SIAM J. CONTROL OPTIM.
Vol. 61, No. 4, pp. 2407–2424

LOG-CONCAVE POSTERIOR DENSITIES ARISING IN CONTINUOUS FILTERING AND A MAXIMUM A POSTERIORI ALGORITHM*

JIAYI KANG[†], ANDREW SALMON[‡], AND STEPHEN S.-T. YAU[§]

Abstract. Nonlinear filtering is fundamental to many engineering problems, as it involves inferring the state of a system given complicated dynamics and noisy observations. Some approaches to nonlinear filtering use the analysis of the underlying PDE or stochastic PDE governing the evolution of the posterior probability distribution over time, one approach, in particular, being the Yau–Yau method. In this paper, we give a maximum a posteriori (MAP) framework for the Yau–Yau method. Furthermore, we propose convex filtering, intermediate between linear and nonlinear filtering, which gives criteria under which the posterior preserves log-concavity. The key tool from the PDE is a result from Korevaar, giving criteria under which a quasilinear parabolic PDE preserves convexity. A bridge between the MAP estimator and the structure of the posterior is explained. Finally, we propose a novel numerical method based on iteration and apply this method to a tracking problem.

Key words. nonlinear filtering problems, convex optimization, DMZ equation, Yau–Yau method

MSC codes. 35K15, 49M99, 60G35, 93E11

DOI. 10.1137/22M1508352

1. Introduction.

The central problem of filtering is to estimate the dynamical system with some related and noisy observations. Kalman [27] proposed the first analytical recursive filtering algorithm, thus opening the door to nonlinear filtering algorithms. Nonlinear filtering plays an important role in state estimation, and it is a significant part of modern control. Filtering can be analyzed using either a continuous or a discrete model for the evolution of both state and observations. While a continuous model for the evolution of state is often closer to the underlying ODE model used to describe the dynamics of a system, a discrete model for the observation process is often closer to the behavior of a sensor that gives a discrete rather than continuous sequence of measurements.

After decades of accumulation and development, there are three major frameworks for nonlinear filtering.

- Minimum mean square error (MMSE) [23, 41],
$$\hat{x}_t^{\mathrm{MMSE}} = \arg\min_{\phi_t} E[\|x_t - \phi_t(y_t)\|_2^2 | \mathcal{Y}_t] \quad \text{for } t \in [0,T].$$
- Maximum a posteriori (MAP) [31, 38],
$$\hat{x}_t^{\mathrm{MAP}} = \arg\max_{x_t} p(t, x_t|\mathcal{Y}_t) \quad \text{for } t \in [0,T],$$
where the $p(t,\cdot|\mathcal{Y}_t)$ is the posterior density at time t.

*Received by the editors July 11, 2022; accepted for publication (in revised form) March 27, 2023; published electronically August 4, 2023.
https://doi.org/10.1137/22M1508352
Funding: This work is supported by Tsinghua University Education Foundation fund (042202008).
[†]Department of Mathematical Sciences, Tsinghua University, Beijing, 100084, China (kangjy19@mails.tsinghua.edu.cn).
[‡]Mathematics Department, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (asalmon@mit.edu).
[§]Yanqi Lake Beijing Institute of Mathematical Sciences and Applications, Beijing, 101400, China, and Department of Mathematical Sciences, Tsinghua University, Beijing, 100084, China (yau@uic.edu).

- Variational inference (VI),

$$\hat{q}_\phi(t,x) = \arg\min_{q_\phi} KL(q_\phi(t,x)\|p(t,x|\mathcal{Y}_t)) \quad \text{for } t \in [0,T],$$

  where the $p(t,\cdot|\mathcal{Y}_t)$ is the posterior density at time t.

These three frameworks are not completely independent and they are equivalent to each other under some certain conditions such as the linear MMSE model [32, 33].

Since 1966, Duncan [20], Mortensen [37], and Zakai [54] have independently derived the stochastic partial differential equation of the unnormalized conditional density function for a continuous state equation with continuous observation, which is the so-called DMZ equation. The DMZ equation can be further transformed to the robust form [16]. In the 1970s. Brockett and Clark [6], Brockett [7], and Mitter [36] independently proposed to construct finite-dimensional filters by using the estimation algebra method. Since the 1990s, in a series of works [13, 11, 10, 9, 8, 51], Yau and his collaborators have completely classified finite-dimensional estimation algebras (FDEA) with maximal rank with arbitrary state space dimension, which includes both Kalman–Bucy and Benés filtering systems as special cases. Since the 2000s, Yau and coworkers [49, 18, 43, 45, 46, 26] have made significant progress in nonmaximal rank FDEA.

