

Article

Information Geometric Approach to Recursive Update in Nonlinear Filtering

Yubo Li *, Yongqiang Cheng, Xiang Li, Xiaoqiang Hua and Yuliang Qin

School of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China; nudtyqcheng@gmail.com (Y.C.); lixiang01@vip.sina.com (X.L.); hxq712@yeah.net (X.H.); yuliang.qin@gmail.com (Y.Q.)

* Correspondence: liyubo@nudt.edu.cn; Tel.: +86-731-8457-4452

Academic Editor: Geert Verdoolaege

Received: 25 November 2016; Accepted: 20 January 2017; Published: 26 January 2017

Abstract: The measurement update stage in the nonlinear filtering is considered in the viewpoint of information geometry, and the filtered state is considered as an optimization estimation in parameter space has been corresponded with the iteration in the statistical manifold, then a recursive method is proposed in this paper. This method is derived based on the natural gradient descent on the statistical manifold, which constructed by the posterior probability density function (PDF) of state conditional on the measurement. The derivation procedure is processing in the geometric viewpoint, and gives a geometric interpretation for the iteration update. Besides, the proposed method can be seen as an extended for the Kalman filter and its variants. For the one step in our proposed method, it is identical to the Extended Kalman filter (EKF) in the nonlinear case, while traditional Kalman filter in the linear case. Benefited from the natural gradient descent used in the update stage, our proposed method performs better than the existing methods, and the results have showed in the numerical experiments.

Keywords: information geometry; nonlinear filtering; Kalman filter; Bayesian filtering; natural gradient descent

1. Introduction

Nonlinear filtering is a significant issue in the field of signal processing, such as target tracking, navigation and audio signal processing. However, there is no close form solution for nonlinear filtering similar to the Kalman filter for linear and Gaussian scenarios. How to deal with the nonlinear state propagation or nonlinear measurement is crucial in the research of nonlinear filtering. Various nonlinear filtering algorithms have been proposed and widely used in the practical applications, such as Extended Kalman filter (EKF) [1], Unscented Kalman filter (UKF) [2], Cubature Kalman filter (CKF) [3], and Particle filter (PF) [4]. Among these methods, EKF, UKF and CKF can be classified into the same catalog, which also has been called as KF-type method that have the certain nonlinear approximation to utilize the common frame of traditional Kalman filter to achieve the goal. Some relations among these KF-type method are given in [5]. The EKF utilizes the Taylor Series Expansion to exploit the analytical structure of nonlinear functions, which also be called as analytical linearization. Similarly, the UKF or CKF which exploit the statistical properties of Gaussian variables that undergo nonlinear transformations, which known as statistical linearization. While the PF is another catalog, which uses the Monte Carlo technique to approximate the PDF. It obtains a weighted sample from the posterior PDF, and provides an asymptotically exact approximation of the posterior PDF as the sample size tends to infinity [6]. In particular, all of these methods can be unified in the framework of Bayesian filtering [7] to induce the optimal estimation of state by means of computing the posterior PDF, i.e., the density of state conditional on the corresponding measurements.

As for the Bayesian filtering, the procedure can be split into two steps: time propagation and measurement update, which also have been called prediction and correction steps in some articles. In the nonlinear and non-Gaussian case, the state propagation step is relatively simple as it can be obtained by approximating the first two moments of the state variable that undergoes a transformation along the state transition function. But the measurement update is more difficult and challenging. Firstly, the measurement is proceeded not easily in itself, with nonlinear and/or non-Gaussian conditions. Secondly, the error which brings by the approximation of state propagation becomes more seriously under the nonlinear measurements. Thirdly, the measurement update step has associated state with measurement, and the posterior PDF of state has to be computed conditional on the measurement with noisy. Therefore, the posterior PDF is the key way to tackle this issue, and it attracted extensive attention in the research. Usually, Gaussian approximation to posterior PDF is used in the measurement update because of the analytically intractable in non-Gaussian scenario [8]. There are two most popular estimators being utilized in the estimation for posterior PDF as the criterion. The one is the linear minimum mean square error (LMMSE) estimator, which approximates the posterior mean and covariance matrix by its estimator and its mean square error matrix, respectively. Based on the LMMSE, Zanetti [9,10] proposes a recursive update method for the measurement update, which overcomes some of the limitations of the EKF. In addition, the famous methods, such as EKF, UKF and CKF can be classified in this class [11]. Another is the maximum a posteriori (MAP) estimator, which estimates the posterior mean and obtain the covariance matrix by linearizing the measurement function around the MAP estimate. As the most well-known iterated EKF (IEKF), which based on the Gauss-Newton optimization [12] or Levenberg-Marquardt (LM) method [13], is categorized in this class. Usually, these iterative methods for nonlinear filter have the better performance. As Lefebvre [14] has shown that IEKF outperform the EKF, unscented Kalman filter (UKF) for implementing the measurement update of the covariance matrix. Besides, there are some other methods to analysis and approximate the posterior PDF recently, such as the Kullback-Leibler divergence (KLD) as metric to analysis the performance from the true joint posterior PDF of the state conditional on the measurement to approximation posterior PDF [15]. This metric can be used to devise new algorithms, such as the iterated posterior linearization filter (IPLF) [8] which can be seen as an approximate recursive KLD minimization procedure.

In particular, when we consider the posterior PDF parameterized by the estimators in itself, some natural characters may attract our attentions and provides a new viewpoints on the estimation. During the posterior PDF approximated recursively by the MAP estimator, a family of the posterior PDFs parameterized by the estimators can construct a statistical manifold, which a Riemannian manifold of probability distributions. Thus, the better approximation of the true posterior PDF can be viewed as the search for the optimum in the statistical manifold. Usually, the search along the direction of conventional gradient can obtain the optimum with the faster speed and better convergence in Euclidean space, but not for the statistical manifold. For the sake of utilizing gradient in statistical manifold, Amari [16,17] has proposed the natural gradient descent for the search direction traditionally motivated from the perspective of information geometry, and it has been proved that steps along the direction of natural gradient descent is the steepest descent in the Riemannian manifold [18]. Based on the direction of steepest descent, the natural gradient descent can obtain the optimum in manifold with faster speed and better convergence. Because of these properties, it has been acted a new tool to analysis the nonlinear in the statistical problems with an optimization perspective. Further, it provides some geometric information, such as Riemannian metrics, distances, curvature and affine connections [19], which may be improve the performance of estimation. To characterize this geometric information, there are two important elements should be considered. The one is that considering the set of PDF as a statistical manifold, and another is the Fisher Information Matrix as the metric of the statistical manifold. With the defining metric by information geometry, the natural gradient descent can be constructed as the product of inverse of Fisher Matric and the conventional steepest. After it has been proposed, natural gradient descent works well for many applications as

an alternative to stochastic gradient descent, such as neural network [20,21], Blind Separation [22], evolution strategy [23,24], and stochastic distribution control systems [25].

Motivated by the recursively estimation for nonlinear measurement update in the Bayesian filtering, we propose a method by using the natural gradient descent method in the statistical manifold constructed by the posterior PDF for measurement update in the nonlinear filtering. After the procedure of the state propagation, the prediction state can be as a prior information in the Bayesian framework. The statistical manifold will be constructed by the posterior PDF. We consider a geometric structure of a this manifold equipped with the Fisher metric, and construct an alternate recursive process on the manifold. Given an initial point in the manifold based on the prior information, the iterated method to seek the optimum can be achieved along the natural gradient steepest descent. At each iteration, the algorithm moves from the current estimate to a new estimate along the geodesic in the direction of the natural gradient descent with a step size. Our method has the better performance compared with the update of EKF, which occur an overshooting in some nonlinear case. In addition, it has also given the theoretical justification for the nonlinear measurement update in the filtering from the information geometry insight and a mathematical interpretation of this method. There are two advantages in the estimation procedure benefitting from the natural gradient descent method. Firstly, based on the fact that the Fisher information matrix (FIM) is the inverse of Cramer-Rao Limit Bounds(CRLB) with respect to the estimation, the natural gradient descent, which has used FIM as metric in the statistical manifold, may achieve the better performance of estimation. Secondly, the natural gradient descent is asymptotically Fisher efficient and often converges faster than the conventional gradient descent shown in [16]. With the better performance and fast convergence, the measurement update with natural gradient descent may be achieved better and faster. Furthermore, based on this measurement update stage, we can construct the different nonlinear filtering method with different state prediction method.

The paper is organized as follows. In Section 2, we give a description for estimation in the statistical manifold and derive the iteration estimation procedure in the viewpoint of information geometry. Then the measurement update using natural gradient descent has deduced in the Section 3. The differences between our proposed method and other existing methods have be discussed in the Section 4. In Section 5, the numerical experiments are presented to illustrate the performance compared with other existing methods. Finally, conclusions are made in Section 6.

2. Information Geometry and Natural Gradient Descent

Information geometry [17] is a new mathematical tool for the study of manifolds of probability distributions. It opens a new prospective to study the geometric structure of information theory and provides a new way to deal with existing statistical problems. It has the powerful ability to handle the non-Gaussian PDF case, such as Weibull distributions [19,26], gamma distributions [27], multivariate generalized Gaussian distribution [28], and non-Euclidean case, for example, the statistical manifold [29], the morphogenetic system [30]. For the optimization problems in statistical signal processing, the Natural gradient descent traditionally motivated from the perspective of information geometry and gradient descent method, has become an alternative to recursive estimation in the statistical manifold. It has a better performance and faster convergence.

Usually in the Euclidean space, the recursive method for estimation can be along the direction of steepest descent in which is relatively straightforward, but it is not the straightforward direction in the Riemannian geometry because of the curved coordinate system. In this case, the curvature of the manifold should be considered, and the steepest gradient descent should follow the curvature of the manifold. This is the important factor for estimation in the manifold, which has been considered in the natural gradient descent method.

In the view of information geometry, the initial point and the target point in the Riemannian manifold are corresponding with the initial and final estimation in the parameters space, respectively. Thus the estimation of parameters can be converted to the recursive procedure seeking optimum in the

manifold. With the direction provided by the natural gradient descent, the recursive in the Riemannian manifold can obtain the optimization estimation of parameters.

In a Riemannian manifold, the Riemannian metric g_p is used at every point p , which describes the relationship between the tangent space and its neighborhood tangent space. If the tangent vectors at point p base on the basis v_p^1, \dots, v_p^n , then the Riemannian metric can be obtained as $g_p^{ij} = \langle v_p^i, v_p^j \rangle$, where $\langle \cdot, \cdot \rangle$ denotes an inner product defined on the tangent space as the same as in the Euclidean space. With these definitions, the inner product of tangent space at point p can be induced as $\langle w_p, w_q \rangle_g = w_p^T g w_q$, where $w_p \in T_p \mathcal{M}$ and $w_q \in T_p \mathcal{M}$. Given the description of tangent basis $w_p = \sum_{i=1}^n \alpha_i v_p^i$ and $w_q = \sum_{j=1}^n \beta_j v_p^j$, their inner product will be $\langle w_p, w_q \rangle_g = \sum_{i,j=1}^n g_p^{ij} \alpha_i \beta_j$.

