Why Are the ARIMA and SARIMA not Sufficient

Shixiong Wang*†, Chongshou Li, and Andrew Lim
Department of Industrial Systems Engineering and Management,
National University of Singapore, Singapore 117576

December 20, 2019

Abstract

The autoregressive moving average (ARMA) model takes the significant position in time series analysis for a wide-sense stationary time series. Facing the trend and seasonal component of a time series, the difference operator and seasonal difference operator, which are bases of ARIMA and SARIMA, respectively, are introduced to remove the trend and seasonal component so that the original non-stationary time series could be transformed into a wide-sense stationary time series, which could then be handled by Box-Jenkins methodology (ARMA). However, such difference operators are more practical experiences than exact theories by now. In this paper, we investigate the power of the (seasonal) difference operator from the perspective of spectral analysis, linear system theory and digital filtering, and point out the characteristics and limitations of (seasonal) difference operator. Besides, the general method that transforms a non-stationary (the non-stationarity in the mean sense) stochastic process to be wide-sense stationary will be presented.

Keywords: Time Series Analysis, Difference Operator, Spectral Analysis, Digital Filtering, Linear System

*Corresponding Author. Email: s.wang@u.nus.edu
†This work is supported by National Research Foundation of Singapore grant NRF-RSS2016-004 and Ministry of Education Academic Research Fund Tier 1 grants R-266-000-096-133, R-266-000-096-731, R-266-000-100-646 and R-266-000-119-133.

1 Glossaries

In order not to confuse readers from different communities, we mention some basic glossaries of interest in this paper.

- ARMA: Autoregressive Moving Average;
- ARIMA: Autoregressive Integrated Moving Average;
- SARIMA: Seasonal Autoregressive Integrated Moving Average;
- Spectrum: The spectrum of a time function is its Fourier transform which describes a time series in Fourier frequency domain;
- Spectral Analysis: The analysis and processing for a time series in Fourier frequency domain. The Fourier transform and inverse Fourier transform connect the time domain and Fourier frequency domain;
- (Discrete) Linear System: A discrete linear operator, defined by a difference equation (see (1)) in autoregressive moving average form, that transforms a time series into another time series (for example, the difference operator);
- Digital Filter: A linear system that takes effect of transforming a time series in Fourier frequency domain. However, it works for a time series in time domain. For example, the moving average method, although it is directly applied in time domain, in essence, it removes the low-frequency component (i.e., the trend) in Fourier frequency domain;
- SADFA: Spectral Analysis and Digital Filtering Approach used to analyze and process a time series;
- DFT/DFS: Discrete Fourier Transform/Series;
- FFT: Fast Fourier Transform;
- DTFT: Discrete Time Fourier Transform;

2 Introduction

One of the intriguing topics of time series analysis is to physically analyse the internal mechanism/dynamics of a system generating the focused time series, and subsequently build a proper mathematical model to describe the dynamics of this system so that we can predict the future with satisfying accuracy. Generally such a time series is a stochastic process rather than a deterministic one which makes the problem more complex.

When it comes to the stochastic process modeling, the reputed Box-Jenkins methodology (Box et al., 2015), also known as ARMA and ARIMA model, stands out. The philosophy of the ARMA model is from the Wold's Decomposition theorem (Papoulis and Pillai, 2002). The theorem supports that the ARMA model is mathematically sufficient to describe a wide-sense stationary stochastic process. After the modeling, the least square method, maximum likelihood method and spectral estimation method et al. could be utilised to estimate the parameters of the model based on the collected time series samples. As a result, we could make use of the past information (collected samples) to reconstruct the underlying dynamics of the focused stochastic process, and further make satisfying prediction. As complements to ARMA, the ARIMA(/SARIMA) aims to transform the focused non-stationary (in the mean sense) stochastic process to be stationary by difference (/seasonal difference) operator with proper order so that they can follow the philosophy of the ARMA model. For notation breifness, we collectively refer to ARIMA and SARIMA as S-ARIMA.

However, such difference operators are more empirical experiences than exact philosophies. Therefore, we do not know why they work well somewhere and ineffectively elsewhere. To this end, in this paper, we aim to investigate the power of the (seasonal) difference operator from the perspective of spectral analysis, linear system theory and digital filtering, and point out the characteristics and limitations that (seasonal) difference operator and S-ARIMA hold. Besides, the general operator that works for transforming a non-stationary (in the mean sense) stochastic process to be wide-sense stationary will be presented. At last, we will show the overall methodology for predicting a non-stationary stochastic process, which is the generalization of S-ARIMA and termed as ARMA-SIN.

As by-products derived from our methodology, we investigate the nature, philosophy, effectiveness, and insufficiency of the time series smoothing methods like exponential smoothing and moving average method (Hyndman and Athanasopoulos, 2018) by showing they are special cases of ARMA-SIN. We show that the traditional exponential smoothing and moving average are problematic due to their innate time-delay property.

In Section 3, we firstly give some examples to show the insufficiencies of the S-ARIMA and the advantages of ARMA-SIN over S-ARIMA, just as intuitive understandings for readers. In Section 4, we will explain the mathematical reason why S-ARIMA makes sense and point out its theoretical insufficiency. Finally, in Section 5, the general ARMA-SIN methodology will be presented, and in Section 6 we will analytically explain and derive the methods we presented in the warming-up Section 3. As a supplement and closing note, we in Appendix A explain the nature of the canonical time series smoothing methods (exponential smoothing and moving average method) from the perspective of the SADFA, and figure out their advantages and shortages.

Our discusses will be based on Linear System Theory, Spectral Analysis, and Digital Filtering. For more on those topics, please see Appendix B and Diniz et al. (2010); Yang (2009); Chaparro and Akan (2018).

Remark 1 *When we mention the insufficiency of S-ARIMA, we actually mean its theoretical insufficiency instead of the prediction accuracy in some particular cases. This is because the exact model only outperforms other models when the problem is exact. For example, if the data is generated from a linear function with sufficiently small Gaussian white noise, then the linear regression model should be better than any other high-order polynomial regression models. However, we cannot assert the linear model is best all the time. Note that the focused time series, namely the exact problem we study in this paper, is a wide-sense stationary stochastic process. Thus, the ARMA model, according to Wold's Decomposition theorem (Papoulis and Pillai, 2002), is the corresponding exact model, meaning the operator that makes the original time series exact as wide-sense stationary is better, because the ultimate issue is to train an ARMA model. It is in this sense that we assert the S-ARIMA model is insufficient.*

3 Scenarios of Warming-up

As intuitive understandings for readers, we in this section provide some simulation scenarios to illustrate the theoretical insufficiency that S-ARIMA holds. Together with, the counterpart solutions given by ARMA-SIN will be also demonstrated. Following this warming-up, in the subsequent sections, we will progressively detail the motivations, philosophies, mathematics and methodologies of generating such useful ARMA-SIN solutions.

3.1 Notations

Before we start, we should define some useful notations here.

1. Let $\boldsymbol{v} = a : l : b$ define a vector $\boldsymbol{v}$ having the lower bound a, upper bound b and step length l. For example, $\boldsymbol{v} = 0 : 0.1 : 0.5$ means $a = 0$, $b = 0.5$, and $l = 0.1$. Thus $\boldsymbol{v} = [0, 0.1, 0.2, 0.3, 0.4, 0.5]^T$;

2. Let the function $length(\boldsymbol{x})$ return the length of the vector $\boldsymbol{x}$. For example, if $\boldsymbol{x} = [1, 2, 3]^T$, we have $length(\boldsymbol{x}) = 3$;

3. Let $\boldsymbol{t}$ denote the continuous time variable, and $\boldsymbol{n}$ its corresponding discrete time variable. For example, if $\boldsymbol{t} = 0 : 0.5 : 100$ (the time span is $100s$, and the sampling time is $T_s = 0.5s$), we will have $\boldsymbol{n} = \boldsymbol{t}/T_s = 0 : 1 : length(t) - 1 = 0 : 1 : 200$; let $N = length(\boldsymbol{n})$;

4. Let $\boldsymbol{x}$ denote a time series of interest;

5. Let $randn(N)$ return a Gaussian white series (mean is zero, variance is one) with length of N;

6. Let $ARMA(p, q|\boldsymbol{\varphi}, \boldsymbol{\theta})$ define an ARMA process with autoregressive order of p and moving average order of q. Besides, the coefficient vectors $\boldsymbol{\varphi}$ and $\boldsymbol{\theta}$ are for autoregressive part and moving average part, respectively;

7. Let the operator $\boldsymbol{y} = H(\boldsymbol{x}|\boldsymbol{a}, \boldsymbol{b})$ define a difference equation as follows

$$a_0 y_k + a_1 y_{k-1} + a_2 y_{k-2} + ... + a_p y_{k-p} = b_0 x_k + b_1 x_{k-1} + b_2 x_{k-2} + ... + b_q x_{k-q}, \quad (1)$$

   where $\boldsymbol{a}$, $\boldsymbol{b}$ are vectors and $length(\boldsymbol{a}) = p + 1$, $length(\boldsymbol{b}) = q + 1$.

3.2 ARMA Series of Ground Truth

We generate a ARMA series for analysis later. Without loss of generality, we arbitrarily set

$$\begin{aligned} \boldsymbol{\theta} &= [13, 5, 6]^T \\ \boldsymbol{\varphi} &= [40, 2, 3, 6, 9]^T, \end{aligned} \tag{2}$$

meaning the time series generated from a Gaussian white series ϵ_k is given as

$$40x_k^0 + 2x_{k-1}^0 + 3x_{k-2}^0 + 6x_{k-3}^0 + 9x_{k-4}^0 = 13\epsilon_k + 5\epsilon_{k-1} + 6\epsilon_{k-2}, \tag{3}$$

where k denotes the discrete time index (namely $k \in \boldsymbol{n}$) and $x_k^0 := 0, \epsilon_k := 0$ if $k < 0$. However, this discrete linear system is guaranteed to be minimum-phase stable (Diniz et al., 2010). That is, it is stable and inversely stable.

Remark 2 *If the focused raw time series is as $\boldsymbol{x} = f(\boldsymbol{x}^0)$ (for example $\boldsymbol{x} = \boldsymbol{x}^0 + \boldsymbol{n}$, linear trend), the operator that exactly transforms $\boldsymbol{x}$ to its wide-sense stationary counterpart $\boldsymbol{x}^0$ should be the best. This is because we finally aim to use the ARMA model to fit the transformed series. Let $\hat{\boldsymbol{x}}^0$ be the transformed series from $\boldsymbol{x}$. The nearer between $\hat{\boldsymbol{x}}^0$ and $\boldsymbol{x}^0$, the better. For more, see Remark 1.*

3.3 The Case of Variant Mean

In this subsection, we investigate a scenario being with variant mean, implying it is non-stationary in the mean sense. Let $\boldsymbol{t} = 0 : 0.5 : 100$ ($T_s = 0.5$), $\boldsymbol{n} = \boldsymbol{t}/T_s$, and $\boldsymbol{x} = \boldsymbol{x}^0 + 0.1\boldsymbol{t}$. It means the trend component is a linear function.

If we follow the standard modelling procedure with Box-Jenkins (ARIMA) method (Hyndman and Athanasopoulos, 2018; Calheiros et al., 2015; Box et al., 2015; Hamilton, 1995), we have the estimated ARIMA(4,1,2) model to handle this problem, meaning the operator used is first order differencing. Instead if we use ARMA-SIN method, we have

$$\hat{\boldsymbol{x}}^0 = H(\boldsymbol{x}|\boldsymbol{a}, \boldsymbol{b}), \tag{4}$$

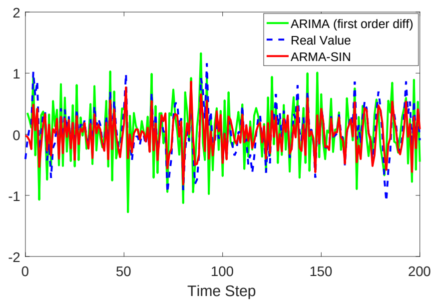
where

$$\boldsymbol{a} = [1, -1.7101, 1.3712, -0.3152]^T,$$

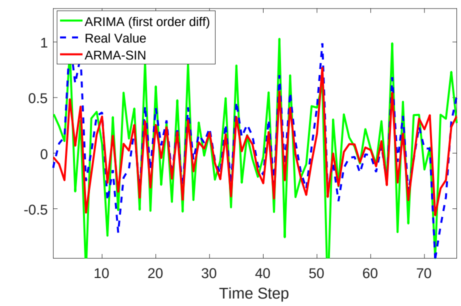
and

$$\boldsymbol{b} = [0.6226, -1.5757, 1.5757, -0.6226]^T,$$

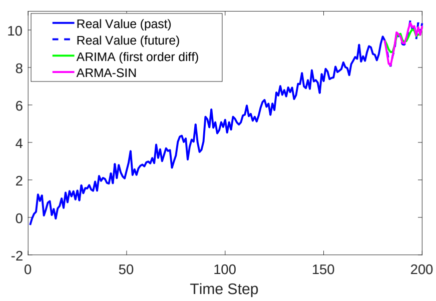
and ARMA part for approximating $\boldsymbol{x}^0$ as ARMA(4,1). The results of two methods are showed in Figure 1.



