International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 17 (2015) pp 37616-37618

An Application Connected with an Approximation of Stochastic Difference Equations

M. Reni Sagayaraj
Department of Mathematics, Sacred Heart College (Autonomous), Tirupattur, India reni.sagaya@gmail.com

P. Manoharan,
Department of Mathematics, Sacred Heart College (Autonomous), Tirupattur, India

Abstract
In this paper we present new modeling and analysis technique for characterizing some initial results stochastic difference equation, where we seek a variation Bayesian treatment of the dynamic data assimilation problem which builds upon our variation Bayesian Gaussian process treatment of the static data assimilation problem. In particular we focus on the issue of defining a Gaussian process approximation to the temporal evolution of the solution of a general stochastic difference equation with positive values.

Keywords: stochastic difference equation, new approach model, Gaussian process approximation.

Introduction

In this paper, we discuss the prior distribution is approximated as a space only Gaussian process, the observations are often given Gaussian likelihoods, and thus inference of the posterior proceeds using well known Gaussian process methods. It is possible to view almost all data assimilation methods as approximations to the Kalman filter new approach[2]. Much early work in data assimilation focused on the static case where the prior distribution was assumed to have a climatological covariance structure whose form was dictated in many cases from the model dynamics, and a mean given by a deterministic forecast from the previous data assimilation cycles posterior mean[3]. At each time step, the covariance of the state was thus ignored and only the mean was propagated forward in time.

Approximations of Stochastic Difference Equations with Gaussian Process Method

We assume that the variables divide into two groups

$$x = \begin{pmatrix} x^2 \\ x^1 \end{pmatrix}$$

such that the positive level is σ^2 on the x^2 variables assumed to be k of the n, while for the remainder it is σ^1.

Assuming the positive to be equal variance and uncorrelated we have

$$\Delta x_k^2 = x_{k+1}^2 - x_k^2 = f^2(x_k)\Delta t + \sigma^2 z^2 \sqrt{\Delta t},$$

$$\Delta x \frac{1}{k} = x \frac{1}{k+1} - x \frac{1}{k} = f^1(x_k)\Delta t + \sigma^1 z^1 \sqrt{\Delta t},$$

Where z^2 (z^1) is a vector of dimension k (n − k) drawn from a multivariate Gaussian with identity covariance. Hence, the true probability of a sequence $\{xi\}_{i=1}^{N}$ is given by

$$P\left(\{x_i\}_{i=1}^{N}\right) = \prod_{i=1}^{N-1} \frac{1}{(2\pi\sigma^{\|2}\Delta t)^{k/2}(2\pi\sigma^{\perp 2}\Delta t)^{(n-k)/2}}$$

$$\exp\left(-\frac{\|\Delta x_i^2 - f^2(x_i)\Delta t\|^2}{2\sigma^{\|2}\Delta t} - \frac{\|\Delta x_i^1 - f^1(x_i)\Delta t\|^2}{2\sigma^{\perp 2}\Delta t}\right)$$

We will approximate this distribution by a distribution Q which we assume has the following form

$$Q\left(\{x_i\}_{i=1}^{N}\right) = \prod_{i=1}^{N-1} \frac{1}{(2\pi\sigma^{\|2}\Delta t)^{k/2}(2\pi\sigma^{\perp 2}\Delta t)^{(n-k)/2}}$$

$$\exp\left(-\frac{\|\Delta x_i^2 - (A_i^2 x_i + b_i^2)\Delta t\|^2}{2\sigma^{\|2}\Delta t} - \frac{\|\Delta x_i^1 - (A_i^1 x_i + b_i^1)\Delta t\|^2}{2\sigma^{\perp 2}\Delta t}\right)$$

where the matrices A_i^2, A_i^1 and vectors b_i^2, b_i^1 are parameters that will be adjusted to minimise the KL divergence KL(Q||P) between the two distributions.

We present here an approximation of the exact optimization of the KL divergence that extends the Kalman filter new approach. The detailed derivations are given we summarize the results here.

If we define

$$A_i = -\langle (b_i - f(x_i))x_i^{`}\rangle \langle x_i x_i^{`}\rangle^{-1} = A(m_i, \sum_i) \text{ and } b_i = -\langle f(x_i)\rangle - A_i\langle x_i\rangle = b(m_i, \sum_i)$$

where averages are over the Gaussian approximation at stage i with mean m_i and covariance $\sum_i$. Taking limits as the interval tends to zero we obtain differential equations for these quantities which are now all functions of t:

$$\frac{dm}{dt} = A(m, \sum)m + b(m, \sum) = \langle f(x)\rangle,$$

and

$$\frac{d\sum}{dt} = diag(\sigma) + A(m, \sum)\sum + \sum A(m, \sum)^{`}$$

$$= diag(\sigma) + \langle \frac{\partial f(x)}{\partial x}\rangle \sum + \sum \langle \frac{\partial f(x)}{\partial x}\rangle^{`}$$

These equations define the evolution of our approximation of the Q process.

