Supplementary Materials to "A Model for Non-Stationary Time Series and Its Applications in Filtering and Anomaly Detection"

Shixiong Wang, Chongshou Li, and Andrew Lim

***Abstract*—In this supplement, we detail: 1) the other possible application scenarios of the proposed method, 2) the proofs (e.g., algebraic manipulations) of lemmas and theorems, and 3) the additional remarks for the paper.**

***Index Terms*—Supplementary Materials, Page Limit**

I. Revisit the State-Space Model

In the paper, we have established the following state-space linear system model for a non-stationary time series (i.e., a non-stationary signal).

$$\begin{cases} \boldsymbol{X}(n+1) &= \boldsymbol{\Phi}\boldsymbol{X}(n) + \boldsymbol{G}\boldsymbol{W}(n) \\ \boldsymbol{Y}(n) &= \boldsymbol{H}\boldsymbol{X}(n) + \boldsymbol{V}(n), \end{cases} \tag{1}$$

where $\boldsymbol{W}(n)$ denotes the modeling error.

II. Application in Time Series Forecasting and Change Point Detection

Note that the Level model and Holt's method (also known as Linear Trend model) mentioned in [1] for time series forecasting are special cases of TVLAP. When $K = 0$, TVLAP becomes the recursive-form Level model. If $K = 1$, TVLAP degenerates into the recursive-form Holt's method. Besides, Algorithm 1 illustrates the change point detection method based on TVLAP-KF.

Algorithm 1 Change Point Detection Method Based on TVLAP-KF

Definition: $\boldsymbol{P}$ as state estimate covariance in Kalman filter; $\boldsymbol{I}$ as identity matrix with proper dimension; ∞ as a big number; ϵ as a small number; abs(x) as the absolute function which return the absolute value of a real number; $\emptyset$ as an empty set

Reservation: Set $\mathbb{E}_m$ to record minima, and Set $\mathbb{E}^m$ to record maxima

Initialize: $\infty \leftarrow 10^5$, $\epsilon \leftarrow 10^{-6}$, $\boldsymbol{X} \leftarrow \boldsymbol{0}$, $\boldsymbol{P} \leftarrow \infty \times \boldsymbol{I}$, $\boldsymbol{Q}$, $\boldsymbol{R}$, $\mathbb{E}_m \leftarrow \emptyset$, $\mathbb{E}^m \leftarrow \emptyset$

Input: $x(n)$, $n = 0, 1, 2, 3, ...$

while true **do**
  $n \leftarrow n + 1$
  // Estimate the Deep Patterns
  $\hat{\boldsymbol{X}}(n)$ = Kalman_Filter[$x(n)$] // See [2] (Chapter 5.1)
  // Obtain the Estimated Mean Function
  $\hat{f}(n) \leftarrow \hat{X}_0(n)$
  // Turning Point Detection
  if abs$(\hat{X}_1(n-1)) < \epsilon$ and $\hat{X}_1(n) > 0$ **then**
    The time series starts to increase
  else if abs$(\hat{X}_1(n-1)) < \epsilon$ and $\hat{X}_1(n) < 0$ **then**
    The time series starts to decrease
  end if
  // Extrema Detection
  if abs$(\hat{X}_1(n)) < \epsilon$ and $\hat{X}_2(n) > 0$ **then**
    $\mathbb{E}_m \leftarrow \{n\} \cup \mathbb{E}_m$ // Minimum reached
  else if abs$(\hat{X}_1(n)) < \epsilon$ and $\hat{X}_2(n) < 0$ **then**
    $\mathbb{E}^m \leftarrow \{n\} \cup \mathbb{E}^m$ // Maximum reached
  end if
end while

Output: estimated mean $\hat{f}(n)$; minima set $\mathbb{E}_m$; maxima set $\mathbb{E}^m$

III. Detailed Proofs

In this section, we detail the proofs of lemmas and theorems.

Lemma 1: $\boldsymbol{\Phi}^K(T) = \boldsymbol{\Phi}(KT)$.

