等価重み粒子フィルタによる時変パラメータの推定

佐藤　峰斗　　総合研究大学院大学　統計科学専攻　博士課程（5年一貫制）4年

Introduction

Motivation: To improve robustness and efficiency for estimating the state and time-varying parameters in nonlinear high-dimensional systems.

Difficulty: Typical state augmentation technique tend to fail to detect parameter changes unless there are strong enough correlations.

Solution: Introduce correlated perturbation and use an efficient particle filter "IEWPF" [1] combined with adaptive moment estimation (Adam [2]) based optimization technique from machine learning.

Methods

State augmentation with correlated perturbation

Time evolution of state in augmented state-space model:

$$z^n = \tilde{f}\left(z^{n-1}\right)+\tilde{\beta}^n = \begin{pmatrix} f\left(x^{n-1},\theta^n\right) \\ \theta^{n-1} \end{pmatrix} + \begin{pmatrix} \beta^n \\ \eta^n \end{pmatrix}$$

$$y^n = H\left(z^n\right)+\varepsilon^n$$

can approximately express by using Taylor series expansion:

$$f\left(x^{n-1},\theta^n\right) = f\left(x^{n-1},\theta^{n-1}\right)+\frac{\partial f}{\partial \theta}\left(\theta^n-\theta^{n-1}\right).$$

Hence,

$$z^n = \begin{pmatrix} x^n \\ \theta^n \end{pmatrix} = \begin{pmatrix} f\left(x^{n-1},\theta^{n-1}\right) \\ \theta^{n-1} \end{pmatrix} + \tilde{\beta}^n \quad \text{, where } p\left(\tilde{\beta}^n\right) = N\left(0, Q^n_{\text{aug.}}\right),$$

$$Q^n_{\text{aug.}} = \begin{pmatrix} \operatorname{cov}\left[x^n,x^n\right] & \operatorname{cov}\left[x^n,\theta^n\right] \\ \operatorname{cov}\left[\theta^n,x^n\right] & \operatorname{cov}\left[\theta^n,\theta^n\right] \end{pmatrix} = \begin{pmatrix} Q_\beta + \frac{\partial f}{\partial \theta} Q_\eta \frac{\partial f}{\partial \theta}^T & \frac{\partial f}{\partial \theta} Q_\eta \\ \left(\frac{\partial f}{\partial \theta} Q_\eta\right)^T & Q_\eta \end{pmatrix}.$$

Combination of IEWPF and parameter optimization

- The Implicit Equal-Weights Particle Filter (IEWPF) [1] can avoid filter degeneracy by using a proposal density which the weight of each particle can be equal to the target weight.
- Assume typical geophysical systems which time step interval T exist between observations, and introduce new proposal distribution during this period as shown in Fig. 1.

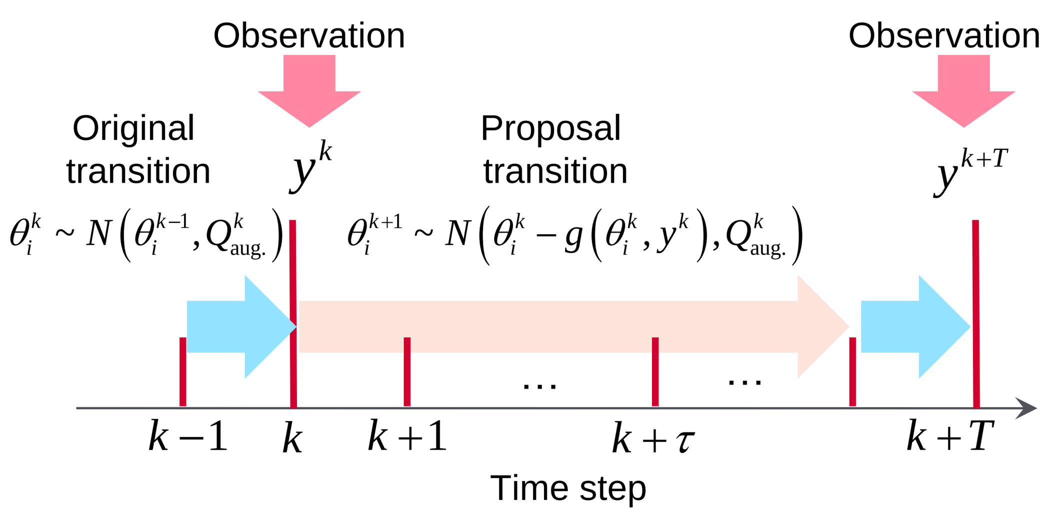


Fig.1. Parameter updating diagram from one step before observed step k to next observation $k+T$. Function g denotes parameter correction term calculated by using the last observation and τ is the number of time steps after the observation.