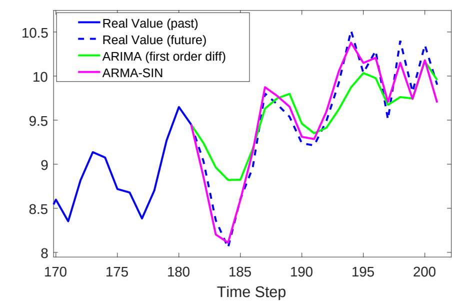
(a) Transformed series



(b) Transformed series (locally enlarged)



(c) Prediction results



(d) Prediction results (locally enlarged)

Figure 1: Transformed series and prediction results for variant mean

Besides, we have 100 times of monte carlo simulation and the averaged prediction MSE is given in Table 1. Clearly, the transformed series $\hat{\boldsymbol{x}}^0$ of ARMA-SIN is nearer to its ground truth $\boldsymbol{x}^0$ than that of ARIMA. Thus the prediction accuracy is more satisfactory.

Table 1: Averaged prediction MSE of ARIMA and ARMA-SIN for variant mean

	ARIMA	ARMA-SIN
MSE	0.0977	0.0203

3.4 The Case of Using SARIMA

In this subsection, we investigate a scenario that is suitable for using SARIMA. Let $\boldsymbol{t} = 0 : 0.1 : 50$ $(T_s = 0.1)$, $\boldsymbol{n} = \boldsymbol{t}/T_s$, and $\boldsymbol{x} = \boldsymbol{x}^0 + sin(5\boldsymbol{t})$. It means the trend component is a sine function.

If we follow the standard modelling procedure with Box-Jenkins (ARIMA) method, we have the estimated SARIMA(4,1,1)(12,1,0,0) model to handle this problem, meaning the operator used is 12-lag seasonal differencing. Instead if we use ARMA-SIN method, we have

$$\hat{\boldsymbol{x}}^0 = H(\boldsymbol{x}|\boldsymbol{a}, \boldsymbol{b}), \tag{5}$$

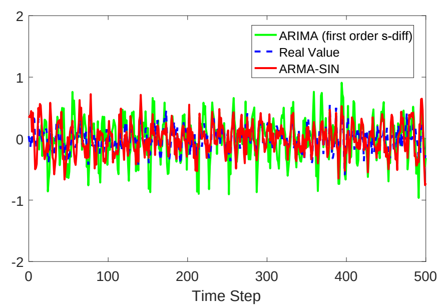
where

$$\boldsymbol{a} = [1, -5.1801, 11.8864, -15.30778, 11.6563, -4.9815, 0.9430]^T,$$

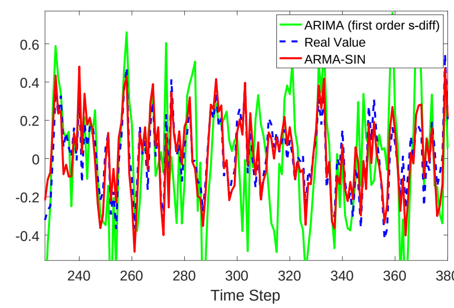
and

$$\boldsymbol{b} = [0.9713, -5.0804, 11.7716, -15.3086, 11.7716, -5.0804, 0.9713]^T,$$

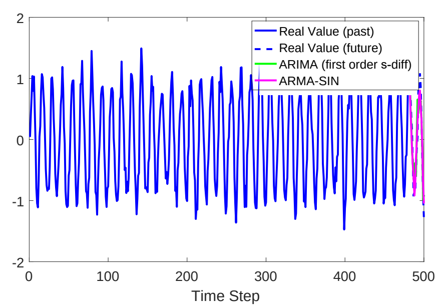
and ARMA part for approximating $\boldsymbol{x}^0$ as ARMA(4,1). The results of two methods are showed in Figure 2.



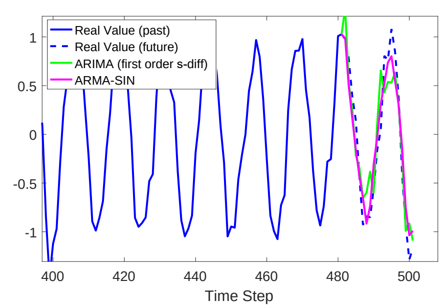
(a) Transformed series



(b) Transformed series (locally enlarged)



(c) Prediction results



(d) Prediction results (locally enlarged)

Figure 2: Transformed series and prediction results of SARIMA and ARMA-SIN (s-diff: seasonal difference)

Besides, we have 100 times of monte carlo simulation and the averaged prediction MSE is given in Table 2. Clearly, the transformed series $\hat{\boldsymbol{x}}^0$ of ARMA-SIN is nearer to its ground

Table 2: Averaged prediction MSE of SARIMA and ARMA-SIN

	ARIMA	ARMA-SIN
MSE	0.1022	0.0516

truth $\boldsymbol{x}^0$ than that of SARIMA. Thus the prediction accuracy is more satisfactory.

3.5 The Case of Directly Estimating the Seasonal Component

In this subsection, we investigate a scenario that we can directly figure out the seasonal Component. Let $\boldsymbol{t} = 0 : 0.1 : 10$ ($T_s = 0.1$), $\boldsymbol{n} = \boldsymbol{t}/T_s$, and $\boldsymbol{x} = \boldsymbol{x}^0 + f(\boldsymbol{t})$, where $f(\boldsymbol{t}) = sin(2\boldsymbol{t})$. Our purpose is to estimate out the function $sin(2t)$ directly from the collected history data.

If we use our ARMA-SIN method, we can know that $f(x)$ is with the form as

$$\begin{aligned} \hat{f}(t) &= 0.9721cos(0.2001n - 1.5440) \\ &= 0.9721cos(0.2001t/T_s - 1.5440) \\ &= 0.9721cos(2.001t - 0.4945\pi) \end{aligned} \tag{6}$$

It is amazingly close to its ground truth of $f(t) = sin(2t) = cos(2t - \pi/2) = cos(2t - 0.5\pi)$.

If we follow the standard modelling procedure with Box-Jenkins methodology, we have the estimated SARIMA(4,0,1)(31,0,0,0) model to handle this problem, meaning the operator used is 31-lag seasonal differencing.

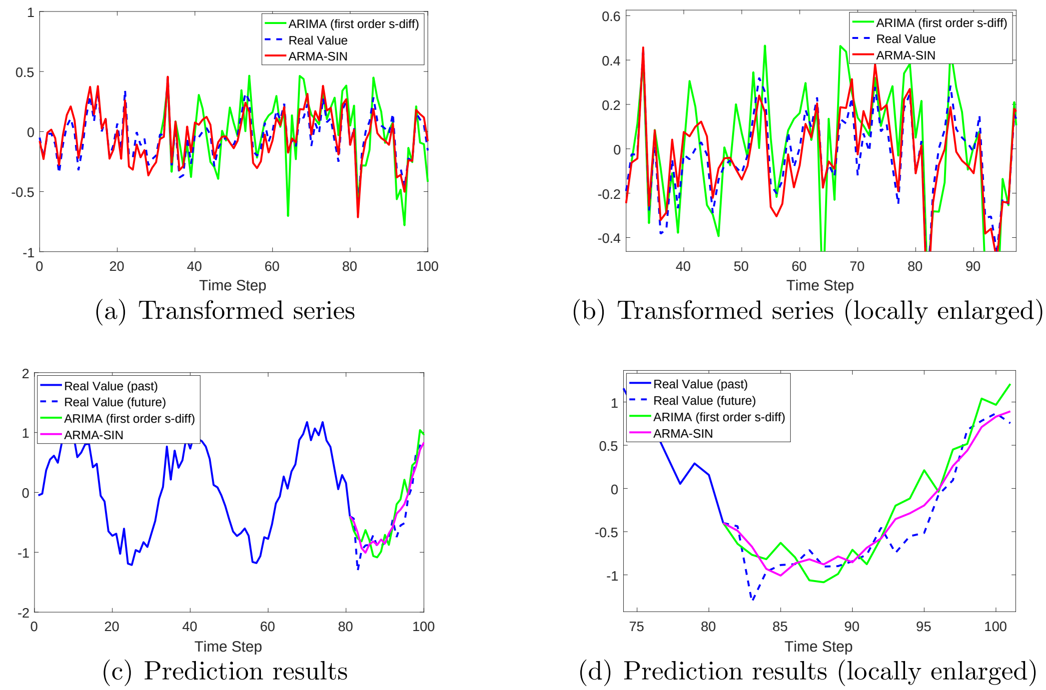
The results of two methods are showed in Figure 3.

Besides, we have 100 times of monte carlo simulation and the averaged prediction MSE is given in Table 3.

Table 3: Averaged prediction MSE of direct estimation

	SARIMA	ARMA-SIN
MSE	0.0844	0.0355

Clearly, the transformed series $\hat{\boldsymbol{x}}^0$ of ARMA-SIN is nearer to its ground truth $\boldsymbol{x}^0$ than that of SARIMA. Thus, the prediction accuracy is more satisfactory.



(a) Transformed series (b) Transformed series (locally enlarged)

(c) Prediction results (d) Prediction results (locally enlarged)

Figure 3: Transformed series and prediction results of direct estimation

Intuitively, we can see from the examples above that the ARMA-SIN method is indeed interesting over the S-ARIMA model. In the following sections, we will progressively explain the philosophy behind and detail the ARMA-SIN method.

4 Secret Behind the ARIMA and SARIMA

4.1 Nature of Difference Operator and Seasonal Difference Operator

As mentioned in the previous section, the ARIMA and SARIMA attempt to make stationary a stochastic process by difference operator and seasonal difference operator. In this section we will investigate the nature of ARIMA and SARIMA from the perspective of SADFA. Specifically, we should pay our attention to the nature of the difference operator and seasonal difference operator.

Theorem 1 *The nature of d-order difference operator is actually a high-pass digital filter that denies the low-frequency components of a time series. Since the trend of a time series* *is generally the low-frequency components (see Discrete Fourier Transform (Diniz et al., 2010)), the d-order difference operator makes sense to make stationary a non-stationary (in the mean sense) stochastic process.*

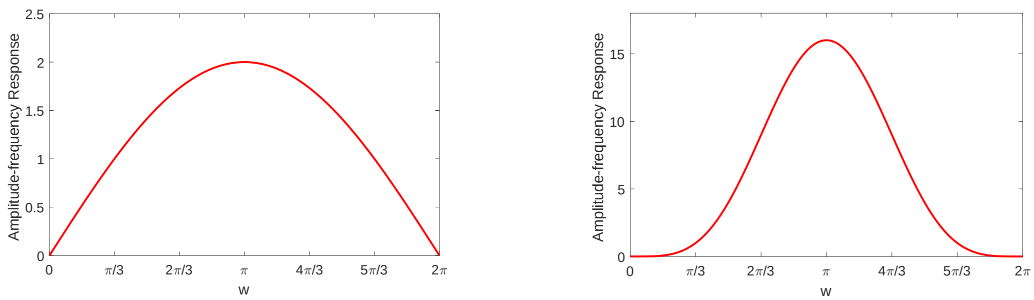
Proof 1 *The transfer function of the d-order difference operator is given as*

$$H(z) = (1 - z^{-1})^d, \tag{7}$$

and the amplitude-frequency response is as

$$|H(e^{jw})| = |(1 - e^{-jw})^d| = [\sqrt{2 - 2\cos(w)}]^d. \tag{8}$$

Equation (8) immediately admits the theorem, since in the interval $[0,\ \pi]$, $|H(e^{jw})|$ *is increasing from zero. Intuitively, see Figure 4.*



(a) 1-order difference operator (b) 4-order difference operator

Figure 4: Amplitude-frequency responses of difference operator

Theorem 2 *The nature of L-lag seasonal difference operator (L-SDO) is actually a comb digital filter that denies some frequency components with* $w = 2k\pi/L, k = 0, 1, 2, ..., L-1$. *Since the seasonal components (namely periodic components) with period L has period L in its spectra as well (meaning the frequencies of seasonal components are* $w = 2k\pi/L, k = 0, 1, 2, ..., L-1$, *see Discrete Fourier Series (Diniz et al., 2010)) the L-lag difference operator makes sense to make stationary a non-stationary (in the mean sense) stochastic process.*

