System Identification with Student’s $t$-Process Dynamical Model

Ayumu Nono\textsuperscript{1}, and Yusuke Uchiyama\textsuperscript{2},

\textsuperscript{1}Graduate School of Engineering, The University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan
\textsuperscript{2}MAZIN Inc.
3-29-14 Nishi-Asakusa, Taito-ku, Tokyo 111-0035, Japan
E-mail: nono-ayumu303@g.ecc.u-tokyo.ac.jp

Abstract

Knowing the exact location of a spacecraft is one of the most important tasks in the operation of a satellite. However, it is difficult to accurately determine the position and velocity of a satellite far from the ground. In order to meet such requirements, various estimation methods have been developed. In this study, we propose a $t$-process dynamic estimation model, which is based on the $t$-distribution, and it is suggested that it is more robust than the Gaussian model. In the present study, we have successfully developed a particle filter-based $t$-process dynamic estimation model.

1 Introduction

Given a set of inputs and outputs we need to infer and identify the behaviour of system considered for the problem of information and controls. In the literature of the system identifications, various methodologies have been developed and proposed. One of the most famous methodology is the family of state space modeling, in particular, linear state space models have been applied to many problems of the system identifications because of its mathematical simplicity and tractability. In addition, the rigorous treatment of filtering scheme for the linear state space model has been developed by Kalman. Then the Kalman filter has been used in the area of motion control, robotics and data assimilation and so on.

Despite the usefulness of the linear state space modeling, nonlinear behaviour is often observed in real systems. To identify the behaviour of nonlinear systems, kernel method has been developed. In this framework, the domain of inputs is mapped from a finite dimensional vector space to a infinite dimensional Hilbert space, and then linear modeling is implemented on the mapped domain. Thus, the kernel method is considered to be a linear modeling on the infinite dimensional Hilbert space.

Adding to the Gaussian noise, the kernel method is extended to stochastic systems as the Gaussian process regression (GPR), which is classified into category of non-parametric Bayesian modeling. As with the linear state state model, the GPR incorporates state variables and then its filtering scheme is designed.

Because of the definition of the the GPR, it is restricted to be utilized to observed data under Gaussian noise. However, non-Gaussian fluctuations are also observed in the real world, such as fluid turbulence, financial time series, and cosmic rays. To identify the system under non-Gaussian noise, the Student’s $t$-process regression (TPR) has been proposed in the literature of the non-parametric Bayesian modeling. As with the GPR, the TPR also consists of the kernel functions to map the finite dimensional input data onto infinite dimensional Hilbert space. In addition, a latent variable model of the TPR has been proposed in the area of mathematical finance recently, and state space modeling on the form of an ordinary differential equation has been considered.

In this study, we propose a Student’s $t$-process dynamical model (TPDM) as a state space model of the TPR. Further, a simultaneous online estimation for state variables and parameters of the TPDM with the aid of particle filtering. Section 2 provides a brief introduction of the TPR, where the basic concept of the GPR is also explained. In Sec. 3, we propose the TPDM as a natural extension of the TPR with latent variables. Followed by this section, we test the TPDM for a problem of state estimation for satellite orbits in Sec. 4. Finally, section 5 is devoted to conclusions of this study and future perspectives.

2 Basis of the Student’s $t$-process regression

2.1 Gaussian process regression

In the problem of system identifications, the relations between input $x \in \mathcal{X} \subset \mathbb{R}^D$ and output $y \in \mathbb{R}$ are assumed to be described by

$$y = \phi(x) + \omega,$$  \hspace{1cm} (1)

where $\phi : \mathcal{X} \to \mathbb{R}$ is an appropriate smooth function and $\omega$ is a Gaussian white noise with zero mean and variance.
Suppose \( \phi(\cdot) \) be a random function with mean
\[
m(\bm{x}) = \mathbb{E}[\phi(\bm{x})],
\]
and covariance
\[
k(\bm{x}, \bm{x}') = \mathbb{E}[\phi(\bm{x})\phi(\bm{x}')] = \mathbb{E}[\phi(\bm{x})\phi(\bm{x}')] = \mathbb{E}[\phi(\bm{x})\phi(\bm{x}')] = \mathbb{E}[\phi(\bm{x})\phi(\bm{x}')],
\]
where \( m : \mathcal{X} \rightarrow \mathbb{R} \) is a mean function and \( k : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R} \) is a kernel function which is symmetric positive definite. The random function \( \phi(\cdot) \) is a Gaussian process if for a set of input \{\( x_1, x_2, \ldots, x_N \)\} with arbitrary integer \( N \) the set of mapped values \{\( \phi(x_1), \phi(x_2), \ldots, \phi(x_N) \)\} follow the Gaussian distribution
\[
\mathcal{N}(m_X, K_{X,X}) = \frac{1}{(2\pi)^{N/2}|K_{X,X}|} \exp\left(-\frac{1}{2}(\bm{y} - m_X)^T K_{X,X}^{-1} (\bm{y} - m_X)\right),
\]
where \( m_X \) and \( K_{X,X} \) are respective, function vector, mean vector and covariance matrix. They are defined by
\[
\phi_X = \begin{bmatrix} \phi(x_1) \\ \phi(x_2) \\ \vdots \\ \phi(x_N) \end{bmatrix}, \quad m_X = \begin{bmatrix} m(x_1) \\ m(x_2) \\ \vdots \\ m(x_N) \end{bmatrix},
\]
and
\[
K_{X,X} = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_N) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_N) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_N, x_1) & k(x_N, x_2) & \cdots & k(x_N, x_N) \end{bmatrix},
\]
with \( X \) being a data matrix \( X = [x_1, x_2, \ldots, x_N]^T \). The mean function \( m(\cdot) \) can set to be zero by \( \phi(x) \rightarrow \phi(x) - m(x) \) without loss of generality.