Proof: Actually, there exists a matrix

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}, \tag{2}$$

such that $\boldsymbol{\Phi}(T) = e^{\boldsymbol{A}T}$. Thus, $\boldsymbol{\Phi}^K(T) = e^{K\boldsymbol{A}T} = \boldsymbol{\Phi}(KT)$. That is,

$$\boldsymbol{\Phi}^K(T) = \begin{bmatrix} 1 & KT & \frac{(KT)^2}{2} & \cdots & \frac{(KT)^K}{K!} \\ 0 & 1 & KT & \cdots & \frac{(KT)^{K-1}}{(K-1)!} \\ 0 & 0 & 1 & \cdots & \frac{(KT)^{K-2}}{(K-2)!} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{3}$$

□

S. Wang, C. Li, and A. Lim are with the Department of Industrial Systems Engineering and Management, National University of Singapore (E-mail: s.wang@u.nus.edu; {iselc, isealim}@nus.edu.sg).

This work is supported by National Research Foundation of Singapore under grant NRF-RSS2016-004.

$$\begin{aligned}\boldsymbol{C}_{\boldsymbol{\Phi},\boldsymbol{G}_1} &= \left[\boldsymbol{G}_1, \boldsymbol{\Phi}\boldsymbol{G}_1, \cdots, \boldsymbol{\Phi}^K\boldsymbol{G}_1\right] \\ &= \begin{bmatrix} \frac{T^K}{K!} & \sum_{i=0}^{K}\frac{(1T)^i}{i!}\frac{(T)^{K-i}}{(K-i)!} & \cdots & \sum_{i=0}^{K}\frac{(KT)^i}{i!}\frac{(T)^{K-i}}{(K-i)!} \\ \frac{T^{K-1}}{K-1!} & \sum_{i=0}^{K-1}\frac{(1T)^i}{i!}\frac{(T)^{K-1-i}}{(K-1-i)!} & \cdots & \sum_{i=0}^{K-1}\frac{(KT)^i}{i!}\frac{(T)^{K-1-i}}{(K-1-i)!} \\ \vdots & \vdots & \ddots & \vdots \\ T & \sum_{i=0}^{1}\frac{(1T)^i}{i!}\frac{(T)^{1-i}}{(1-i)!} & \cdots & \sum_{i=0}^{1}\frac{(KT)^i}{i!}\frac{(T)^{1-i}}{(1-i)!} \\ 1 & 1 & \cdots & 1 \end{bmatrix}. \end{aligned} \tag{7}$$

Lemma 2: The Vandermonde matrix defined as

$$V = \begin{bmatrix} 1 & \alpha_1 & \alpha_1^2 & \dots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \alpha_2^2 & \dots & \alpha_2^{n-1} \\ 1 & \alpha_3 & \alpha_3^2 & \dots & \alpha_3^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_m & \alpha_m^2 & \dots & \alpha_m^{n-1} \end{bmatrix}, \tag{4}$$

is of full rank if $\forall i \neq j$, we have $\alpha_j \neq \alpha_i$.

Proof: Since $\det(V) = \prod_{1\leq i<j\leq n}(\alpha_j - \alpha_i)$ (see [3], Chapter 6.1), the lemma stands. □

Lemma 3: The linear time-invariant system defined in (1) is uniformly completely observable, if K is finite.

Proof:

$$\boldsymbol{O} = \begin{bmatrix} \boldsymbol{H} \\ \boldsymbol{H}\boldsymbol{\Phi} \\ \vdots \\ \boldsymbol{H}\boldsymbol{\Phi}^K \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & T & \frac{(T)^2}{2} & \cdots & \frac{(T)^K}{K!} \\ 1 & 2T & \frac{(2T)^2}{2} & \cdots & \frac{(2T)^K}{K!} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & KT & \frac{(KT)^2}{2} & \cdots & \frac{(KT)^K}{K!} \end{bmatrix}. \tag{5}$$