Kalman Filter New Approach

A standard approach for modeling dynamical systems with additive Gaussian measurement noise is the Kalman Filter New Approach (KFNA)[1]. When the system is nonlinear, one of its variants, for example the extended Kalman Filter New Approach (EKFNA), can be used. KFNA and EKFNA

are concerned with propagating the two first moments of the filtering distribution $p(x_{i+1}|y_{0:i+1})$, which are given by

$$\bar{x}_{i+1} = \mathbb{E}\{x_i|y_{0:i+1}\}, \quad (1)$$

$$\bar{S}_{i+1} = \mathbb{E}\{(x_i - \bar{x}_{i+1})(x_{i+1} - \bar{x}_{i+1})^T|y_{0:i+1}\}, \quad (2)$$

where $y_{0:i+1} \equiv \{y_0, \ldots, y_{i+1}\}$ are the observations up to time t_{i+1}. In order to propagate these quantities, they proceed in two steps. In the prediction step, the two moments are estimated given the observations up to time t_i:

$$p(x_{i+1}|y_{0:i+1}) = \int p(x_{i+1}|x_i)p(x_i|y_{0:i})dx_i \quad (3)$$

This allows computing the predicted state $\mathbb{E}\{x_{i+1}|y_{0:i+1}\}$ and the predicted state covariance $\mathbb{E}\{(x_{i+1}|\bar{x}_{i+1})(x_{i+1}|\bar{x}_{i+1})^T|y_{0:i}\}$. Next, the correction step consists in updating these estimates based on the new observation y_{i+1}:

$$p(x_{i+1}|y_{0:i+1}) \propto p(y_{i+1}|x_{i+1})p(x_{i+1}|y_{0:i}) \quad (4)$$

which leads to the desired conditional moments (1) and (2). KFNA is particularly attractive for online learning as it is not required to keep trace of the previous conditional expectations. Unfortunately, the integral in (3) is in general intractable when the system is nonlinear or when the state transition probability p($x_{i+1}|x_i$) is non-Gaussian. Therefore, approximations are required.

A common approach is to linearize the system, which corresponds to EKFNA. This leads to a Gaussian approximation of the transition probability [6]. Moreover, if the likelihood $p(y_{i+1}|x_{i+1})$ is assumed to be Gaussian, then the filtering density is a Gaussian one at each ti+1. An alternative approach is to resume the past information by the marginal Q(x_{i+1}) and make predictions as follows

$$p(x_{i+1}|y_{0:i}) = \mathcal{N}(x_{i+1}|m_{i+1}, \sum_{i+1}) \quad (5)$$

Where

$$m_{i+1} = m_i + (Am_i + b_i)\Delta t \quad (6)$$

$$\sum_{i+1} = \sum_i + (2A\sum_i + diag\{\sigma\})\Delta t \quad (7)$$

This approximation is expected to be better than EKFNA, as the parameters A and b of the linear approximation are adjusted at each iteration. If we further assume that the likelihood $p(y_{i+1}|x_{i+1})$ is of the form $\mathcal{N}(y_{i+1}|x_{i+1}, R)$, then the correction step (4) is given by

$$p(y_{i+1}|x_{i+1}) = \mathcal{N}(y_{i+1}|\bar{x}_{i+1}, \bar{S}_{i+1}) \quad (8)$$

With

$$\bar{x}_{i+1} = S_{i+1}(\sum_{i+1}^{-1} m_{i+1} + R^{-1}y_{i+1}) \quad (9)$$

$$\bar{S}_{i+1} = (\sum_{i+1}^{-1} + R^{-1}) \quad (10)$$

Note that in this approach, only the filtering density (and its associated moments) are propagated through time. In contrast, GP framework allows us to define a distribution over the entire function space (i.e., over time). It is expected that this will have a smoothing effect and will lead to a better tracking of the state transitions.