Furthermore, Yau and Yau [52] used a different way to solve the DMZ equation from the Lie algebra method. Motivated by Yau and Yau [52], Yau and Hu [50, 25] first proposed a novel direct method to solve the explicit solutions of the DMZ equation with arbitrary initial conditions. The explicit solutions of the DMZ equation can provide important guidance for the design of numerical algorithms. Shi, Yang, and Yau [44] designed an effective numerical method for time-invariant filtering problems by using the direct method and the Gaussian approximation algorithm. Chen, Shi, and Yau extended the numerical method to time-varying filtering problems in [12]. In practice, many numerical algorithms use the idea of linearization, such as the extended Kalman filter (EKF; 1979) [34], the iterated extended Kalman filter (IEKF; 1993) [5], and the unscented Kalman filter (UKF; 2000) [47]. More and more algorithms based on linearization of MMSE have been proposed which have created prosperity for algorithms based on MMSE.

In the meantime, the projection filter [39] motivated by the VI was proposed. In the 1990s, the particle filter [22, 2] was proposed to avoid calculating the posterior densities by simulating the distributions directly. A variety of sampling techniques [21, 30] which are based on the framework of VI develop the different filtering algorithms. Combining with the linearization method, the Gaussian particle filter [29] was developed. While the filtering model based on VI frameworks has quickly developed, Yau and Luo proposed an algorithm, called the Hermite spectral method [35], which directs parameterized posterior densities for the Yau–Yau method [53]. From then on, many different methods including the proper orthogonal decomposition method [48] and the Legendre Galerkin method [19] have been used in the Yau–Yau algorithm.

Accordingly, less attention has been paid to MAP estimation techniques in this field. Differently from the MMSE, the MAP estimator approximates the point with the maximum likelihood, not the mean of the posterior state. Importantly, MAP estimators can be used for target tracking problems, communications, radar tracking, sonar ranging, and satellite navigation [4]. Under the assumption that posterior densities are Gaussians, the well-known IEKF, induced by using the Gauss–Newton optimization [5], can be interpreted as a MAP estimator. And generally speaking, the complexity of the MAP estimators is smaller [33]. So, it is valuable to find the

MAP estimator for the nonlinear filtering problems. From history, we can see the DMZ equation and the Yau–Yau method are central concepts for nonlinear filtering problems in the sense of not only MMSE but also VI. One natural question is whether PDE methods, such as the Yau–Yau method, can be used to develop MAP estimators. For the MAP framework, the core problem is to calculate the maximum point of posterior density, which means that we should be able to solve the maximum points in the posterior densities of PDE. Unfortunately, it is impossible to construct an explicit equation for the maximum point in the posterior densities, and usually, it is not well defined by considering the process where the initial density is unimodal and terminal density is bimodal. In order to overcome such an issue, we need to investigate the case that the posterior densities are all unimodal, or equivalently, the densities are all logarithmic concave. Korevaar established a theory to describe the convexity in a posteriori density function of the parabolic PDE [28], which provides the foundation for our work.

In this paper, the main contributions are as follows:

- We propose the concept of convex filtering and explain the reason why linear filtering maintains a log-concave distribution when the initial distribution is log-concave.
- Convex filtering can be considered as a natural generalization of linear filtering and provides a mathematical foundation for continuous MAP models.
- We prove that the Yau-type FDEA system is a convex filter under appropriate conditions.
- Motivated by the concept of convex filtering, we proposed a novel iteration method for solving the general filtering problems which is called the iteration optimization Kalman filter (IOKF).

Bearings-only tracking (BOT) is the benchmark scenario that we use to evaluate the performance of IOKF and compare the result to that of UKF and IEKF.

2. Background.

The problem of filtering is to optimally estimate the state of a system evolving in time according to a stochastic differential equation by using noisy observations. Such a system can be given by the signal observation model.

