i\in\{1,\cdots,l\}}\left\{\sum_{j=0}^{\infty}\frac{\Gamma(d_i+j)}{\Gamma(d_i)j!}\mathfrak{B}^j\right\}\boldsymbol{\epsilon}_t$$

$$= \sum_{j=0}^{\infty}\mathrm{diag}_{i\in\{1,\cdots,l\}}\left\{\frac{\Gamma(d_i+j)}{\Gamma(d_i)j!}\right\}\boldsymbol{\epsilon}_{t-j},$$

where $\Gamma(\cdot)$ is the Gamma function.

d is called memory parameter.

For $d_i > 0$ and $i \in [l]$,

$$\gamma_k \propto k^{2d_i-1} \text{ as } k \to \infty, \text{ and } f_X(\lambda) \propto \lambda^{-2d_i} \text{ as } \lambda \to 0^+.$$

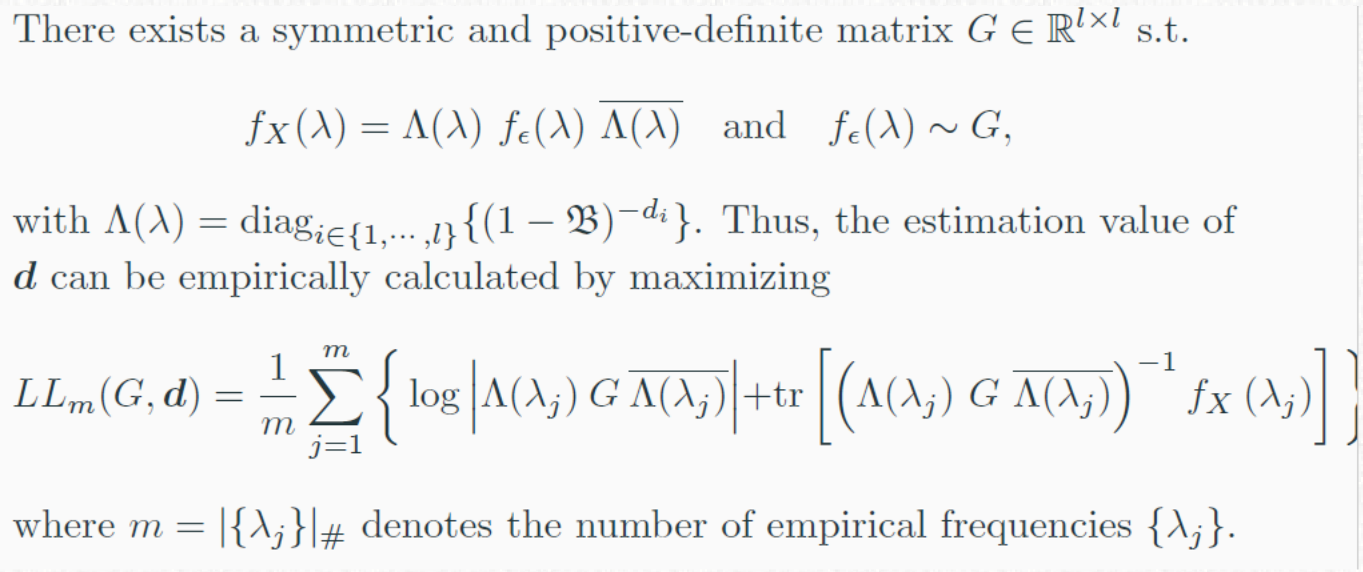
In contrast, $d_i < 0$ leads to an anti-persistent process.

has the power and potential of mimicking the data with long-range dependency.

Solving the Memory Parameter

$$\begin{pmatrix} (1-\mathfrak{B})^{d_1} & & 0 \\ & \ddots & \\ 0 & & (1-\mathfrak{B})^{d_l} \end{pmatrix} \begin{pmatrix} \mathbf{X}_{1t} \\ \vdots \\ \mathbf{X}_{lt} \end{pmatrix} = \begin{pmatrix} \epsilon_{1t} \\ \vdots \\ \epsilon_{lt} \end{pmatrix}$$

A trivial solution way: maximum likelihood estimation.



There exists a symmetric and positive-definite matrix $G \in \mathbb{R}^{l \times l}$ s.t.

$$f_X(\lambda) = \Lambda(\lambda)\ f_\epsilon(\lambda)\ \overline{\Lambda(\lambda)} \quad \text{and} \quad f_\epsilon(\lambda) \sim G,$$

with $\Lambda(\lambda) = \mathrm{diag}_{i\in\{1,\cdots,l\}}\{(1-\mathfrak{B})^{-d_i}\}$. Thus, the estimation value of $\boldsymbol{d}$ can be empirically calculated by maximizing

$$LL_m(G, \boldsymbol{d}) = \frac{1}{m}\sum_{j=1}^{m}\left\{\log\left|\Lambda(\lambda_j)\ G\ \overline{\Lambda(\lambda_j)}\right| + \mathrm{tr}\left[\left(\Lambda(\lambda_j)\ G\ \overline{\Lambda(\lambda_j)}\right)^{-1} f_X(\lambda_j)\right]\right\}$$

where $m = |\{\lambda_j\}|_\#$ denotes the number of empirical frequencies $\{\lambda_j\}$.

Highlights:

G indicates the covariance matrix.

f_X is consistent, correspondingly, ε_t is weakly-stationary.

otherwise, ε_t must be Gaussian.

Aperiodic Spectrum Estimation

Spectrum Density Estimation for f_X.

Parametric: Inconsistent.

DFT and FFT.

Semi-Parametric: Inconsistent.

windows functions, taper function, and kernel smooth.

Non-parametric: Consistent, provide appropriate thresholds.

Wavelet (via threshold): plug-in estimation of the variance and log-transformation of the periodogram.