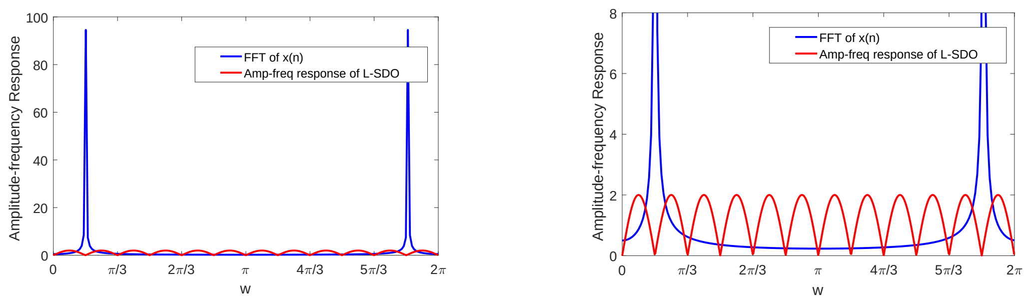
Proof 2 *The transfer function of the L-lag seasonal difference operator is given as*

$$H(z) = 1 - z^{-L}, \tag{9}$$

and the amplitude-frequency response is as

$$\begin{aligned} |H(e^{jw})| &= |1 - e^{-jLw}| \\ &= |[1 - \cos(Lw)] + j\sin(Lw)| \\ &= \sqrt{2 - 2\cos(Lw)}. \end{aligned} \quad (10)$$

Equation (10) immediately admits the theorem, since in the interval $[0,\ \pi]$, $|H(e^{jw})|$ *are zero-valued at* $w = 2k\pi/L, k = 0, 1, 2, 3, ..., L-1$. *Intuitively, we consider a sine series* $x(n)$ *with period* $L = 12$ *(*$t = 0 : 2\pi/L : 100$*) and take* 12*-lag seasonal difference over it. The results are illustrated in Figure 5.*



(a) Full picture (b) Locally enlarged

Figure 5: FFT of $x(n)$ and Amplitude-frequency responses of L-lag seasonal difference operator

Remark 3 *Recall that in discrete Fourier frequency domain, every* $|H(e^{jw})|$ *is periodic over* w *and the period is* 2π. *Since the highest frequency is indicated by* $w = \pi$ *and lowest is by* $w = 0$ *and/or* $w = 2\pi$, *we should only pay attention on the interval* $[0,\ 2\pi]$. *The trend of a time series is usually in low-frequency interval (around* $w = 0$*) with large values (see Figure 6 (a),* w *below* $\pi/6$*). Besides, according to Discrete Fourier Series (Diniz et al., 2010), the seasonal component has impulses uniformly distributed (uniformly spaced) in the all frequency domain* $w \in [0,\ \pi]$ *(see Figure 7 (a), around* $w = \pi/6$*). Therefore, in frequency domain, we expect to remove all these outliers so that the transformed time series would become stationary.*

Remark 4 *Note that if the seasonal component of a time series is perfectly a sine function, it only has one impulse in its spectra because the other impulses are zero-valued. For more,* *see Discrete Fourier Series (Diniz et al., 2010). This is why Figure 7 (a) only has one impulse other than many.*

4.2 Why Are the ARIMA and SARIMA Not Sufficient

By now, Theorem 1 and Theorem 2 support the effectiveness of S-ARIMA somewhere. However, it is obvious that except the unwanted frequencies, d-order difference operator of ARIMA and L-lag seasonal difference operator of SARIMA also negatively impact the desired frequencies which should remain unchanged, for example, the frequencies higher than $w = \pi/3$ in Figure 4 have been significantly and unwantedly amplified, and the frequencies at and around $w = 2k\pi/12, k = 0, 2, 3, ...11,$(that is , $k \neq 1$) in Figure 5 have also been unwantedly wiped away. This raises the insufficiency of S-ARIMA. Thus we have Theorem 3.

Theorem 3 *The ARIMA and SARIMA model are theoretically insufficient since they also negatively impact (distort) the desired frequency components (points or intervals).*

Proof 3 *Due to Theorem 1, Theorem 2 and the statement itself in this theorem, the conclusion stands.*

Remark 5 *Although the ARIMA and SARIMA are theoretically insufficient, they are easiest methods to handle the non-stationary problems in practice. It means instead of designing some more proper (but more complicated) operators to make stationary a time series, we can for simplicity choose (seasonal) difference operator, if the performances are satisfactory for some specific problems in engineering. It is simplicity of such difference operators that makes them popular in engineering over years.*

5 The General ARMA-SIN Model

Our ARMA-SIN is based on SADFA (spectral analysis and digital filtering approach) and Box-Jenkins methodology. In detail, when we have a general time series, we first use SADFA to transform it to be wide-sense stationary and then use Box-Jenkins methodology to model the transformed series. Thus, SIN is for SADFA, while ARMA is for Box-Jenkins methodology.

In Algorithm 1, we detail the methodology of ARMA-SIN.

Algorithm 1 ARMA-SIN

1: **Spectral Analysis:** transform the focused time series into Fourier frequency domain and identify the impulses frequency points (or interval) in frequency domain. Usually, points for the seasonal components and interval (low-frequency region around $w = 0$) for the trend;

2: **Digital Filtering:** design a proper digital filter with proper cut-off frequencies such that the focused time series could be made stationary. Ideally, such proper digital filter should only remove the unwanted frequency points or interval while keeping the rest unchanged;

3: **ARMA:** follow the standard Box-Jenkins methodology to model the stationary remainder (transformed time series) as a ARMA one;

4: **Forecast:** predict the future with two independent process, ARMA part and the trend (and/or seasonal) part, respectively, and then integrate the results together. Note that the raw time series subtract the ARMA remainder gives the trend (or seasonal) part.

Remark 6 *As we can see, the ARIMA and SARIMA are special cases of Algorithm 1. For ARIMA, the digital filter we use is just difference operator (special case of high-frequency-pass filter, that is, low-frequency-stop) to remove (stop) the low-frequency trend part. For SARIMA, the digital filter we use is just seasonal difference operator (special case of comb filter, that is, fixed-point-frequency-stop filter) to remove (stop) the frequency impulses in frequency domain.*

6 The Derivation for Solutions of Scenarios of Warming-up

In this section, we will detail the derivation for solutions of scenarios of warming-up given in Section 3. For briefness, we will not repeat the standard modelling procedures of Box-

Jenkins methodology (Box et al., 2015). All the filters this paper designed are based on IIR (infinite impulse response) model and Elliptic method (Diniz et al., 2010).

Remark 7 *Usually it is hard to decide the best parameters like cut-off frequencies at the first glance for a general problem. However, we could try some times with plausible parameters and choose the best among them. This seems tedious but unavoidable in practice because there is no free lunch, advanced method means complex coding, complex parameters selecting and/or complex calculation burden. In this sense, the use of (seasonal) difference operator seems really delightful for all of us. Nevertheless, the trade-off is the unsatisfactory performances and the powerlessness somewhere.*

6.1 The Case of Variant Mean

The time series is showed in Figure 1 (c). The trend is linear. According to Discrete Fourier Transform theory, the linear trend could be represented by the finite sum of sine functions with acceptable approximation error. In order to separate (extract) the linear trend, we should design a high-pass filter which denies the low-frequency component (trend) to pass and allows the high-frequency components to go through without any changes. Since we use the Elliptic method to design a IIR filter (the function *ellipord* in MATLAB, for more, see its reference page in MathWorks (2018b)), we set the parameters in Table 4.

Table 4: The filter parameters for the case of variant mean

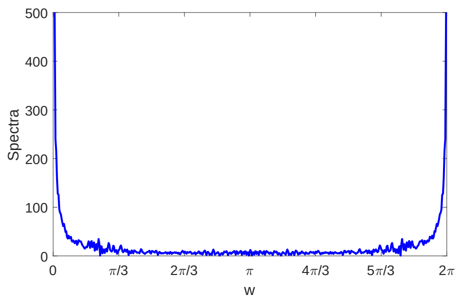
	Wp	Ws	Rp	Rs
Value	0.25	0.2	1	10

The specific meanings and detailed usage of parameters in Table 4 should be found in Diniz et al. (2010) or the reference page of the function *ellipord* in MathWorks (2018b).

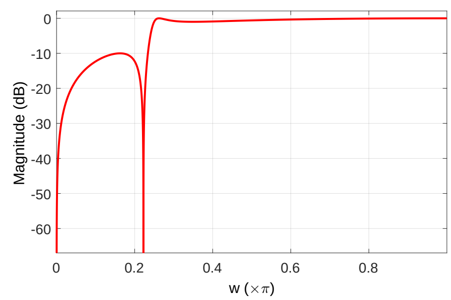
This high-pass filter actually defines the system (4).

The spectral analysis results are given in Figure 6.

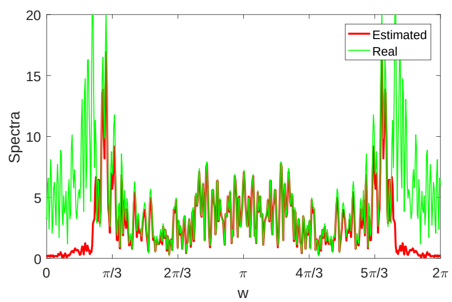
From Figure 6, we can see the 1-order difference operator significantly distorts (amplifies) the spectra in high frequency interval, although it is powerful to detrend. However, our ARMA-SIN keeps 1 ($20 \log(1) = 0dB$) in higher frequency area, making no changes to



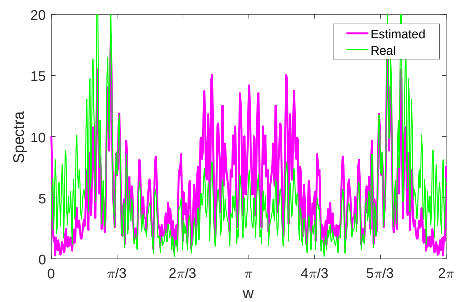
(a) Spectrum of $x(n)$



(b) Digital filter to detrend the $x(n)$



(c) Estimated spectra of x^0 using ARMA-SIN



(d) Estimated spectra of x^0 using 1-order difference

Figure 6: Spectra analysis for the case of variant mean

desired components. This is why our ARMA-SIN outperforms ARIMA. Note that Figure 6 (b) is given in logarithmic unit of decibel (dB). For more on decibel, see Diniz et al. (2010).

6.2 The Case of Using SARIMA

This case is concerned with the periodic time series which is suitable for SARIMA. The time series is showed in Figure 2 (c). In Figure 7, we can check the spectra of the time series. We can see that there is an outstanding line-spectra (impulse, outlier) around $w = 0.5$. Thus we want to design a band-stop digital filter to deny the outstanding frequency component. Since we use the Elliptic method to design a IIR filter (the function *fdesign.bandstop* in MATLAB, for more, see its reference page in MathWorks (2018a)), we set the parameters in Table 5.

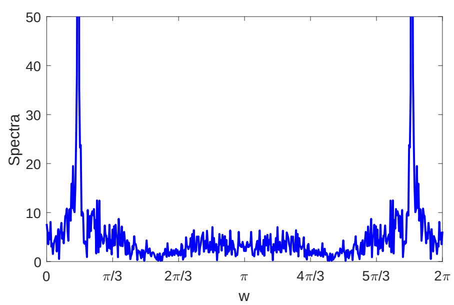
This band-stop filter actually defines the system (5).

The spectral analysis results are given in Figure 7.

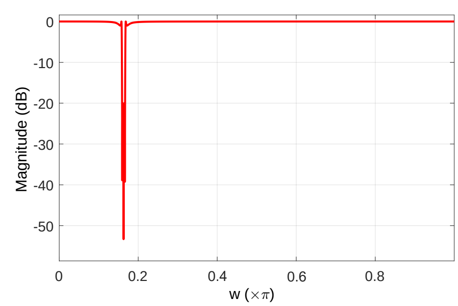
The spectra of the 12-lag seasonal difference operator could be found in Figure 5 as a

Table 5: The filter parameters for the case of using SARIMA

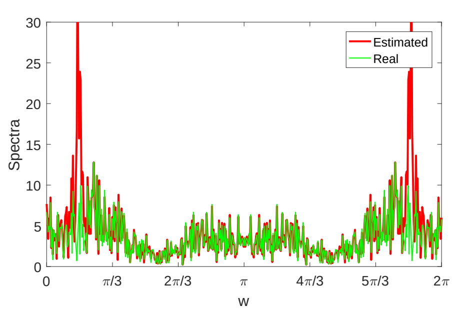
	Value	Comments
Fpass1	0.158	First Passband Frequency ($\times\pi$)
Fstop1	0.16	First Stopband Frequency ($\times\pi$)
Fstop2	0.165	Second Stopband Frequency ($\times\pi$)
Fpass2	0.168	Second Passband Frequency ($\times\pi$)
Apass1	1	First Passband Ripple (dB)
Astop	20	Stopband Attenuation (dB)
Apass2	1	Second Passband Ripple (dB)



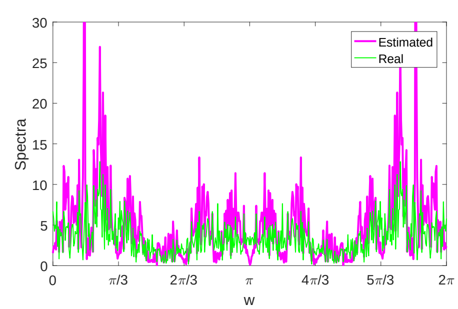
(a) Spectrum of $x(n)$



(b) Digital filter to detrend the $x(n)$



(c) Estimated spectra of x^0 using ARMA-SIN



(d) Estimated spectra of x^0 using 12-lag seasonal difference

Figure 7: Spectra analysis for the case of using SARIMA

comparison to Figure 7.