Given pairs of inputs and outputs \( \mathcal{D} = \{ (x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N) \} \), the GPR for Eq. (1) is implemented by the Gaussian process used as prior distribution:
\[
y \sim \mathcal{N}(m_X, K_{X,X} + \sigma_0^2 I),
\]
where \( y = [y_1, y_2, \ldots, y_N]^T \) is an output vector and \( I \) is the \( N \)-dimensional identity matrix. The law of summation for Gaussian distributed random variables yields Eq. (8).

The prediction task of the GPR is implemented by its conditional probability density function. For an additional input data \( \bm{x}^* \) one can obtain predicted output \( y^* \) by the identified system model. In the Bayesian sense, the predicted output \( y^* \) is sampled from the conditional probability density function
\[
y^* \sim p(y^* | \bm{x}^*, X, y).
\]
The conditional probability density function \( p(y^* | \bm{x}^*, X, y) \) is also the Gaussian distribution as
\[
y^* \sim \mathcal{N}(m^*, k^*),
\]
where the mean and variance are derived as
\[
m^* = m_{x^*} + K_{X,x^*}^T (K_{X,X} + \sigma_0^2 I)^{-1}(y - m_X),
\]
\[
k^* = K_{X,x^*} - K_{X,x^*}^T (K_{X,X} + \sigma_0^2 I)^{-1} K_{X,x^*}.
\]
It is confirmed that Eqs. (11) and (12) map the additional input to the predicted value.

### 2.2 Student's t-process regression

In what follows we assume the mean function to be zero for the sake of brevity. Suppose the regression model in Eq. (1) is parameterized by \( \lambda \equiv \sigma_0^2 \) and \( K_{X,X} \rightarrow \lambda^{-1} K_{X,X} \). The \( \lambda \)-parameterized GPR is given as
\[
y \sim \mathcal{N}(\lambda(K_{X,X} + \sigma_0^2 I)|\lambda). \tag{13}
\]
Note here that the notation of the mean functions is omitted in Eq. (13). The parameter \( \lambda \) is assumed to be a random variable being subjected to a one-parameter family of the gamma distribution defined by
\[
p(\lambda) = \frac{\nu/2)^\nu}{\Gamma(\nu/2)} \lambda^{\nu/2 - 1} e^{-\lambda/2}, \tag{14}
\]
with \( \nu \) being a positive real parameter called as degree of freedom. Marginalizing the conditional Gaussian distribution in Eq. (13) with respect to \( \lambda \) by the gamma distribution in Eq. (14), we obtain
\[
\int_0^\infty \mathcal{N}(\lambda(K_{X,X} + \sigma_0^2 I)|\lambda)p(\lambda)d\lambda = \mathcal{T}(I + K_{X,X}; \nu), \tag{15}
\]
where \( \mathcal{T}(I + K_{X,X}; \nu) \) is the Student’s t-distribution defined by
\[
\mathcal{T}(I + K_{X,X}; \nu) = \frac{\Gamma(\frac{\nu+N}{2})}{(\pi \nu)^{N/2} \Gamma(\frac{\nu}{2})} \left[ 1 + \frac{1}{\nu} y^T(I + K_{X,X})^{-1} y \right]^{-\frac{\nu+N}{2}}. \tag{16}
\]
In this way, the GRP leads to the Student’s t-distribution with the kernel functions in Eq. (16), which is known as the TPR.

Through the same procedure of the GPR, the prediction task of the TPR is implemented by its conditional probability density function. For an additional input \( \bm{x}^* \), the conditional probability density function for the predicted output \( y^* \) is given as
\[
y^* \sim \mathcal{T}(m^*, k^*; \nu^*), \tag{17}
\]
where $m^*$, $k^*$, and $\nu^*$ are derived as

$$m^* = k_T^x x^* (I + K_{X,X})^{-1} y,$$  \hspace{1cm} (18)

$$k^* = \frac{\nu - \beta^* - 2}{\nu - N - 2} [k_{x^* x^*} - k_T^x x^* (I + K_{X,X})^{-1} k_{X,x^*}].$$  \hspace{1cm} (19)

$$\beta^* = y_T^T (I + K_{X,X})^{-1} y,$$  \hspace{1cm} (20)

$$\nu^* = \nu + N.$$  \hspace{1cm} (21)

It is notably confirmed that $k^*$ and $\nu^*$ include correction terms with respect to the size of given data set.