Threshold

$$\mathbf{J}_T(\lambda_j) = \sum_{(i,\kappa)\in\mathfrak{J}_T} \tau(\cdot\,; \alpha_{i,\kappa}, \rho_{i,\kappa})\varphi_{i,\kappa}(\lambda_j).$$

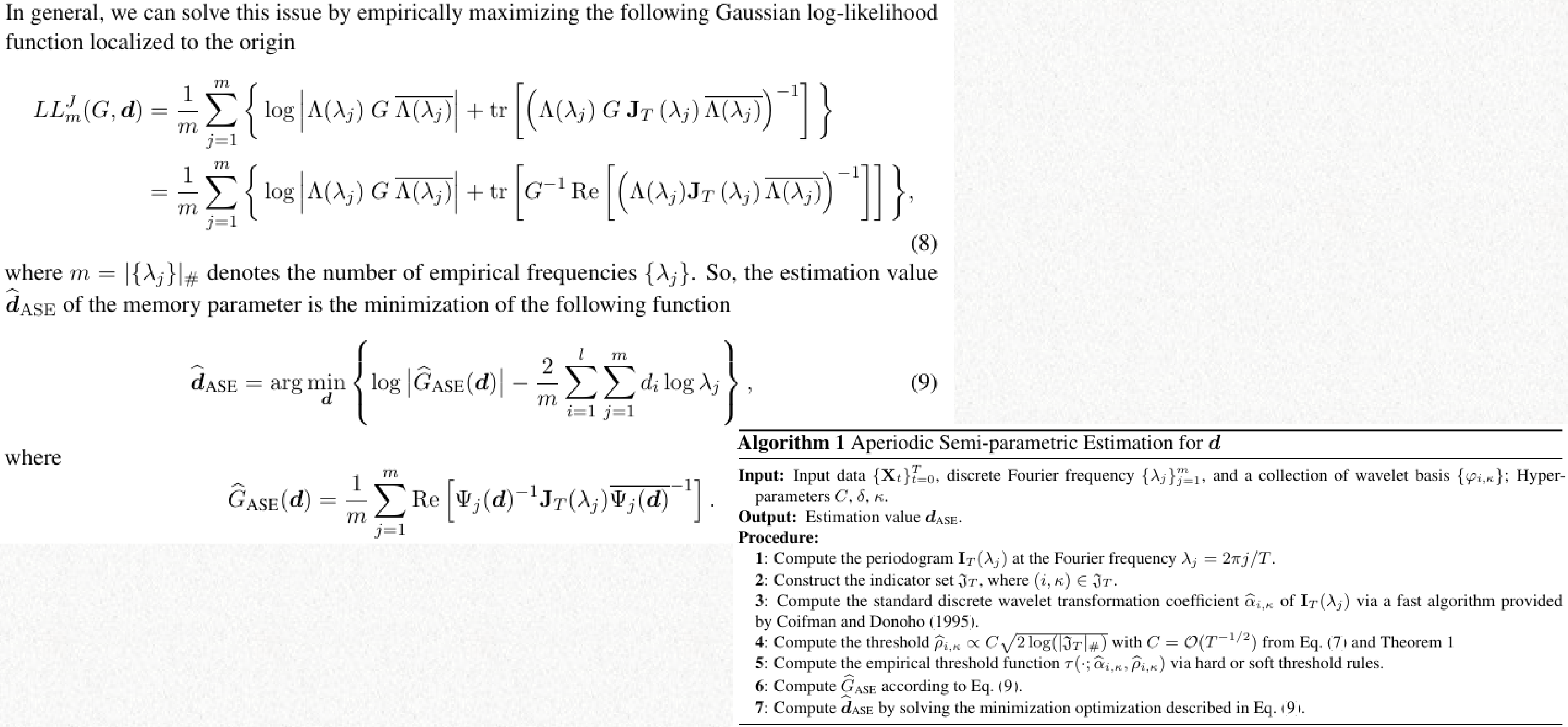
Threshold function (soft or hard)

Basis function

$$\widehat{\rho}_{i,\kappa} = \Theta(T^{-1/2}) \cdot \left(\int_{-\pi}^{\pi} \varphi_{i,\kappa}(\lambda)\mathbf{I}_T(\lambda)\,\mathrm{d}\lambda\right) \cdot \sqrt{2\log(|\mathfrak{J}_T|_{\#})}.$$

Key Idea: local weighted l_1 norms of the periodogram.

Recall ARISE Process



In general, we can solve this issue by empirically maximizing the following Gaussian log-likelihood function localized to the origin

$$LL_m^J(G,\boldsymbol{d}) = \frac{1}{m}\sum_{j=1}^{m}\left\{ \log\left|\Lambda(\lambda_j)\, G\, \overline{\Lambda(\lambda_j)}\right| + \operatorname{tr}\left[\left(\Lambda(\lambda_j)\, G\, \mathbf{J}_T(\lambda_j)\, \overline{\Lambda(\lambda_j)}\right)^{-1}\right]\right\}$$

$$= \frac{1}{m}\sum_{j=1}^{m}\left\{ \log\left|\Lambda(\lambda_j)\, G\, \overline{\Lambda(\lambda_j)}\right| + \operatorname{tr}\left[G^{-1}\operatorname{Re}\left[\left(\Lambda(\lambda_j)\mathbf{J}_T(\lambda_j)\, \overline{\Lambda(\lambda_j)}\right)^{-1}\right]\right]\right\}, \tag{8}$$

where $m = |\{\lambda_j\}|_{\#}$ denotes the number of empirical frequencies $\{\lambda_j\}$. So, the estimation value $\widehat{\boldsymbol{d}}_{\text{ASE}}$ of the memory parameter is the minimization of the following function

$$\widehat{\boldsymbol{d}}_{\text{ASE}} = \arg\min_{\boldsymbol{d}}\left\{\log\left|\widehat{G}_{\text{ASE}}(\boldsymbol{d})\right| - \frac{2}{m}\sum_{i=1}^{l}\sum_{j=1}^{m} d_i \log\lambda_j\right\}, \tag{9}$$

where

$$\widehat{G}_{\text{ASE}}(\boldsymbol{d}) = \frac{1}{m}\sum_{j=1}^{m}\operatorname{Re}\left[\Psi_j(\boldsymbol{d})^{-1}\mathbf{J}_T(\lambda_j)\overline{\Psi_j(\boldsymbol{d})}^{-1}\right].$$

