State Estimation using Gaussian Process Regression for Colored Noise Systems

Kyuman Lee
School of Aerospace Engineering
Georgia Institute of Technology
Atlanta, GA 30332
404-422-3697
klee400@gatech.edu

Eric N. Johnson
School of Aerospace Engineering
Georgia Institute of Technology
Atlanta, GA 30332
404-385-2519
eric.johnson@ae.gatech.edu

Abstract—**The goal of this study is to use Gaussian process (GP) regression models to estimate the state of colored noise systems. The derivation of a Kalman filter assumes that the process noise and measurement noise are uncorrelated and both white. In relaxing those assumptions, the Kalman filter equations were modified to deal with the non-whiteness of each noise source. The standard Kalman filter ran on an augmented system that had white noises and other approaches were also introduced depending on the forms of the noises. Those existing methods can only work when the characteristics of the colored noise are perfectly known. However, it is usually difficult to model a noise without additional knowledge of the noise statistics. When the parameters of colored noise models are totally unknown and the functions of each underlying model (nonlinear dynamic and measurement functions) are uncertain or partially known, filtering using GP-Color models can perform regardless of whatever forms of colored noise. The GPs can learn the residual outputs between the GP models and the approximate parametric models (or between actual sensor readings and predicted measurement readings), as a member of a distribution over functions, typically with a mean and covariance function. Lastly, a series of simulations, including Monte Carlo results, will be run to compare the GP based filtering techniques with the existing methods to handle the sequentially correlated noise.**

TABLE OF CONTENTS

1. INTRODUCTION

Estimating the state of a dynamic system is a fundamental problem at modern GNC areas. The most successful techniques for state estimation are Bayesian filters such as a Kalman filter. The key assumption of the standard Kalman filter is that the process noise and measurement noise are uncorrelated and both white. These are often too restrictive assumptions, and applications with non-white noise frequently arise in practice.

Many practical systems exist in which the correlation times of the random measurement errors are not short compared to the times of interest in the system; for brevity such errors are called "colored" noise [1]. In audio engineering, electronics, physics, and many other fields, the color of a noise signal is generally understood to be some broad characteristic of its power spectrum. Different colors of noise have significantly different properties: for example, as audio signals they will sound different to human ears, and as images they will have a visibly different texture. In image processing, the quality of images will directly influence the accuracy of locating landmarks. During sampling and transmission, images are often degraded by noises which may originate from a multiplicity of sources. These noises are colorful and their variances are not known beforehand. The original images having the colored noise must be filtered to ensure the accuracy of location and measurement [2]. Furthermore, in radar and sonar signal processing, various colored noise are: external spurious signals that, intentionally or not, jam reception; echoes from various reflectors in the landscape; and, in sonar, reverberation [3]. Therefore, each application using such signals typically requires the noise of a specific color and it is necessary to extend conventional Kalman filtering approach in order to solve the problem of the color noise effectively.

To deal with colored noise over several decades, the standard Kalman filter ran on an augmented system that had white noises. Bryson et al [1] introduced the augmented filter for colored measurement noise first. For example, due to the non-stationary nature of the speech signal, augmented Kalman filtering of colored noise for speech enhancement was also described [4], [5]. Recently, Liu et al [6] presented a self-tuning Kalman filter for auto regressive or moving average (ARMA) signals with colored noise. Also, if the process noise is colored, then it is straightforward to modify the Kalman filter equations and obtain an equivalent but high-order system with white process noise [7]. On the other hand, the augmented state procedure may lead to ill-conditioned computations in constructing the data processing filter. It is a singular problem as the case of no noise in some of the measurements because the correlation matrix of the measurement noise is singular, i.e., R^{-1} does not exist, in the notation of a Kalman filter. Instead, another filtering, measurement differencing [8], [9], that was capable of converting the received colored noise into white noise, was established. The contribution of the colored part of the noise to the received noisy signal was first estimated, and this estimate was then subtracted from the received signal. This approach, by subtraction of the colored part of the noise, was strictly equivalent to the whitening approach and had lower dimension than the augmented state filters. Simon [10] provided a survey of all previously noted methods. Details will be described in Section 3. as preliminaries.

Those existing methods can only work when the characteris-

tics (power and spectral density) of colored noise are assumed to be known. However, it is typically difficult to model a noise without a priori knowledge of the noise statistics or a supplementary measurement.

A Gaussian process (GP) is a function approximation which can be thought of as a distribution over functions. The true function at any given point in the domain exists in the Gaussian distribution modeled by mean and covariance functions evaluated at the same point [11]. GPs are used in several disciplines when some underlying function is unknown but needed. In implementation, a GP can be formulated as the regression of training data (points in the state space and function evaluations) with respect to a basis function. A GP can be dynamically updated with new function evaluations or modeled once at the onset of an algorithm.

The machine learning community has applied GP to both controls and estimation processes in the past. When a system has unknown dynamics and measurement models, a GP can be used to learn them. The GP mean provides approximates of the state transition matrix and measurement model, while the GP covariance provides estimates of the process and measurement white noise. Estimation methods have ranged from Bayesian filtering with GP to extended and unscented Kalman filtering with GP [12], [13]. Likely, when the parameters of colored noise models are totally unknown and the functions of each underlying model (nonlinear dynamic and measurement functions) are uncertain or partially known, filtering using enhanced GP-Color models can perform regardless of whatever forms of colored noise. The GPs can learn the residual outputs between the GP models and the approximate parametric models (or between prior estimated states by the GPs and the outputs of parametric models), as a member of a distribution over functions, typically with a mean and covariance function. The Q and R noise covariance matrices in the filter can also be learned by the GP-Color prediction model and observation model, respectively. Input-dependent noise was described in the article of Kersting et al [14], but we are not aware of previous work with colored noise functions using GPs.