Hidden state process:

$$dx_t = f(t,x_t)dt + g(t,x_t)dV_t. \tag{2.1}$$

Observation process:

$$\begin{aligned} dy_t &= h(t,x_t)dt + dW_t \quad \text{(Continuous)},\\ z_{t_k} &= H(t_k,x_{t_k}) + B_{t_k} \quad \text{(Discrete)}. \end{aligned} \tag{2.2}$$

The stochastic processes x_t, V_t, y_t, and W_t are valued in $\mathbb{R}^n$, $\mathbb{R}^p$, $\mathbb{R}^m$, and $\mathbb{R}^m$, respectively. V_t, W_t, x_0 are assumed to be mutually independent and $y_0 = 0$. The V_t and W_t processes are Brownian motions and their covariance matrices are $E[dV_t dV_t^\top] = I_n dt$ and $E[dW_t dW_t^\top] = S(t)dt$. B_{t_k} is given by a Gaussian distribution with covariance matrix $\tilde{S}_{t_k}$. The functions $f(t,x)$, $h(t,x)$, $g(t,x)$, and $H(t,x)$ are all smooth functions with suitable dimension of input and output.

In this section, we will recall the DMZ equation for continuous state evolution and continuous observations and the Bayesian framework for the discrete observation system. Finally, we will recall a theorem of Korevaar, which is the main technical tool we will use to study the convexity properties of the parabolic PDE we consider.

2.1. Review of the Yau–Yau method for the DMZ equation. At the beginning of this subsection, let us review the well-known DMZ [20, 37, 54] equation of (2.1) with continuous observation in (2.2). For simplicity of discussion, we assume that $g(t,x)$ in (2.1) satisfies $g(t,x)g(t,x)^\dagger \equiv I_n$.

The unnormalized density function $\sigma(t,x)$ of X_t conditioned on the observation history $\mathcal{Y}_t$ satisfies as follows:

$$\begin{cases} d\sigma(t,x) = L_0\sigma(t,x)dt + h^\top(t,x)S^{-1}(t)\sigma(t,x)dy_t, \\ \sigma(0,x) = \sigma_0(x), \end{cases} \tag{2.3}$$

where

$$L_0 := \frac{1}{2}\sum_{i=1}^n \frac{\partial^2}{\partial x_i^2} - \sum_{i=1}^n f_i \frac{\partial}{\partial x_i} - \sum_{i=1}^n \frac{\partial f_i}{\partial x_i} - \frac{1}{2}h^\top(x,t)S^{-1}h(x,t). \tag{2.4}$$

The DMZ equation can be transformed into a deterministic PDE by using the following robust exponential transformation:

$$\begin{aligned} \rho(t,x) &= \exp[-K(t,x)]\sigma(t,x), \\ K(x,t) &= h^\top(x,t)S^{-1}(t)y_t. \end{aligned} \tag{2.5}$$

By using the robust transformation, we can obtain the robust DMZ equation as

$$\begin{cases} \frac{\partial \rho}{\partial t} = \frac{1}{2}\sum_{i=1}^n \frac{\partial^2 \rho}{\partial x_i^2} \\ +\sum_{i=1}^n (\frac{\partial K}{\partial x_i} - f_i)\frac{\partial \rho}{\partial x_i} \\ +(\frac{\partial}{\partial t}(h^\top S^{-1})y_t + \frac{1}{2}\sum_{i=1}^n [\frac{\partial^2 K}{\partial x_i^2} + (\frac{\partial K}{\partial x_i})^2] \\ -\sum_{i=1}^n f_i \frac{\partial K}{\partial x_i}(t,x) \\ -\sum_{i=1}^n \frac{\partial f_i}{\partial x_i} - \frac{1}{2}(h^\top S^{-1}h))\rho(t,x), \\ \rho(0,x) = \sigma_0(x). \end{cases} \tag{2.6}$$