Algorithm 1 Aperiodic Semi-parametric Estimation for $\boldsymbol{d}$

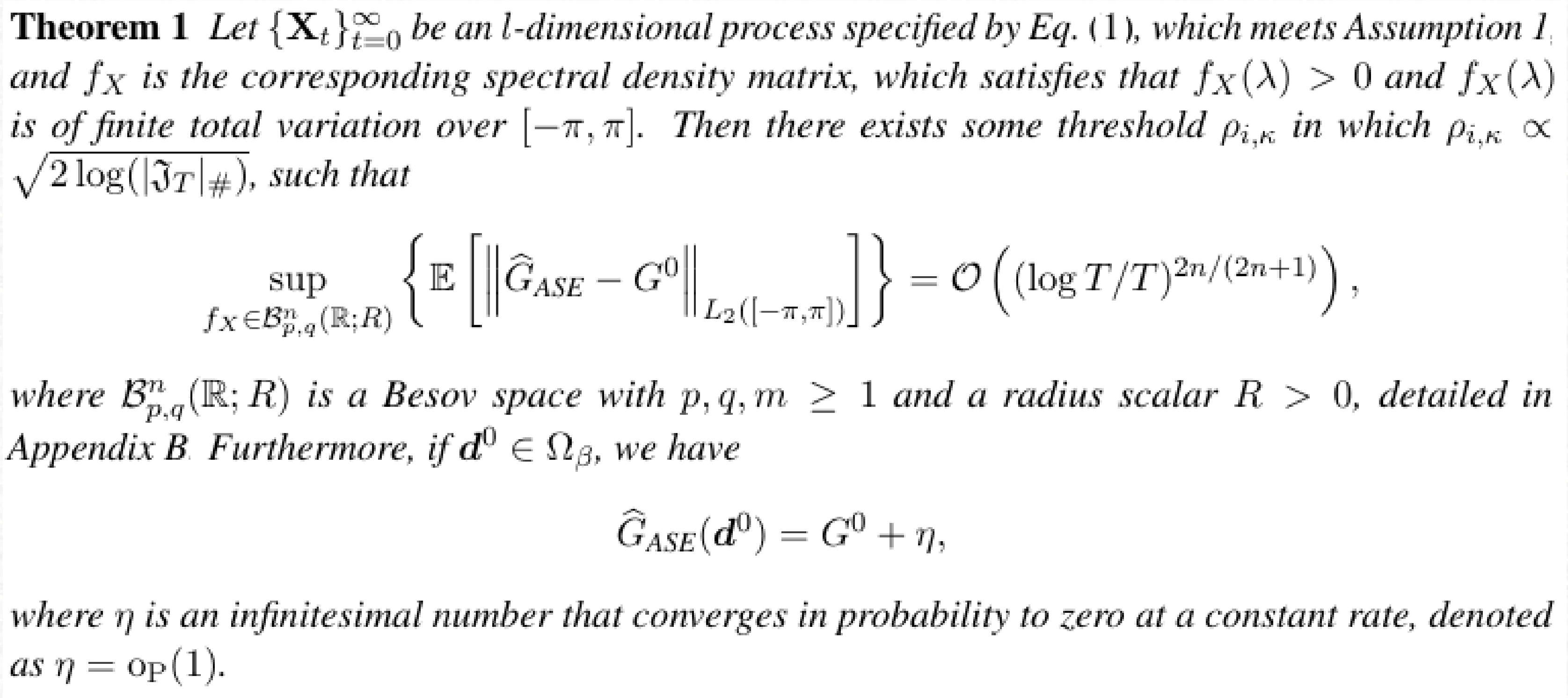
Input: Input data $\{\mathbf{X}_t\}_{t=0}^{T}$, discrete Fourier frequency $\{\lambda_j\}_{j=1}^{m}$, and a collection of wavelet basis $\{\varphi_{i,\kappa}\}$; Hyper-parameters C, δ, κ.

Output: Estimation value $\boldsymbol{d}_{\text{ASE}}$.

Procedure:

```
1: Compute the periodogram $\mathbf{I}_T(\lambda_j)$ at the Fourier frequency $\lambda_j = 2\pi j/T$.
2: Construct the indicator set $\mathfrak{J}_T$, where $(i,\kappa) \in \mathfrak{J}_T$.
3: Compute the standard discrete wavelet transformation coefficient $\widehat{\alpha}_{i,\kappa}$ of $\mathbf{I}_T(\lambda_j)$ via a fast algorithm provided by Coifman and Donoho (1995).
4: Compute the threshold $\widehat{\rho}_{i,\kappa} \propto C\sqrt{2\log(|\mathfrak{J}_T|_{\#})}$ with $C = \mathcal{O}(T^{-1/2})$ from Eq. (7) and Theorem 1
5: Compute the empirical threshold function $\tau(\cdot;\widehat{\alpha}_{i,\kappa},\widehat{\rho}_{i,\kappa})$ via hard or soft threshold rules.
6: Compute $\widehat{G}_{\text{ASE}}$ according to Eq. (9).
7: Compute $\widehat{\boldsymbol{d}}_{\text{ASE}}$ by solving the minimization optimization described in Eq. (9).
```

Theoretical Guarantee



Theorem 1 *Let $\{\mathbf{X}_t\}_{t=0}^{\infty}$ be an l-dimensional process specified by Eq. (1), which meets Assumption 1, and f_X is the corresponding spectral density matrix, which satisfies that $f_X(\lambda) > 0$ and $f_X(\lambda)$ is of finite total variation over $[-\pi, \pi]$. Then there exists some threshold $\rho_{i,\kappa}$ in which $\rho_{i,\kappa} \propto \sqrt{2\log(|\mathfrak{J}_T|_{\#})}$, such that*

$$\sup_{f_X \in \mathcal{B}_{p,q}^{n}(\mathbb{R};R)} \left\{ \mathbb{E}\left[\left\| \widehat{G}_{ASE} - G^0 \right\|_{L_2([-\pi,\pi])} \right] \right\} = \mathcal{O}\left((\log T/T)^{2n/(2n+1)} \right),$$