From Figure 7, we can see the 12-lag seasonal difference operator significantly distorts (amplifies) the spectra at and around $w = 2k\pi/12, k = 0, 2, 3, ..., 11$, although it is powerful to wipe out the frequency at $w = 2\pi/12$ ($k = 1$). This is why our ARMA-SIN outperforms ARIMA.

6.3 The Case of Directly Estimating the Seasonal Component

In this section we will use the information presented in spectra of a time series to directly estimate the seasonal (impulses in frequency domain) component. The philosophy is based on DFT (discrete Fourier transform) and Inverse DFT (Diniz et al., 2010; Yang, 2009; Chaparro and Akan, 2018).

Suppose the outstanding value of FFT of $x(n)$ is at $n = K$ ($K \in \{1, 2, 3, ..., length(\boldsymbol{n})\}$) and its value is H_K, then we have $w = 2\pi K/N$. Besides, by DFT (and/or FFT), the amplitude in time domain is given as $A = 2|H_K|/N$ and phase is as $\varphi = \varphi(H_k)$, where $|H_K|$ denotes the modulus of H_K and $\varphi(H_k)$ denotes the argument (angle) of H_K. Note that H_K is complex-valued. Therefore the line-spectra (impulse in spectra) component of $x(n)$ with frequency of $w = 2\pi K/N$ has its expression in time domain as

$$x_p(n) = A\cos(wn + \varphi). \tag{11}$$

Specifically in this section, the spectra of the interested time series is as Figure 8.

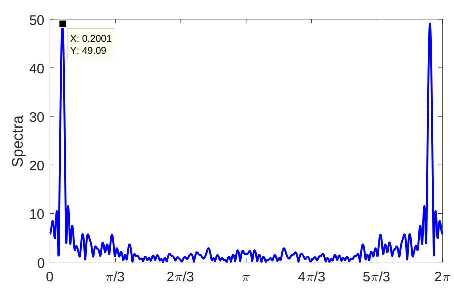


Figure 8: Spectra analysis for the case of direct estimation

From the figure we can directly know that $w = 0.2001$, and $A = 2 \times 49.09/N = 0.9721$ ($N = 101$). Besides, by checking the corresponding phase at this frequency $w = 0.2001$, we have $\varphi = -1.5440$. Thus (11) admits (6).

7 Conclusions

We have in this paper discussed the natures, philosophies, effectiveness, insufficiencies and improvements of ARIMA and SARIMA from the perspectives of Linear System Analysis, Spectra Analysis and Digital Filtering. We show that S-ARIMA also distorts desired frequencies when making stationary a time series. Fortunately, our ARMA-SIN could remain the innocent frequencies unchanged. However, we should admit that although the displayed ARMA-SIN is powerful, general and interesting, it is relatively hard for beginners to make use of, compared to the simple exponential smoothing, moving average, ARIMA, SARIMA and the like. Applying this method requires enough experience to identify the spectral components of interest and setup the proper parameters like cut-off frequencies to extract or discard them by designing a proper digital filter. Although depressive mentioning this, we should make clear that it is still a bright future if we can design some powerful algorithms to help us automatically select the proper parameters to design the useful digital filters, just like the cross-validation method in training a LASSO model. However, this challenging work should be jointly handled with inspired scholars in this area.

References

Blanchard, B. S., Fabrycky, W. J., and Fabrycky, W. J. (1990). *Systems engineering and analysis*, volume 4. Prentice Hall Englewood Cliffs, NJ.

Bloomfield, P. (2004). *Fourier analysis of time series: an introduction*. John Wiley & Sons.

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Calheiros, R. N., Masoumi, E., Ranjan, R., and Buyya, R. (2015). Workload prediction using arima model and its impact on cloud applications qos. *IEEE Transactions on Cloud Computing*, 3(4):449–458.

Chaparro, L. and Akan, A. (2018). *Signals and Systems using MATLAB*. Academic Press.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2009). *Introduction to algorithms*. MIT press.

Diniz, P. S., Da Silva, E. A., and Netto, S. L. (2010). *Digital signal processing: system analysis and design*. Cambridge University Press.

Dorf, R. C. and Bishop, R. H. (2011). *Modern control systems*. Pearson.

Granger, C. W. J. and Hatanaka, M. (1964). *Spectral Analysis of Economic Time Series.(PSME-1)*. Princeton university press.

Hamilton, J. D. (1995). Time series analysis. *Economic Theory. II, Princeton University Press, USA*, pages 625–630.

Helfenstein, U. (1991). The use of transfer function models, intervention analysis and related time series methods in epidemiology. *International journal of epidemiology*, 20(3):808–815.

Hyndman, R., Koehler, A. B., Ord, J. K., and Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach*. Springer Science & Business Media.

Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

Koopmans, L. H. (1995). *The spectral analysis of time series*. Elsevier.

Kreyszig, E. (1978). *Introductory functional analysis with applications*, volume 1. wiley New York.

Liu, H., Zhu, L., Pan, Z., Bai, F., Liu, Y., Liu, Y., Patel, M., Farantatos, E., and Bhatt, N. (2017). Armax-based transfer function model identification using wide-area measurement for adaptive and coordinated damping control. *IEEE Transactions on Smart Grid*, 8(3):1105–1115.

Mani, R., Oppenheim, A. V., Willsky, A. S., and Nawab, S. H. (1997). *Solutions Manual [for] Signals and Systems, [by] Alan V. Oppenheim, Alan S. Willsky with S. Hamid Nawab*. Prentice Hall.

MathWorks (2018a). Bandstop filter specification object. `https://www.mathworks.com/help/dsp/ref/fdesign.bandstop.html`. Accessed on March 14, 2019.

MathWorks (2018b). Minimum order for elliptic filters. `https://www.mathworks.com/help/signal/ref/ellipord.html?searchHighlight=ellipord&s_tid=doc_srchtitle`. Accessed on March 14, 2019.

Mitra, S. K. and Kuo, Y. (2001). *Digital signal processing*. McGraw-Hill New York.

Papoulis, A. and Pillai, S. U. (2002). *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education.

Priestley, M. B. (1981). *Spectral analysis and time series*. Academic press.

Robbins, A. H. and Miller, W. C. (2012). *Circuit analysis: Theory and practice*. Cengage Learning.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423.

Von Bertalanffy, L. (1968). General system theory. *New York*, 41973(1968):40.

Wang, S., Li, C., and Lim, A. (2019). Optimal joint estimation and identification theorem to linear gaussian system with unknown inputs. *Signal Processing*, 161:268–288.

Wiener, N. (1965). *Cybernetics or Control and Communication in the Animal and the Machine*, volume 25. MIT press.

Yang, W. Y. (2009). *Signals and Systems with MATLAB*. Springer Science & Business Media.

Appendices

A Investigation on Exponential Smoothing and Moving Average

A.1 The Nature of Exponential Smoothing and Moving Average

In this section, we aim to discuss the nature of the canonical time series analysis methods like Exponential Smoothing method and Moving Average method (Hyndman et al., 2008; Hyndman and Athanasopoulos, 2018).

Let's firstly consider the exponential smoothing method.

Theorem 4 *The nature of the exponential smoothing (ETS) method is a low-pass filter, meaning it allows the low-frequency components of a time series to pass and denies the high-frequency components. Thus ETS can extract the trend which is the low-frequency components, and discard the high-frequency fluctuating components. Mathematically, the exponential smoothing is given as* $y_n = \alpha x_n + (1-\alpha) y_{n-1}$*, where* α *denotes the smoothing coefficient.*

Proof 4 *The transfer function of the system defined by exponential smoothing should be*

$$H(z) = \frac{\alpha}{1-(1-\alpha)z^{-1}}. \tag{12}$$

Its frequency response is given as

$$H(e^{jw}) = \frac{\alpha}{1-(1-\alpha)e^{-jw}}. \tag{13}$$

Thus the amplitude-frequency response is

$$|H(e^{jw})| = \frac{\alpha}{\sqrt{1+(1-\alpha)^2 - 2(1-\alpha)\cos(w)}}. \tag{14}$$

Clearly, in the interval $w \in [0,\ 2\pi]$*, the curve of (14) is convex and bell-like, and the valley value is reached at* $w = \pi$*. Note that* $|H(e^{jw})|$ *is periodic over* w *and the period is* 2π*.*

As an illustration and without loss of generality, we consider the case of $\alpha = 0.1$. We have

$$H(z) = \frac{0.1}{1 - 0.9z^{-1}}. \tag{15}$$

Its frequency response is given as

$$H(e^{jw}) = \frac{0.1}{1 - 0.9e^{-jw}}. \tag{16}$$

We in Figure 9 illustrate the amplitude-frequency response $|H(e^{jw})|$ of (16). Figure 9 indicates that the exponential smoother (ETS) is actually a low-pass digital filtering, meaning it allows the low frequency components to pass and denies the high frequency components. Since the trend of a time series is a low-frequency time function (meaning it changes slowly), the ETS system allows it to pass; while the fluctuating components of a time series is a high-frequency time function (meaning it changes quickly), the ETS attenuates it. Thus the output of the such ETS system is the trend component of a time series.

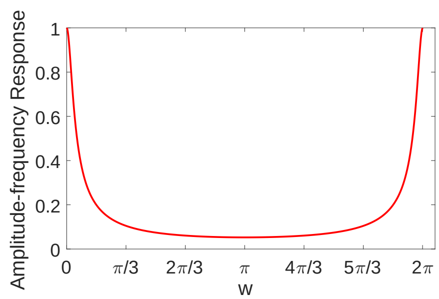
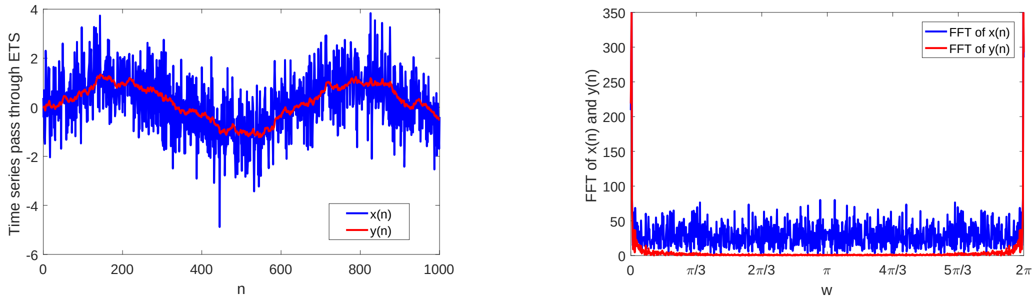


Figure 9: The amplitude-frequency response of exponential smoothing method

Together with Theorem 4, let's figure out what happens to a time series $x(n)$ passing the ETS. Consider $t = 0 : 0.01 : 10$ and $T_s = 0.01$, meaning $n = 0 : 1 : 1001$. Let $x(n) = \sin(nT_s) + randn(length(n))$. Then Figure 10 gives the original time series $x(n)$ and the generated time series $y(n)$ by the ETS, in both time domain and frequency domain.

Figure 10 explicitly upholds the rightness of Theorem 4, meaning all fluctuating (high-frequency) components have been denied to pass through, but the trend part (low-frequency) has been preserved.

Then let's investigate the nature of Moving Average.



(a) Time domain

(b) Frequency domain

Figure 10: Time series before and after the ETS

Theorem 5 *The nature of the moving average (MA) method is a low-pass filter, meaning it allows the low-frequency components of a time series to pass and denies the high-frequency components. Thus MA can extract the trend which is the low-frequency components, and discard the high-frequency fluctuating components. Mathematically, the moving average is given as*

$$y_n = \frac{1}{N}\sum_{i=0}^{N-1} x_{n-i}, \tag{17}$$

where N denotes the order of MA.

Proof 5 *The transfer function of the system defined by moving average should be*

$$H(z) = \frac{1}{N}\sum_{i=0}^{N-1} z^{-i} = \frac{1}{N}\frac{1-z^{-N}}{1-z^{-1}}. \tag{18}$$

Its frequency response is given as

$$H(e^{jw}) = \frac{1}{N}\frac{1-e^{-jwN}}{1-e^{-jw}}. \tag{19}$$

Thus the amplitude-frequency response is

$$|H(e^{jw})| = \frac{1}{N}\frac{\sin(wN/2)}{\sin(w/2)}. \tag{20}$$

Clearly, in the interval $w \in [0,\ 2\pi]$, the curve of (20) is high near $w = 0$ and low around $w = \pi$. Thus the moving average is a low-pass digital filter.

As an illustration and without loss of generality, we consider the case of $N = 8$. We have its amplitude-frequency response as

$$|H(e^{jw})| = \frac{1}{8}\frac{\sin(8w/2)}{\sin(w/2)}. \tag{21}$$

We in Figure 11 illustrate the amplitude-frequency response $|H(e^{jw})|$ of (21).