Note that, if K tends to infinity, many entries of $\boldsymbol{O}$ would tend to zeroes. Thus, if K is finite, by Lemma 2, we have

$$\begin{aligned} rank(\boldsymbol{O}) &= rank\left(\begin{bmatrix} 0^0 & 0^1 & 0^2 & \cdots & 0^K \\ 1^0 & 1^1 & 1^2 & \cdots & 1^K \\ 2^0 & 2^1 & 2^2 & \cdots & 2^K \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ K^0 & K^1 & K^2 & \cdots & K^K \end{bmatrix}\right) \\ &= K+1, \end{aligned} \tag{6}$$

meaning $\boldsymbol{O}$ is of full rank. According to the definition of observability, this lemma stands. □

Lemma 4: The linear time-invariant system defined in (1) is uniformly completely controllable, if K is finite and $\boldsymbol{G}$ is given as one of the following cases:

(a) $\boldsymbol{G}_1 = [\frac{T^K}{K!}, ..., T, 1]'$;
(b) $\boldsymbol{G}_2 = diag\{\frac{T^K}{K!}, ..., T, 1\}$;
(c) $\boldsymbol{G}_3$ as an identity matrix $\boldsymbol{I}$ with proper dimensions.

Proof: Let $\boldsymbol{C}_{\boldsymbol{\Phi},\boldsymbol{G}}$ denotes the controllability matrix defined by the pair $[\boldsymbol{\Phi}, \boldsymbol{G}]$. Since $\boldsymbol{C} = [\boldsymbol{G}, \boldsymbol{\Phi}\boldsymbol{G}, ..., \boldsymbol{\Phi}^K\boldsymbol{G}]$, it is easy to check that $rank(\boldsymbol{C}_{\boldsymbol{\Phi},\boldsymbol{G}_3}) = K+1$ (full rank). Due to $rank(\boldsymbol{C}_{\boldsymbol{\Phi},\boldsymbol{G}_2}) = rank(\boldsymbol{C}_{\boldsymbol{\Phi},\boldsymbol{G}_3})$, $rank(\boldsymbol{C}_{\boldsymbol{\Phi},\boldsymbol{G}_2}) = K+1$ also holds. As for $\boldsymbol{C}_{\boldsymbol{\Phi},\boldsymbol{G}_1}$, we have (7).

By the binomial theorem, the entry of $\boldsymbol{C}_{\boldsymbol{\Phi},\boldsymbol{G}_1}$ at $(I+1, J+1)$ is therefore

$$\begin{aligned} \boldsymbol{C}_{\boldsymbol{\Phi},\boldsymbol{G}_1}(I+1, J+1) &= \sum_{i=0}^{K-I}\frac{(JT)^i T^{K-I-i}}{i!(K-I-i)!} \\ &= \frac{1}{(K-I)!}(JT+T)^{K-I}, \end{aligned} \tag{8}$$

where $I, J = 0, 1, 2, ..., K$, giving $\boldsymbol{C}_{\boldsymbol{\Phi},\boldsymbol{G}_1}$ further as

$$\boldsymbol{C}_{\boldsymbol{\Phi},\boldsymbol{G}_1} = \begin{bmatrix} \frac{T^K}{K!} & \frac{(2T)^K}{K!} & \frac{(3T)^K}{K!} & \cdots & \frac{[(K+1)T]^K}{K!} \\ \frac{T^{K-1}}{K-1!} & \frac{(2T)^{K-1}}{K-1!} & \frac{(3T)^{K-1}}{K-1!} & \cdots & \frac{[(K+1)T]^{K-1}}{K-1!} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ T & 2T & 3T & \cdots & (K+1)T \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix}. \tag{9}$$

Note that, if K tends to infinity, many entries of $\boldsymbol{C}_{\boldsymbol{\Phi},\boldsymbol{G}_1}$ would tend to zeroes. Thus, if K is finite, by Lemma 2, we have

$$\begin{aligned} & rank(\boldsymbol{C}_{\boldsymbol{\Phi},\boldsymbol{G}_1}) \\ &= rank\left(\begin{bmatrix} 1^K & 2^K & \cdots & (K+1)^K \\ 1^{K+1} & 2^{K-1} & \cdots & (K+1)^{K-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1^1 & 2^1 & \cdots & (K+1)^1 \\ 1^0 & 2^0 & \cdots & (K+1)^0 \end{bmatrix}\right) \\ &= K+1. \end{aligned} \tag{10}$$