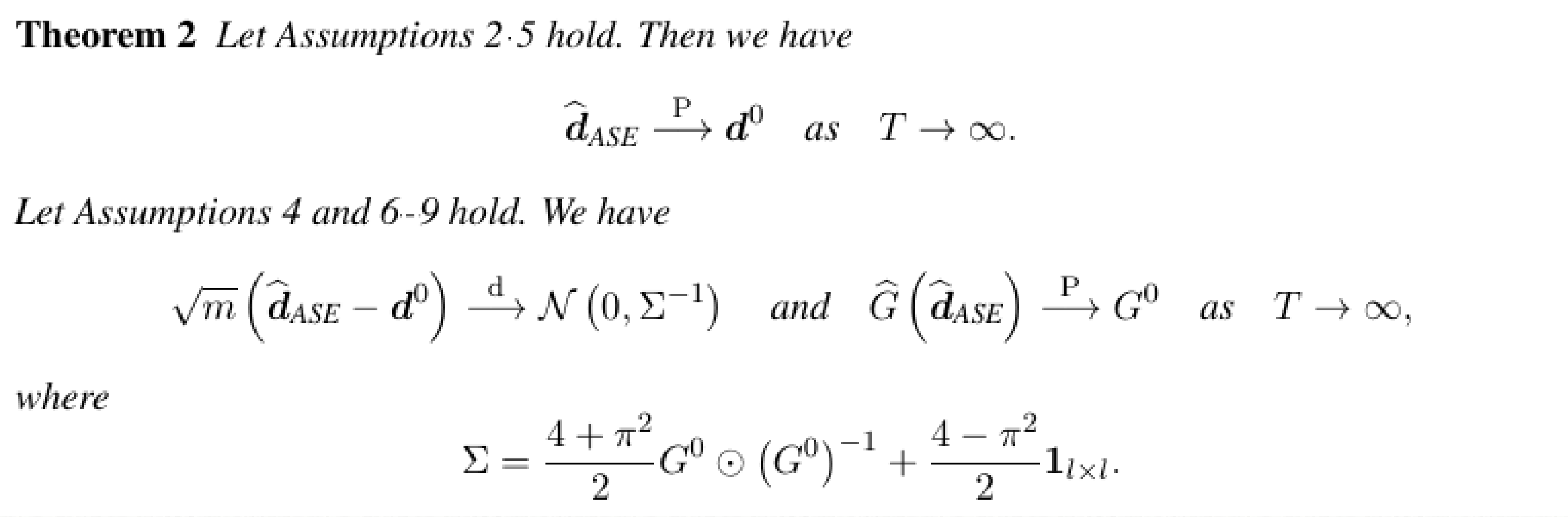
where $\mathcal{B}_{p,q}^{n}(\mathbb{R}; R)$ is a Besov space with $p, q, m \geq 1$ and a radius scalar $R > 0$, detailed in Appendix B. Furthermore, if $\mathbf{d}^0 \in \Omega_\beta$, we have

$$\widehat{G}_{ASE}(\mathbf{d}^0) = G^0 + \eta,$$

where η is an infinitesimal number that converges in probability to zero at a constant rate, denoted as $\eta = o_P(1)$.

Theorem 1 establishes the consistency of $\hat{G}_{\text{ASE}}$, including a guarantee that $\hat{G}_{\text{ASE}}$ has the near-optimal rate of mean-square convergence and a consistent approximation in probability. There optimal rate of mean-square convergence, alternatively known as minimax rate is $T^{-2n/(2n+1)}$.

Theoretical Guarantee



Theorem 2 *Let Assumptions 2·5 hold. Then we have*

$$\widehat{\boldsymbol{d}}_{\mathrm{ASE}} \xrightarrow{\mathrm{P}} \boldsymbol{d}^0 \quad as \quad T \to \infty.$$

Let Assumptions 4 and 6-9 hold. We have

$$\sqrt{m}\left(\widehat{\boldsymbol{d}}_{\mathrm{ASE}} - \boldsymbol{d}^0\right) \xrightarrow{\mathrm{d}} \mathcal{N}\left(0, \Sigma^{-1}\right) \quad and \quad \widehat{G}\left(\widehat{\boldsymbol{d}}_{\mathrm{ASE}}\right) \xrightarrow{\mathrm{P}} G^0 \quad as \quad T \to \infty,$$

where

$$\Sigma = \frac{4+\pi^2}{2} G^0 \odot \left(G^0\right)^{-1} + \frac{4-\pi^2}{2} \mathbf{1}_{l\times l}.$$

Theorem 2 establishes the consistency and asymptotic normality of $\hat{d}_{\mathrm{ASE}}$.

Notice that both Theorems 1 and 2 hold without Periodogram and Gaussianity assumptions.

Brief Summary

Part I: Parametric Integrated Process

$$\begin{pmatrix} (1-\mathfrak{B})^{d_1} & & 0 \\ & \ddots & \\ 0 & & (1-\mathfrak{B})^{d_l} \end{pmatrix} \begin{pmatrix} \mathbf{X}_{1t} \\ \vdots \\ \mathbf{X}_{lt} \end{pmatrix} = \begin{pmatrix} \epsilon_{1t} \\ \vdots \\ \epsilon_{lt} \end{pmatrix}$$

Part II: Aperiodic Spectrum Estimation

$$\mathbf{J}_T(\lambda_j) = \sum_{(i,\kappa)\in\mathfrak{I}_T} \tau(\cdot;\alpha_{i,\kappa},\rho_{i,\kappa})\varphi_{i,\kappa}(\lambda_j).$$

Characteristics of Data:

- Long-range dependency.
- Non-stationarity.
- Aperiodic spectrum.