The rest of this paper is organized as follows. The next two sections demonstrate the formulation of the problem and the outline of the approach. Section 4 explains how each model with colored noise can be learned by GPs and section 5 presents a novel filtering algorithm using GPs for colored noisy systems. Next, the simulation environment is described and results are presented. Finally, conclusion and future work are discussed.

2. Models and Setup

We consider discrete-time stochastic dynamics systems of the form

$$x_k = f(x_{k-1}) + w_{k-1} \tag{1}$$
$$z_k = h(x_k) + v_k \tag{2}$$

where $x_k \in \mathbb{R}^n$ is the state and $z_k \in \mathbb{R}^m$ is the measurement at time step $k = 1, \cdots, T_0$. f is the nonlinear dynamic function and h is the nonlinear measurement function. Both are partially known or uncertain.

Now suppose that we have colored process and measurement noises. The process noise is itself the output of the dynamic system and the measurement noise is itself the output of the observation system.

$$w_k = \phi(w_{k-1}) + \eta_k \tag{3}$$
$$v_k = \psi(v_{k-1}) + \zeta_k \tag{4}$$
$$\eta_k \sim \mathcal{N}(0, Q_k)$$
$$\zeta_k \sim \mathcal{N}(0, R_k)$$
$$\mathbb{E}[\eta_k \eta_j^T] = Q_k \delta_{k-j}$$
$$\mathbb{E}[\zeta_k \zeta_j^T] = R_k \delta_{k-j}$$
$$\mathbb{E}[\eta_k \zeta_j^T] = 0$$

where η_k is zero-mean white noise that is uncorrelated with w_{k-1} and ζ_k is zero-mean white noise that is uncorrelated with v_{k-1}. ϕ is the nonlinear colored process noise function and ψ is the nonlinear colored measurement noise function. Both functions are totally unknown, and we don't even know if each signal has colored noise in the form of Eqs. (3) and (4). The covariances of the noise functions are given as

$$\begin{aligned}\mathbb{E}[w_k w_{k-1}^T] &= \mathbb{E}[\phi(w_{k-1}) w_{k-1}^T + \eta_k w_{k-1}^T] \\ &= \mathbb{E}[\phi(w_{k-1}) w_{k-1}^T] \approx \left(\frac{\partial \phi}{\partial w}|_{w_{k-1}}\right) \mathbb{E}[w_{k-1} w_{k-1}^T] \neq 0 \\ \mathbb{E}[v_k v_{k-1}^T] &= \mathbb{E}[\psi(v_{k-1}) v_{k-1}^T + \zeta_k v_{k-1}^T] \\ &= \mathbb{E}[\psi(v_{k-1}) v_{k-1}^T] \approx \left(\frac{\partial \psi}{\partial v}|_{v_{k-1}}\right) \mathbb{E}[v_{k-1} v_{k-1}^T] \neq 0\end{aligned}$$

We see that w_k or v_k is colored noise because it is correlated with itself at other time steps.

3. Preliminaries

Gaussian Process Regression

A Gaussian process is a nonparametric tool for learning regression functions from sample data. Consider now the case where we have measurements of the observation which are corrupted with white noise

$$y_i = h(x_i) + \nu_i = h_i + \nu_i, \ \ \forall\, i = 1, \cdots, N$$

where $\nu_i \sim \mathcal{N}(0, \beta^{-1})$. Since the white noise is independent of each data point, we have that

$$\begin{aligned} p(y_{1:N}|h_{1:N}) &= \mathcal{N}(h_{1:N},\ \beta^{-1} I_{N\times N}) \\ p(h_{1:N}) &= \mathcal{N}(0, K) \\ \Rightarrow p(y_{1:N}) &= \mathcal{N}(0, C_N) \end{aligned} \tag{5}$$

where the covariance matrix $C_N \in \mathbb{R}^{N\times N}$ and it is defined as

$$C_N = K + \beta^{-1} I_{N\times N}$$

Therefore every element of the covariance matrix C will have the form

$$C(x_i, x_j) = k(x_i, x_j) + \beta^{-1}\delta_{i,j}$$

The most widely used kernel function is the squared exponential, which has the form

$$k(x_i, x_j) = \theta_0 \exp(-\frac{\theta_1}{2}\|x_i - x_j\|^2)$$

The hyperparameters, $\theta = [\theta_0, \theta_1, \beta^{-1}]$, can be learned by maximizing the log marginal likelihood of the training outputs given the inputs:

$$\theta_{max} = \arg\max_{\theta}\{\ln p(y_{1:N}|x_{1:N}, \theta)\} \tag{6}$$

The goal in regression is to predict y_{N+1} for a new input point x_{N+1} given the set of training data, $D = \langle x_{1:N}, y_{1:N}\rangle$. From Eq. (5),

$$p(y_{1:N+1}) = \mathcal{N}(0, C_{N+1})$$

where

$$C_{N+1} = \begin{pmatrix} C_N & k_* \\ k_*^T & c \end{pmatrix}$$

where $c = k(x_{N+1}, x_{N+1}) + \beta^{-1}$ and then $p(y_{N+1}) = \mathcal{N}(0, c)$. Now we can claim the conditional distribution is a Gaussian distribution with mean and covariance specified as follows

$$\text{GP}_{\mu}(y_{N+1}|D, \theta) = k_*^T C_N^{-1} y_{1:N} \tag{7}$$

$$\text{GP}_{\Sigma}(y_{N+1}|D, \theta) = c - k_*^T C_N^{-1} k_* \tag{8}$$

where $k_* \in \mathbb{R}^N$ and it has elements $k(x_1, x_{N+1}), k(x_2, x_{N+1})$, ..., $k(x_N, x_{N+1})$.