Consider a family of probability distributions \mathcal{S} on \mathbb{R}^d parameterized by n real-valued variables $\theta = [\theta_1 \cdots \theta_n]^T$ as

$$\mathcal{S} = \{p_\theta = p(y; \theta) \mid \theta \in \Theta\} \tag{1}$$

where $y \in \mathbb{R}^d$ is a random variable and Θ is a open subset of \mathbb{R}^n . The mapping $\varphi : \mathcal{S} \rightarrow \mathbb{R}^n$ defined by $\varphi(p_\theta) = \theta$ can be viewed as a coordinate system of \mathcal{S} . With the Fisher information matrix (FIM)

$$\mathbf{F}(\theta) = \int \frac{\partial \log p(y; \theta)}{\partial \theta} \left(\frac{\partial \log p(y; \theta)}{\partial \theta} \right)^T p(y; \theta) dy \tag{2}$$

as the Riemannian metric $\mathbf{G}(\theta)$ in the view of information geometry, which also termed Fisher metric, the \mathcal{S} can be considered as a Riemannian manifold. Thus the \mathcal{S} can be called an n -dimensional statistical manifold on \mathbb{R}^d , which the parameter θ plays the role of the coordinate system for \mathcal{S} . The element of the Riemannian metric $\mathbf{G}(\theta) = [g_{ij}(\theta)]$ can be written as the follow formation

$$g_{ij}(\theta) = \mathbf{E} \left[\frac{\partial \log p(y; \theta)}{\partial \theta_i} \frac{\partial \log p(y; \theta)}{\partial \theta_j} \right] \tag{3}$$

where \mathbf{E} denotes the expectation with respect to y . The Fisher metric is the only invariant metric to be given to the statistical manifold [31]. Because of the fact that

$$\begin{aligned} \mathbf{E} \left[\frac{\partial \log p(y; \theta)}{\partial \theta_i} \right] &= \int \frac{\partial \log p(y; \theta)}{\partial \theta_i} p(y; \theta) dy \\ &= \frac{\partial}{\partial \theta} \int p(y; \theta) dy = 0 \end{aligned} \tag{4}$$

the FIM can also be expressed in terms of the expectation of Hessian matrix of log-likelihood through algebraic manipulations, which may be easier to compute for certain problems.

$$\mathbf{E} \left[\frac{\partial \log p(y; \theta)}{\partial \theta_i} \frac{\partial \log p(y; \theta)}{\partial \theta_j} \right] = -\mathbf{E} \left[\frac{\partial^2 \log p(y; \theta)}{\partial \theta_i \partial \theta_j} \right] \tag{5}$$

Consider the smooth mapping $\ell : \mathcal{S} \rightarrow \mathbb{R}$, then a set of basis of tangent space of \mathcal{S} can be given as $\{\partial \ell(\theta) / \partial \theta_i\}_{i=1, \dots, n}$. For example, when the usual log-likelihood functions $\log : \mathcal{S} \rightarrow \mathbb{R}$ as the mapping, the set of basis of tangent space of \mathcal{S} is $\{\partial \log p(y; \theta) / \partial \theta_i\}_{i=1, \dots, n}$. This definition has established the local coordinate systems of statistical manifold \mathcal{S} . Let P and Q be two close points on statistical manifold \mathcal{S} corresponding to the coordinates $\ell(\theta^p)$ and $\ell(\theta^q) = \ell(\theta^p + \Delta\theta)$. Assume that the vector $\vec{PQ} \in T_P \mathcal{S}$ has a fixed length, namely

$$\|\vec{PQ}\|^2 = \epsilon^2 \tag{6}$$

where ε is a sufficiently small positive constant, and a neighborhood ball can be defined according to the constant ε . Then one can obtain the relationship in the tangent space as

$$\overrightarrow{PQ} = \varepsilon \mathbf{v} \tag{7}$$

where $\mathbf{v} = \sum_{i=1}^n a_i \frac{\partial \ell(\theta)}{\partial \theta_i} |_{\theta=\theta^p} \in T_P \mathcal{S}$. This relationship can be considered as a tangent vector of $T_P \mathcal{S}$ at the point P in the view of geometry. For normalizing the tangent vector, the tangent vector satisfies the constraint $\|\mathbf{v}\|^2 = 1$. In the statistical manifold, the tangent vector is with respect to the random variable y . As the description $\mathbf{a} = (a_1, \dots, a_n)^T$ is denoted and the expectation is used in the constraint, we can get the relationship

$$\begin{aligned} \|\mathbf{v}\|^2 &= \mathbf{a}^T \mathbb{E} \begin{bmatrix} \frac{\partial \ell(\theta)}{\partial \theta_1} & \frac{\partial \ell(\theta)}{\partial \theta_1} & \dots & \frac{\partial \ell(\theta)}{\partial \theta_1} & \frac{\partial \ell(\theta)}{\partial \theta_n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial \ell(\theta)}{\partial \theta_n} & \frac{\partial \ell(\theta)}{\partial \theta_1} & \dots & \frac{\partial \ell(\theta)}{\partial \theta_n} & \frac{\partial \ell(\theta)}{\partial \theta_n} \end{bmatrix} \mathbf{a} \\ &= \mathbf{a}^T \mathbf{G} \mathbf{a} = 1 \end{aligned} \tag{8}$$

where \mathbf{G} is the Fisher metric of the statistical manifold \mathcal{S} .

Considering two points P and Q in \mathcal{S} , which have been mapped into the Euclidean space as $\ell(\theta^p)$ and $\ell(\theta^q)$, we can obtain the approximation formation as

$$\begin{aligned} \ell(\theta^q) - \ell(\theta^p) &= (\theta^p + \Delta\theta - \theta^p)^T \nabla \ell(\theta^p) \\ &= \Delta\theta^T \nabla \ell(\theta^p) \end{aligned} \tag{9}$$

where

$$\nabla \ell(\theta^p) = \left(\frac{\partial \ell(\theta)}{\partial \theta_1}, \dots, \frac{\partial \ell(\theta)}{\partial \theta_n} \right)^T \Bigg|_{\theta=\theta^p} \tag{10}$$

For simplifying the description, the tangent vector \mathbf{v} can be rewritten as

$$\mathbf{v} = \sum_{i=1}^n a_i \frac{\partial \ell(\theta)}{\partial \theta_i} = \mathbf{a}^T \nabla \ell(\theta) \tag{11}$$

In the tangent space of the manifold \mathcal{S} at the point P , the two points satisfy the relationship

$$\begin{aligned} \overrightarrow{PQ} &= \ell(\theta^q) - \ell(\theta^p) \\ &= \ell(\theta^p + \Delta\theta) - \ell(\theta^p) \end{aligned} \tag{12}$$

Combining the Equations (7), (9) and (12), we can get the relationship between the parameter space and the tangent space of manifold

$$\Delta\theta = \varepsilon \mathbf{a} \tag{13}$$

To obtain the parameter \mathbf{a} in the tangent space under the constraint (8), we use the Lagrange function method

$$F(\mathbf{a}, \lambda) = \ell(\theta^p) + \varepsilon \mathbf{a}^T \nabla \ell(\theta^p) + \lambda(1 - \mathbf{a}^T \mathbf{G} \mathbf{a}) \tag{14}$$

where λ is the Lagrange multiplier. Then we have

$$\frac{\partial}{\partial \mathbf{a}_i} \{ \varepsilon \mathbf{a}^T \nabla \ell(\theta^p) + \lambda(1 - \mathbf{a}^T \mathbf{G} \mathbf{a}) \} = 0 \tag{15}$$

Solving this equation, one obtains

$$\mathbf{a} = \frac{\varepsilon}{2\lambda} \mathbf{G}^{-1} \nabla \ell (\theta^p) \quad (16)$$

Substituting the Equation (16) into (8), we obtain the conclusion in equation

$$\begin{aligned} 1 = \mathbf{a}^T \mathbf{G} \mathbf{a} &= \left(\frac{\varepsilon}{2\lambda} \mathbf{G}^{-1} \nabla \ell (\theta^p) \right)^T \mathbf{G} \left(\frac{\varepsilon}{2\lambda} \mathbf{G}^{-1} \nabla \ell (\theta^p) \right) \\ &= \frac{\varepsilon^2}{4\lambda^2} \nabla \ell (\theta^p)^T \mathbf{G}^{-1} \nabla \ell (\theta^p) \end{aligned} \quad (17)$$

Since $\nabla \ell (\theta^p)^T \mathbf{G}^{-1} \nabla \ell (\theta^p)$ is positive, the value of multiplier λ can be computed as

$$\lambda = \frac{\varepsilon}{2} \sqrt{\nabla \ell (\theta^p)^T \mathbf{G}^{-1} \nabla \ell (\theta^p)} \quad (18)$$

Then, the relationship between point P and Q in the tangent space of manifold S can be obtained.

$$\ell (\theta^p + \Delta\theta) = \ell (\theta^p) + \varepsilon \mathbf{a}^T \nabla \ell (\theta^p) \quad (19)$$

Simultaneously, the correspondence relationship in the parameter space $\Theta \subset \mathbb{R}^n$ is

$$\begin{aligned} \theta^q &= \theta^p + \Delta\theta = \theta^p + \varepsilon \mathbf{a} \\ &= \theta^p + \frac{\varepsilon}{\sqrt{\nabla \ell (\theta^p)^T \mathbf{G}^{-1} \nabla \ell (\theta^p)}} \mathbf{G}^{-1} \nabla \ell (\theta^p) \end{aligned} \quad (20)$$

where the \mathbf{G}^{-1} is the inverse of the Fisher metric \mathbf{G} . In particular, the representation $\mathbf{G}^{-1} \nabla \ell (\theta)$ is called natural gradient of the mapping ℓ in the Riemannian manifold. Compared with the steepest direction $\nabla \ell (\theta)$ in the Euclidean space, the natural gradient introduces the inverse of Fisher metric \mathbf{G} to describe the curvature of Riemannian manifold, and it is invariant under the choice of coordinate system.

As the simplification form of Amari suggested in the article [16], the Equation (20) can be written as

$$\theta^q = \theta^p + \eta \mathbf{G}^{-1} (\theta^p) \nabla \ell (\theta^p) \quad (21)$$

where the parameter satisfies $0 < \eta \leq 1$ and it controls the convergence speed.