Adam-based parameter correction term

- Using gradient of log-likelihood at the last observation step, update momentum m and norm (squared gradient) v as following equation. μ and ρ denote hyper-parameter that control these moving average.

$$\nabla_\theta L_i^k \equiv -2\nabla_\theta \ln\left[p\left(y^k \middle| x_i^{k-1},\theta_i^{k-1}\right)\right],$$

$$\text{where } p\left(y^k \middle| x_i^{k-1},\theta_i^{k-1}\right) \propto \exp\left[-\frac{1}{2}\left(y^k - Hf\left(x_i^{k-1},\theta_i^{k-1}\right)\right)^T R^{-1}\left(y^k - Hf\left(x_i^{k-1},\theta_i^{k-1}\right)\right)\right]$$

$$m_i^k = \mu m_i^{k-1}+(1-\mu)\nabla_\theta L_i^k \;,\; v_i^k = \rho v_i^{k-1}+(1-\rho)\left(\nabla_\theta L_i^k\right)^2$$

- Then parameter correction term g is expressed as follows. λ_0 and δ denote step-size and a term to avoid division by 0, respectively.

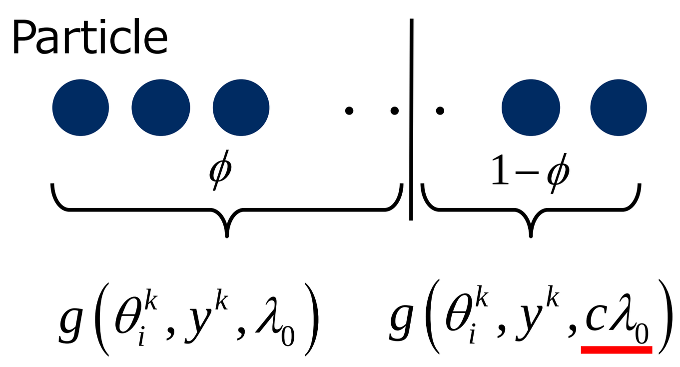
$$g\left(\theta_i^k, y^k\right) = \frac{\lambda_0}{\tau}\frac{\hat{m}^k}{\sqrt{\hat{v}^k}+\delta}\theta^{k-1} \quad \text{, where } \hat{m}^k = \frac{m^k}{1-\mu},\ \hat{v}^k = \frac{v^k}{1-\rho}$$

References

1. Zhu M., van Leeuwen PJ, Amezcua J., 2016, 'Implicit equal-weights particle filter', Q. J. R. Meteorol. Soc. 142: 1904–1919.
2. P. Kingma D. and Lei Ba J., 'Adam: A method for stochastic optimization', conference paper at ICLR 2015

Mixture proposal density

Remaining problem has been the choice of the step-size factor. Especially in proposed method combined with Adam-based correction term, step-size plays a critical role in balancing a stable estimation with fast tracking to the abrupt changes. Therefore, in order to improve both stability and follow-up capability for uncertain environment, mixture proposal density is introduced as follows.



$$q\left(z_i^{n+1}\middle|z_i^n\right) = \phi\mathcal{N}\left(\begin{pmatrix} f\left(x_i^n,\theta_i^n\right) \\ \theta_i^n - g\left(\theta_i^n, y^n, \lambda_0\right) \end{pmatrix}, \mathbf{Q}^n_{aug.}\right) + (1-\phi)\mathcal{N}\left(\begin{pmatrix} f\left(x_i^n,\theta_i^n\right) \\ \theta_i^n - g\left(\theta_i^n, y^n, c\lambda_0\right) \end{pmatrix}, \mathbf{Q}^n_{aug.}\right).$$

Case study : Lorenz-96 with parameterized forcing

Proposed method is tested with twin experiment using Lorenz-96 model whose forcing is parameterized as follows:

$$\frac{d}{dt}x_j = \left(x_{j+1}-x_{j-2}\right)x_{j-1}-x_j+F_j \;,\quad F_j = f_0+\theta_1\sin\left(\frac{2\pi}{\theta_2}j\right).$$

x_j (j=1,…,N_j) is the state variables at position j of dimension N_j and f_0, θ_1, θ_2 is unknown or time-varying parameter. In the following experiments, the true model error covariance matrix Q is chosen as tridiagonal matrices, the main diagonal value is 0.20 and both sub- and superdiagonal values are 0.05. The observation error matrix R is diagonal with main diagonal value 0.01. For the assimilation, we choose same matrix Q and R as when observation was generated.