The most important method for solving the robust DMZ equation is the Yau–Yau algorithm [53]. Denote the sequence of observation times as $\mathcal{P}_N = \{0 = \tau_0 < \tau_1 < \cdots < \tau_N = T\}$. Let ρ^k be the solution of the robust DMZ equation with $y_t = y_{\tau_{k-1}}$ on the time interval $\tau_{k-1} \le t < \tau_k$, so that for $k = 0, 1, \ldots, N$, ρ^k satisfies the following equation:

$$\begin{cases} \frac{\partial \rho^k}{\partial t} = \frac{1}{2}\sum_{i=1}^n \frac{\partial^2 \rho^k}{\partial x_i^2}(t,x) \\ +\sum_{i=1}^n (\frac{\partial \tilde{K}_k}{\partial x_i} - f_i)\frac{\partial \rho^k}{\partial x_i} \\ +(\frac{\partial}{\partial t}(h^\top S^{-1})y_{\tau_{k-1}} + \frac{1}{2}\sum_{i=1}^n [\frac{\partial^2 \tilde{K}_k}{\partial x_i^2} + (\frac{\partial \tilde{K}_k}{\partial x_i})^2] \\ -\sum_{i=1}^n f_i \frac{\partial \tilde{K}_k}{\partial x_i}(t,x) \\ -\sum_{i=1}^n \frac{\partial f_i}{\partial x_i} - \frac{1}{2}(h^\top S^{-1}h))\rho^k(t,x), \\ \rho^1(0,x) = \sigma_0(x), \\ \rho^k(\tau_{k-1},x) = \rho^{k-1}(\tau_{k-1},x), k = 2, 3, \ldots, N, \\ \tilde{K}_k = h^\top(x,t)S^{-1}(t)y_{\tau_{k-1}}. \end{cases} \tag{2.7}$$

Define the norm of $\mathcal{P}_N$ by $|\mathcal{P}_N| = \sup_{1\le k\le N}(\tau_k - \tau_{k-1})$. The convergence of the Yau–Yau algorithm has been proved in both the pointwise sense and the L^2 sense in [53], i.e.,

$$\lim_{|\mathcal{P}_N|\to 0} \rho^k(t,x) = \rho(t,x). \tag{2.8}$$

The key proposition of the Yau–Yau method is given as follows.

PROPOSITION 2.1 (Proposition 2.1 in [35]). *For each* $\tau_{k-1} \le t \le \tau_k$, $k = 1, 2, \dots, N$, $\rho^k(t,x)$ *satisfies* (2.7) *if and only if*

$$\varrho^k(t,x) = \exp(h^\top(t,x) S^{-1} y_{\tau_{k-1}}) \rho^k(t,x) \tag{2.9}$$

satisfies the PDE

$$\frac{\partial \varrho^k}{\partial t} = L_0 \varrho^k, \tag{2.10}$$

where L_0 *is defined in* (2.4), *while the initial data is updated as follows:*

$$\begin{cases} \varrho^1(0,x) = \sigma_0(x) \\ or \\ \varrho^k(\tau_{k-1},x) = \exp(h^\top(\tau_{k-1},x) S^{-1}(\tau_{k-1})(y_{\tau_{k-1}} - (y_{\tau_{k-2}})) \cdot \varrho^{k-1}(\tau_{k-1},x) \\ \ for\ k \ge 2. \end{cases} \tag{2.11}$$

Remark 2.2. Equation (2.10) is defined as the Yau–Yau PDE in [3].

We can summarize the Yau–Yau method as the following diagram:

$$\text{Numerical solver} \longleftarrow \varrho^k(t,x) \underset{\text{normalizing}}{\overset{\|\mathcal{P}_N\| \to 0}{\longrightarrow}} p(t,x|\mathcal{Y}_t). \tag{2.12}$$

2.2. The Bayesian framework for the continuous system with discrete observation. The setting of continuous state space and discrete observations is known as the continuous-discrete system. The probability density of x_t in (2.1), denoted $\tilde{p}(t,x)$, satisfies a forward Kolmogorov equation, also known as a Fokker–Planck equation,

$$\begin{gathered} \frac{\partial}{\partial t}\tilde{p}(t,x) = \frac{1}{2}\sum_{i=1}^n \frac{\partial^2}{\partial x_i^2}(\tilde{p}(t,x)) - \sum_{i=1}^n \frac{\partial}{\partial x_i}(f_i \tilde{p}(t,x)), \\ \tilde{p}(0,x) \text{ is given as initial condition } t \in [0,\infty). \end{gathered} \tag{2.13}$$

For a sequence of observations $\{z_{t_k}\}_{k=1}^N$ on the time series $\{0 = t_0 < t_1 < \cdots < t_N \le T\}$, we define that $\tilde{p}^k(t,x) := \tilde{p}(t,x|\mathcal{Y}_k)$, where $\mathcal{Y}_k$ is σ-algebra formed by observations at the first k steps.