The geometric interpretation of this method is that the process path has moved from point P to point Q in the manifold S along the steepest direction, while the parameter has moved from θ^p to θ^q in parameter space. As the process path has been moving in the manifold S , the parameter has been estimated recursively. The recursive formation as follow

$$\theta^{t+1} = \theta^t + \eta \mathbf{G}^{-1} (\theta^t) \nabla \ell (\theta^t) \quad (22)$$

where t denotes the times of iterative procedure in which correspond with the iterated points in the statistical manifold. This iteration procedure has been illustrated in Figure 1.

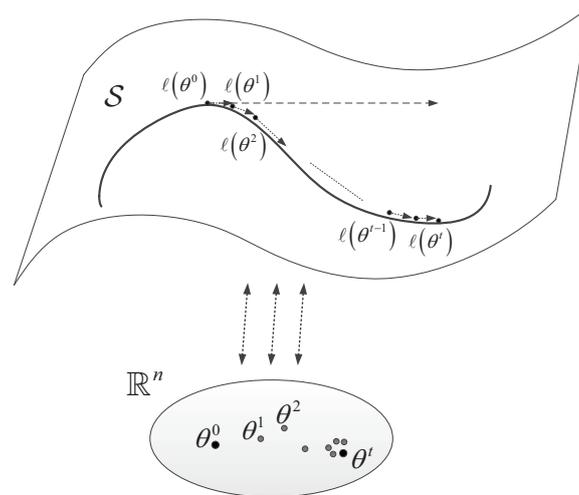


Figure 1. The iterative procedure for estimation in parameter space can be converted to the recursive estimation in the statistical manifold from the information geometric viewpoint.

For the recursive process, we can use KLD, which is often used to show the discrepancy between two probability distributions, as the stopping criterion. It is defined by

$$D[\theta^p : \theta^q] = \int p(y; \theta^p) \log \frac{p(y; \theta^p)}{p(y; \theta^q)} dy \tag{23}$$

where $p(y; \theta^p)$ and $p(y; \theta^q)$ as the specification with respect to two parameters θ^p and θ^q , respectively. When the two probability distributions are infinitesimally close, corresponding with $\theta^q = \theta^p + d\theta$, the KLD between two nearby distributions can be expanded as

$$D[\theta^p : \theta^p + d\theta] = \frac{1}{2} d\theta^T \mathbf{G}(\theta) d\theta \tag{24}$$

Motivated by the relationship between the statistical manifold and the parameter space, the estimation of state conditional on the measurement in state parameter space can be converted to the iteration procedure on the manifold in which constructed by the posterior PDF. This method provides a new alternative for the measurement update in nonlinear filtering based on the Bayesian framework.

3. Filtering Based on Information Geometry

In this section, we present the ideas of measurement update based on the natural gradient descent. In Bayesian framework, the predicted state acts as the prior information with respect to the state to deduce the posterior distribution under the condition of measurements. Then the natural gradient descent method has been used to estimation the final state recursively. Here, we consider the usual scenario which discrete-time state space as follow

$$x_k = f(x_{k-1}) + w_k \tag{25}$$

$$y_k = h(x_k) + v_k \tag{26}$$

where $x_k \in \mathbb{R}^n$ and $y_k \in \mathbb{R}^m$ denote the system discrete-time state and measurement at the instant k respectively, f and h denote the state transition and measurement functions. w_k and v_k are the noise correspondence with state and measurement respectively. Generally the additive white Gaussian noise (AWGN) has been studied for simplicity.

In Bayesian filtering, the time propagation can be processed according to the state transition function, and the prediction of state can be achieved in this step. Based on the Bayesian principle, the prediction of state is considered as the prior information about the state.

$$\begin{aligned} p(\hat{x}_k^-) &= p(x_k | \mathbf{y}_{k-1}) \\ &= \int p(x_k | x_{k-1}) p(x_{k-1} | \mathbf{y}_{k-1}) dx_{k-1} \\ &= \int p(x_k | x_{k-1}) p(x_{k-1} | y_{k-1}) dx_{k-1} \end{aligned} \quad (27)$$

where $\mathbf{y}_{k-1} = \{y_{k-1}, y_{k-2}, \dots, y_1\}$. The relationship $p(x_{k-1} | \mathbf{y}_{k-1}) = p(x_{k-1} | y_{k-1})$ means that the current state only depends on current measurement and is conditionally independent from all past measurements. The PDF $p(x_k | x_{k-1})$ and $p(x_{k-1} | y_{k-1})$ denote the state transition PDF and posterior PDF of state for last time, respectively. To aim at obtaining the posterior PDF of current time under the condition of current measurement, we can use the Bayesian principle for deducing. The procedure of the measurement update in the form of Bayesian filtering as

$$\begin{aligned} p(\hat{x}_k^+) &= p(x_k | \mathbf{y}_k) \\ &= \frac{p(x_k, y_k | \mathbf{y}_{k-1}) p(\mathbf{y}_{k-1})}{p(y_k | \mathbf{y}_{k-1}) p(\mathbf{y}_{k-1})} \\ &= \frac{p(y_k | x_k, \mathbf{y}_{k-1}) p(x_k | \mathbf{y}_{k-1})}{\int p(y_k | x_k, \mathbf{y}_{k-1}) p(x_k | \mathbf{y}_{k-1}) dx_k} \\ &= \frac{p(y_k | x_k) p(x_k | y_{k-1})}{\int p(y_k | x_k) p(x_k | y_{k-1}) dx_k} \end{aligned} \quad (28)$$

since the likelihood $p(y_k | x_k, \mathbf{y}_{k-1}) = p(y_k | x_k)$, i.e., the measurement y_k is conditionally independent of all past measurements. The denominator of the posterior PDF is the normalizing constant independent of the state. Thus the posterior PDF can be rewrite as

$$\begin{aligned} p(\hat{x}_k^+) &= p(x_k | y_k) \\ &\propto p(y_k | x_k) p(x_k | y_{k-1}) \\ &= p(y_k | x_k) p(\hat{x}_k^-) \end{aligned} \quad (29)$$

where \propto means "is proportional to". $p(\hat{x}_k^-)$ and $p(y_k | x_k)$ are the prior and the likelihood respectively. In this paper, we assume that they are all Gaussian distribution for simplicity, which the Gaussian approximation has been used for non-Gaussian distribution and beyond our paper. Based on the procedure of state propagation, the prior distribution can be described as

$$p(\hat{x}_k^-) = \frac{\exp\left[-\frac{1}{2}(x_k - \hat{x}_k^-)^T \mathbf{P}^{-1}(x_k - \hat{x}_k^-)\right]}{\sqrt{(2\pi)^n |\mathbf{P}|}} \quad (30)$$

Usually, the likelihood with respect to the measurement is chosen as

$$p(y_k | x_k) = \frac{\exp\left[-\frac{1}{2}(y_k - h(x_k))^T \mathbf{R}^{-1}(y_k - h(x_k))\right]}{\sqrt{(2\pi)^m |\mathbf{R}|}} \quad (31)$$

where $|\cdot|$ is the determinant, n and m are the number of dimension of state and measurement, respectively.

In the stage of measurement update, the goal is to obtain the optimal estimation of state under the conditions of the corresponding measurement and the prior information of state. From the Bayesian perspective, it is the maximum a posterior PDF of the prior and the likelihood. This MAP method has been used in the estimation for statistical signal processing extensively. With the optimization view,

the solution that maximizes $p(\hat{x}_k^+)$ is equivalent to minimizing its negative log likelihood function. The negative log operation of posterior distribution is

$$L(x_k) = -\log p(\hat{x}_k^+) = \frac{1}{2} (x_k - \hat{x}_k^-)^T \mathbf{P}^{-1} (x_k - \hat{x}_k^-) + \frac{1}{2} (y_k - h(x_k))^T \mathbf{R}^{-1} (y_k - h(x_k)) + C \tag{32}$$

where C is a constant which not effect the estimation of state x_k . Defining the negative log likelihood function

$$\ell(x_k) = \frac{1}{2} (x_k - \hat{x}_k^-)^T \mathbf{P}^{-1} (x_k - \hat{x}_k^-) + \frac{1}{2} (h(x_k) - y_k)^T \mathbf{R}^{-1} (h(x_k) - y_k) \tag{33}$$

thus the object function can be convert into

$$\max_{x_k} p(\hat{x}_k^+) = \min_{x_k} \ell(x_k) \tag{34}$$

Based on the posterior PDF, the statistical manifold \mathcal{S} can be defined as

$$\mathcal{S} = \{p(x_k|y_k)\} \tag{35}$$

Meanwhile, the negative log function as the mapping from \mathcal{S} to \mathbb{R} , i.e., $\ell : \mathcal{S} \rightarrow \mathbb{R}$ has constructed the coordinate system.

Let ℓ_1 and ℓ_2 stand for the first and second part of the Equation (33), $e_y = h(x_k) - y_k$ and $e_x = x_k - \hat{x}_k^-$ for the error of measurement and the state estimation respectively. Computing the first derivative of ℓ_1 and ℓ_2 about the x_k and x_k^T respectively

$$\frac{\partial \ell_1}{\partial x_k} = (\partial h)^T \mathbf{R}^{-1} e_y \tag{36}$$

$$\frac{\partial \ell_1}{\partial x_k^T} = e_y^T \mathbf{R}^{-1} (\partial h) \tag{37}$$

$$\frac{\partial \ell_2}{\partial x_k} = \mathbf{P}^{-1} e_x \tag{38}$$

$$\frac{\partial \ell_2}{\partial x_k^T} = e_x^T \mathbf{P}^{-1} \tag{39}$$

where $\partial h = \frac{\partial h(x_k)}{\partial x_k} \in \mathbb{R}^m \times \mathbb{R}^n$ denotes the Jacobian of $h(x_k)$ about x_k . The covariance matrix \mathbf{R}^{-1} and \mathbf{P}^{-1} are symmetry. Thus the gradient of the negative log likelihood is

$$\nabla \ell(x_k) = (\partial h)^T \mathbf{R}^{-1} e_y + \mathbf{P}^{-1} e_x \tag{40}$$