Evaluation results of mixture method (N_j = 40)

In order to investigate the ability of the proposed mixture method to estimate time-varying parameters, a trajectory comparing results of different step-sizes (a)(b) and mixture (c) is shown in Fig.2. Compared to the small step-size (a), a reduction of about 20% of the time-averaged RMSE is achieved with the proposed mixing method (c).

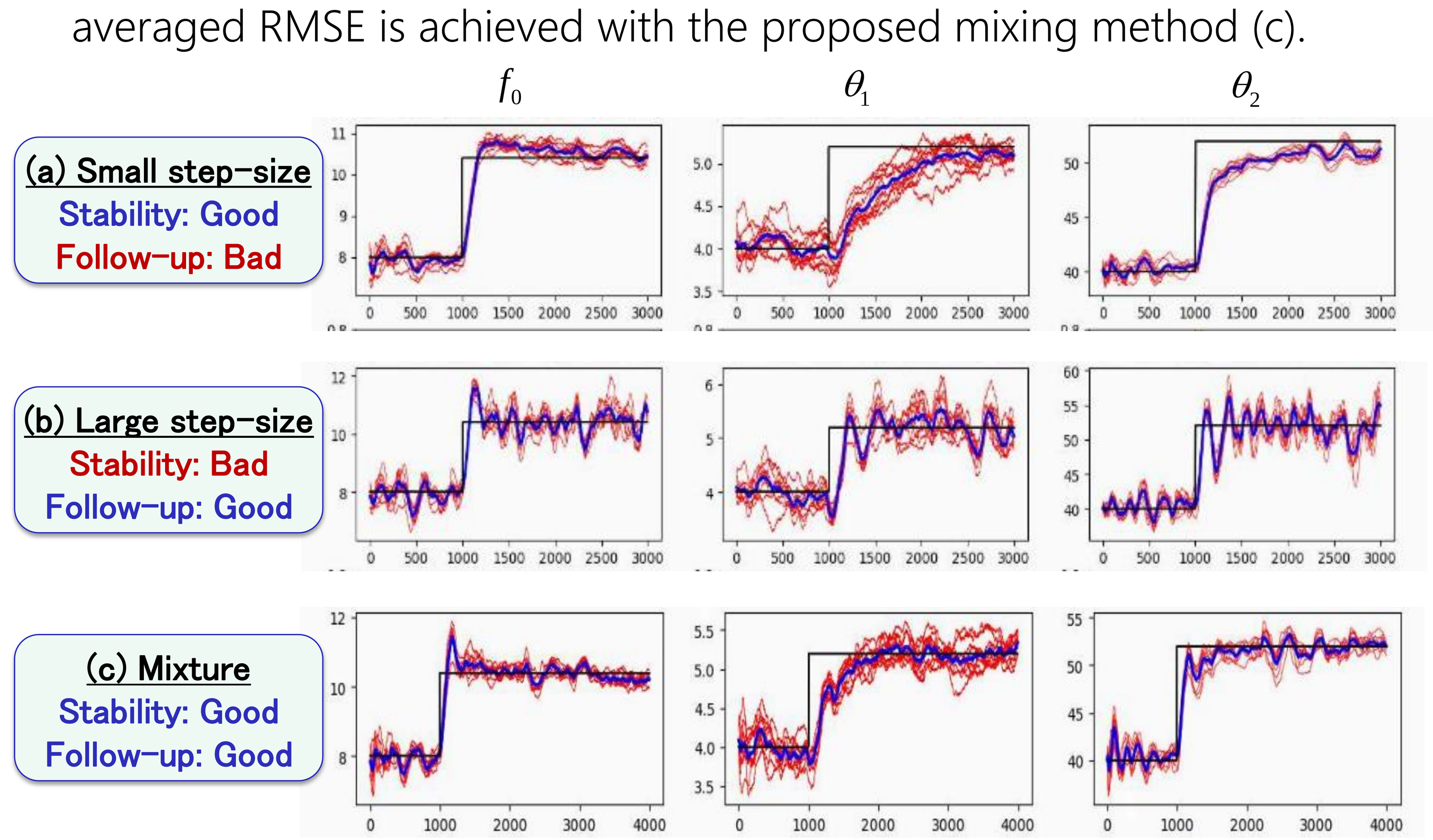


Fig.2. Time-series results of 40-dimensional Lorenz-96 with 10 ensemble members. All variables are observed by every 4th step and the observations are generated from the true parameter value [f_0, θ_1, θ_2] = [8.0, 4.0, 40.0] changed 30% bigger at the 1000th step.

Conclusion

Proposed method has capability to track the change of the model characteristics (e.g. parameters) by extremely fewer ensemble even in case of high-dimensional application.