Enhanced GP Models

The GP for regression has a zero mean assumption in Eq. (7), and if a query state is far away from the training states, then the output of GP quickly tends towards zero. This makes the choice of training data for the GP very important. A parametric model is one which attempts to represent a particular phenomenon with physical equations. The disadvantage of parametric models is that substantial domain expertise is required to build these models, and even then they are often simplified representations of the actual systems.

Higher prediction accuracy can be obtained by combining GP models with parametric models. We call the combination of GP and parametric models enhanced-GP (EGP) models [12], [13]. The EGP models alleviate some of the problems found when using either model alone. Using this idea of EGP method, we can develop novel GP models for colored noise systems. Details will be described in Section 4.

Existing Methods for Colored Noise Models

There are a couple of ways to solve the colored noise problem when each parameter of Eqs. (1), (2), (3), and (4) is perfectly known. Here we will solve the discrete-time problem by augmenting the state. First, we augment the original dynamic model as follows:

$$F_{k-1}^{\text{aug}} = \begin{bmatrix} \frac{\partial f}{\partial x}|_{\hat{x}_{k-1}^+} & I & 0 \\ 0 & \frac{\partial \phi}{\partial w}|_{\hat{w}_{k-1}^+} & 0 \\ 0 & 0 & \frac{\partial \psi}{\partial v}|_{\hat{v}_{k-1}^+} \end{bmatrix},$$

$$Q_{k-1}^{\text{aug}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & Q_{k-1} & 0 \\ 0 & 0 & R_{k-1} \end{bmatrix}$$

where the hat " ˆ " denotes a estimate or an approximate. Then, perform the time update of the state estimate and estimation-error covariance from time $(k-1)^+$ to time k^-

$$\hat{x}_k^- = f(\hat{x}_{k-1}^+) + \phi(\hat{w}_{k-2}^+)$$

$$P_k^- = F_{k-1}^{\text{aug}} P_{k-1}^+ (F_{k-1}^{\text{aug}})^T + Q_{k-1}^{\text{aug}}$$

where superscript $-$ or $+$ denote the a priori or a posteriori estimates, respectively. At time k^-, compute the following Jacobian matrix and augment the original measurement model as follows:

$$H_k^{\text{aug}} = \begin{bmatrix} \frac{\partial h}{\partial x}|_{\hat{x}_k^-} & 0 & I \end{bmatrix}$$

Then, perform the measurement update of the state estimate and estimation-error covariance

$$\hat{z}_k = h(\hat{x}_k^-) + \psi(\hat{v}_{k-1}^+)$$

$$K_k = P_k^- (H_k^{\text{aug}})^T (H_k^{\text{aug}} P_k^- (H_k^{\text{aug}})^T)^{-1}$$

$$\begin{bmatrix} \hat{x}_k^+ \\ \hat{w}_k^+ \\ \hat{v}_k^+ \end{bmatrix} = \begin{bmatrix} \hat{x}_k^- \\ \hat{w}_k^- \\ \hat{v}_k^- \end{bmatrix} + K_k(z_k - \hat{z}_k)$$

$$P_k^+ = P_k^- - K_k H_k^{\text{aug}} P_k^-$$

The colored noise problem ends up being solved with the above augmented extended Kalman filter (EKF). However, this method can only work when the characteristics of colored noise are assumed to be known.

4. Learning GP Models for Colored Noise Systems (GPC)

Training Data

The training data is a sampling from the dynamics and observations of the system. The training data for each GP consists of a set of input-output relations.

Given $x_{1:T_0}$ and $z_{1:T_0}$, let the approximate parametric prediction and observation models be denoted $\hat{f}$ and $\hat{h}$. These are assumed to be known but not perfect based on partial knowledge of actual systems in Eqs. (1) and (2). The approximate process and measurement residuals are denoted $\hat{w}$ and $\hat{v}$ and defined as

$$\hat{w}_{1:T_0-1} = x_{2:T_0} - \hat{f}(x_{1:T_0-1})$$

$$\hat{v}_{1:T_0} = z_{1:T_0} - \hat{h}(x_{1:T_0})$$

where $x_{1:T_0} = [x_1, x_2, ..., x_{T_0}]$ and $z_{1:T_0} = [z_1, z_2, ..., z_{T_0}]$.

By Eqs (3) and (4), each residual is the function of itself at previous one time step. The training data sets of each residual are given as

$$\hat{D}_W = \langle \hat{w}_{1:T_0-2},\ \hat{w}_{2:T_0-1}\rangle$$

$$\hat{D}_V = \langle \hat{v}_{1:T_0-1},\ \hat{v}_{2:T_0}\rangle$$

By optimizing similar to Eq. (6), the hyperparameters θ_w and θ_v can be learned.