Solved by Part I about modeling.

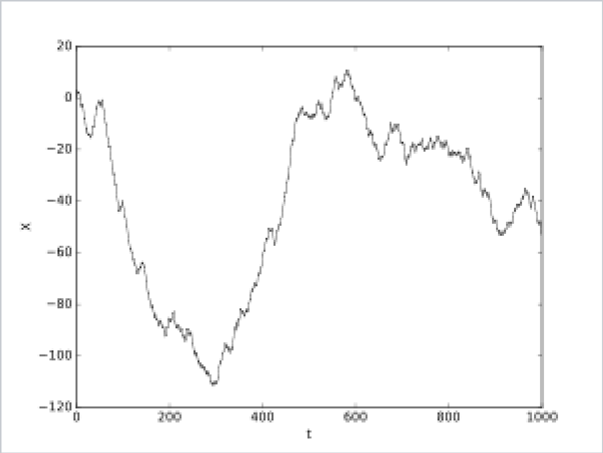
Solved by Part II about handling time series.

Theoretical Guarantee:

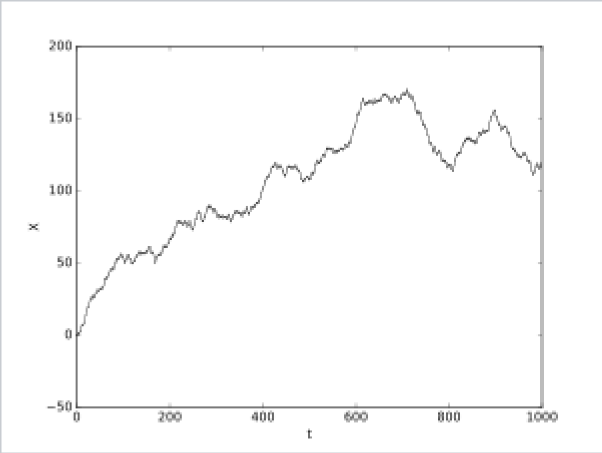
- Near-optimal convergence.
- Consistency.
- Asymptotic normality .

Both Theorems 1 and 2 hold without Periodogram and Gaussianity assumptions.

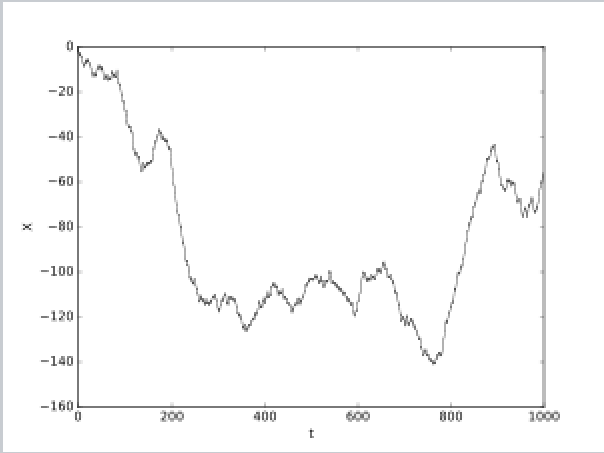
此处是用 ARISE-AR(1) 生成的数据, 图中的 H=d+1/2.



"H" = 0.75 realisation 1

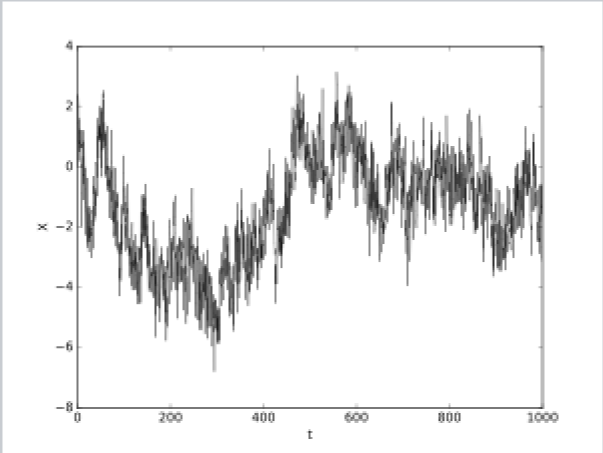


"H" = 0.75 realisation 2

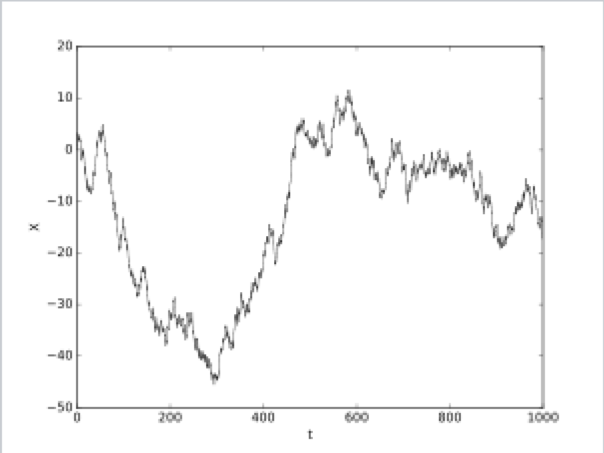


"H" = 0.75 realisation 3

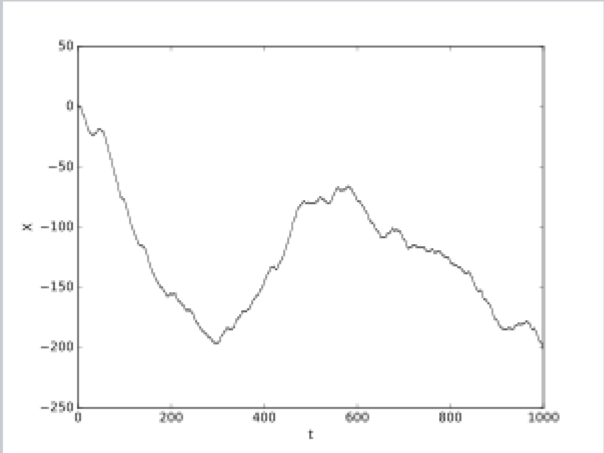
H>0.5时,
抖动小,
具有长趋势.