### 2.3 Student’s $t$-process latent variable model

For the situation that only output data are observed, the Student’s $t$-process latent variable model (TPLVM) has been proposed as a method of unsupervised learning in the sense of non-parametric Bayesian models. Without input data the matrix $X$ in Eq. (16) is regarded as a parameter which should be estimated. To estimate the hyper parameters of the TPLVM including $X$, variational inference has been installed by the Kullback-Leibler (KL) divergence as

$$\text{KL}[q(X) || p(X|y)] = - \int q(X) \log \frac{p(y|X)p(X)}{q(X)} dX$$

$$+ \log p(y),$$  \hspace{1cm} (22)

where $p(X)$ and $q(X)$ are a prior and approximated distributions for $X$. Minimization for the KL divergence yields a plausible inference for $q(X)$.

### 3 Student’s $t$-process dynamical model

Incorporating dynamical latent variables into the TPLVM, we develop a Student’s $t$-process dynamical model (TPDM): Suppose $(x_t, y_t)$ be a pair of state and observed variables at time $t$, the TPDM is defined by the following state space model:

$$x_{t+1} \sim T(K_{x x}, \nu_x)$$  \hspace{1cm} (23)

$$y_{t+1} \sim T(K_{x y}, x_{t+1}, \nu_y).$$  \hspace{1cm} (24)

The dimensions of the state and observed variables are allowed to be different. As with other state space models, filtering, smoothing, and predicting schemes for the TPDM are able to be designed. In general the kernel functions contain its intrinsic parameters. For instance, the radial basis function, one of the most used kernel functions, is of the form:

$$k_{\text{RBF}}(x, x') = \alpha^2 \exp \left(-\frac{|x - x'|^2}{l^2}\right),$$  \hspace{1cm} (25)

where $\alpha$ and $l$ are positive real parameters. Within the degree of freedoms $\nu_x$ and $\nu_y$, the parameters of the kernel functions is composed of parameter vector $\theta$. The specific value of $\theta$ can be estimated by the method of maximum likelihood within the set of observed data $\{y_1, y_2, \cdots, y_T\}$, or other estimation methods.

It is pointed out that point estimations such as the method of maximum likelihood is likely to fall in over-estimation. To overcome this problem, we employ particle filtering as an online Bayesian estimation. On the processes of the particle filtering we treat the parameter vector $\theta$ as a dynamical variable whose evolution is described by

$$\theta_{t+1} \sim N(a\theta_t + (1 - a)\bar{\theta}, \bar{V}_{\theta}),$$  \hspace{1cm} (26)

where $a$ is a positive real parameter in the open interval $(0,1)$, $\bar{\theta}$ and $\bar{V}_{\theta}$ are ensemble mean and covariance of $M$ particles, respectively. By this method, the state variables and parameters are estimated simultaneously.

### 4 Application for satellite orbit estimation

Here, we adopt the TPDM to the problem of satellite orbit estimation. In this problem state variables represent two-dimensional position and velocity elements[3];

$$X = [x, y, v_x, v_y]$$  \hspace{1cm} (27)

In addition, the motion of a satellite is assumed to be a simple circular motion of a qualitative point and to be given by

$$\frac{d\vec{v}}{dt} = -G \frac{mM}{r^3} \vec{r},$$  \hspace{1cm} (28)

where $G$ is the gravitational constant, $m$ is the mass of the satellite, $M$ is the mass of the earth, and $r$ is the position vector of the satellite when the center of the earth is at the origin.

In this setup, the true position of the satellite is not able to be observed, and the information available is the observed quantity $Y$, which is composed of the state vector and a white noise $w$:

$$Y = X + w$$  \hspace{1cm} (29)

Representative simulated results are shown in Fig. 1. The green line shows the true position of the satellite, the gray line is the observed position information with the white noise, and the red line is the position of the satellite estimated from the $t$-process dynamic model from the observed information. The value on the vertical axis represents the $x$-coordinate of the two-dimensional coordinates. There are some deviations, but generally the true values are captured.

### 5 Conclusions

In this study, we have developed the TPDM as an extension of the TPLVM for state space modeling. As an example of the TPDM, we estimated the true satellite orbit from the observed data consist of the position of the satellite and the observed noise.
We will continue to work on increasing the dimension of the state variables and adapting the TPDM to more complicated models, and compare the model with other filtering methods to find out its advantageous points.

References

[1] R. E. Kalman and R. S. Bucy: New Results in linear filtering and prediction, *Trans. ASME, J. Basic Eng.*, 82 D, pp. 95–108, 1960.

[2] A. H. Jazwinsky: *Stochastic Process and Filtering Theory*, Academic Press, N.Y., 1970.

[3] Xiaolin Ning and Xin Ma: *Analysis of Filtering Methods for Satellite Autonomous Orbit Determination Using Celestial and Geomagnetic Measurement*, Hindawi Publishing Corporation Mathematical Problems in Engineering, Volume 2012, 2012.