GPC Models

The idea of GPC method is to use a GP to learn the residual output between the GP model and the approximate parametric model. In particular, the residuals of prediction model can be approximate outputs between prior estimated states by the GPs and the outputs of parametric models. Also, the residuals of observation model can be approximate outputs between actual sensor readings and predicted measurement readings. Then, from Eqs. (1) and (2), GPC models at time step $k = T_0 + 1, \cdots, T$ become

$$\begin{aligned} x_k &= f(x_{k-1}) + \phi(w_{k-2}) + \eta_{k-1} \\ &= \hat{f}(x_{k-1}) + [f(x_{k-1}) - \hat{f}(x_{k-1})] + \phi(w_{k-2}) + \eta_{k-1} \\ \Rightarrow\ x_k &= \hat{f}(x_{k-1}) + \text{GP}_\mu\left(\hat{w}_{k-2}, \hat{D}_W\right) + \hat{\eta}_{k-1} \end{aligned} \tag{9}$$

$$\begin{aligned} z_k &= h(x_k) + \psi(v_{k-1}) + \zeta_k \\ &= \hat{h}(x_k) + [h(x_k) - \hat{h}(x_k)] + \psi(v_{k-1}) + \zeta_k \\ \Rightarrow\ z_k &= \hat{h}(x_k) + \text{GP}_\mu\left(\hat{v}_{k-1}, \hat{D}_V\right) + \hat{\zeta}_k \end{aligned} \tag{10}$$

where

$$\hat{\eta}_k \sim \mathcal{N}\left(0,\ \text{GP}_\Sigma(\hat{w}_{k-1}, \hat{D}_W)\right)$$
$$\text{and}\quad \hat{\zeta}_k \sim \mathcal{N}\left(0,\ \text{GP}_\Sigma(\hat{v}_{k-1}, \hat{D}_V)\right).$$

where all parameters of ϕ, ψ, η, and ζ of colored noise are unknown since it is typically difficult to model a noise without a priori knowledge of the noise statistics or a supplementary measurement.

The key advantages of GPC are their modeling flexibility, their ability to provide uncertainty estimates, and their ability to learn noise and smoothness parameters from training data.

5. GPC-BASED BAYES FILTERS

Algorithm of GPC-EKF

The following describes the integration of GPC prediction and observation models into the EKF. In addition to the GPC mean and covariance estimates used in Eqs. (9) and (10), the incorporation of GPC models into the EKF requires a linearization of the GPC prediction and observation model in order to propagate the state and observation, respectively. For the EKF, this linearization is computed by taking the first term of the Taylor series expansion of the GP function.

By the derivation of Ko et al [12], the Jacobian of the GP mean function (7) can be expressed as

$$\frac{\partial \text{GP}_\mu(x_*, D)}{\partial(x_*)} = \frac{\partial(k_*)}{\partial(x_*)}^T C_N^{-1} y_{1:N} \tag{11}$$

$$\text{where}\quad \frac{\partial(k_*)}{\partial(x_*)} = \begin{bmatrix} \frac{\partial(k(x_1,x_*))}{\partial(x_*[1])} & \cdots & \frac{\partial(k(x_1,x_*))}{\partial(x_*[d])} \\ \vdots & \ddots & \vdots \\ \frac{\partial(k(x_N,x_*))}{\partial(x_*[1])} & \cdots & \frac{\partial(k(x_N,x_*))}{\partial(x_*[d])} \end{bmatrix}$$

$$\frac{\partial(k(x,x_*))}{\partial(x_*[i])} = -\theta_1(x_*[i] - x[i])\, k(x,x_*),\ \ i = 1, \cdots, d$$

where k_* is the vector of kernel values between the query input, $x_* = \hat{w}_{k-1}$ or $\hat{v}_{k-1}$, and the training inputs, $\hat{w}_{1:T_0-2}$ or $\hat{v}_{1:T_0-1}$, respectively. $N(= T_0 - 2$ or $T_0 - 1)$ is the number of training data, and $d(= n$ or $m)$ is the dimensionality of the input space.

Eq. (11) defines the d-dimensional Jacobian vector of the GP mean function for a single output dimension. The full $n \times n$ Jacobian of a prediction model or $m \times n$ Jacobian of a observation model is determined by stacking n or m Jacobian vectors together, one for each of the output dimensions.

We are now prepared to incorporate the GPC models into augmented EKF, as shown in Algorithm 1.