"H" = 0.15



"H" = 0.55



"H" = 0.95

H<0.5时,
抖动非常大,
剧烈震荡.

Generalized ARISE Models

Next, we proceed to develop some generalized formations of the ARISE process.

ARISE-ARMA(p,d,q):

ARMA(p,q)

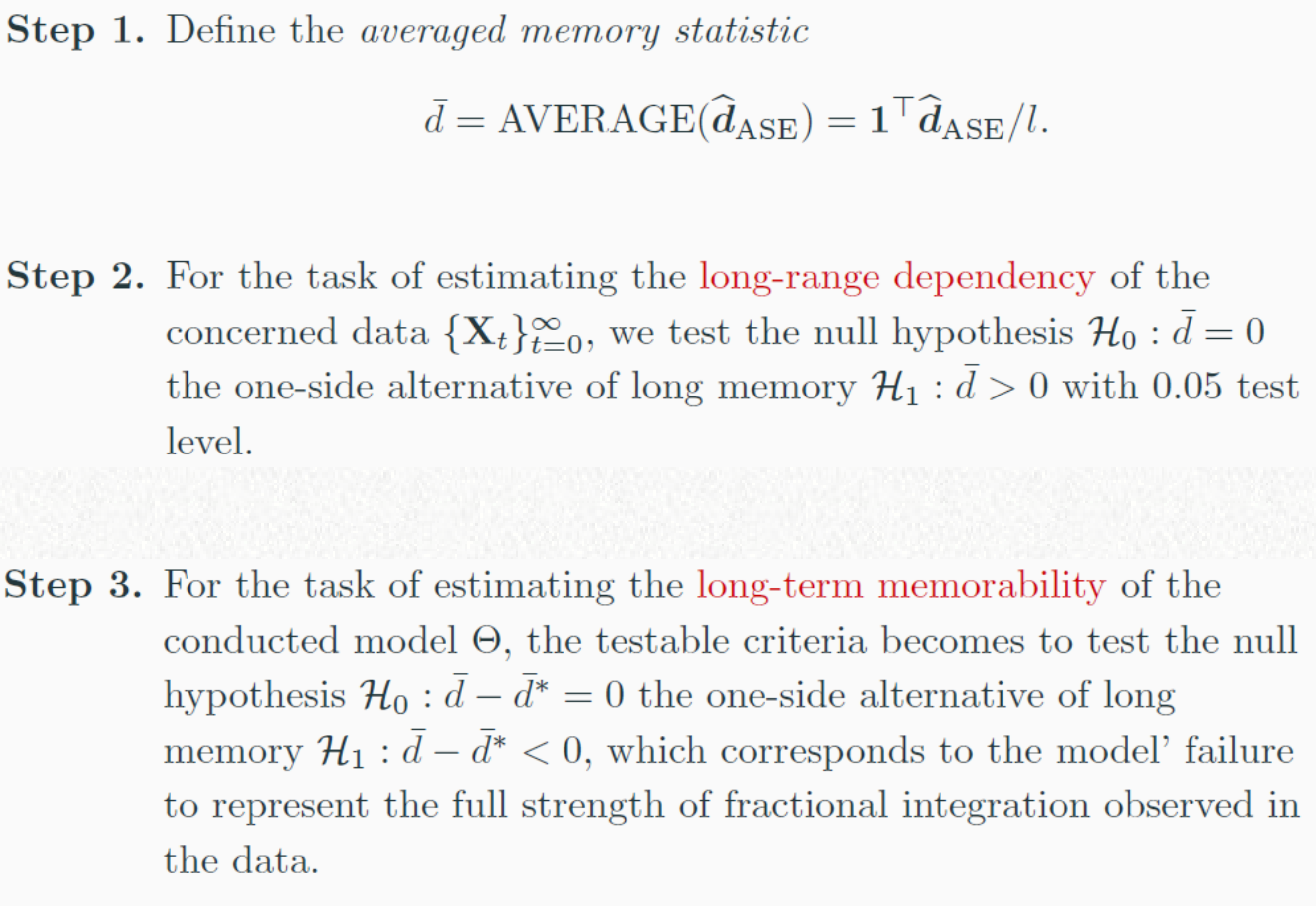
$$\mathrm{diag}_{i\in\{1,\cdots,l\}}\left\{(1-\mathfrak{B})^{d_i}\right\}\mathbf{X}_t=\phi^{-1}(\mathfrak{B})\ \boldsymbol{\epsilon}_t\ \psi(\mathfrak{B}).$$

ARISE-Θ(p,d,q):

Machine learning model

$$\mathrm{diag}_{i\in\{1,\cdots,l\}}\left\{(1-\mathfrak{B})^{d_i}\right\}\mathbf{X}_t=\Theta(\boldsymbol{\epsilon}_t).$$

Investigation as the Memorability Indicator



Step 1. Define the *averaged memory statistic*

$$\bar{d} = \mathrm{AVERAGE}(\widehat{\boldsymbol{d}}_{\mathrm{ASE}}) = \mathbf{1}^{\top}\widehat{\boldsymbol{d}}_{\mathrm{ASE}}/l.$$

Step 2. For the task of estimating the long-range dependency of the concerned data $\{\mathbf{X}_t\}_{t=0}^{\infty}$, we test the null hypothesis $\mathcal{H}_0 : \bar{d} = 0$ the one-side alternative of long memory $\mathcal{H}_1 : \bar{d} > 0$ with 0.05 test level.

Step 3. For the task of estimating the long-term memorability of the conducted model Θ, the testable criteria becomes to test the null hypothesis $\mathcal{H}_0 : \bar{d} - \bar{d}^* = 0$ the one-side alternative of long memory $\mathcal{H}_1 : \bar{d} - \bar{d}^* < 0$, which corresponds to the model' failure to represent the full strength of fractional integration observed in the data.