Approximations to KF New Approach

If we use the variable X to denote the complete sequence $\{x_i\}_{i=1}^N$, we can compute this as

$$KL(Q \| P) = \mathrm{E}_{X \square Q}\left[\log \frac{Q(X)}{P(X)}\right]$$

$$= \sum_{i=1}^{N-1} \mathrm{E}_{X \square Q}\left[\log \frac{Q(X_{i+1} \mid X_i)}{P(X_{i+1} \mid X_i)}\right]$$

$$= \sum_{i=1}^{N-1} \int dX_i Q(X_i) \int dX_i Q(X_{i+1} \mid X_i)$$

$$\left[\frac{\left\|\Delta X_i^2 - f^2(X_i)\Delta t\right\|^2 - \left\|\Delta X_i^2 - (A_i^2 X_i + b_i^2)\Delta t\right\|^2}{2\sigma^{\|2}\Delta t} + \frac{\left\|\Delta X_i^1 - f^1(X_i)\Delta t\right\|^2 - \left\|\Delta X_i^1 - (A_i^1 X_i + b_i^1)\Delta t\right\|^2}{2\sigma^{\perp 2}\Delta t}\right]$$

$$= \sum_{i=1}^{N-1} \int dX_i Q(X_i) \left[\frac{\Delta t}{2\sigma^{\|2}}\left\|(A_i^2 X_i + b_i^2) - f^2(X_i)\right\|^2 + \frac{\Delta t}{2\sigma^{\perp 2}}\left\|(A_i^1 X_i + b_i^1) - f^1(X_i)\right\|^2\right]$$

Where the last equality follows from using the equality

$$\mathrm{E}_{X \square N(\mu,\sigma)}\left[(x-a)^2\right] = \sigma^2 + (\mu - a)^2$$

for the k components of the first vector and n − k components of the second.

In order to minimize the NA divergence, we must take the derivative with respect to the parameters, set to zero and solve. We obtain

$$2A_i^2 \langle X_i X_i^{`} \rangle + 2\langle b_i^2 - f^2(X_i) X_i^{`} \rangle = 0$$

$$2\langle A_i^2 X_i \rangle + b_i^2 - \langle f^2(X_i) \rangle = 0$$

$$2A_i^1 \langle X_i X_i^{`} \rangle + 2\langle b_i^1 - f^1(X_i) X_i^{`} \rangle = 0$$

$$2\langle A_i^1 X_i \rangle + b_i^1 - \langle f^1(X_i) \rangle = 0$$

where the angle brackets indicate expectations with respect to the marginal distribution Q(x_i), which is Gaussian with mean and covariance mi and $\sum_i$ respectively (note that these are over the full variable set). We obtain the equations for the parameters as

$$A_i = \langle b_i - f(X_i) X_i^{`} \rangle \langle X_i X_i^{`} \rangle^{-1} = A(m_i, \Sigma_i)$$

$$and \quad b_i = \langle f(X_i) - A_i(X_i) = b(m_i, \Sigma_i)$$

where the matrix A_i is formed by concatenating A_i^2 and A_i^1 and similarly b_i. The expressions for A(m, Σ) and b(m, Σ) can be given as

$$A(m,\Sigma)(\Sigma + mm^{|}) = \langle f(X_i) X^{|} \rangle - bm^{|}$$

$$b(m,\Sigma) = \langle f(X) \rangle - Am.$$

Substituting **b** from the second equation in the first gives

$$A(m,\Sigma)(\Sigma + mm^{|}) = \langle f(X_i) X^{|} \rangle - \langle f(X) \rangle m^{|} + Amm^{|}$$

$$\Rightarrow A(m,\Sigma)\Sigma = \langle f(X)(X^{|} - m^{|}) \rangle$$

$$= \left\langle \frac{\partial f(X)}{\partial X} \right\rangle \Sigma \quad (11)$$

with the last equality following from an integration by parts. Writing expressions for the generation of x_{i+1} we have

$$x_{i+1}^2 = x_i^2 + (A^2 x_i + b^2)\Delta t + \sigma^2 u^2 \sqrt{\Delta t}$$
$$x_{i+1}^1 = x_i^1 + (A^1 x_i + b^1)\Delta t + \sigma^1 u^1 \sqrt{\Delta t}$$

where **u** is an n-dimensional vector of independent zero mean Gaussian variables with unit variance. We can combine these into the single equation

$$x_{i+1} = x_i + (Ax_i + b)\Delta t + diag(\sigma) u \sqrt{\Delta t}$$

where diag(σ) denotes the diagonal matrix whose first k entries are σ^2σl and remaining entries σ^1. We also know that x_i is generated by a Gaussian with mean mi and covariance Σ_i, so that

$$x_i = m_i + \sqrt{\sum_i} v$$

where **v** is a vector of zero mean unit variance Gaussian variables. Hence, we obtain the following expression for x_{i+1}

$$x_{i+1} = m_i + \sqrt{\sum_i} v + A m_i \Delta t + A\sqrt{\sum_i} v \Delta t + b\Delta t + diag(\sigma) u \sqrt{\Delta t}$$