Then the Fisher metric \mathbf{G} of manifold \mathcal{S} can be obtained

$$\begin{aligned} \mathbf{G}(x_k) &= \mathbf{E} \left[\frac{\partial \ell}{\partial x_k} \left(\frac{\partial \ell}{\partial x_k} \right)^T \right] = -\mathbf{E} \left[\frac{\partial^2 \ell}{\partial x_k \partial x_k^T} \right] \\ &= -\mathbf{E} \left[\frac{\partial}{\partial x_k} \left(e_y^T \mathbf{R}^{-1} (\partial h) + e_x^T \mathbf{P}^{-1} \right) \right] \\ &= -\mathbf{E} \left[e_y^T \mathbf{R}^{-1} (\partial^2 h) + (\partial h)^T \mathbf{R}^{-1} (\partial h) + \mathbf{P}^{-1} \right] \\ &= -(\partial h)^T \mathbf{R}^{-1} (\partial h) - \mathbf{P}^{-1} \end{aligned} \tag{41}$$

where $\partial^2 h = \frac{\partial^2 h(x_k)}{\partial x_k \partial x_k^T} \in \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^n$ and E denotes the expectation with respect to the measurement y_k . In the above deducing procedure, the follow principle has been used

$$D_x \left(\mathbf{f}(x)^T \mathbf{A} \mathbf{g}(x) \right) = \mathbf{g}(x)^T \mathbf{A}^T D_x \mathbf{f}(x) + \mathbf{f}(x)^T \mathbf{A} D_x \mathbf{g}(x) \tag{42}$$

With the computed gradient and Fisher metric, the natural gradient of manifold is

$$\mathbf{G}(x_k)^{-1} \nabla \ell(x_k) = - \left((\partial h)^T \mathbf{R}^{-1} (\partial h) + \mathbf{P}^{-1} \right)^{-1} \left((\partial h)^T \mathbf{R}^{-1} e_y + \mathbf{P}^{-1} e_x \right) \tag{43}$$

Meanwhile, considering the denominator of Equation (20), it can be computed as

$$\begin{aligned} & \nabla \ell(x_k)^T \mathbf{G}(x_k)^{-1} \nabla \ell(x_k) \\ &= \text{tr} \left[\mathbf{G}(x_k)^{-1} \nabla \ell(x_k) \nabla \ell(x_k)^T \right] \\ &= \text{tr} \left[(\partial h^T \mathbf{R}^{-1} \partial h + \mathbf{P}^{-1})^{-1} (\partial h^T \mathbf{R}^{-1} e_y + \mathbf{P}^{-1} e_x) (\partial h^T \mathbf{R}^{-1} e_y + \mathbf{P}^{-1} e_x)^T \right] \\ &= \text{tr} \left[(\partial h^T \mathbf{R}^{-1} \partial h + \mathbf{P}^{-1})^{-1} (\partial h^T \mathbf{R}^{-1} e_y e_y^T \mathbf{R}^{-1} \partial h + \partial h^T \mathbf{R}^{-1} e_y e_x^T \mathbf{P}^{-1} + \mathbf{P}^{-1} e_x e_y^T \mathbf{R}^{-1} \partial h + \mathbf{P}^{-1} e_x e_x^T \mathbf{P}^{-1}) \right] \end{aligned} \tag{44}$$

For simplicity, we use the expectation value to substitute the value which computed in each iterative procedure. Thus

$$\begin{aligned} & E \left[\mathbf{G}(x_k)^{-1} \nabla \ell(x_k) \nabla \ell(x_k)^T \right] \\ &= \left(\partial h^T \mathbf{R}^{-1} \partial h + \mathbf{P}^{-1} \right)^{-1} \left(\partial h^T \mathbf{R}^{-1} \mathbf{R} \mathbf{R}^{-1} \partial h + \mathbf{P}^{-1} \mathbf{P} \mathbf{P}^{-1} \right) \\ &= \left(\partial h^T \mathbf{R}^{-1} \partial h + \mathbf{P}^{-1} \right)^{-1} \left(\partial h^T \mathbf{R}^{-1} \partial h + \mathbf{P}^{-1} \right) \\ &= \mathbf{I} \end{aligned} \tag{45}$$

where e_y and e_x are independent, and \mathbf{I} is the n -dimension square matrix which correspondence to the n -dimension state. Based on the natural gradient descent, we can construct the recursive procedure to get the optimal estimation in the measurement update step

$$\begin{aligned} x_k^t &= x_k^{t-1} + \frac{\varepsilon}{\sqrt{\nabla \ell(x_k^{t-1})^T \mathbf{G}^{-1} \nabla \ell(x_k^{t-1})}} \mathbf{G}^{-1}(x_k^{t-1}) \nabla \ell(x_k^{t-1}) \\ &\approx x_k^{t-1} + \frac{\varepsilon}{\sqrt{\text{tr}(\mathbf{I})}} \mathbf{G}^{-1}(x_k^{t-1}) \nabla \ell(x_k^{t-1}) \\ &\approx x_k^{t-1} + \frac{\varepsilon}{\sqrt{n}} \mathbf{G}^{-1}(x_k^{t-1}) \nabla \ell(x_k^{t-1}) \end{aligned} \tag{46}$$

where n is the dimension of the state. Replacing the parameter as $\eta = \frac{\varepsilon}{\sqrt{n}}$, the natural gradient descent method becomes the simplification form suggested by Amari. Given the above, the iterative procedure for estimating the state can be constructed

$$\begin{aligned} x_k^t &= x_k^{t-1} + \eta \mathbf{G}^{-1}(x_k^{t-1}) \nabla \ell(x_k^{t-1}) \\ &= x_k^{t-1} + \eta \left(\partial h^T \mathbf{R}^{-1} \partial h + \mathbf{P}_{t-1}^{-1} \right)^{-1} \left(\partial h^T \mathbf{R}^{-1} (y_k - h(x_k)) + \mathbf{P}_{t-1}^{-1} (x_k^{t-1} - x_k) \right) \Big|_{x_k=x_k^{t-1}} \\ &= x_k^{t-1} + \eta \left(\partial h_{t-1}^T \mathbf{R}^{-1} \partial h_{t-1} + \mathbf{P}_{t-1}^{-1} \right)^{-1} \left(\partial h_{t-1}^T \mathbf{R}^{-1} \right) \left(y_k - h(x_k^{t-1}) \right) \end{aligned} \tag{47}$$

where $\partial h_{t-1} = \frac{\partial h(x_k)}{\partial x_k} \Big|_{x_k=x_k^{t-1}}$, and \mathbf{P}_{t-1} is the covariance matrix after the $t - 1$ iterative procedure.

For the iterative procedure, a stopping criterion should be set for final estimation. In the statistical manifold, the KLD is a better criterion for measurement the distance between two PDF. Because of

the estimated state corresponding with the posterior PDF, the convergence in statistical manifold also means the better estimation in parameter space. Besides, Morelande [15] has shown that the lower KLD value to select the optimum parameter for the approximation posterior PDF has better performance. Thus, we choose the KLD as the stopping criterion in the iteration procedure of measurement update following the statistical manifold.

$$D[x_k^{t-1} : x_k^t] = \frac{1}{2}(x_k^t - x_k^{t-1})^T \mathbf{G} (x_k^t - x_k^{t-1}) \leq \delta \quad (48)$$

Also, we can choose a stopping criterion as the follow form from the view of parameter space in which has been used in the IEKF or other iterated estimation problems

$$\|\hat{x}_k^t - \hat{x}_k^{t-1}\|_2^2 \leq \gamma \quad (49)$$

Based on this measurement update method, we can construct nonlinear filtering by combining the existing method for state propagation. Here, we use the Ensemble Kalman filter (EnKF) for state propagation, which uses the Monte Carlo technique for integral operation in the Bayesian filtering. For the last procedure ensemble has N_s elements as

$$S_{k-1|k-1} = \{\hat{x}_{k-1|k-1}^i, i = 1, \dots, N_s\} \quad (50)$$

We can get the prediction ensemble as

$$S_{k|k-1} = \{\hat{x}_{k|k-1}^i = f(\hat{x}_{k-1|k-1}^i) + w_k^i, i = 1, \dots, N_s\} \quad (51)$$

where w_k^i is generated according the state process noise. Then, the mean and the covariance matrix can be computed

$$\bar{x}_{k|k-1} = \frac{1}{N_s} \sum_{i=1}^{N_s} \hat{x}_{k|k-1}^i \quad (52)$$

$$\mathbf{P}_{k|k-1} = \frac{1}{N_s - 1} \sum_{i=1}^{N_s} (\hat{x}_{k|k-1}^i - \bar{x}_{k|k-1})(\hat{x}_{k|k-1}^i - \bar{x}_{k|k-1})^T \quad (53)$$

In the Bayesian framework, these prediction means and covariance will be incorporated in the procedure as prior information of state to propel the measurement update by the natural gradient method. With the stopping criterion of iterative procedure, we can get the final estimation of state. Meanwhile, the iterative estimation will converge to its final value because of the convergence of the natural gradient descent, and the iterative procedure can process some iterative estimation value of state. Here, we consider these values as the ensemble of measurement update for next period of nonlinear filter in the EnKF.

$$S_{k|k} = \{\hat{x}_{k|k}^i, i = 1, \dots, N_s\} \quad (54)$$

Thus, the complete nonlinear filter method has proposed based on state propagation by EnKF and measurement update by natural gradient descent method, and has been shown in Algorithm 1.

Algorithm 1 Natural Gradient Descent Method for Measurement Update in EnKF

Input: Measurement $\mathbf{y} = \{y_1, y_2, \dots, y_T\}$, ensemble N_s , parameter $\eta \in (0, 1]$.

Initialize: $x_0, \mathbf{P}_0, \hat{x}_0^i = x_0 + \sqrt{\mathbf{P}_0} \mathcal{N}(0, 1)$, ensemble $S_{0|0} = \{\hat{x}_0^i, i = 1, \dots, N_s\}$.