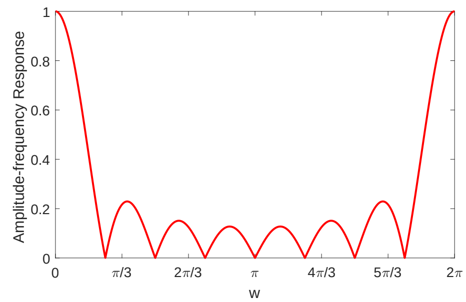


Figure 11: The amplitude-frequency response of moving average method

A.2 Why Are ETS and MA Problematic

Up to now, we have made clear the nature of ETS and MA (see Theorem 4, Theorem 5). They are powerful in many application scenarios. However, we should still point out their theoretical insufficiency:

- Potential to negatively impact (attenuate or amplify) the other desired frequencies, and cannot completely wipe out the unwanted components (the attenuation to high-frequency is not perfectly zero);

- The Time Delay issue.

Compared to the Time Delay issue, the first one is trivial because practically we do not expect the most ideal algorithm performances, as long as the bias is acceptable. As an intuitive understanding, we look again the example in Subsection A.1. We know that the truth of $y(n)$ should be $x_p(n) = sin(nT_s)$. Thus we need to figure out how close between $y(n)$ and $x_p(n)$, see Figure 12.

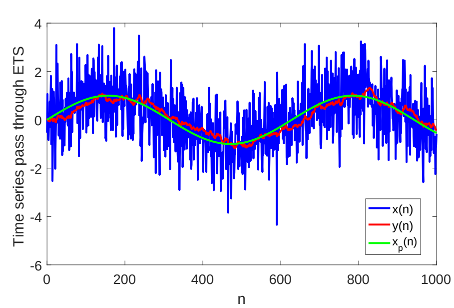


Figure 12: The time delayed ETS

Obviously, there is a time delay in $y(n)$ compared to its truth $x_p(n)$. However, this is easy to investigate.

Recall that in Fourier frequency analysis, besides the amplitude-frequency response, there also exits the phase-frequency response, meaning a component $e^{j(w'n)}$ (complex exponential time series) must accept a phase delay of $\varphi(w')$ when passing through the system $H(e^{jw})$ so that the generated $y(n)$ is not only with the change of $|H(e^{jw'})|$ in amplitude, but also with the change of $\varphi(w')$ in phase. This is the nature of a LTI (linear-time shift invariant) system so that it is unavoidable. Thus the real-time filtering method, meaning no phase delay, is nonexistent. The term real-time here means the times series is in sequential, not in block. That means if n denotes the current discrete time index, it is impossible for us to know the data at $n + 1$.

However, it is still possible to design a zero-phase (no-phase-delay) LTI Digital Filter for block data.

Theorem 6 (Mitra and Kuo (2001)) *Suppose $x(n)$ is the focused time series and $H(z)$ is a designed LTI Digital Filter. Let $H(z)$ be shorted as $\mathcal{H}$, and $\mathcal{H}x(n)$ denote that $x(n)$ passes through the system $H(z)$. Let the operator $z(n) = \mathcal{W}[x(n)]$ generate the time series $z(n)$ such that $z(n) = x(-n)$. If $y(n)$ is generated by the following four procedures:*

1. $u(n) = \mathcal{H}x(n)$;
2. $v(n) = \mathcal{W}u(n)$;
3. $w(n) = \mathcal{H}v(n)$;
4. $y(n) = \mathcal{W}w(n)$,

then $y(n)$ has no phase delay compared to $x(n)$.

Proof 6 *Easy to show that*

$$\begin{aligned} Y(e^{jw}) &= W^*(e^{jw}) = [V(e^{jw})H(e^{jw})]^* \\ &= [U^*(e^{jw})H(e^{jw})]^* = U(e^{jw})H^*(e^{jw}) \\ &= X(e^{jw})H(e^{jw})H^*(e^{jw}) \\ &= X(e^{jw})|H(e^{jw})|^2, \end{aligned} \tag{22}$$

where H^* denotes the complex conjugate of a complex variable H. Equation (22) implies the generating procedure only introduces the changes to the amplitude of $x(n)$, doing nothing to the phase of $x(n)$. Thus there is no time delay between $y(n)$ and $x(n)$. Note that if the DTFT of $x(n)$ is $X(e^{jw})$, the DTFT of $x(-n)$ will be $X^*(e^{jw})$ (easy to show by the Definition of DTFT). Note again that the term $x(-n)$ admits the non-causality (namely, non real-time).

As an intuitive understanding to Theorem 6, see Figure 13, which is the zero-delay counterpart of Figure 12.

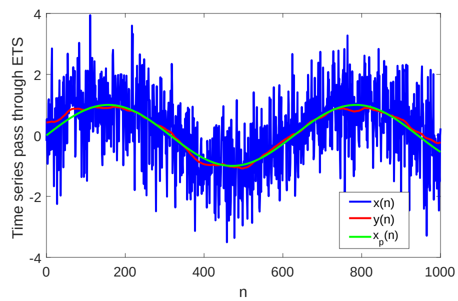


Figure 13: The no-time-delay ETS

Surely, the same story keeps in Moving Average method. For briefness, we dismiss the detailed discussion to MA.

A.3 Improved Exponential Smoothing and Moving Average

Theorem 6 suggests the improved ETS and MA method having no time delay, which is more useful in time series analysis and processing for block data which has already been completely collected before doing analysis.

B New Perspectives of Time Series Analysis and Processing

In this section, we need to present the theory bases to generate the ARMA-SIN methodology. Plus, we will also differentiate the concepts of **Time Series Analysis** and **Time Series Processing**. This idea is taken from the Signal Processing community. In the following, we use the term Time Series Analysis to denote the procedure of analyzing the mathematical properties of a time series, and Time Series Processing to denote the procedure of generating a new more interesting time series from the given one. By doing so, we can find the counterpart concepts of Time Series Analysis and Time Series Processing as Spectral Analysis and Digital Filtering, respectively. The terms of Spectral Analysis and Digital Filtering are from Signal Processing community. Since the signal is philosophically a kind of time series, they should mutually share some analytical and transformational methodologies, rather than independently develop themselves in separate ways. Thus, to bridge the gap between the Signal Processing community and Time Series Analysis community is also one of our academic objectives in this paper.

In order not to confuse readers unfamiliar with the targeted topics in this section, we will try our best to deliver more philosophies instead of mathematics. However, the theoretical sufficiency will be fully guaranteed.

B.1 Linear System Theory

What the term *system* means is a philosophical issue. When it is brought to engineering, the system theory booms. The emergence of system theory (Von Bertalanffy, 1968), information theory (Shannon, 1948) and cybernetics (Wiener, 1965) indicates the coming of a new era to both philosophy, management and engineering.

Mathematically, a system is a function mapping an input x to an output y. The input and output are generally time-dependent variables. The time variable could be continuous or discrete. Thus the input and output could also be continuous or discrete ones. Accordingly we have $x(t)$, $y(t)$ for continuous case and $x(n)$, $y(n)$ for discrete. If the time variable is discrete, the system will be called as discrete system, and accordingly the input and output will be **Time Series**.

In management science, one typical instance of a system is human population problem. Imagining the whole society as a system, the total population of human is the system output, we should have government policy, availability of fresh water and food and so on as system inputs. The issue of adjusting the system output by changing the system input is termed as System Control (Dorf and Bishop, 2011), philosophically also known as Cybernetics, and by changing the system structure and parameters is termed as Systems Engineering (Blanchard et al., 1990).

In system analysis community, one type of the popular and important systems is Linear Time-shift Invariant (LTI) system, see Definition 1.

Definition 1 *A system $y = H(x)$ is called Linear Time-shift Invariant (LTI) system if it satisfies*

- **Homogeneity:** $ky = H(kx)$, $k \in \boldsymbol{R}$;
- **Additivity:** $y_1 + y_2 = H(x_1 + x_2)$, *where* $y_1 = H(x_1)$ *and* $y_2 = H(x_2)$;
- **Time-shift Invariance:** $y(t - t_0) = H[x(t - t_0)]$, *where t is time variable and t_0 is a fixed time point.*

The homogeneity and additivity collectively define the ***Linearity***

$$k_1 y_1 + k_2 y_2 = H(k_1 x_1 + k_2 x_2), k_1, k_2 \in \boldsymbol{R}.$$

The time-shift invariance actually indicates the system structure and parameters remain unchanged over time so that the delayed input brings the delayed output, and the two delays are same. Besides, what's more interesting, as long as the input keeps same just with some delay, the output will also keeps same with same delay. According to Wang et al. (2019), linear system theory takes dominant position in practice in consideration of system analysis and system refactor. For more information on this point, please see *Introduction* of Wang et al. (2019).

One canonical category of LTI system is given in (23) as differential equation for continuous case $y = H_c(x|a, b)$, and (24) as difference equation for discrete case $y = H_s(x|a, b)$.

$$a_0 \frac{d^p y}{dt^p} + a_1 \frac{d^{p-1} y}{dt^{p-1}} + a_2 \frac{d^{p-2} y}{dt^{p-2}} + ... + a_p y = b_0 \frac{d^q x}{dt^q} + b_1 \frac{d^{q-1} x}{dt^{q-1}} + b_2 \frac{d^{q-2} x}{dt^{q-2}} + ... + b_q x, \quad (23)$$

where a, b are vectors and $length(a) = p + 1$, $length(b) = q + 1$, $p \geq 0,\ q \geq 0$.

$$a_0 y_k + a_1 y_{k-1} + a_2 y_{k-2} + ... + a_p y_{k-p} = b_0 x_k + b_1 x_{k-1} + b_2 x_{k-2} + ... + b_q x_{k-q}, \tag{24}$$

where y_k stands for $y(k)$.

We emphasize that the left-hand-side of (24) as Auto-Regression (AR) part, and the right-hand-side as Moving Average (MA) part.

For notation simplicity, we use $y = H(x)$ to exclusively denote the discrete case $y = H_s(x)$. Because in this paper we pay more attention to discrete system case so that we could model the time series problem as a discrete system problem.

Remark 8 *Many practical systems dynamics, namely the system functions $H(\cdot)$, are with the mathematical form of (23). For example, if we let $p = 2$, $a_0 = 1$, $a_i = 0$ $(i = 1, 2)$, $q = 0$, $b_0 = 1$, $y := s$ denote the displacement of a moving object, and $x := a$ be the acceleration, we should have $\frac{d^2 s}{dt^2} = a$. This admits the reputed Newton's second law in physics, or in theoretical mechanics. Another example is from electrical engineering. Actually the dynamics of circuits consisting of inductors, capacitances and resistances could be described by differential equation (23) or differential equations, in which each equation has the form of (23) (Robbins and Miller, 2012).*

In fact, the continuous system (23) and the discrete system (24) are highly related. The gap is bridged by the discretization theory (Diniz et al., 2010; Chaparro and Akan, 2018; Yang, 2009). For more on how to convert a continuous system to its discrete counterpart, please see Diniz et al. (2010); Chaparro and Akan (2018); Yang (2009). In this case, we have $t = T_s n$, where T_s denotes the sampling period. Note that the discrete counterpart of a continuous system actually has different coefficients as it. That we in (23) and (24) use the same notation $a_0,\ a_1\ , ...,\ b_0,\ b_1, ...$ is just for notation simplicity.

Actually the system (23) is theoretically sufficient to describe many different systems holding different dynamics. Since y is a function of x, also a function of t. We should have $y = y(t)$. Suppose y is continuous over a compact set $T, t \in T$, we could find a polynomial function $p(t)$ such that $y(t)$ and $p(t)$ are as close as desired. This is by

Weierstrass approximation theorem (Kreyszig, 1978). Thus (23) could take a special form of

$$a_0 \frac{d^p y}{dt^p} = b_0 t, \tag{25}$$

meaning we require x also take its special case of $x = t$. Since (23) is more general than polynomial family, it is theoretically sufficient to describe many practical systems mathematically. Besides, discretization theory asserts the discrete system could be analytically derived from its continuous counterpart given the sampling period T_s. Thus we can know that the discrete representation of a system is also practically sufficient. This is the intuitive (but not strict) philosophy why the linear system theory, for example the discrete linear system (24), is practically sufficient to many real problems.

Remark 9 *It is possible that although the data we collected is a sequence, the philosophy behind may be continuous, meaning the nature of a discrete system we concern, (24), is actually a continuous system. Thus studying the continuous system theory is still meaningful for time series problem. Because in some cases the discrete system (24) is in fact a refactor of a continuous system (23). For example, when we study the human population problem, we may build a linear time series model (24). However, the population system is actually a continuous system. It evolves all the time rather than merely at the time points we pay attention on, namely the time points we collect the data.*

Remark 10 *From the perspective of statistics, the time series model (24) is actually a successful application of linear regression model. The philosophy is that we may use the past information and causal variable of the interested variable to linearly represent it. Thus the discrete model (24) may have different philosophies from different field viewpoints.*

The three interesting properties of a system is Causality, Stability and Inverse-system Stability. The Inverse-system Stability is equivalent to Minimum Phase. The former is usually referred to in System Control community, and the latter is frequently mentioned in Signal Processing field. Specifically we have the following Definition 2.