Algorithm 1 The GPC-EKF

$(\hat{w}^+_{T_0-1}, \hat{v}^+_{T_0}, \hat{x}^+_{T_0}, P^+_{T_0}, z_{T_0+1})$ are given:
Require: $X^{\text{aug}} = [\hat{x} \quad \hat{w} \quad \hat{v}]^T$
$\hat{w}^+_{T_0} = \text{GP}_\mu(\hat{w}^+_{T_0-1}, \hat{D}_W)$
for $k = (T_0 + 1) : T$ **do**
Time Updates:
$\hat{w}^-_k = \text{GP}_\mu(\hat{w}^+_{k-1}, \hat{D}_W)$
$\hat{Q}_{k-1} = \text{GP}_\Sigma(\hat{w}^+_{k-1}, \hat{D}_W)$
$\hat{\Phi}_{k-1} = \frac{\partial \text{GP}_\mu(\hat{w}^+_{k-1}, \hat{D}_W)}{\partial \hat{w}^+_{k-1}}$
$\hat{v}^-_k = \text{GP}_\mu(\hat{v}^+_{k-1}, \hat{D}_V)$
$\hat{R}_{k-1} = \text{GP}_\Sigma(\hat{v}^+_{k-1}, \hat{D}_V)$
$\hat{\Psi}_{k-1} = \frac{\partial \text{GP}_\mu(\hat{v}^+_{k-1}, \hat{D}_V)}{\partial \hat{v}^+_{k-1}}$
$\hat{x}^-_k = \hat{f}(\hat{x}^+_{k-1}) + \hat{w}^+_{k-1}$
$F^{\text{aug}}_{k-1} = \begin{bmatrix} \frac{\partial \hat{f}}{\partial x}\big|_{\hat{x}^+_{k-1}} & I_{n\times n} & 0_{n\times m} \\ 0_{n\times n} & \hat{\Phi}_{k-1} & 0_{n\times m} \\ 0_{m\times n} & 0_{m\times n} & \hat{\Psi}_{k-1} \end{bmatrix}$
$Q^{\text{aug}}_{k-1} = \text{diag}(0_{n\times n}, \hat{Q}_{k-1}, \hat{R}_{k-1})$
$P^-_k = F^{\text{aug}}_{k-1} P^+_{k-1} (F^{\text{aug}}_{k-1})^T + Q^{\text{aug}}_{k-1}$
Measurement Updates:
$\hat{z}_k = \hat{h}(\hat{x}^-_k) + \hat{v}^-_k$
$H^{\text{aug}}_k = \begin{bmatrix} \frac{\partial \hat{h}}{\partial x}\big|_{\hat{x}^-_k} & 0_{m\times n} & I_{m\times m} \end{bmatrix}$
$K_k = P^-_k (H^{\text{aug}}_k)^T (H^{\text{aug}}_k P^-_k (H^{\text{aug}}_k)^T)^{-1}$
$\begin{bmatrix} \hat{x}^+_k \\ \hat{w}^+_k \\ \hat{v}^+_k \end{bmatrix} = \begin{bmatrix} \hat{x}^-_k \\ \hat{w}^-_k \\ \hat{v}^-_k \end{bmatrix} + K_k(z_k - \hat{z}_k)$
$P^+_k = P^-_k - K_k H^{\text{aug}}_k P^-_k$
end for

$\hat{Q}_k$ is the additive white noise part of the process noise, which corresponds directly to the GPC uncertainty. $\hat{\Phi}_k$, the linearization of the prediction model, is the Jacobian of the GPC mean function found in Eq. (11). Similarly, the noise covariance, $\hat{R}_k$, and the linearization of the observation model, $\hat{\Psi}_k$, are computed by using the GPC observation model.

GPs are typically defined for scalar outputs, and GPC based Bayes Filters represent models for vectorial outputs by learning a separate GPC for each output dimension. This forces the resulting noise covariances, $\hat{Q}_k$ and $\hat{R}_k$, to be modeled as independent diagonal matrices.

6. Simulation Results

A variety of cases have been simulated to examine the influence of the color noise and to test the effectiveness of the GPC-EKF algorithm for colored-noise systems. The results will be given by means of Monte Carlo methods.

GPC only

To verify the GPC mean and variance functions given in Eqs. (9) & (10) as well as the hyperparameter learning via optimization, one example of GPC regression is given as

$$\begin{aligned}
\text{True:}\ & y_k = \sin(x_k) + 0.99 v_{k-1} + \zeta_k \text{ where } \zeta_k \sim \mathcal{N}(0,\ 0.1) \\
\text{GP:}\ & y_k = \text{GP}_\mu\left(x_k, D\right) + \nu_k \text{ where } \nu_k \sim \mathcal{N}(0, \text{GP}_\Sigma\left(x_k, D\right)) \\
\text{GPC:}\ & y_k = 0.9\sin(x_k) + \text{GP}_\mu(\hat{v}_{k-1}, \hat{D}_V) + \hat{\zeta}_k \\
& \qquad \text{where } \hat{\zeta}_k \sim \mathcal{N}(0, \text{GP}_\Sigma(\hat{v}_{k-1}, \hat{D}_V))
\end{aligned}$$

Figure 1 contains a sine function with color noise (black), noisy samples drawn from the function (red plus), the resulting GPC(blue)/GP(green) mean function estimate, and the GPC(dotted blue)/GP(dotted green) uncertainty sigma bounds. The GPC hyperparameters are determined by optimization of the data likelihood. The uncertainty gets wide where the data points are sparse.

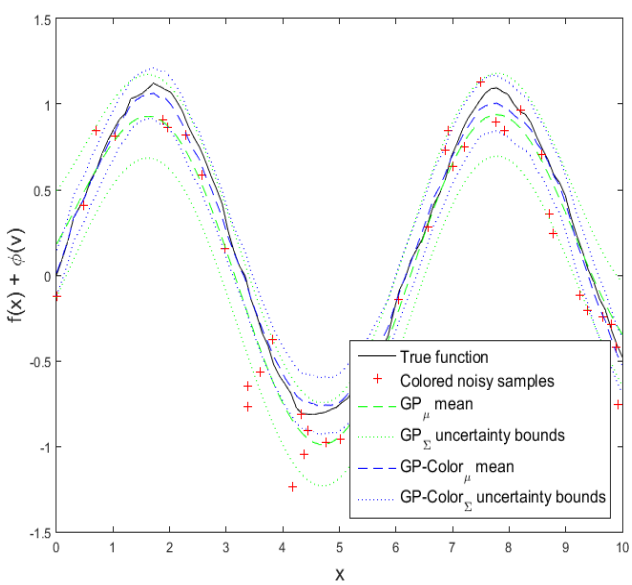


Figure 1. GP only vs. GPC with improper parametric model

	GP	GPC
Mean RMS error	0.5989	0.2865

Even though the GPC are learned by using an improper parametric model that has a 10% modeling error, the RMS error of the GPC is even smaller than that of the GP only.