For $k = 1 : T$

 % compute prediction ensemble

$S_{k|k-1} = \{\hat{x}_{k|k-1}^i = f(\hat{x}_{k-1|k-1}^i) + w_{k'}^i, i = 1, \dots, N_s\}$;

 % compute mean and covariance of prediction ensemble

$\bar{x}_{k|k-1} = \frac{1}{N_s} \sum_{i=1}^{N_s} \hat{x}_{k|k-1}^i$;

$\mathbf{P}_{k|k-1} = \frac{1}{N_s-1} \sum_{i=1}^{N_s} (\hat{x}_{k|k-1}^i - \bar{x}_{k|k-1})(\hat{x}_{k|k-1}^i - \bar{x}_{k|k-1})^T$;

 % natural gradient descent for measurement update

$x_{k|k}^0 = \bar{x}_{k|k-1}, \mathbf{P}_{k|k}^0 = \mathbf{P}_{k|k-1}, t = 1$;

For $t = 1 : N_s$

$\mathbf{H}_{t-1} = \left. \frac{\partial h(x_k)}{\partial x_k} \right|_{x_k = x_{k|k}^{t-1}}$;

$\mathbf{G} = \mathbf{H}_{t-1}^T \mathbf{R}^{-1} \mathbf{H}_{t-1} + (\mathbf{P}_{k|k}^{t-1})^{-1}$;

$\nabla \ell = (\mathbf{H}_{t-1}^T \mathbf{R}^{-1}) (y_k - h(x_{k|k}^{t-1}))$;

$x_{k|k}^t = x_{k|k}^{t-1} + \eta \mathbf{G}^{-1} \nabla \ell$;

$\mathbf{P}_{k|k}^t = \frac{1}{N_s-1} \sum_{i=1}^{N_s} (\hat{x}_{k|k-1}^i - x_{k|k}^t)(\hat{x}_{k|k-1}^i - x_{k|k}^t)^T$;

$\mathbf{D}[x_k^{t-1} : x_k^t] = \frac{1}{2} (x_{k|k}^t - x_{k|k}^{t-1})^T \mathbf{G} (x_{k|k}^t - x_{k|k}^{t-1})$;

End for

 % Select final estimation

For $t = 1 : N_s$

If $\mathbf{D}[x_k^{t-1} : x_k^t] \leq \delta$ and $\|\hat{x}_{k|k}^t - \hat{x}_{k|k}^{t-1}\|_2 \leq \gamma$

$\hat{x}_{k|k} = x_{k|k}^t$;

$\mathbf{P}_{k|k} = \mathbf{P}_{k|k}^t$;

End if

End for

$S_{k|k} = \{\hat{x}_{k|k}^i, i = 1, \dots, N_s\}$;

End for

Output: State $\mathbf{x} = \{\hat{x}_{k|k}, k = 1, \dots, T\}$

4. Comparing with the Existing Methods

As the natural gradient descent method for measurement update, we can get some conclusions comparing with the existing methods. When we choose the parameter as $\eta = 1$ with a single iteration, the iterative procedure (47) will be simplified as

$$x_k = x_{k-1} + \left(\partial h_{k-1}^T \mathbf{R}^{-1} \partial h_{k-1} + \mathbf{P}^{-1} \right)^{-1} \left(\partial h_{k-1}^T \mathbf{R}^{-1} \right) (y_k - h(x_{k-1})) \quad (55)$$

Thus, the update has deduced to the traditional Kalman filter and the EKF. When the measurement function is linear, i.e., $y_k = h(x_k) = \mathbf{H}x_k$, the Jacobian of $h(x_k)$ can be obtained easily as $\frac{\partial h(x_k)}{\partial x_k} = \mathbf{H}$, $\frac{\partial h(x_k)}{\partial x_k^T} = \mathbf{H}^T$. As for the nonlinear measurement function, the first-order Taylor expansion has been used as the same as the linear case. The procedure becomes

$$\begin{aligned} x_k &= x_{k-1} + \left(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{P}^{-1} \right)^{-1} \left(\mathbf{H}^T \mathbf{R}^{-1} (y_k - h(x_{k-1})) \right) \\ &= x_{k-1} + \left(\mathbf{P} \mathbf{H}^T \right) \left(\mathbf{H} \mathbf{P} \mathbf{H}^T + \mathbf{R} \right)^{-1} (y_k - h(x_{k-1})) \end{aligned} \quad (56)$$

where the matrix inversion lemma

$$(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{P}^{-1})^{-1} \mathbf{H}^T \mathbf{R}^{-1} = \mathbf{P} \mathbf{H}^T (\mathbf{H} \mathbf{P} \mathbf{H}^T + \mathbf{R})^{-1} \quad (57)$$

has been used in the simplification. In comparison, this is the same as the measurement update in traditional Kalman filter and EKF.

Besides, there are two iteration methods existing before. The one is IEKF [12,32], which also has been derived based on the MAP method. Its update step has the form as

$$x_k^t = \hat{x}_k^- + \left(\mathbf{P} \mathbf{H}_{t-1}^T \right) \left(\mathbf{H}_{t-1} \mathbf{P} \mathbf{H}_{t-1}^T + \mathbf{R} \right)^{-1} \left(y_k - h(x_k^{t-1}) - \mathbf{H}_{t-1} (\hat{x}_k^- - x_k^{t-1}) \right) \quad (58)$$

where $\mathbf{H}_{t-1} = \left. \frac{\partial h(x_k)}{\partial x_k} \right|_{x_k = x_k^{t-1}}$.

Comparing with IEKF and our proposed method, there are some differences. Firstly, the initial estimation value in each iteration is the same as the prior estimation in the IEKF, while our proposed method bases on the last iterated estimation. Secondly, the parameter η of the natural gradient descent controls the update step of state, but the IEKF has no parameters to control the update step. Thirdly, the innovations are used in the natural gradient descent method, while the IEKF adds the prediction error of current estimation state and the prior estimation state. Further from the formulation of the IEKF update, the first step of IEKF is identified to the EKF, and later steps are from the direction of overestimation convergence to the true state.

Another iteration method is RUF [9], which derived based on the LMMSE method. It defines the cross-covariance between the measurement noise and the estimation error for each steps. Then the LMMSE method has been used to deduce the update step. Also, it has considered the fraction of update, which selected as the reciprocal of the number of iteration steps. In comparison with our proposed method, the RUF uses the manual setting number of iteration steps, which also influences the update step, and has no criterion for evaluation the estimation performance. In some case, the other criterions may provide guarantee for convergence and performance in the procedure. Moreover the detailed differences among these methods will be discussed in the numerical examples in the next section.

5. Numerical Examples

In this section, we present three numerical examples to demonstrate the advantages of our proposed method and illustrate the differences between the proposed method and the existing method. The first scalar example provides a good geometrical interpretation in the convergence speed and the iteration procedure. The second example, which has been used as the common test platform for nonlinear filtering in scalar, is provided to validate the performance. The third example is a multidimensional filtering problem in which has tracked a vehicle entering atmosphere.

5.1. Iteration and Convergence for Measurement Update

Consider a sample scalar nonlinear measurement as

$$y = x^5 + v \quad (59)$$

where x denotes the state, and y denotes the measurement with noise $v \sim \mathcal{N}(0, 0.1^2)$. Only for the measurement update step, we assume the mean and variance of prediction state, which as prior information in the Bayesian framework, are $\hat{x}^- = 2.5$ and $\mathbf{P} = 0.5^2$, and the true state value is set at $x = 4$. With noisy measuring, the measurement $y = 1024.4$. We compare our proposed method with EKF, IEKF, and RUF. The results are shown in the Figure 2. The EKF only has one step in the procedure for estimation the posterior, while the other methods have more than one steps. For the IEKF method, the results illustrate that the first step of IEKF is the same as the EKF, which also can be

deduced from the expression of two methods. However, the more iteration steps in the IEKF make the estimation converge to the true posterior state. In detail, the second and more iteration steps in the IEKF have the same beginning state for estimation, but the ending state after estimation is more close to true state. Completing the procedure, the more steps in the IEKF, the more accurate estimation than the EKF. Considering the RUF, the error of estimation is incorporated in the to refine the estimation. After some iterated steps, the error would be approximately reduced to zero. Thus the final estimation is convergence to the true state. For our proposed method, the parameter η controls the convergence behaviours to true state. When the parameter $\eta = 0.8$, the first step for estimation is bigger than the true state as same as the EKF, then the later iteration steps convergence to true state in the contrary direction of RUF. As the parameter $\eta = 0.2$, the convergence path of iteration steps is the same direction of RUF, but has the fast convergence speed. Moreover, we consider the convergence of the natural gradient descent method. We use the random technique to generate the initial state and parameter $\eta = 0.5$. The results are show in Figure 3 after running 100 Monte Carlo. It has showed that the iteration steps, which the natural gradient descent convergence to the true state, always need less 30 times.

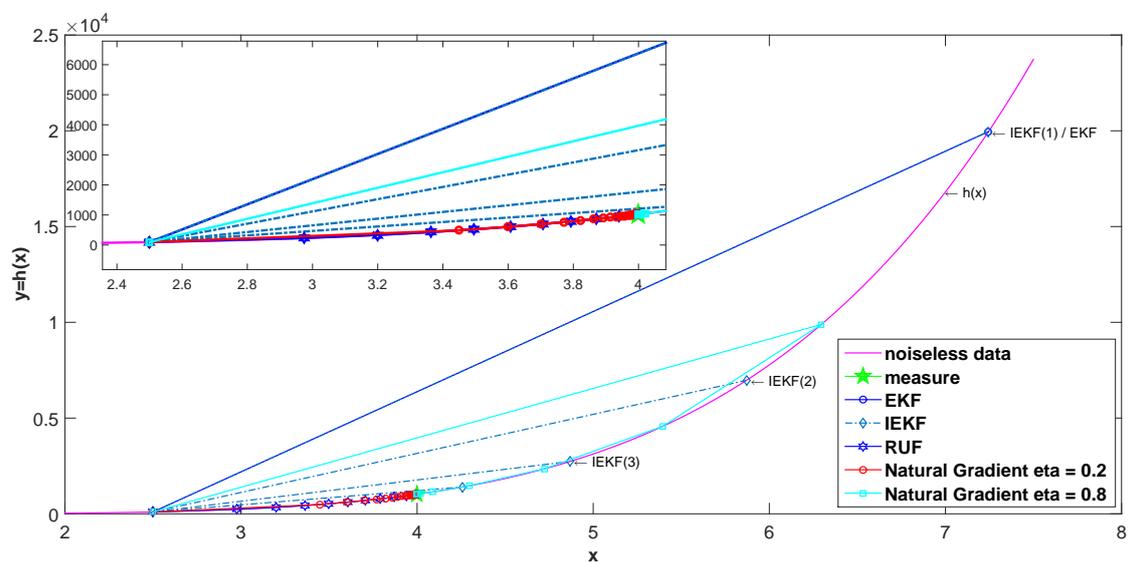


Figure 2. Comparison of the iteration procedure.