Definition 2 *A system is (physically) causal if the output only depends on the present and past values of the input, meaning the output of a causal system to an input appears only* *while or after the input is applied to the system (Yang, 2009); A system is stable if the output is bounded given a bounded input (Chaparro and Akan, 2018); A system $H(\cdot)$ is inversely stable if its inverse system (namely $1/H(\cdot)$) exists, is causal and is stable. When a system is inversely stable, it is conceptually equivalent to Minimum Phase system (Diniz et al., 2010).*

The mentioned three properties are critical as well for time series analysis. For example, the nature of unit-root test in time series analysis is actually the stability of a discrete system. This point will be illustrated later.

Among all of the system analysis methodologies, the Laplace Transform and Z Transform standout (Yang, 2009; Diniz et al., 2010; Chaparro and Akan, 2018). The Laplace transform is defined for continuous time system (23) and Z Transform for discrete time system (24). Actually the Z transform is also known as the discrete version of Laplace transform. For motivations of having those transforms and detailed theories, please see Yang (2009); Diniz et al. (2010); Chaparro and Akan (2018). We in this paper just display some important conclusions without exploring more.

The Laplace transform to the continuous system (23) is as

$$H(s) = \frac{b_0 s^q + b_1 s^{q-1} + b_2 s^{q-2} + ... + b_q}{a_0 s^p + a_1 s^{p-1} + a_2 s^{p-2} + ... + a_p}, \tag{26}$$

where s is a complex variable consisting of the real and image part, defined in real-image coordinate.

The Z transform to the discrete system (24) is as

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + ... + b_q z^{-q}}{a_0 + a_1 z^{-1} + a_2 z^{-2} + ... + a_p z^{-p}}, \tag{27}$$

where z is also a complex variable defined in polar coordinate. In the following, we use $H(z|a, b)$ to denote a system transfer function defined in (27).

Note that (26) and (27) are mathematically meaningful only given some constraints like convergence region and so on. For more, see Yang (2009); Diniz et al. (2010); Chaparro and Akan (2018).

The function of the variable z in (27) seems really resemble to the backward shift operator, also known as lag operator (Hamilton, 1995). However, note that they are naturally different. The former is a complex variable sufficient to do complex analysis, while the latter is merely a algebraic operator having no analytical properties. Interested readers are invited to refer to Yang (2009); Diniz et al. (2010); Chaparro and Akan (2018).

Equation (27) is also known as Transfer Function describing a discrete system (Liu et al., 2017; Helfenstein, 1991; Dorf and Bishop, 2011; Diniz et al., 2010; Chaparro and Akan, 2018; Yang, 2009).

Remark 11 *It is worthy of mentioning that the ARMA process, in system analysis sense, is a linear system driven by a sequence of Gaussian white noise, meaning the system input is Gaussian white noise (Papoulis and Pillai, 2002). Besides, according to Linearity defined in Definition 1, a ARMAX process (Liu et al., 2017) is also a linear system. However, it is driven by two inputs. The one is Gaussian white noise, and another is an exogenous variable x.*

For notation simplicity, we use $\text{ARMA}[z|(p,q),(\varphi,\theta)]$, $\text{ARMA}(z|p,q)$ to denote the transfer function of a ARMA process.

Lemma 1 *A linear discrete system (24) is a casual system. Besides, if the modulus of all the roots of its denominator characteristic polynomial are larger than one, the system is stable.*

Proof 7 *The casuality is sure. For stability, see Yang (2009).*

By Lemma 1, we should mention the nature of unit-root test (Hamilton, 1995; Box et al., 2015) in time series analysis. It is actually require the system to be stable so that the bounded input would not lead to a unbounded output. Obviously a unbounded-output system is meaningless to practical problems (Chaparro and Akan, 2018). As for the requirement of unit-root test that the modulus of all the roots of its numerator characteristic polynomial are larger than one, it in fact requires the inverse system of (24) and (27) to be stable, meaning it is a minimum phase system. The property of minimum phase of a system is important in control system designing for perfect tracking problem, and signal processing for minimum time delay problem.

The inverse system of a ARMA system is actually a whitening filter, meaning it could convert a colored noise (autocorrelated noise, ARMA process) sequence to a white noise (uncorrelated noise) sequence. Thus the invertibility of ARMA system is also meaningful. Because we want the whitening filter to be useful (stable).

When the two subsystems are combined together, the integrated one is given in Lemma 2.

Lemma 2 *If $y = H_1(x|a_1, b_1)$ and $w = H_2(y|a_2, b_2)$, meaning the system H_1 cascades the system H_2, we have a equivalent virtual system $w = H_3(x|a_3, b_3)$ from input x to output w such that $H_3(z|a_3, b_3) = H_1(z|a_1, b_1)H_2(z|a_2, b_2)$. Intuitively see Fig. 14.*

Proof 8 *See Diniz et al. (2010); Chaparro and Akan (2018).*

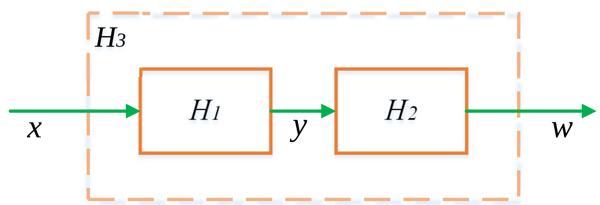


Figure 14: Two systems cascaded together

Lemma 2 upholds the validity of mathematical expression of the SARIMA model (Hyndman and Athanasopoulos, 2018; Box et al., 2015). As an example, if we let $H_1(z) = 1 - z^{-L}$ (the lag-L seasonal difference operator), and let $H_2(z) = \text{ARMA}(z|p_2, q_2)$ (the ARMA modelling part), we can see the conclusion.

Plus, Lemma 2 also indicates the nature of time series processing from the perspective of system analysis. Specifically, the nature is in fact that we let the interested time series x pass through a system H so that x could be transformed into y. For instance, the system $H(z) = 1 - z^{-1}$ means the general first order difference, and $H(z) = 1 - z^{-L}$ means the seasonal lag-L difference. Later we will show the nature of exponential smoothing, moving average and so on in such pattern.

Actually, all the properties of a LTI system could be uniquely determined by its impulse response function $h(n)$. The impulse response of a LTI system means the response of the system driven by a impulse function $\delta(n)$, where $\delta(0) = 1$ and $\delta(i) = 0,\ i \neq 0$.

Lemma 3 *A system is causal if $h(n) = 0$ when $n < 0$. A system is stable if and only if $\sum_{n=-\infty}^{\infty} |h(n)| < \infty$.*

Proof 9 *See Diniz et al. (2010); Chaparro and Akan (2018); Yang (2009).*

Lemma 4 *$H(z) = \sum_{n=0}^{\infty} h(n)z^{-n}$; $y(n) = h(n) * x(n) = x(n) * h(n)$, where $*$ denotes the convolution operator.*

Proof 10 *See Diniz et al. (2010); Chaparro and Akan (2018); Yang (2009).*

The Z transform could also be defined for a time series $x(n),\ \ -\infty < n < \infty$.

Definition 3 *The Z transform of a time series $x(n)$ is given as $X(z) = \sum_{n=-\infty}^{\infty} x(n)z^{-n}$. For more information like convergence region and so on, see Diniz et al. (2010); Chaparro and Akan (2018); Yang (2009).*

Then we could have the relation in Z domain among the system property, the input and the output.

Lemma 5 *Suppose a LTI system is with a transfer function $H(z)$, and the Z transform of the input $x(n)$ and $y(z)$ are $X(z)$ and $Y(z)$, respectively. We have $Y(z) = H(z)X(z)$.*

Proof 11 *See Diniz et al. (2010); Chaparro and Akan (2018); Yang (2009).*

Lemma 4 and 5 admit that the complex convolution computing in time domain could be transformed into a simple algebraic calculation in Z domain.

Lemma 5 again indicates the nature of the time series analysis and processing. Specifically, the nature of time series analysis is that we focus and investigate a time series $x(n)$ in its transformed domain $X(z)$; the nature of time series processing is that the interested time series pass through a purpose-based and custom-designed system (particularly the system is a LTI one) and generate a new time series, which is more interesting (intuitively, as an example, recall an original time series and the one that exponential smoothing generates from the original).

When we let z take its special form $z = e^{jw}$, meaning a complex variable in polar coordinate with modulus of 1 all the time, where j is complex unit and w physically denotes a variable in Fourier frequency domain. We can see from Lemma 5 that $Y(e^{jw}) = H(e^{jw})X(e^{jw})$. $H(e^{jw})$ here means the Frequency Response of a system and $Y(e^{jw})$, $X(e^{jw})$ actually denote the DTFT, Discrete Time Fourier Transform (if x and y are aperiodic), or Discrete Fourier series (if x and y are periodic), of the time $y(n)$ and $x(n)$, respectively.

Motivation 1 *The reason why we focus on Fourier frequency domain to do spectral analysis is that a time series x actually consists of many Fourier frequency components. If we can properly design a LTI system H such that x can pass through, the interested part of x will be remained and transformed while the rest will be discarded and/or attenuated. This is the nature of time series analysis and processing in the system analysis sense, just as moving average, exponential smoothing do. The designed system to time series is generally a Digital Filter which can split the x into different parts, just as the time series decomposition methods like STL, X11 et al. do.*

When we investigate a time series and a LTI system in Fourier frequency domain, the spectral analysis issue arises.

B.2 Spectral Analysis

The Spectral Analysis, also known as Fourier Frequency Domain Analysis, to time series starts from Granger and Hatanaka (1964), followed by many other monographs (Priestley, 1981; Koopmans, 1995; Bloomfield, 2004). However, reviewing the current main-streams of time series analysis, the spectral analysis failed to draw enough attentions from researchers in time series analysis community (Hamilton, 1995; Hyndman et al., 2008; Box et al., 2015; Hyndman and Athanasopoulos, 2018). The exclusive part thought highly of is Fourier Series Expansion (also known as harmonic analysis). That is, if a interested time series is periodic, it can be represented by the summation combinations of trigonometric functions (Bloomfield, 2004). However we should point out that the Fourier series expansion is merely the special case of spectral analysis for periodic series. Concerning the aperiodic series, spectral analysis could also give beautiful conclusions, just like Fourier series expansion does.

There are two main reasons for such a dilemma: (a) The spectral analysis approach is not a traditional methodology for time series community, and if without basis knowledge on complex analysis and concepts of system analysis, it is hard to understand and further apply to practical problems; (b) **Digital Filtering** technology is not sufficiently introduced to time series community, since spectral analysis is only powerful for time domain problems if accompanied with digital filtering.

More interestingly, we should mention that the moving average method, exponential smoothing method and even the Holt's method (Hamilton, 1995; Hyndman et al., 2008; Hyndman and Athanasopoulos, 2018) are merely the special cases of SADFA, spectral analysis and digital filtering approach. Plus, it is also powerful for time series decomposition just like STL, X11 and so on do. We will make clear this later.

To this end, in this subsection we aim to review the issues of Spectral Analysis, and in the next subsection we will give a brief on Digital Filtering.

The term *spectrum* is coined by one of the greatest scientists, Sir Isaac Newton (1642-1726), when he studied the decomposition problem of light, declaring a white light could be split into many different lights with different colors by a prism. The color of a light beam is determined by its frequency, red for lower frequency and blue for higher frequency. The conceptual philosophy was borrowed by Joseph Fourier (1768-1830), a great French engineer, in 1822. He pointed out that a periodic continuous function $x(t)$ with some other proper properties, namely Dirichlet's Conditions (Mani et al., 1997), could be decomposed into the sum of infinite-many sine functions with different frequencies. For each frequency, we term it as a spectral line in the spectrum $X(\Omega)$ of $x(t)$, where $X(\Omega)$ is the Fourier transform (or specially, the Fourier series) of $x(t)$, and Ω denotes the continuous frequency like $2\pi \times 5Hz$, $2\pi \times 20Hz$, $2\pi \times 8KHz$, $2\pi \times 600MHz$, $2\pi \times 2.4GHz$ and so on.

The Spectral Analysis aims to build the close (or equivalent) relation between the time domain analysis and the (Fourier) frequency domain analysis. Since the time series is in time domain, it is suitable to be transformed to frequency domain to have further compound analysis (spectral analysis) and processing (digital filtering).

The Spectral Analysis consists of two part: (a) the spectral analysis for a time series $x(n)$, and (b) the spectral analysis for a discrete system $H(z)$. The spectral analysis for time series aims to figure out the spectral properties of a time series, while the spectral analysis for a discrete system aims to decide which type of system to be designed such that it can transform a time series $x(n)$ into a desired new time series $y(n)$, which is more interesting. For example, please recall an original time series $x(n)$ and the new time series $y(n)$ after difference operator defined by the system $H(z) = 1 - z^{-1}$. Actually it is the spectral properties of $x(n)$ that admits the power of the system $H(z)$. Specifically, according to the spectral properties of $H(z)$ (namely $H(e^{jw})$), the system $H(e^{jw})$ can in the frequency domain wipe away the undesired part (namely the trend) of $x(n)$. Note that it is the trend that makes $x(n)$ a non-stationary process in the variant mean sense. This is the philosophy of Digital Filtering which will be detailed later.