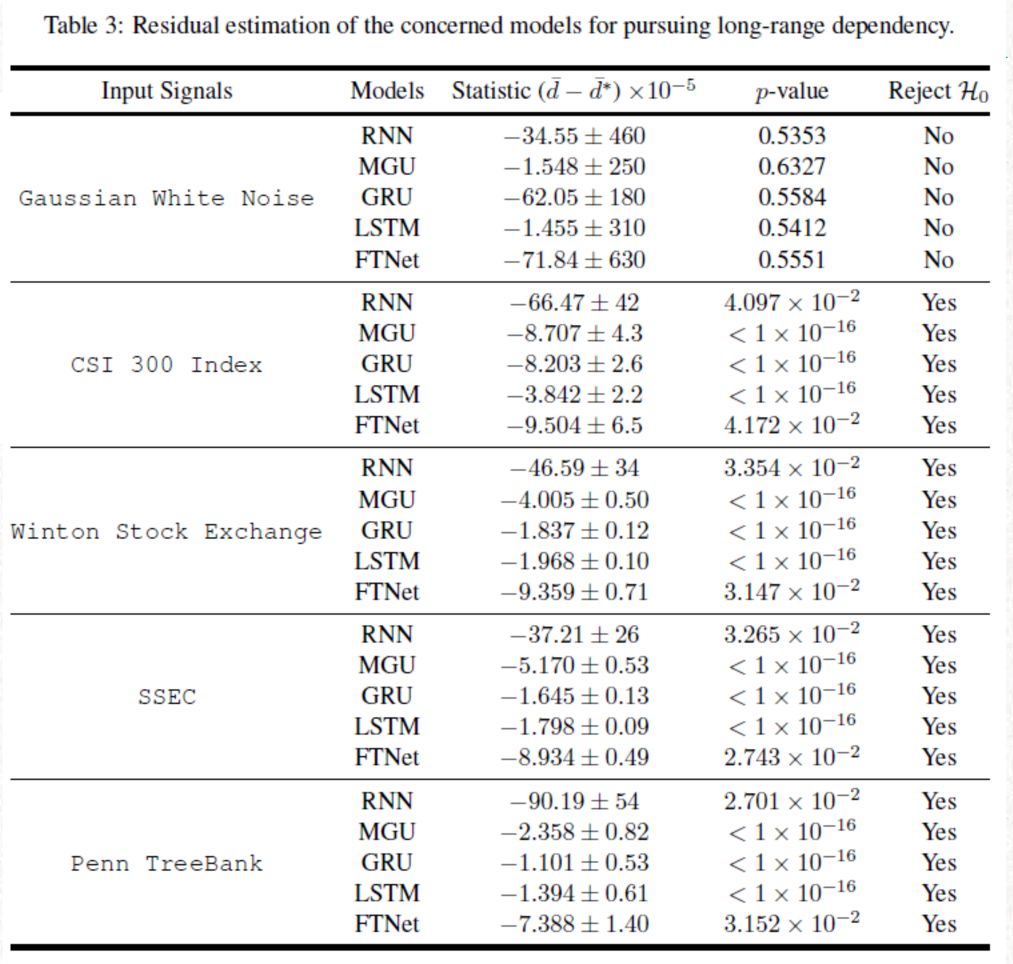
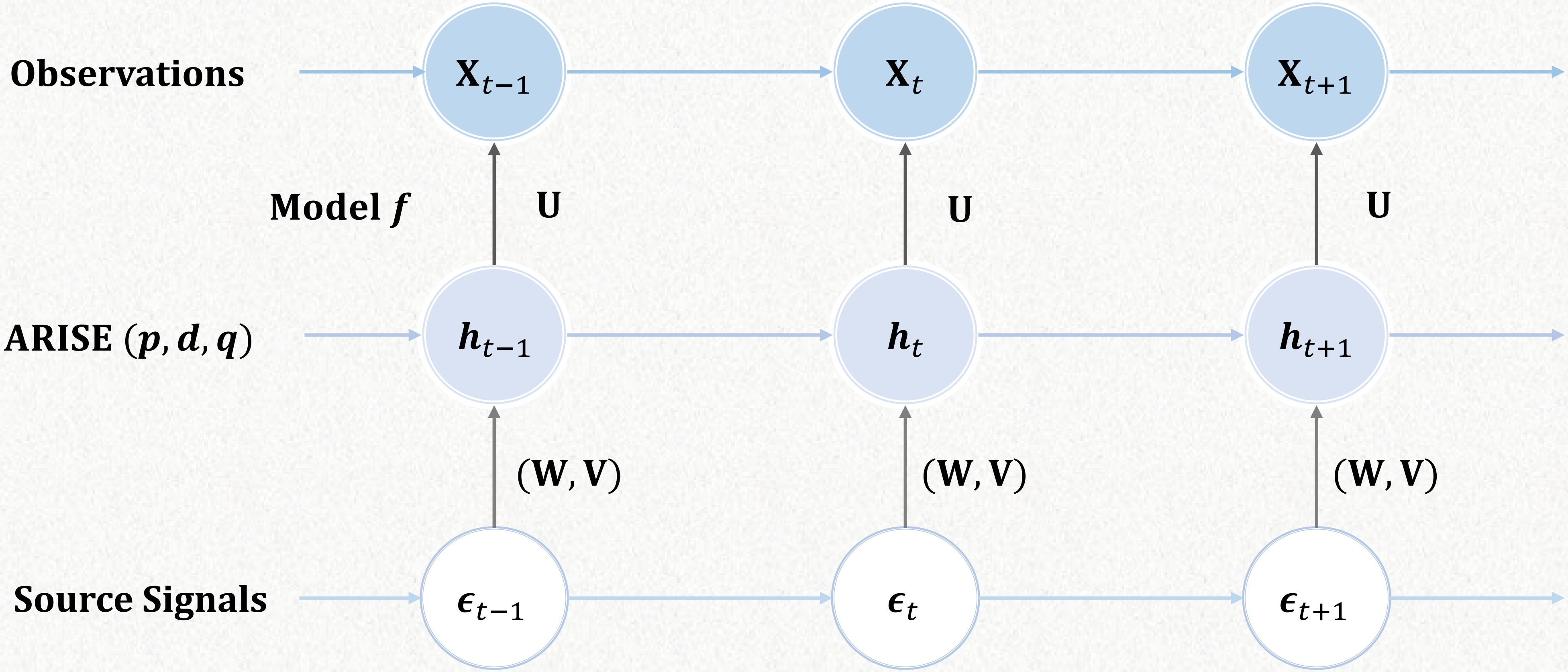