In the continuous-discrete filtering, the condition densities only update at separate time t_k. There are two different steps as follows:

- The first step is the evolution between observations.
- The second step is to update by using the new observation.

The evolution of the conditional densities between observations satisfies the Fokker–Planck equation (2.13). That is,

$$\begin{gathered} \frac{\partial}{\partial t}\tilde{p}^k(t,x) = \frac{1}{2}\sum_{i=1}^n \frac{\partial^2}{\partial x_i^2}(\tilde{p}^k(t,x)) - \sum_{i=1}^n \frac{\partial}{\partial x_i}(f_i \tilde{p}^k(t,x)), \\ \text{and } \tilde{p}^k(t_k,x) \text{ is given as initial condition, } t \in [t_k, t_{k+1}). \end{gathered} \tag{2.14}$$

Then, we need to update the $\tilde{p}^k(t_{k+1},x)$ to $\tilde{p}^{k+1}(t_{k+1},x)$ by using the new observation $z_{t_{k+1}}$. The well-known Bayes rule governs such an update step. Combining it with (2.2), the update step is given by

$$\tilde{p}^{k+1}(t_{k+1},x) = c_{k+1}\tilde{p}^k(t_{k+1},x) e^{-\frac{1}{2}(H(t_{k+1},x)-z_{k+1})^\top S^{-1}(t_{k+1})(H(t_{k+1},x)-z_{k+1})}, \tag{2.15}$$

where the c_k is some normalizing factor.

2.3. The convexity in parabolic PDEs. We now recall a special case of a result of Korevaar [28, Theorems 1.6, 2.5], which we recall in the modified form below.

THEOREM 2.3 (Korevaar [28]). *Let Ω be a smooth, bounded, strongly convex domain (that is, all principal curvatures are positive) and let $u \in C^2([0,T]\times\Omega)$ be a function satisfying*

$$\frac{\partial}{\partial t}u(t,x) = \sum_{i,j=1}^{n} a_{i,j}(\nabla u)\frac{\partial^2 u}{\partial x_i \partial x_j} - b(x,\nabla u), \tag{2.16}$$

as well as the boundary condition $u|_{\partial\Omega} = \infty$. Suppose furthermore that $u(t,x) = -\log v_t(x)$, where v_t satisfies

$$\nabla v_t(x)\cdot n(x) > 0$$

for any point $x \in \partial\Omega$ with interior normal $n(x)$. If $b(x,y)$ is concave in x, for a fixed y and $a_{i,j}(\nabla u)$ is positive definite as a matrix for any fixed ∇u, and if $u(0,x)$ is convex, then $u(t,x)$ is convex for all t.

For readers' convenience, the proof is given in the appendix.

3. The continuous filtering framework via convexity and MAP. In this section, we will introduce a novel MAP framework for a continuous filtering system by using Theorem 2.3, and we will define a new type of filtering system, the convex filtering system. This class of filtering systems will ensure the logarithmic concavity of the posterior density function under appropriate conditions so that the posterior distributions are always unimodal. This ensures that the MAP estimate should be continuous in time, thus providing a good theoretical foundation for reducing the computational complexity of numerical methods.

As in the diagram (2.12), we shall study the log-concavity of $\rho^k(t,x)$ through the log-concavity of $\varrho^k(t,x)$. It is easy to see that if h is a linear function, then $\rho^k(t,x)$ is log-concave if and only if $\varrho^k(t,x)$ is log-concave.

3.1. The log-concavity in posterior densities.

3.1.1. Convex filtering with continuous observations. First, we need to transform for the densities $\varrho^k(t,x)$ into $u^k(t,x)$ by using a Hopf–Cole transformation [24, 14],

$$u^k(t,x) := -\log \varrho^k(t,x), \tag{3.1}$$

and it is easy to find that (3.1) is well-defined for the support set of the function $\varrho^k(t,x)$,

For $x \in \operatorname{supp}(\varrho^k(t,x))$, the Yau–Yau PDE system in Proposition 2.1 is converted to the following form after the change of variables:

$$\frac{\partial u^k}{\partial t} = \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\partial^2 u^k}{\partial x_i^2} - \frac{\partial u^k}{\partial x_i}\frac{\partial u^k}{\partial x_i}\right) - \sum_{l=1}^{n} f_l \cdot \frac{\partial u^k}{\partial x_l} + \nabla\cdot f + \frac{1}{2}h^\top S^{-1}h, \tag{3.2}$$

with initial condition

$$u^k(t_{k-1},x) = u^{k-1}(t_{k-1},x) - h^T(t_{k-1},x)S^{-1}(t_{k-1})(y_{t_{k-1}} - y_{t_{k-2}})$$

and $u^0(0,x) = -\log\sigma_0(x)$.