EKF only - Standard and Augmented

Let the state vector be the position and velocity of a vehicle, $x(t) = [\text{pos, vel}]^T$. The vehicle flies with following linear time-invariant dynamics with colored process noise, and a simple position sensor with colored measurement noise is used as follows:

$$x_k = F_{k-1}x_{k-1} + \Phi_{k-2}w_{k-2} + \eta_{k-1} \tag{12}$$
$$y_k = H_k x_k + \Psi_{k-1}v_{k-1} + \zeta_k \tag{13}$$

where

$$F_k = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}, \quad H_k = [1 \quad 0]$$
$$\Phi_k = \begin{bmatrix} 0.99 & 0 \\ 0 & 0.99 \end{bmatrix}, \quad \Psi_k = 0.99$$
$$\eta_k \sim \mathcal{N}\left(0, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right), \quad \zeta_k \sim \mathcal{N}(0,\ 1).$$

Initialize a filter

$$x_{T_0} = [0, 0]^T, \quad \hat{x}^+_{T_0} = \mathbb{E}(x_{T_0})$$
$$P^+_{T_0} = \mathbb{E}[(x_{T_0} - \hat{x}^+_{T_0})(x_{T_0} - \hat{x}^+_{T_0})^T] = I_{2\times 2}$$

The first simulation compares a standard EKF and an augmented EKF, introduced in Section 3., of given the colored noise system, Eqs. (12) & (13). For an ideal (but not realistic) case in this simulation only, let's assume that we know perfect underlying functions, F_k and H_k, as well as exact values of all parameters(Φ_k, Ψ_k, etc.) of each colored noise. Since the augmented EKF is complicated or requires more computation, there is a question regarding why the augmented EKF is considered here. See Figure 2 for an answer.

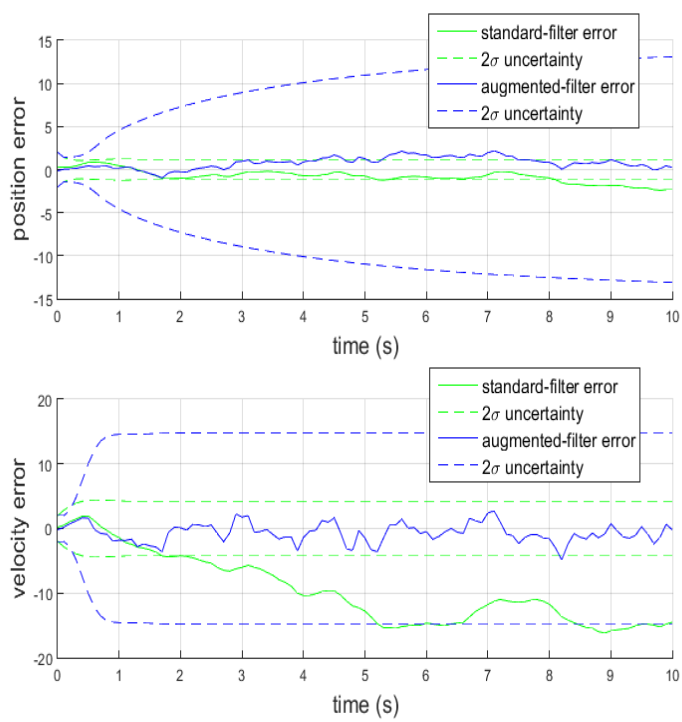


Figure 2. Standard EKF vs. Augmented EKF

The estimate errors of the standard EKF are not bounded within $2-\sigma$ uncertainties, and they are much bigger than the estimate errors of augmented EKF. Furthermore, the figure can give information on the existence of color-noise as well as on the significance of the color of noisy signals. For example, if the color noise is not remarkable, the two filters may perform similarly. Thus, even if there is no modeling error and all parameters are perfectly known, then the augmented EKF is necessary in this case.

GPC-EKF with Improper Parametric Models

This simulation presents an execution of the GPC-EKF to the example system of the colored noise problem.

If an engineer is uncertain of the underlying functions and ignores the color noise, since the engineer is not even aware of whether each signal has colored noise, EGP-based EKF, introduced in Section 3., can be run for the given system. We can call this approach EGP-White. However, the GPC-EKF can deal with the improper underlying models as well as the totally unknown parameters of colored noise in the same system.

Let's assume that the approximate parametric model, $\hat{F}$, has 10% modeling errors at each time step. From Eq. (12),

$$F_k = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \Rightarrow \hat{F}_k = \begin{bmatrix} 1 & \Delta t \\ 0 & 0.9 \end{bmatrix}$$

Figure 3 illustrates, under this circumstance, that the GPC-EKF outperforms EGP-White. In other words, the GPC-EKF will in fact result in a significant increase in estimation performance. The resulting estimation errors of GPC-EKF method are bounded by $2-\sigma$ uncertainties, and unfortunately, the estimate errors of EGP-White diverge as the simulation progresses. By comparing Figure 3(a) with Figure 2, the GPC-EKF is shown here to be suitable to the colored-noise problem with unknown parameters.

Now we are curious about whether the GPC-EKF method can work as well in white-noise systems since we do not typically know if a system has colored noise. Let's assume a system where Φ and Ψ are zeros, i.e., the system has white noises. The next simulation shows how the GPC-EKF method is robust and suitable to both the color-noise system and white-noise system.