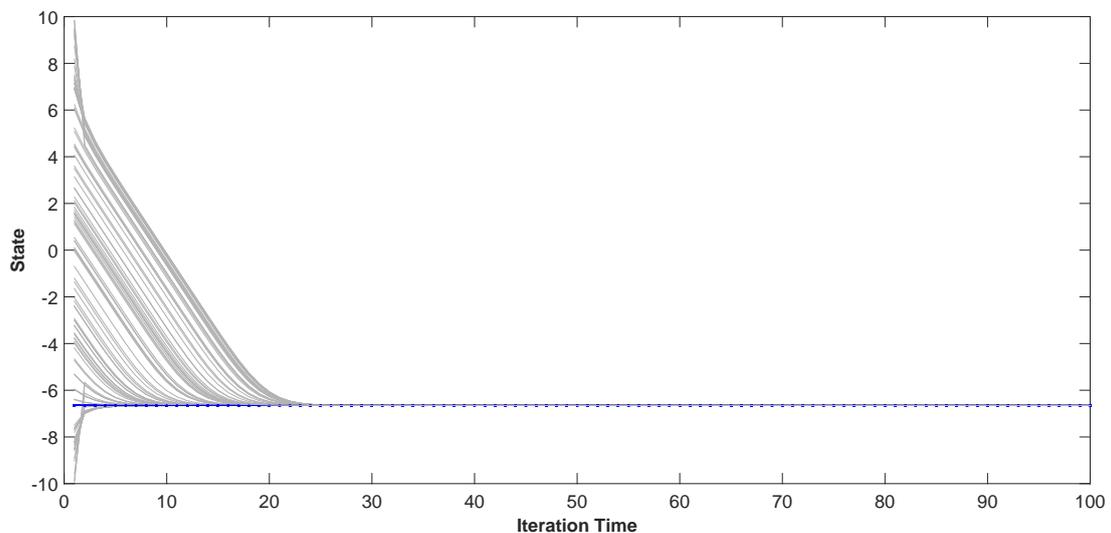


Figure 3. The convergence of the natural gradient descent method.

5.2. Univariate Nonstationary Growth Model (UNGM)

This model has been used extensively to validate the performance of nonlinear filtering because of its high nonlinearity. The model can be formulated as [33]

$$x_k = \frac{x_{k-1}}{2} + \frac{25x_{k-1}}{1+x_{k-1}^2} + 8 \cos(1.2k) + u_k \quad (60)$$

$$y_k = \frac{x_k^3}{20} + v_k \quad (61)$$

where u_k and v_k denote the additive white Gaussian noise corresponding with state and measurement, as $u_k \sim \mathcal{N}(0, Q)$ and $v_k \sim \mathcal{N}(0, R)$. In this experiment, we set the variance as $Q = 10$ and $R = 1$, respectively. We assume the initiate state as $p(x_0) = \mathcal{N}(x_0; 0, 1)$ before the filtering procedure. The true state that we used in the simulations with 100 time steps has a severe fluctuation, which has showed in the Figure 4. The state varies frequently, and the first order Taylor series expanded can not depict fitting in the procedure. Thus, the EKF has the bad performance in the experiment. In addition, the state will make the nonlinear measurement more difficult.

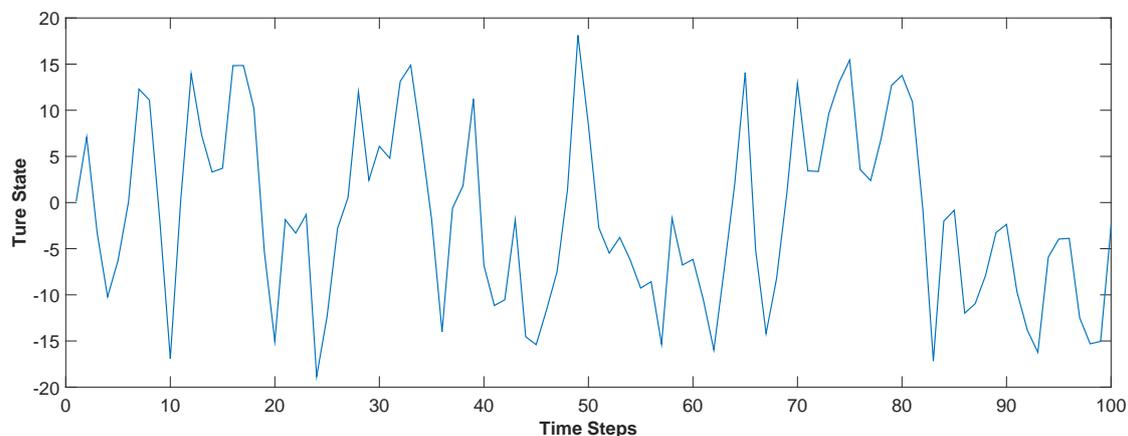


Figure 4. True state in the experiment. It varies in short time, and brings some difficult for the measurement update.

As a performance metric, we use Root Mean Square Error (RMSE) in the example. Let \hat{x}_k^i and x_k^i denote the estimated state and true state at time k , respectively. Then the RMSE is defined as

$$\text{RMSE}_k = \sqrt{\frac{1}{M} \sum_{i=1}^M (\hat{x}_k^i - x_k^i)^2}$$

where M is the total number of Monte Carlo runs.

In our experiment, the RMSE is obtained by using 100 Monte Carlo runs, averaging over different realizations of the measurements and the true states. The comparison is achieved between our proposed method, which called natural gradient descent Kalman filter (NgdKF) here, and the four existing methods, i.e., EnKF, IEKF, EKF, and RUF. In the EnKF and our proposed method, the number of ensemble used each time is 200, and the initial ensemble generated randomly with $p(x_0)$. During the procedure of IEKF, the stopping criterion is $\|\hat{x}_{k|k}^t - \hat{x}_{k|k}^{t-1}\|_2^2 \leq 10^{-4}$, while $\|\hat{x}_{k|k}^t - \hat{x}_{k|k}^{t-1}\|_2^2 \leq 10^{-4}$ and $\mathbf{D}[x_k^{t-1} : x_k^t] \leq 10^{-5}$ in our proposed method. The number of iteration steps in the RUF is set as 200. The parameter $\eta = 0.5$ is selected in our proposed method for the iterated step by natural gradient descent method. The comparison results measured by the RMSE are showed in the Figure 5. As we can see that our proposed method performs much better than EKF, EnKF, RUF, and IEKF. As aforementioned in the Figure 4, the EKF becomes poor because of first order Taylor series expanded

is not fit the measurement function. In comparing the iteration methods, i.e., IEKF, RUF and NgdKF, with no iteration method, i.e., EKF and EnKF, the front class methods have the better performance than the later. This conclusion also has been mentioned by Lefebvre [14]. However, the later is computed faster than the front. In fact, from the update formulation of each method, the iteration methods have more than one steps, while the no iteration only has one step. In the nonlinear filter, we have to take more times or steps to improve the filtered precision. Based on the comparison results, our proposed method has the better performance.

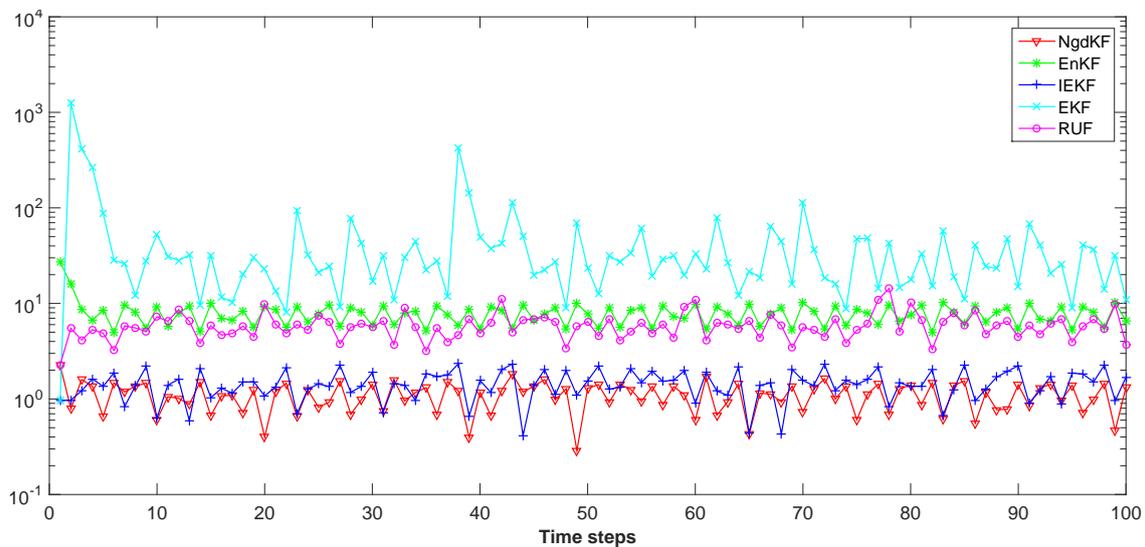


Figure 5. RMSE Comparison. Our proposed method (NgdKF) is better perform than other existing methods.

5.3. Tracking of Vehicle Entering Atmosphere

In the previous example, we have shown the performance in scalar filtering problem. In this example, we will study a multidimensional dynamic system. We analyze target tracking problem of a vehicle entering atmosphere at high altitude with a high speed [34]. It has the continuous state, which constructed by position (x_1 and x_2), velocity (x_3 and x_4) and drag coefficient (x_5), and the discrete measurement constructed by range (r_k) and bearing (b_k). This experiment is a better platform to measure the filtering algorithms. Firstly, the number of dimension between the state and the measurement is difference. Secondly, the noise is partial in the state propagation. Thirdly, the measurement function is nonlinear. Fourthly, the drag coefficient is a constant but affected by the noise, which can give a good criterion for measuring the filtering performance.

With the vector description of the continuous state as $X(t) = [x_1(t), x_2(t), x_3(t), x_4(t), x_5(t)]^T \in \mathbb{R}^5$, and the discrete-time measurement as $Y(k) = [r(k), b(k)]^T \in \mathbb{R}^2$, the state equation of dynamic is characterized as

$$\dot{X}(t) = F(t)X(t) + W(t) \tag{62}$$

where $\dot{X}(t) = [\dot{x}_1(t), \dot{x}_2(t), \dot{x}_3(t), \dot{x}_4(t), \dot{x}_5(t)]^T$ is the the derivative of $X(t)$, $W(t) = [0, 0, w_1, w_2, w_3]^T$ denotes the random noise. In above equation, the state propagation function is formulated as

$$F(t) = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ G(t) & 0 & D(t) & 0 & 0 \\ 0 & G(t) & 0 & D(t) & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{63}$$

where the parameter $G(t) = \frac{Gm_0}{R^3(t)}$ and $D(t) = \beta(t)\exp\{\frac{R_0-R(t)}{H_0}\}V(t)$ denote gravity-related force term and drag-related force term, respectively. The $R(t) = \sqrt{x_1^2(t) + x_2^2(t)}$ is the distance from the center of the Earth, $\beta(t)$ is the ballistic coefficient with $\beta(t) = \beta_0\exp(x_5(t))$, and $V(t) = \sqrt{x_3^2(t) + x_4^2(t)}$ is the speed. The constant parameters in our experiment is set as $Gm_0 = 3.9860 \times 10^5$, $\beta_0 = -0.59783$, $H_0 = 13.406$, and $R_0 = 6347$.