The Fourier Spectral Analysis for time series contains four parts (Yang, 2009; Chaparro and Akan, 2018; Diniz et al., 2010): (a) Fourier Transform for a continuous time function $x(t)$ (FT); (b) Fourier Series Expansion for a continuous periodic time function $x_\sim(t)$ (FS); (c) Fourier Transform for a discrete time function, namely time series, $x(n)$ (DTFT, discrete time Fourier transform); and (d) Fourier Series Expansion for a discrete periodic time function, periodic time series, $x_\sim(n)$ (DFS, discrete Fourier series). For notation simplicity, we in the following still use x to uniformly denote a periodic time function $x_\sim$.

Suppose $x(t)$ is a periodic continuous time function, and the period of it is T, meaning the angular frequency is $\Omega_0 = 2\pi/T$. If it satisfies the Dirichlet's conditions (Mani et al., 1997), we can decompose $x(t)$ into the sum of infinite-many complex-valued continuous sine functions. Namely

$$x(t) = \sum_{n=-\infty}^{\infty} X(n\Omega_0)e^{jn\Omega_0 t}, \tag{28}$$

where $X(n\Omega_0)$, Fourier Series, denotes the coefficient (or proportion) that the complex sine component $e^{jn\Omega_0 t}$ takes to compose a function $x(t)$. The $X(n\Omega_0)$ is mathematically given as

$$X(n\Omega_0) = \frac{1}{T}\int_{-T/2}^{T/2} x(t)e^{-jn\Omega_0 t}dt. \tag{29}$$

Clearly, $X(n\Omega_0) =: X(n) =: X_n$ here is a complex variable. Thus $X_n = |X_n|e^{j\varphi_n}$, where $|X_n|$ denotes the modulus and φ_n denotes the argument.

Remark 12 *Note that the FS defined in (28) is actually equivalent to the FS defined by the sum of real trigonometric functions (Chaparro and Akan, 2018),*

$$\begin{aligned} x(t) &= A_0 + 2\sum_{n=1}^{\infty}[A_n\cos(n\Omega_0 t) + B_n\sin(n\Omega_0 t)] \\ &= X_0 + 2\sum_{n=1}^{\infty}|X_n|cos(n\Omega_0 t + \varphi_n), \end{aligned} \tag{30}$$

where $|X_n| = \sqrt{A_n^2 + B_n^2}$ *and* $\varphi_n = -\tan^{-1}(B_n/A_n)$*. The definitions of* A_n *and* B_n *could be found in Chaparro and Akan (2018). The story holds due to Euler's formula of* $e^{jn\Omega_n t} = \cos(n\Omega_n t) + j\sin(n\Omega_n t)$*, meaning we can use this trick to simplify the definition of (30) as (28). However, the result is that this trick brings us to complex domain. The rewards is better analytical properties.*

Note that a **continuous periodic** function in time domain admits a **discrete spectrum** in frequency domain.

When it comes to a continuous aperiodic time function, the story becomes a bit more complex.

Suppose $x(t)$ is a continuous aperiodic time function satisfying $x(t) \in L_2$ and Dirichlet's conditions, we could have the Fourier transform of $x(t)$ as

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(j\Omega)e^{j\Omega t}d\Omega, \tag{31}$$

where $X(j\Omega)$ is FT of $x(t)$ and is defined as

$$X(j\Omega) = \int_{-\infty}^{\infty} x(t)e^{-j\Omega t}dt, \tag{32}$$

where $\Omega = 2\pi f$ (f is frequency with the unit of Hz) is angular frequency with the unit of rad/s.

Note that a **continuous aperiodic** function in time domain admits a **continuous spectrum** in frequency domain.

If we investigate the Fourier Transform of a periodic function $x(t)$ instead of Fourier Series, we have

$$\begin{aligned} X(j\Omega) &= \int_{-\infty}^{\infty} \left[\sum_{n=-\infty}^{\infty} X(n\Omega_0) e^{jn\Omega_0 t} \right] e^{-j\Omega t} dt \\ &= \sum_{n=-\infty}^{\infty} X(n\Omega_0) \int_{-\infty}^{\infty} e^{j(\Omega - n\Omega_0)t} dt \\ &= 2\pi \sum_{n=-\infty}^{\infty} X(n\Omega_0) \delta(\Omega - n\Omega_0), \end{aligned} \tag{33}$$

where $\delta(\cdot)$ denotes the dirac function.

Equation (33) indicates that the Fourier Transform of a continuous periodic function is infinity at some discrete frequency points while zero elsewhere. This kind of spectrum distribution is what we called **Line Spectra**. The regular FT of a aperiodic function is thus a **Continuous Spectra**. Particularly, the dirac function has **Flat Spectra** since its FT keeps constant over the frequency domain.

After sampling (Chaparro and Akan, 2018), a continuous function will become a discrete one, which is a time series. Instead, we can also directly collect data at some discrete time points as a time series. Based on, and similar to the TF and TS for continuous cases, we can define the Discrete Time FT, DTFT, for a aperiodic time series.

Let $w := \Omega T_s$ (Ts is sampling period in time axis, and thus $f_s = 1/T_s$ is sampling frequency) denote the **Digital Frequency** with the unit of rad, we have the DTFT for aperiodic $x(n)$ as

$$x(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{jw}) e^{jwn} dw, \tag{34}$$

where $X(e^{jw})$ is DTFT of $x(n)$ and given as

$$X(e^{jw}) = \sum_{n=-\infty}^{\infty} x(n) e^{-jwn}. \tag{35}$$

Since we have $\Omega t = 2\pi f t = 2\pi f n T_s = \Omega n T_s = wn$. Thus w is the discrete counterpart of Ω. Thus we term it as Digital Frequency.

Proposition 1 *In digital frequency domain, $w = \pi$ denotes w is the highest frequency and $w = 0$ denotes w is the lowest frequency. Since $H(e^{jw})$ is a periodic function over w with period of 2π, we have $w = \pi + 2k\pi$ as the highest frequency and $w = 2k\pi$ as the lowest* *frequency, where $k = 0, \pm 1, \pm 2, \pm 3, \ldots$. The frequency at $w = \pi/2$ is thus the intermediate frequency. A frequency w is higher if it is closer to $\pi + 2k\pi$, and lower if closer to $2k\pi$.*

Proposition 1 stands by the sampling theory (Yang, 2009; Chaparro and Akan, 2018; Diniz et al., 2010). Intuitively, since $\Omega_s \geq 2\Omega, \forall\Omega$ *defined in* (32), by Sampling Theorem (Yang, 2009; Chaparro and Akan, 2018; Diniz et al., 2010), we have $w = \Omega T_s \leq \Omega_s T_s/2 = 2\pi f_s T_s/2 = \pi$. Thus the $w = \pi$ is the highest frequency.

Note that a **discrete aperiodic** function in time domain admits a **continuous but periodic spectrum** in frequency domain. The periodicity over w is easy to show by definition that $X(e^{j(w+2\pi)}) = X(e^{jw})$, meaning the period is 2π. The periodicity is a significant nature of the spectrum over digital frequency domain.

Going here, it is interesting to mention the relationship between (35) and Definition 3 of $X(e^{jw}) = X(z)|_{z=e^{jw}}$.

Now suppose $x(n)$ is periodic with the period of N, we then have the DFS as

$$x(n) = \frac{1}{N} \sum_{k=-\infty}^{\infty} X(k) e^{j\frac{2\pi}{N}nk}, \tag{36}$$

where $X(k)$ is DFS of $x(n)$ and given as

$$X(k) = \sum_{n=-\infty}^{\infty} x(n) e^{-j\frac{2\pi}{N}nk}. \tag{37}$$

Note that a **discrete periodic** function in time domain admits a **discrete and periodic spectrum** in frequency domain. The periodicity is also N (Chaparro and Akan, 2018).

Remark 13 *Interestingly, the periodicity in time domain leads to the discreteness in frequency domain, and the discrete spectrum is sampled from the continuous spectrum of the aperiodic function which is a period of the periodic time function. The conclusion stands both for continuous case and for discrete case.*

Among FT, FS, DTFT and DFS, only the DFS seems reasonable to be numerically calculated and digitally saved in a digital computer, since only it has both discrete time-domain expression and discrete frequency-domain expression. However, this requires $x(n)$

to be periodic, which is impossible in practice since most of $x(n)$ failed to show this property. This is thus the main reason that only the DFS is highly thought of in time series analysis to reconstruct a periodic time series as (30).

The trick here to handle this dilemma is to consider the collected time series as periodic although it maybe not in fact, meaning if the length of collected $x(n)$ is N, we then virtually consider $x(n)$ as a periodic time series with period of N. This trick allows us to use DFS to the interested time series. However, another practical issue is that we need to use a limited memory to save the results of DFS. For DFS, this seems impossible since the sum is taken from $-\infty$ to ∞. In consideration of the periodicity of DFS, we may pay our attention exclusively within a period of DFS, which brings the Discrete Fourier Transform (DFT).

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j\frac{2\pi}{N}nk}, \tag{38}$$

where $X(k)$ is DFT of $x(n)$ and given as

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{N}nk}. \tag{39}$$

Since we aim to analyze the spectral properties of a time series $x(n)$, we should have digital frequency variable w in its transformed domain, just as DTFT (35) does. However, DFT, although powerful for digital calculation, failed to explicitly show this point. Thus we need to figure out the underlying relationship between DTFT and DFT.

Lemma 6 *The DFT $X(k)$ is the uniform sampling and approximation to the DTFT e^{jw}. $X(k) = X(e^{jw})|_{w=\frac{2\pi}{N}k}$, meaning if we take N points within a period of w (namely 2π), $w = 2\pi\frac{k}{N}$ and $k = N\frac{w}{2\pi}$, we have DFT from DTFT, where N denotes the length of the time series $x(n)$.*

Proof 12 *See Chaparro and Akan (2018); Yang (2009); Diniz et al. (2010).*

Note that the calculation burden of DFT is heavy. In order to improve the calculation efficiency, the Fast Fourier Transform (FFT) algorithm is proposed (Cormen et al., 2009; Diniz et al., 2010; Chaparro and Akan, 2018). However, the FFT is not a new transform as DTFT and DFT, just a fast implementation algorithm to DFT.

Now we have already finished the spectral analysis of a time series $x(n)$. Thus the second main topic of spectral analysis is for a discrete system $H(z)$ defined in (27).

As we showed in the end of Subsection B.1, the frequency response of a LTI system $H(z)$ is defined as $H(e^{jw}) = H(z)|_{z=e^{jw}}$. Suppose $X(e^{jw})$ and $Y(e^{jw})$ are DTFT of $x(n)$ and $y(n)$. We have $Y(e^{jw}) = H(e^{jw})X(e^{jw})$ by Lemma 5.

Suppose $x(n) = e^{jw'n}$, it is easy to show that the DTFT $X(e^{jw})$ is as

$$X(e^{jw}) = 2\pi \sum_{k=-\infty}^{\infty} \delta(w - w' + 2\pi k). \tag{40}$$

Thus we have

$$Y(e^{jw}) = 2\pi \sum_{k=-\infty}^{\infty} H(e^{jw})\delta(w - w' + 2\pi k). \tag{41}$$

Equation (41) immediately imply that the output $y(n)$ is also a complex sine function. Its frequency remains same (namely w') with $x(n)$. The only difference is that the modulus and argument have been modulated in frequency domain by $H(e^{jw})$. Thus we have

$$Y(e^{jw}) = |H(e^{jw})|e^{j(w'n+\varphi(w))}, \tag{42}$$

where $|H(e^{jw})|$ denotes the modulus of $H(e^{jw})$ given the frequency w and $\varphi(w)$ denotes the corresponding argument of $H(e^{jw})$. Since w is a variable, $|H(e^{jw})|$ and $\varphi(w)$ are functions, and defines the **Amplitude-frequency Response** and **Phase-frequency Response**.

Here it is now theoretically sufficient to give the motivation and purpose of spectrum-based time series analysis and processing.

Motivation 2 *Recall from the definition of DTFT that $x(n)$ is a sum of infinite-many complex sine functions e^{jwn}, for some n. From (42), we can see if we can carefully design a system $H(e^{jw})$ such that $|H(e^{jw})|$ is small in some frequency points $w_i,\ i = 0, 1, 2, ...$ or a frequency interval $[w_{low},\ w_{up}]$, and remains 1 elsewhere, then the components of $x(n)$ represented by frequencies at $w_i,\ i = 0, 1, 2, ...$ or in $[w_{low},\ w_{up}]$ will be attenuated (or wiped out if small enough). By doing so, if we remain only the interested components, we could finish a time series transformation by spectral analysis approach from $x(n)$ to $y(n)$.*

Note that it is also possible to design a proper $|H(e^{jw})|$ such that every frequency components of $x(n)$ will be properly handled (amplified, kept, and/or attenuated) as desired. Thus the interested time series will be finely transformed into $y(n)$ as expected.

As mentioned before, we are interested in the techniques of Digital Filtering, meaning passing the desired frequency components and stoping the undesired frequency components. Thus we in the following subsection will briefly introduce it.