Hence, we can compute

$$m_{i+1} = \mathbb{E}[x_{i+1}] = m_i + A m_i \Delta t + b\Delta t$$

Taking the difference between means, dividing by Δt and taking limits gives

$$\frac{dm}{dt} = A(m, \sum) m + b(m, \sum) = \langle f(x) \rangle$$

Next consider

$$\sum_{i+1} = \mathbb{E}[(x_{i+1} - m_{i+1})(x_{i+1} - m_{i+1})`]$$
$$= \mathbb{E}[\sqrt{\sum_i} v v` \sqrt{\sum_i}] + \mathbb{E}[\sqrt{\sum_i} v v` \sqrt{\sum_i} A` \Delta t] + \mathbb{E}[A\sqrt{\sum_i} v v` \sqrt{\sum_i} \Delta t] + diag(\sigma)\Delta t + O((\Delta t)^2)$$

Taking the difference between means, dividing by Δt and taking limits gives

$$\frac{d\sum}{dt} = diag(\sigma) + A(m, \sum)\sum + \sum A(m, \sum)`$$
$$= diag(\sigma) + \langle \frac{\partial f(x)}{\partial x} \rangle \sum + \sum \langle \frac{\partial f(x)}{\partial x} \rangle `$$

where we have made use of equation (11). The result of this computation hold for all values of σ^2 and σ^1, so that we can consider the case where we let σ^1 tend to zero. This allows us to encode quite general noise models and covariances as the examples in Appendix C illustrate.

Examples of different Positive Models

General covariance

If we wish to introduce noise with a known fixed covariance $\sum_0$ as for example given by spatial relations between the locations of a climate grid model, we can double the number of variables to n = 2k to obtain

$$\Delta x_k^2 = x_{k+1}^2 - x_k^2 = z\sqrt{\Delta t},$$
$$\Delta x_k^1 = x_{k+1}^1 - x_k^1 = f(x_k^1)\Delta t + \sqrt{\sum_0 x_k^2},$$

where z is k dimensional zero mean unit variance Gaussian random variables and f (x^1) is the system being studied.

Ornstein-Uhlenbeck process

We have seen that for a wiener process {X(t)}, the displacement Δx, in a small interval of the Δt is also small, being of [O($\sqrt{\Delta t}$)]. The velocity which is of O($\sqrt{\Delta t}/\Delta t$) tends to infinity as $\Delta t \to 0$. Thus the wiener process does not provide a satisfactory model for Brownian motion for small values of t, although for moderate and large values of t it does so[6]. An alternative model which holds for small t was proposed by Ornstein and Uhlenbeck in 1930. Here instead of the displacement X (t), the velocity U(t)=X`(t) at time t is considered.

The equation of motion of a Brownian particle can be written as

$$dU(t) = -\beta U(t) dt + dF(t),$$

Where $-\beta U(t)$ represents the systematic part due to the resistance of the medium and $dF(t)$, represents the random component.

Conclusion

In this paper we have developed a new model for the final version of the results for a Gaussian process with the covariance function determined by the approximation of the process Q. A discussion of how the kernel can be obtained from the above computations is included.

References

[1] D Cornford, L Csat´o, D J Evans, and M Opper. Bayesian analysis of the scatterometer wind retrieval inverse problem: some new approaches. Journal of the Royal Statistical Society B, 66:609–626, 2004.

[2] R. E. Kalman. A new approach to linear filtering and prediction problems. Transactions of the ASME, Journal of Basic Engineering, 82:34– 45, 1960.

[3] E Kalnay. Atmospheric Modelling, Data Assimilation and Predictability.Cambridge University Press, Cambridge, 2003.

[4] A. C. Lorenc. Analysis methods for numerical weather prediction. Quarterly Journal of the Royal Meteorological Society, 112:1177–1194, 1986.

[5] A C Lorenc. The potential of the ensemble Kalman filter for NWP? a comparison with 4D- Var. Quarterly Journal of the Royal Meteorological Society, 129:3183–3203, 2003.

[6] Carl Edward Rasmussen and Christopher K.I. Williams. Gaussian Processes for Machine Learning. The MIT Press, Cambridge, Mas- sachusetts, 2006.