As shown in Figure 4, even if true signals do not have color-noise, the GPC-EKF is working here as well. Obviously, GPC-EKF estimation errors are smaller than the position sensor errors. That is why an appropriate estimator is needed to design although some sensors offer noisy signals of state. Lastly, by comparing Figure 4 with Figure 3(a), the GPC-EKF can be considered regardless of whatever forms of noise.

7. Conclusion

This paper describes that filtering using Gaussian process models for colored noise systems (GPC-EKF) is applicable to the color-noise problem when the characteristics of the noise are unknown and the underlying parametric functions are uncertain. Without information on the color noise, the GPC model demonstrates how Gaussian process regression can be used to learn the colored noise systems. Furthermore, the algorithm of GPC-EKF approach to reliably estimate state is introduced. The various simulations, including Monte Carlo results, show why augmented filter is required here and why the color noise should be handled in the GP-based filters. The performance of a standard EKF that ignores the color of noisy signals as well as of an augmented EKF that compensates the colored noise are verified based on each estimation error. As expected, the accuracy of GPC-EKF increases significantly rather than the performance of the enhanced GP-based filter with improper white noise models for colored noise systems. Lastly, the results validate how the GPC-EKF method is robust to an improper model and to even a white-noise system as well.

Future work will examine the application of GPC-EKF approach to an auto regressive or a moving average (ARMA) signal that is considered as the parametric model of color noise. In addition, other novel GPs can directly be incorporated into this approach. For example, heteroscedastic GPs [14] can allow accurate inference for input-dependent noise models and the sparse spectrum GPs [15] can be used to reduce the computational complexity of online-fashion GPs.

References

[1] A. Bryson and D. Johansen, "Linear filtering for time-varying systems using measurements containing colored noise," *IEEE Transactions on Automatic Control*, vol. 10, no. 1, pp. 4–10, 1965.

[2] X. Shenshu, Z. Zhaoying, Z. Limin, X. Changsheng, and Z. Wendong, "Adaptive filtering of color noise using the kalman filter algorithm," in *Instrumentation and Measurement Technology Conference, 2000. IMTC 2000. Proceedings of the 17th IEEE*, vol. 2. IEEE, 2000, pp. 1009–1012.

[3] F. Le Chevalier, *Principles of radar and sonar signal processing*. Artech House, 2002.

[4] J. D. Gibson, B. Koo, and S. D. Gray, "Filtering of colored noise for speech enhancement and coding," *IEEE Transactions on Signal Processing*, vol. 39, no. 8, pp. 1732–1742, 1991.

[5] D. C. Popescu and I. Zeljkovic, "Kalman filtering of colored noise for speech enhancement," in *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, vol. 2. IEEE, 1998, pp. 997–1000.

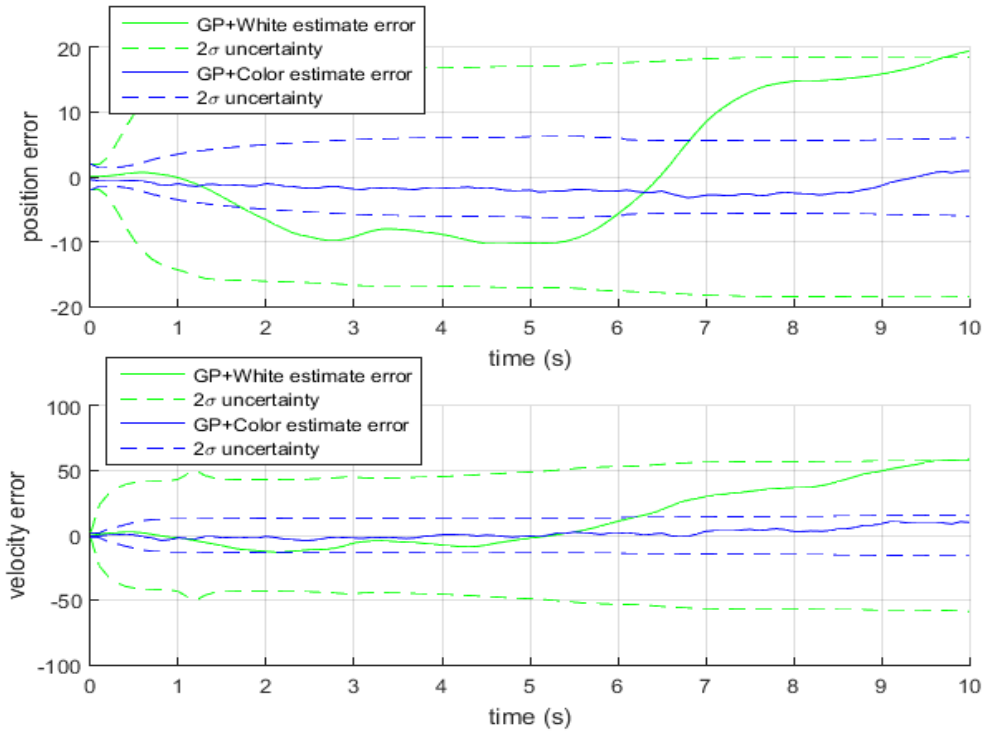
[6] J. Liu and Z. Deng, "Self-tuning weighted measurement fusion kalman filter for arma signals with colored noise," *Applied Mathematics & Information Sciences*, vol. 6, no. 1, pp. 1–7, 2012.

[7] R. S. Bucy and P. D. Joseph, *Filtering for stochastic processes with applications to guidance*. American Mathematical Soc., 1987, vol. 326.