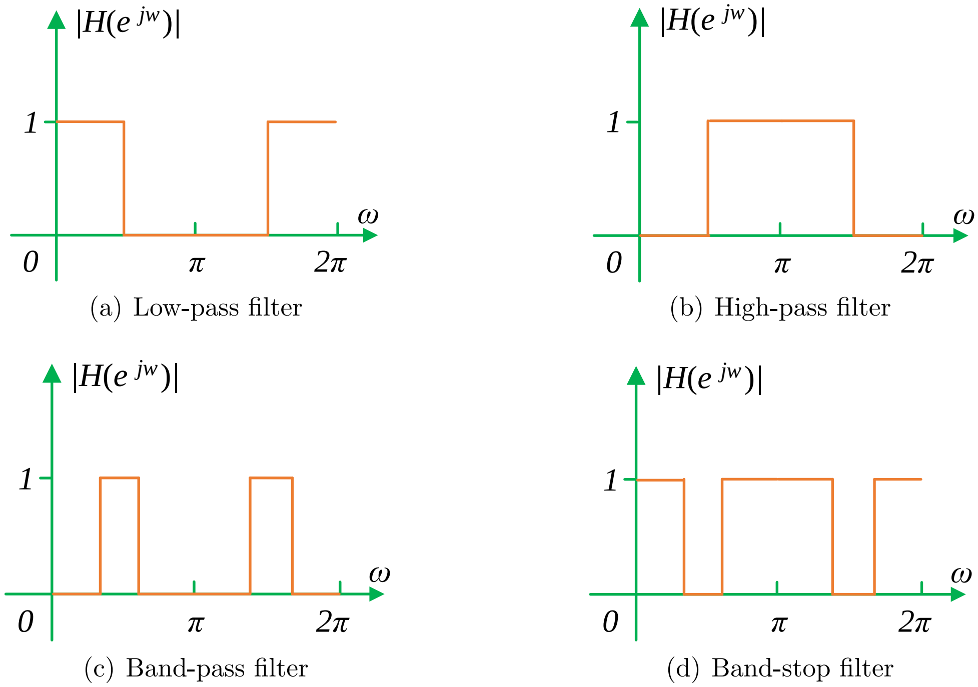
B.3 Digital Filtering

As stated before, a digital filter (DF) is used to process a time series in frequency domain. In practice, the most used DFs are with the form of LTI system, like exponential smoothing method (ETS), moving average method (MA), and difference operator in ARIMA. Details on ETS, MA and difference operator will be showed later. Specifically, the LTI system dynamics of a DF is given as (24) in time domain, (27) in Z domain, and in frequency domain if $z := e^{jw}$. The spectral analysis and digital filtering are mutually complementary. Spectral analysis tells us what to do, meaning which frequency components should be preserved or wiped, and Digital Filtering tells us how to do, meaning which LTI DF could alter the spectra of a time series as desired before. The two are all indispensable for time series analysis and processing.

The digital filter is the discrete counterpart of analog filter. Analog filter works for continuous time function and system (23), while digital filter cares about the discrete time series and system (24).

Typically, the LTI digital filter has four types, called low-pass, high-pass, band-pass and band-stop filter, respectively. The low-pass filter only allows the low frequency components, near $w = 0$, to pass and denies the others. The high-pass filter only allows the high frequency components, near $w = \pi$, to pass and denies the others. The band-pass filter only allows the frequency components within the given frequency subinterval to pass and denies the others. The band-stop filter, opposite to band-pass, denies the frequency components within the given frequency subinterval to pass and allows the others. The **ideal** diagrams of four mentioned digital filters are given in Figure 15.

The breakpoint(s) that $|H(e^{jw})|$ suddenly changes its value from 1 to 0, or vice versa, in Figure 15 is/are called cut-off frequencies. Note that except $|H(e^{jw})|$ is periodic over w with period of 2π, it is symmetrical as well within each period (Diniz et al., 2010; Chaparro and Akan, 2018).



(a) Low-pass filter

(b) High-pass filter

(c) Band-pass filter

(d) Band-stop filter

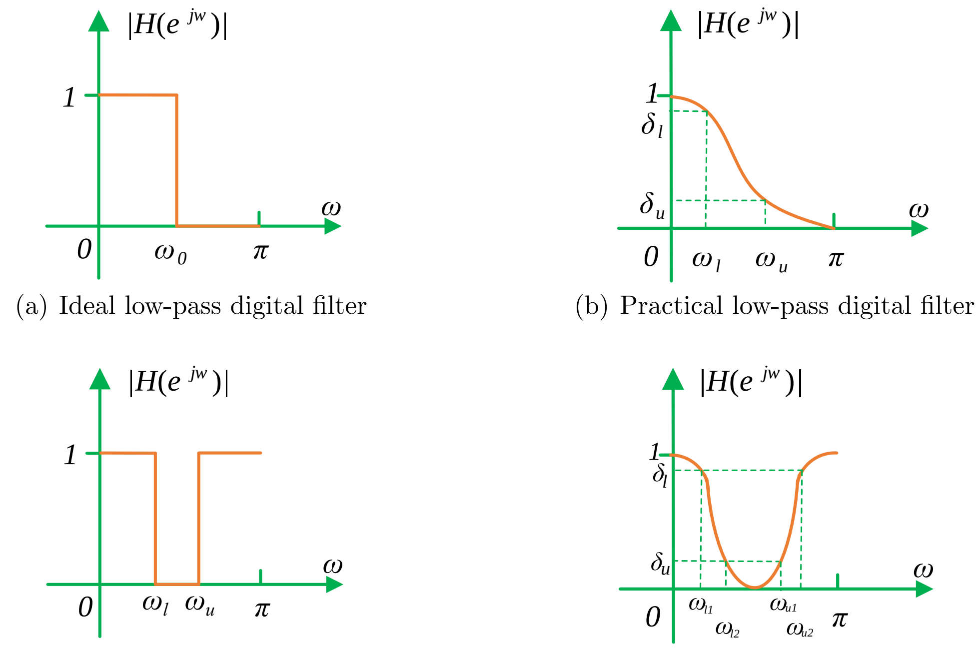
Figure 15: Four types of typical digital filter.

However, unfortunately, the patterns showed in Figure 15 are just mathematically ideal, meaning they are physically impossible to make real. Because the impulse response $h(n)$ of the ideal digital filter, in mathematical sense, requires $h(n) \neq 0$, when $n < 0$. This contradicts the causality claimed in Lemma 3. For more, see Diniz et al. (2010).

Thus in practice, we instead design the approximated scheme to ideal filters. Without loss of generality, we take the low-pass filter and band-stop filter as examples, showed in Figure 16.

From Figure 16 (a),(b), we can know that there exists only one cut-off frequency ideally, but two cut-off frequencies practically. w_l denotes the lower cut-off frequency while w_u denotes the upper one, meaning we guarantee the attenuation to the components with frequency lower than w_l to be no less than δ_l, and the attenuation to the components with frequency higher than w_u to be no larger than δ_u. The logic keeps similar to band-stop filter in Figure 16 (c),(d). In practice, the performances of a designed low-pass filter are satisfying if:

1. The gap between 1 and δ_l is small enough;



(a) Ideal low-pass digital filter

(b) Practical low-pass digital filter

(c) Ideal band-stop digital filter

(d) Practical band-stop digital filter

Figure 16: The ideal and practical low-pass, band-stop digital filters

2. The gap between 0 and δ_u is small enough;

3. The gap between w_l and w_u is small enough, and they all are close enough to w_0;

4. The order of the filter, namely p and q in (27), is small (the smaller, the better).

However, the dilemma is that in order to optimize the requirements $1 \sim 3$, we must accept the order to be high enough.

There are typically two types of digital filters based on LTI system and spectra analysis, the Infinite Impulse Response (IIR) filter and Finite Impulse Response (FIR) filter (Diniz et al., 2010). Mathematically, the LTI system (27) specifies the IIR filter. When a IIR filter takes its special case of without denominator polynomial, it becomes a FIR filter, given in (43).

$$H(z) = b_0 + b_1 z^{-1} + b_2 z^{-2} + \ldots + b_q z^{-q}. \tag{43}$$

IIR and FIR have numerous different natures including the filter performances given fixed order, the designing methods, the linear phase properties and so on. For more, one is referred to Diniz et al. (2010).

Worthy of mentioning is that the IIR and FIR model can not only generate the typical four types of filters showed in Figure 15, but also design some special types of filters like comb filter and difference filter (Diniz et al., 2010). As an example, we show the frequency response of comb filter here, which is also useful in time series processing. Later we will show the nature of L-lag seasonal difference operator is actually also a type of digital comb filter.

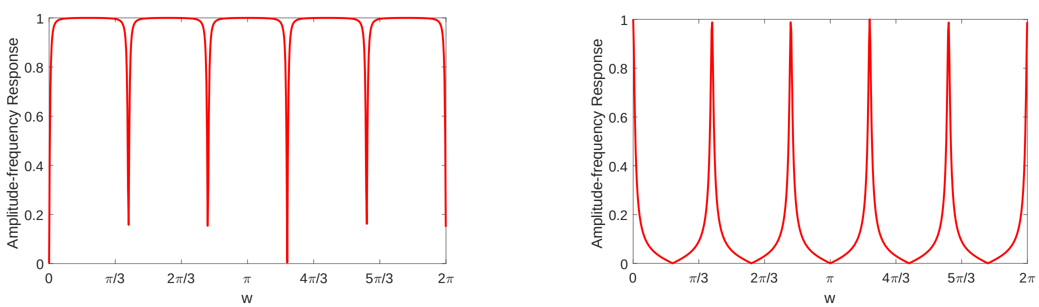
Two special types of comb filter are mathematically given as

$$H(z) = \frac{1+\rho}{2}\frac{1-z^{-N}}{1-\rho z^{-N}}, \tag{44}$$

and

$$H(z) = \frac{1-\rho}{2}\frac{1+z^{-N}}{1-\rho z^{-N}}, \tag{45}$$

where ρ is the filter parameter and N is order that determine the performances. If we set $N = 5$ and $\rho = 0.9$, we have the amplitude-frequency responses in Figure 17.



(a) Amplitude-frequency responses of (44)   (b) Amplitude-frequency responses of (45)

Figure 17: The comb filter

Figure 17 immediately indicates that if we want to wipe out some unwanted frequency components of a time series, we just need to design a comb filter as Figure 17(a), and if we want to preserve some desired frequency components of a time series, we just need to design a comb filter as Figure 17(b), to process it. For more on comb filter, see Diniz et al. (2010).

Up to now, the most intriguing designing methods to IIR are Butterworth method, Chebyshev method, Elliptic method and so on; to FIR are window method (like rectangular window, Kaiser window, and so on), optimization methods (like weighted least-squares method, Chebyshev optimal approximation method, and so on) et al. All the above-mentioned methods are canonical designing ones in signal processing community which have already been implemented by MATLAB. One can find the detailed mathematical derivations in Diniz et al. (2010) and study the sample MATLAB codes shared by Diniz et al.

All the filters this paper designed are based on IIR model and Elliptic method. In order to help readers easily check the powerful ARMA-SIN methodology this paper presents, all the MATLAB source codes we use are available from the corresponding author or online upon the publication of this paper.

B.4 Stochastic Process and Wold's Decomposition Theorem

After introducing the Linear System Theory, Spectral Analysis and Digital Filtering, it is sufficient to pay attention back on S-ARIMA to investigate its nature. Recall that in Introduction we have already point out the philosophy of ARMA and S-ARIMA. They in fact aim to model the dynamics of a stochastic process so that we can use the past information from collected time series to satisfyingly predict the future. Mentioning this, we cannot ignore the reputed Wold's Decomposition Theorem in stochastic process analysis.

Firstly we give the strict mathematical definition of the Wide-sense stationary (WSS) stochastic process.

Definition 4 (Papoulis and Pillai (2002)) *A real-valued stochastic process $x(t)$ is WSS if it satisfies:*

- ***Invariant Mean**: $E\{x(t)\} = \eta$, where η is a constant;*
- ***Invariant Autocorrelation**: $E\{x(t_1)x(t_2)\} = E\{x(t_1+\tau)x(t_1)\} = R(\tau)$, meaning it only depends on $\tau := t_2 - t_1$, having nothing to do with t_1.*

Invariant autocorrelation immediately admits the **Invariant Variance**, since $E\{x(t)\}^2 = R(0)$. Then we should turn to Wold's Decomposition Theorem.

Theorem 7 (Wold's Decomposition Theorem (Papoulis and Pillai, 2002)) *Any WSS stochastic process $x(n)$ could be decomposed into two subprocess: (a) Regular process; and* *(b) Predictable process. Namely*

$$x(n) = x_r(n) + x_p(n), \tag{46}$$

where $x_r(n)$ *is a regular process and* $x_p(n)$ *is a predictable process. Furthermore, the two processes are orthogonal (meaning uncorrelated):* $E\{x_r(n+\tau)x_p(n)\} = 0$.

The detailed concepts of Regular Process, also known as Rational-spectra process from the spectral analysis perspective, and Predictable Process, also known as Line-spectra process, could be found in Papoulis and Pillai (2002).