Regularity theory for linear parabolic PDEs allows us to conclude the smoothness of solutions of the PDE system in Proposition 2.1 and that the other hypotheses on the logarithmic transform in Theorem 2.3 hold. By applying Theorem 2.3, the solutions of (3.2) are convex for any time t if $u^k(t_{k-1},x)$ is convex and $-b(x,\nabla u^k)$ is convex for any fixed ∇u^k, where b is defined as

$$b(x,\nabla u^k) := \frac{1}{2}\sum_{i}^{n}\left(\frac{\partial u^k}{\partial x_i}\right)^2 + \sum_{i=1}^{n} f_i \cdot \frac{\partial u^k}{\partial x_i} - \nabla\cdot f - \frac{1}{2}h^\top S^{-1}h. \tag{3.3}$$

Remark 3.1. If f is linear, the condition that $-b(x,\nabla u^k)$ is a convex function of x is equivalent to the condition that $\frac{1}{2}h^\top(t,x)S^{-1}(x)h(t,x)$ is a convex function. On the other hand, if f is nonlinear, there may be a value of ∇u^k such that $-b(x,\nabla u^k)$ fails to be convex.

Linear observation cases. The novelty of this paper is that many of the Yau filtering systems have the convex filter property as shown below.

THEOREM 3.2. *For a continuous state-observation process with state equation in* (2.1) *and continuous observation process in* (2.2)*, assume $h(x)=Hx$ is linear for H an $m\times n$ matrix. In addition, assume that there exists a smooth function Φ, $n\times n$ matrices J and L, with J symmetric, and an n-dimensional vector l such that the following conditions are satisfied:*

1. *$f(x)=Lx+l-\nabla\Phi(x)$.*
2. *$\Phi(x)+x^\top Jx$ is convex.*
3. *$u(0,x)-\Phi(x)-x^\top Jx$ is convex.*
4. *$x^\top H^\top S^{-1}Hx-\Delta\Phi(x)+|\nabla\Phi(x)|^2+4x^\top(J^\top-L^\top)Jx$ is convex.*

Then, the filtering system is a convex filter in the sense that the solution of (3.2)*, $u(t,x)$ is convex for all t.*

Proof. For any observation sequence on the time series $\mathcal{P}_N=\{0=\tau_0<\tau_1<\cdots<\tau_N=T\}$, let $u^k(t,x)$ denote the solution of the log Yau–Yau PDE defined in (3.2). Consider the change of variables $w^k(\tau_{k-1},x)=u^k(\tau_{k-1},x)-\Phi(x)-x^\top Jx$. Then the third assumption implies that $w^1(0,x)$ is convex. If $v^k(\tau_{k-1},x)=u^k(\tau_{k-1},x)-\Phi(x)$, then v^k satisfies the following PDE:

$$\begin{aligned}\frac{\partial v^k}{\partial t} =& \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\partial^2 v^k}{\partial x_i^2}-\left(\frac{\partial v^k}{\partial x_i}\right)^2\right)-(Lx+l)^\top\nabla v^k\\ &+\frac{1}{2}\left[x^\top H^\top S^{-1}Hx+2\mathrm{Tr}(L)+(-\Delta\Phi(x)+|\nabla\Phi(x)|^2)\right].\end{aligned}\tag{3.4}$$

Therefore, $w^k(t,x)$ satisfies the following PDE:

$$\begin{aligned}\frac{\partial w^k}{\partial t} =& \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\partial^2 w^k}{\partial x_i^2}-\left(\frac{\partial w^k}{\partial x_i}\right)^2\right)-(Lx+l-2Jx)^\top\nabla w^k\\ &+\frac{1}{2}\left(x^\top H^\top S^{-1}Hx+2\mathrm{Tr}(L+J)-\Delta\Phi(x)+|\nabla\Phi(x)|^2+4|Jx|^2-4(x^\top L^\top+l^\top)Jx\right).\end{aligned}$$