The measurement $Y(k) = [r_k, b_k]^T$ is obtained by a radar located at $[x_{1,r}, x_{2,r}]^T = [6374, 0]^T$ with a frequency of 10 Hz. The measurement equation as

$$Y(k) = h(X(k)) + u(k) \tag{64}$$

where

$$h(X(k)) = \begin{bmatrix} \sqrt{(x_1(t_k) - x_{1,r})^2 + (x_2(t_k) - x_{2,r})^2} \\ \tan^{-1} \left(\frac{x_2(t_k) - x_{2,r}}{x_1(t_k) - x_{1,r}} \right) \end{bmatrix} \tag{65}$$

and $u(k) = [u_{1,k}, u_{2,k}]^T$ is the measurement noise. The parameter t_k is the time instant of k -time measurement.

In the experiment, the state noise $\mathbf{w}(t) = [w_1(t), w_2(t), w_3(t)]^T$ with $\mathbf{E}[\mathbf{w}(t + \tau)\mathbf{w}(t)] = \text{diag}([2.4064 \times 10^{-5}, 2.4064 \times 10^{-5}, 0])\delta(t - \tau)$, and the measurement random noise $u(k) = \text{diag}([10^{-6}, 17^2 \times 10^{-6}])$.

For the continuous time state propagation, we have to discrete as a discrete-time state model. Here the Euler approximation method has been used to discrete the model

$$X(k + 1) = X(k) + \tau F(k)X(k) + W(k) \tag{66}$$

where k and τ are the time instant and time interval correspondent with the measurement time, respectively. With this discrete dynamic model, the usual filtering methods can be used. The initial condition of state for filtering has been set as $X(0) = [6500.4, 349.14, -1.8093, -6.7967, 0]^T$ and $\mathbf{P}(0) = [10^{-6}, 10^{-6}, 10^{-6}, 10^{-6}, 1]^T$.

In our experiment, we compare four methods, i.e., RUF, IEKF, EnKF and our proposed method (NgdKF), with time interval $\tau = 0.1$, time steps $K = 100$ and Monte Carlo $M = 100$. For the EnKF, the number of ensemble is chose as $N_s = 200$. In our proposed method, the parameter $\eta = 0.1$, $\delta = 10e - 2$ and $\gamma = 10e - 2$. The position root mean square error (RMSE), velocity RMSE and drag coefficient RMSE considered for measuring the performance. They are defined as

$$\text{MSE}_k^{\text{pos}} = \sqrt{\frac{1}{M} \sum_{i=1}^M (x_1^i(k) - \hat{x}_1^i(k))^2 + (x_2^i(k) - \hat{x}_2^i(k))^2} \tag{67}$$

$$\text{MSE}_k^{\text{vel}} = \sqrt{\frac{1}{M} \sum_{i=1}^M (x_3^i(k) - \hat{x}_3^i(k))^2 + (x_4^i(k) - \hat{x}_4^i(k))^2} \tag{68}$$

$$\text{MSE}_k^{\text{Dra}} = \sqrt{\frac{1}{M} \sum_{i=1}^M (x_5^i(k) - \hat{x}_5^i(k))^2} \tag{69}$$

where $x_1^i(k), x_2^i(k), x_3^i(k), x_4^i(k), x_5^i(k)$ and $\hat{x}_1^i(k), \hat{x}_2^i(k), \hat{x}_3^i(k), \hat{x}_4^i(k), \hat{x}_5^i(k)$ are correspond with the true state and filtered state of k -time instant in i -time simulation, respectively.

The results are showed in the Figures 6–8. In Figure 6, the position RMSE has been compared. The performance can be divided into two levels. The better level constructs by the IEKF and our proposed method, and the poor level contains the RUF and the EnKF. As for Figure 7 in which

comparing the velocity RMSE, our proposed method and the IEKF have the better performance than the RUF and the EnKF, while the EnKF performs better than the RUF. For the performance of estimation the drag coefficient, our proposed method has the same performance with the IEKF, which better than the other methods. The RUF has performed better than the EnKF.

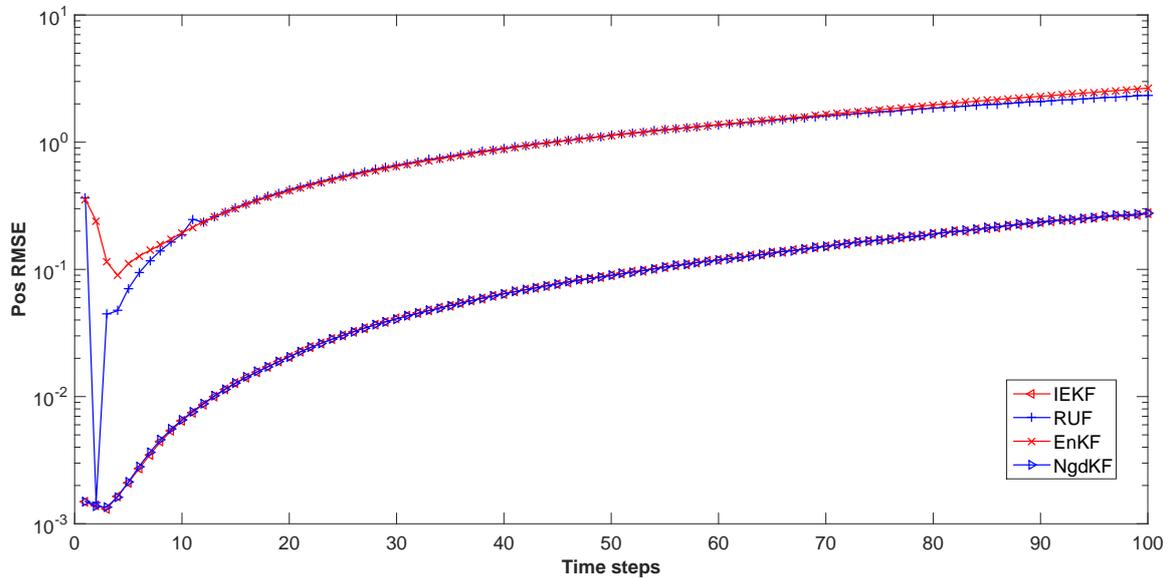


Figure 6. The position RMSE.

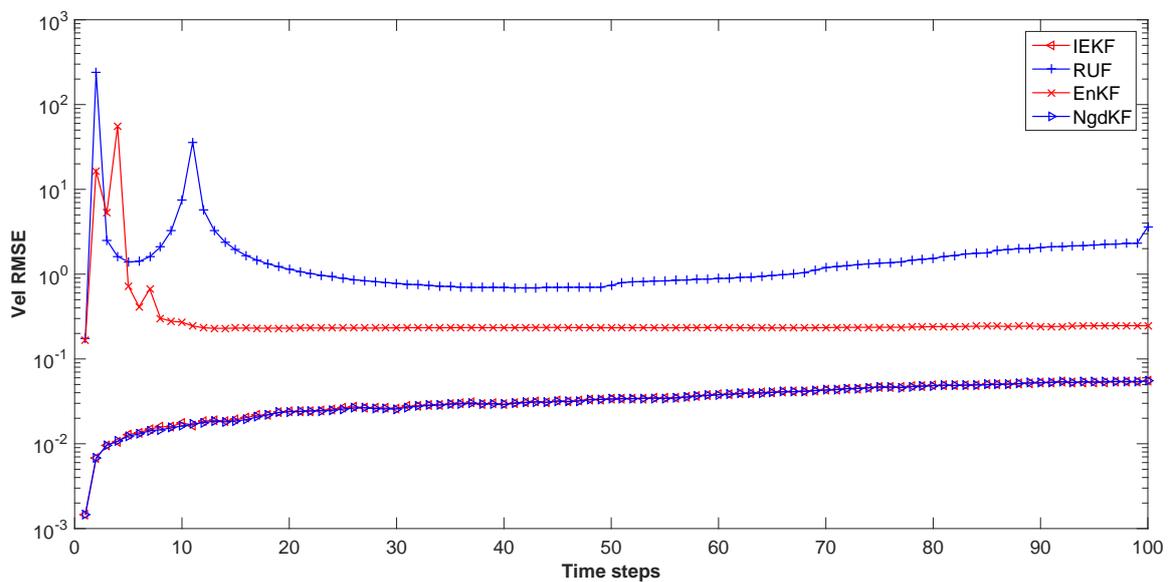


Figure 7. The velocity RMSE.

In the experiment, the positions are affected directly by the measurement, but have no noise. Thus, the better performance in dealing with nonlinear measurement, the better estimation for positions is achieved. Comparing with the estimation of positions, we can make a conclude intuitively that the better performance of our proposed method in dealing with nonlinear measurement. For the estimation procedure of drag coefficient, the noise has affected the estimation primarily, which would be the same level all the times. Thus the performance for estimating the drag coefficient may reflect the robust of the filtering. Accordingly, the robust property of our proposed method and the IEKF have been shown in the experiment. Comparing the three RMSE, our proposed method has the

same performance with the IEKF. The reasons is that our proposed method and the IEKF all have convergence to a stable estimation, and the iteration procedures have made the estimation close to the true state. Also, it has verified the convergence of our proposed method. From Figure 8, we can see the poor performance of the EnKF. This may be affected by the limit numbers of the ensemble used in the filtering. Besides, the RUF is performance moderately may indicated that the convergence has not achieved in the procedure. In all, our proposed method has converged a stable estimation and close to the true state. The performance is the same as the IEKF in multidimensional signal processing, and better than the other methods.

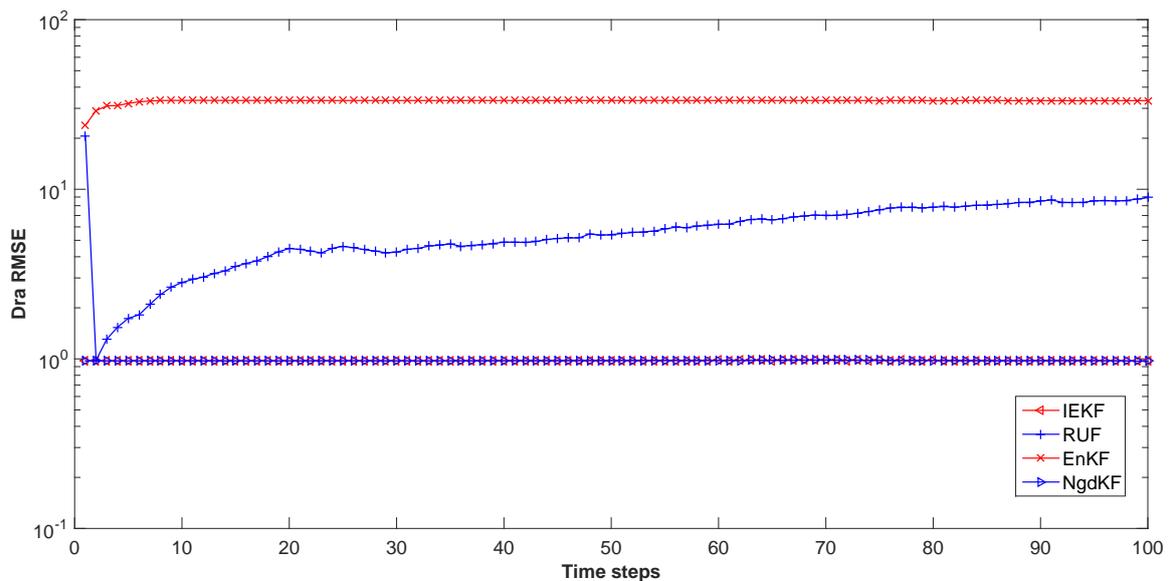


Figure 8. The drag coefficient RMSE.