Since $\boldsymbol{C}_{\boldsymbol{\Phi},\boldsymbol{G}_1}$ defined in (7) is rank-sufficiency, this lemma stands. □

Theorem 1: For any given norm-finite $\hat{\boldsymbol{X}}_{0|0}$, if $\boldsymbol{\Phi}$, $\boldsymbol{G}$ and $\boldsymbol{R}$ are bounded, $[\boldsymbol{\Phi}, \boldsymbol{H}]$ is uniformly completely observable, and $[\boldsymbol{\Phi}, \boldsymbol{G}]$ is uniformly completely controllable, then

$$\hat{\boldsymbol{X}}_{n|n} \to_d \boldsymbol{X}_n, \text{as } n \to \infty, \tag{11}$$

meaning

$$\hat{p}^{(k)}(n) \to_d p^{(k)}(n), \text{as } n \to \infty, \forall k = 0, 1, 2, ..., K. \tag{12}$$

Proof: According to [4], [5], [6] (see Chapter 4.4), with support of our Lemma 3 and Lemma 4, this theorem holds. Note that $rank(\boldsymbol{O}_{\boldsymbol{\Phi},\boldsymbol{H}}) = rank(\boldsymbol{O}_{\boldsymbol{\Phi},\boldsymbol{H}\boldsymbol{R}^{1/2}})$, and $rank(\boldsymbol{C}_{\boldsymbol{\Phi},\boldsymbol{G}}) = rank(\boldsymbol{C}_{\boldsymbol{\Phi},\boldsymbol{G}\boldsymbol{Q}^{1/2}})$, where $\boldsymbol{R}^{1/2}(\boldsymbol{R}^{1/2})' = \boldsymbol{R}$ and $\boldsymbol{Q}^{1/2}(\boldsymbol{Q}^{1/2})' = \boldsymbol{Q}$. Since $\boldsymbol{R}$ and $\boldsymbol{Q}$ are positive definite, the decomposition can be made. $\boldsymbol{O}_{\boldsymbol{\Phi},\boldsymbol{H}}$ denotes the observability matrix defined by the pair $[\boldsymbol{\Phi},\boldsymbol{H}]$. The notation conventions keep the same to $\boldsymbol{C}_{\boldsymbol{\Phi},\boldsymbol{G}}$, $\boldsymbol{O}_{\boldsymbol{\Phi},\boldsymbol{H}\boldsymbol{R}^{1/2}}$ and $\boldsymbol{C}_{\boldsymbol{\Phi},\boldsymbol{G}\boldsymbol{Q}^{1/2}}$. $\square$

IV. Additional Remarks

Remark 1: In target tracking problem [7], the canonical Static model, Constant Velocity (CV) model, and Constant Acceleration (CA) model are special cases of TVLAP. When $K = 0$, TVLAP gives the Static model. If $K = 1$, we have the CV model. If $K = 2$, TVLAP degenerates into the CA model. $\square$

References

[1] J. Durbin and S. J. Koopman, *Time Series Analysis by State Space Methods*. Oxford University Press, 2012.

[2] D. Simon, *Optimal State Estimation: Kalman, H_∞, and Nonlinear Approaches*. John Wiley & Sons, 2006.

[3] H. Roger and R. J. Charles, *Topics in Matrix Analysis*. Cambridge University Press, 1994.

[4] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *Journal of Basic Engineering*, vol. 83, no. 1, pp. 95–108, 1961.

[5] B. Anderson, "Stability properties of Kalman-Bucy filters," *Journal of the Franklin Institute*, vol. 291, no. 2, pp. 137–144, 1971.

[6] B. D. Anderson and J. B. Moore, *Optimal Filtering*. Courier Corporation, 1979.

[7] X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking. part i. dynamic models," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1333–1364, 2003.