Since the initial condition is convex, then the fourth property shows that $w^k(\tau_{k-1},x)$ is convex for all τ_{k-1}. By the second property, this means that u^k is also convex, being the sum of two convex functions. Therefore, the posterior density function $\varrho^k(t,x)$ in the standard Yau–Yau method is log-concave. By taking the limit $\|\mathcal{P}_N\|\to 0$, any posterior density is log-concave for $t\in[0,T]$. ∎

Remark 3.3. In the special case that the linear portion in Theorem 3.2 vanishes, $l=0$ and $L=0$, and in addition the matrix $J=0$, the above theorem above recovers the log-concavity portion of [40, Lemma 5.1], which is the most general log-concavity result that we know of for filtering in the literature. Note that in this special case, the dynamics are required to be a gradient flow and are proven using different techniques. Part of our interest in the more general formulation is that it encompasses additional results for finite-dimensional filters, including the Benes filter. One simple one-dimensional example is when $f(x) = \tanh(x)$, which comes from a nonconvex potential. It is notable that in such examples the posterior distribution will be log-concave even though the Gibbs distribution will not be.

DEFINITION 3.4. *A function $a : \mathbb{R}^n \to \mathbb{R}$ is said to be a semiconcave function of a linear modulus if there exists a constant $C \geq 0$ such that*

$$\lambda a(x) + (1-\lambda)a(y) - a(\lambda x + (1-\lambda)y) \leq \lambda(1-\lambda)C\|x-y\|_2^2 \tag{3.5}$$

for any $x, y \in \mathbb{R}^n$ and for all $\lambda \in [0,1]$. And by Remark 2.1 *in* [1], *the condition* (3.5) *equals to that $a(x) - C\|x\|_2^2$ is concave function.*

COROLLARY 3.5. *For a potential Φ, we assume that $-\Phi$ and $\Delta\Phi(x) - |\nabla\Phi(x)|^2$ are semiconcave functions of the linear modulus; then for sufficiently large H (that is, $H - cI$ is positive definite for sufficiently large c), the continuous filtering system as in Theorem* 3.2 *is convex if the initial distribution is sufficiently log-concave (that is, condition* 3 *in Theorem* 3.2 *is satisfied).*

Proof. By the definition of semiconcavity, we can find a positive definite J such that $\Phi(x) + x^T Jx$ is convex. Fixing such a J, we can also ensure that the initial distribution is sufficiently log-concave so that condition 3 of Theorem 3.2 is satisfied. It remains to check that condition 4 can be made to hold, which follows by choosing H to be sufficiently large. ∎

One example of systems satisfying condition 1 of Theorem 3.2 is FDEA systems of maximum rank. In the FDEA case, the choice of J is often clear—there should be a minimal choice of J in many cases.

Nonlinear observation cases. Next, we need to analyze the initial condition more clearly. The convexity of the $u^k(\tau_{k-1}, x)$ depends on not only the $u^{k-1}(t_{k-1}, x)$ but also

$$-h^T(\tau_{k-1}, x)S^{-1}(\tau_{k-1})(y_{\tau_{k-1}} - y_{\tau_{k-2}}).$$

Intuitively, $(y_{\tau_{k-1}} - y_{\tau_{k-2}})$ should be small enough so that u^k will be convex after it is updated. Also, the nonlinearities of h should be small so that $\varrho^k(t,x)$ maintains log-concavity for any $1 \leq k \leq N$. So, we shall summarize this argument into the following definition.

DEFINITION 3.6 (admissible observation condition for convexity). *The admissible observation condition for convexity of the continuous filtering system with the continuous observation process is*

- $u^{k-1}(\tau_{k-1}, x) - h^\top(\tau_{k-1}, x)S^{-1}(\tau_{k-1})(y_{\tau_{k-1}} - y_{\tau_{k-2}})$ *is convex for any* $2 \leq k \leq N$, *where* $\mathcal{P}_N = \{0 = \tau_0 < \tau_1 < \cdots < \tau_N = T\}$.