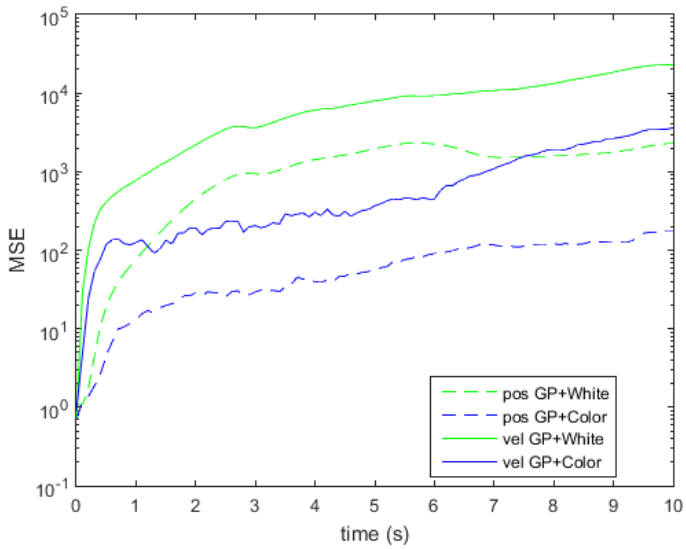
[8] A. BRYSON, JR and L. Henrikson, "Estimation using sampled data containing sequentially correlated noise." *Journal of Spacecraft and Rockets*, vol. 5, no. 6, pp. 662–665, 1968.

[9] E. B. Stear and A. R. Stubberud, "Optimal filtering for gauss-markov noise," *International Journal of Control*, vol. 8, no. 2, pp. 123–130, 1968.

[10] D. Simon, *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons, 2006.



(a) Estimation error by Monte Carlo Method



(b) Mean Square Error

Figure 3. EGP-White vs. GPC-EKF

[11] C. E. Rasmussen, "Gaussian processes for machine learning," 2006.

[12] J. Ko and D. Fox, "Gp-bayesfilters: Bayesian filtering using gaussian process prediction and observation models," *Autonomous Robots*, vol. 27, no. 1, pp. 75–90, 2009.

[13] J. Ko, *Gaussian process for dynamic systems*. University of Washington, 2011.

[14] K. Kersting, C. Plagemann, P. Pfaff, and W. Burgard, "Most likely heteroscedastic gaussian process regression," in *Proceedings of the 24th international conference on Machine learning*. ACM, 2007, pp. 393–400.

[15] M. Lázaro-Gredilla, J. Quiñonero-Candela, C. E. Rasmussen, and A. R. Figueiras-Vidal, "Sparse spectrum gaussian process regression," *The Journal of Machine Learning Research*, vol. 11, pp. 1865–1881, 2010.

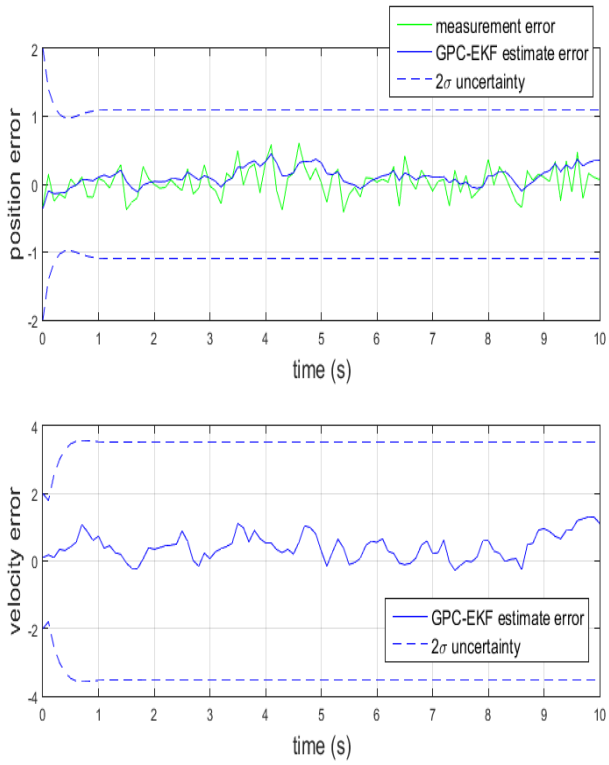


Figure 4. GPC-EKF in white-noise system

BIOGRAPHY

Kyuman Lee *received his B.S. degrees in Aerospace Engineering from Korea Advanced Institute of Science and Technology in 2007 and a M.S. in Aerospace Engineering in 2013 from Georgia Institute of Technology. He is currently a Ph.D student at Georgia Institute of Technology. His current research activities include adaptive estimation for vision-aided navigation of UAVs.*

Eric N. Johnson *received his B.S. degree in Aeronautics and Astronautics Engineering from University of Washington in 1991; a M.S. in Aeronautics at the George Washington University in 1993; a S.M. in Aeronautics and Astronautics from Massachusetts Institute of Technology in 1995; and a Ph.D. in Aerospace Engineering from Georgia Institute of Technology in 2000. He is currently a Lockheed Martin Associate Professor of Avionics Integration at Georgia Institute of Technology. His interests include: Fault Tolerant Estimation and Control as enabled by advancements in digital processing and new sensor technologies; Aerospace Vehicle Design, particularly as enabled by digital avionics (micro-air vehicles, unmanned aerial vehicles (UAVs), autonomous vehicles, and new configurations); Digital Avionics Systems, including architectures and redundancy management; and Modeling/Simulation, including improving aerospace system design and integration processes.*