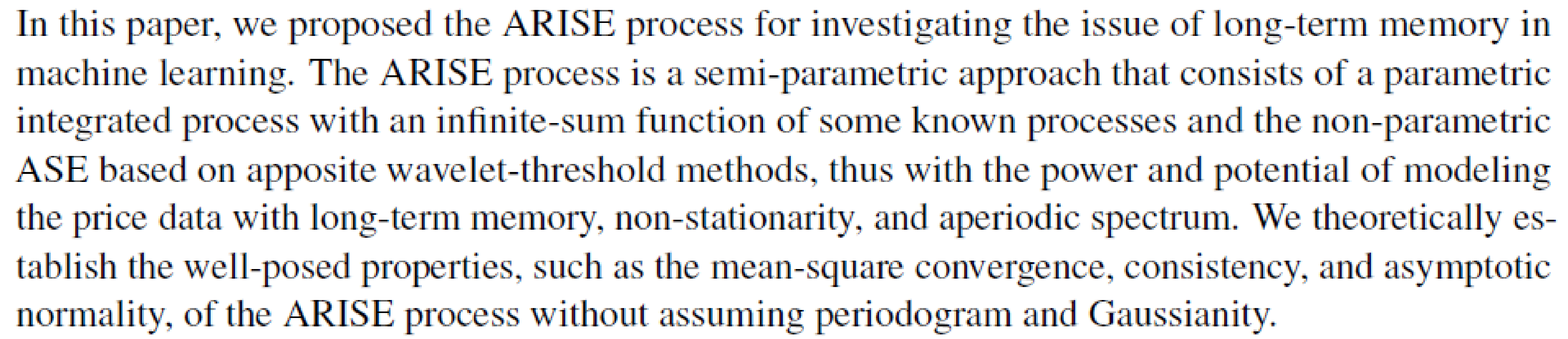
Table 3: Residual estimation of the concerned models for pursuing long-range dependency.

Input Signals	Models	Statistic $(\bar{d} - \bar{d}^*) \times 10^{-5}$	p-value	Reject $\mathcal{H}_0$
Gaussian White Noise	RNN	-34.55 ± 460	0.5353	No
	MGU	-1.548 ± 250	0.6327	No
	GRU	-62.05 ± 180	0.5584	No
	LSTM	-1.455 ± 310	0.5412	No
	FTNet	-71.84 ± 630	0.5551	No
CSI 300 Index	RNN	-66.47 ± 42	4.097×10^{-2}	Yes
	MGU	-8.707 ± 4.3	$< 1 \times 10^{-16}$	Yes
	GRU	-8.203 ± 2.6	$< 1 \times 10^{-16}$	Yes
	LSTM	-3.842 ± 2.2	$< 1 \times 10^{-16}$	Yes
	FTNet	-9.504 ± 6.5	4.172×10^{-2}	Yes
Winton Stock Exchange	RNN	-46.59 ± 34	3.354×10^{-2}	Yes
	MGU	-4.005 ± 0.50	$< 1 \times 10^{-16}$	Yes
	GRU	-1.837 ± 0.12	$< 1 \times 10^{-16}$	Yes
	LSTM	-1.968 ± 0.10	$< 1 \times 10^{-16}$	Yes
	FTNet	-9.359 ± 0.71	3.147×10^{-2}	Yes
SSEC	RNN	-37.21 ± 26	3.265×10^{-2}	Yes
	MGU	-5.170 ± 0.53	$< 1 \times 10^{-16}$	Yes
	GRU	-1.645 ± 0.13	$< 1 \times 10^{-16}$	Yes
	LSTM	-1.798 ± 0.09	$< 1 \times 10^{-16}$	Yes
	FTNet	-8.934 ± 0.49	2.743×10^{-2}	Yes
Penn TreeBank	RNN	-90.19 ± 54	2.701×10^{-2}	Yes
	MGU	-2.358 ± 0.82	$< 1 \times 10^{-16}$	Yes
	GRU	-1.101 ± 0.53	$< 1 \times 10^{-16}$	Yes
	LSTM	-1.394 ± 0.61	$< 1 \times 10^{-16}$	Yes
	FTNet	-7.388 ± 1.40	3.152×10^{-2}	Yes

Latent State-Space Model for Inference and Forecasting



Summary



In this paper, we proposed the ARISE process for investigating the issue of long-term memory in machine learning. The ARISE process is a semi-parametric approach that consists of a parametric integrated process with an infinite-sum function of some known processes and the non-parametric ASE based on apposite wavelet-threshold methods, thus with the power and potential of modeling the price data with long-term memory, non-stationarity, and aperiodic spectrum. We theoretically establish the well-posed properties, such as the mean-square convergence, consistency, and asymptotic normality, of the ARISE process without assuming periodogram and Gaussianity.

Something not mentioned:

- Proof for Theorems 1 and 2, corresponding Monte-Carlo study
- Developed model for inference and forecasting
- Retrieve some physical systems, such as the Lorenz attractor
- Hyper-parameters, training methods, and techniques
- Computational Complexity

Thank you !

Q & A

Shao-Qun Zhang and Zhi-Hua Zhou. ARISE: ApeRIodic SEmi-parametric Process for Efficient Markets without Periodogram and Gaussianity Assumptions. 2021. [arXiv:2111.06222]