6. Conclusions

In this paper, we have derived a new method for measurement update of the nonlinear filtering. In our proposed method, the measurement update has been considered as an optimization estimation in the view of information geometry, and the estimation in the parameter space has been corresponded with the iteration procedure in a statistical manifold. We have constructed the statistical manifold based on the posterior PDF of state under the condition of measurement in the Bayesian framework. Along the statistical manifold, the natural gradient descent method has been used for the optimization estimation. In the iteration procedure, the Fisher information matrix has applied to influence the iteration direction and convergence speed. Thus, it achieves the better performance and faster convergence. In addition, our proposed method has also given the theoretical justification for the measurement update in nonlinear filtering from the information geometric insight and a mathematical interpretation. We also discuss that in some conditions our proposed method is identical to the existing Kalman filter and its variants. In the procedure of deducing our method, the Bayesian framework is the bridge connecting the traditional filtering problems with the statistical manifold, thus the information geometric approach can be used in the nonlinear filtering. Also, the Bayesian framework provides a unifying framework for the dynamical model regardless of discrete-time or continuous-time state functions. That is means that the information geometric approach for nonlinear filtering will no limit to discrete-time or continuous-time dynamical model. Because of the ability of information geometry for processing non-Gaussian PDF, we can induce the filtering algorithms under the condition of non-Gaussian PDF or noise. Furthermore, our method has only consider the measurement update stage in the nonlinear filtering, and we can construct other nonlinear filtering combining with different state propagation methods for some filtering problems. In reviewing our method, there are some points can be considered thoroughly in the future. Firstly, the parameter of update steps η will be considered for selecting

adaptively. Secondly, the covariance matrix of prior estimation of state may have other approaches to obtain. Apart from the covariance matrix of our method which computed by the ensemble, there are two other approaches for the estimation of covariance matrix. In some articles, the mean of estimation is determined by the maximum likelihood estimates, while the covariance matrix can be approximated by the inverse of the Fisher information matrix [35]. The other method for covariance is estimating the mean and the covariance simultaneously, such as evolutionary algorithms [36]. The different approaches have different effects in the filtering. Thus the influence will be analyzed and the suit approach would be selected for the filtering. Thirdly, the number of steps for the convergence of our method will be considered in the future work. The proper number of steps can guarantee the convergence and spend less computation in the iteration procedure.

Acknowledgments: This work was supported by the National Natural Science Foundation of China under grant No. 61302149 and 61601479. The authors gratefully acknowledge the reviewers for their very valuable and insightful comments and suggestions, which have improved the presentation.

Author Contributions: Yubo Li put forward the original ideas and performed the research. Yongqiang Cheng and Xiang Li conceived and designed the simulations comparing with other existing methods. Xiaoqiang Hua makes some discussion on the natural gradient descent method. Yuliang Qin reviewed the paper and provided useful comments. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Haug, A. *Bayesian Estimation and Tracking: A Practical Guide*; John Wiley & Sons: Hoboken, NJ, USA, 2012.
- Julier, S.; Uhlmann, J. Unscented filtering and nonlinear estimation. *Proc. IEEE* **2004**, *92*, 401–422.
- Arasaratnam, I.; Haykin, S. Cubature Kalman filters. *IEEE Trans. Autom. Control* **2009**, *54*, 1254–1269.
- Arulampalam, M.; Maskell, S.; Gordon, N. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Trans. Signal Process.* **2002**, *50*, 174–188.
- Gustafsson, F.; Hendeby, G. Some relations between extended and unscented Kalman filters. *IEEE Trans. Signal Process.* **2012**, *60*, 545–555.
- Crisan, D.; Doucet, A. A survey of convergence results on particle filtering methods for practitioners. *IEEE Trans. Signal Process.* **2002**, *50*, 736–746.
- Stano, P.; Lendek, Z.; Braaksma, J.; Babuška, R.; Keizer, C.; den Dekker, A. Parametric Bayesian Filters for Nonlinear Stochastic Dynamical Systems: A Survey. *IEEE Trans. Cybern.* **2013**, *43*, 1607–1624.
- García-Fernández, Á.; Svensson, L.; Morelande, M.; Sarkka, S. Posterior Linearization Filter: Principles and Implementation Using Sigma Points. *IEEE Trans. Signal Process.* **2015**, *63*, 5561–5573.
- Zanetti, R. Recursive Update Filtering for Nonlinear Estimation. *IEEE Trans. Autom. Control* **2012**, *57*, 1481–1490.
- Zanetti, R. Adaptable Recursive Update Filter. *J. Guid. Control Dyn.* **2015**, *38*, 1295–1299.
- Fatemi, M.; Svensson, L.; Hammarstrand, L.; Morelande, M. A study of MAP estimation techniques for nonlinear filtering. In Proceedings of the 2012 15th International Conference on Information Fusion (FUSION), Singapore, 9–12 July 2012; pp. 1058–1065.
- Bell, B.; Cathey, F.; The iterated Kalman filter update as a Gauss-Newton method. *IEEE Trans. Autom. Control* **1993**, *38*, 294–297.
- Bellaire, R.; Kamen, E.; Zabin, S. A new nonlinear iterated filter with applications to target tracking. *Proc. SPIE Signal Data Process. Small Targets* **1995**, *2561*, 240–251.
- Lefebvre, T.; Bruyninckx, H.; Schutter, J. Kalman filters for nonlinear systems: A comparison of performance. *Int. J. Control* **2004**, *77*, 639–653.
- Morelande, M.; García-Fernández, A. Analysis of Kalman filter approximations for nonlinear measurements. *IEEE Trans. Signal Process.* **2013**, *61*, 5477–5484.
- Amari, S. Natural gradient works efficiently in learning. *Neural Comput.* **2008**, *10*, 251–276.
- Amari, S.; Nagaoka, H. *Methods of Information Geometry*; American Mathematical Society: Providence, RI, USA, 2007.
- Raskutti, G.; Mukherjee, S. The Information Geometry of Mirror Descent. *IEEE Trans. Inf. Theory* **2015**, *61*, 1451–1457.

19. Galanis, G.; Chu, P.C.; Kallos, G.; Kuo Y.H.; Dodson, C.T.J. Wave Height Characteristics in the North Atlantic Ocean: A new approach based on statistical and geometrical techniques. *Stoch. Environ. Res. Risk Assess.* **2012**, *26*, 83–103.
20. Amari, S.; Park, H.; Fukumizu, K. Adaptive method of realizing natural gradient learning formultilayer perceptrons. *Neural Comput.* **2000**, *12*, 1399–1409.
21. Park, H.; Amari, S.; Fukumizu, K. Adaptive natural gradient learning algorithms for various stochastic models. *Neural Netw.* **2000**, *7*, 755–764.
22. Zhang, L.; Cichocki, A.; Amari, S. Natural Gradient Algorithm for Blind Separation of Overdetermined Mixture with Additive Noise. *IEEE Signal Process. Lett.* **1999**, *6*, 293–295.
23. Yi, S.; Wierstra, D.; Schaul, T.; Schmidhuber, J. Stochastic search using the natural gradient. In Proceedings of the 26th International Conference on Machine Learning, Montreal, QC, Canada, 14–18 June 2009; pp. 1161–1168.
24. Akimoto, Y.; Nagata, Y.; Ono, I.; Kobayashi, S. Theoretical Foundation for CMA-ES from Information Geometry Perspective. *Algorithmica* **2012**, *64*, 698–716.
25. Zhang, Z.; Sun, H.; Peng, L.; Jiu, L. A Natural Gradient Algorithm for Stochastic Distribution Systems. *Entropy* **2014**, *16*, 4338–4352.
26. Galanis, G.; Famelis, I.; Liakatas, A. A new Kalman Filter based on Information Geometry techniques for optimizing numerical environmental simulations. *Stoch. Environ. Res. Risk Assess.* **2016**, doi:10.1007/s00477-016-1332-5.
27. Cai, Y.; Dodson, C.T.J.; Doig, A.; Wolkenhauer, O. Information-theoretic analysis of protein sequences shows that amino acids self-cluster. *J. Theor. Biol.* **2002**, *218*, 409–418.
28. Verdoolaege, G.; Scheunders, P. On the Geometry of Multivariate Generalized Gaussian Models. *J. Math. Imaging Vis.* **2012**, *43*, 180–193.
29. Harsha K.V.; Subrahmanian, M.K.S. *F*-Geometry and Amari's α -Geometry on a Statistical Manifold. *Entropy* **2014**, *16*, 2472–2487.
30. Resconi, G. Geometry of risk analysis (morphogenetic system). *Stoch. Environ. Res. Risk Assess.* **2009**, *23*, 425–432.
31. Amari, S. *Differential-Geometrical Methods in Statistics*; Lecture Notes in Statistics 28; Springer: Berlin/Heidelberg, Germany, 1985.
32. Sibley, G.; Sukhatme, G.; Matthies, L. The iterated sigma point Kalman filter with applications to long range stereo. *Robot. Sci. Syst.* **2006**, *8*, 235–244.
33. García-Fernández, Á.; Morelande, M.; Grajal, J. Truncated unscented Kalman filtering. *IEEE Trans. Signal Process.* **2012**, *60*, 3372–3386.
34. Dunik, J.; Ejstraka, O.; Simandl, M.; Blasch, E. Random-Point-Based Filters: Analysis and Comparison in Target Tracking. *IEEE Trans. Aerosp. Electron. Syst.* **2015**, *51*, 1403–1421.
35. Iguzquiza, E.; Chica-Olmo, M. Geostatistical simulation when the number of experimental data is small: An alternative paradigm. *Stoch. Environ. Res. Risk Assess.* **2008**, *22*, 325–337.
36. Beyer, H. Convergence Analysis of Evolutionary Algorithms that Are Based on the Paradigm of Information Geometry. *Evol. Comput.* **2014**, *22*, 679–709.



© 2017 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).