Remark 3.7. We can find that if $h(t,x)$ is a linear function in x, then Definition 3.6 is satisfied and it does not depend on observation y_t. However, if the $h(t,x)$ is a nonlinear function in x, then convex filtering can be obtained only by imposing appropriate conditions on the observation process y_t.

In general, the time steps between observations should be small so that admissible observations conditions for convexity are satisfied. On the contrary, if the time step is relatively large, the continuous observation model is not suitable, and the continuous-discrete system should be used. We now summarize the above analysis on nonlinear observations into the following theorem.

THEOREM 3.8. *For the continuous filtering system, if the system equation f is linear, $\frac{1}{2}h^\top(t,x)S^{-1}(t)h(t,x)$ is a convex function and admissible observation for convexity conditions Definition 3.6 is satisfied for any $\mathcal{P}_N = \{0 = \tau_0 < \tau_1 < \cdots < \tau_N = T\}$, then the log of posterior density $-\log p(t,x|\mathcal{Y}_t)$ is convex.*

Proof. The $u^k(t,x)$ is convex by using Theorem 2.3, ϱ^k is log-concave by using Definition 3.6, and the theorem is proved by taking the limit $\|\mathcal{P}_N\| \to 0$ in (2.12). ◻

Remark 3.9. The linear filtering system with log-concave initial distribution is convex filtering by Theorem 3.8.

3.1.2. Convex filtering for a continuous-discrete system. Starting with (2.14) and (2.15) and applying the substitution

$$\tilde{u}^k(t,x) = -\log \tilde{p}(t,x|\mathcal{Y}_{t_k}), \text{ for } t \in [t_k, t_{k+1}), \tag{3.6}$$

we get $\tilde{u}^k(t_k,x) = \tilde{u}^{k-1}(t_k,x) + \frac{1}{2}(H(t_k,x) - z_k)^\top S^{-1}(H(t_k,x) - z_k)$,

$$\frac{\partial \tilde{u}^k}{\partial t} = \frac{1}{2}\sum_i^n \left(\frac{\partial^2 \tilde{u}^k}{\partial x_i^2} - \left(\frac{\partial \tilde{u}^k}{\partial x_i}\right)^2 \right) - \sum_{l=1}^n f_l \cdot \frac{\partial \tilde{u}^k}{\partial x_l} - \nabla \cdot f.$$

THEOREM 3.10. *For such a continuous-discrete filtering model, if the $\frac{1}{2}(H(t_k,x) - z_k)^\top S^{-1}(H(t_k,x) - z_k)$, $\tilde{u}^0(0,x)$ and f are convex in x for any fixed $\nabla\tilde{u}^k$, then the $\tilde{u}^k(t,x)$ for any k is a convex function.*

Proof. The theorem is true by using Theorem 2.3. ◻

By using Korevaar's theorem, we shall see several examples of convex filters coming from nonlinear continuous dynamics with continuous observation in the next section. It needs to be pointed out that it is hard to construct a convex filter from a continuous-discrete filtering system. However, for the continuous-discrete filtering system, there is an important lemma to ensure that the nonconvexity will be effectively controlled under some suitable assumptions.

LEMMA 3.11. *Consider a continuous-discrete setting with linear dynamics so that $f(t,x) = f_0(t) + f_1(t)x$. Suppose that $A(t)$ is a matrix-valued function differentiable except with jump discontinuities at t_k. Suppose $A(t)$ satisfies a Riccati-type equation*

$$\frac{d}{dt}A(t) = -A^\top(t)A(t) - f_1(t)^\top A(t) - A(t)f_1(t) \tag{3.7}$$

on all intervals (t_{k-1}, t_k).

Suppose furthermore that $\sigma_0(0,x)e^{\frac{1}{2}x^\top A(0)x}$ is log-concave and our observation process z_k is such that

$$\frac{1}{2}(H(\tau_k,x) - z_{t_k})^\top S(t_k)(H(t_k,x) - z_{t_k}) - \frac{1}{2}x^\top(A(t_k^+) - A(t_k^-))x \tag{3.8}$$

is convex for all k. Then $p_k(t,x)e^{\frac{1}{2}x^\top A(t)x}$ is log-concave on all intervals (t_{k-1}, t_k).

Proof. First, let $\tilde{v}^k(t,x) = -\log(p_k(t,x)e^{\frac{1}{2}x^\top A(t)x})$ with $t \in (t_{k-1}, t_k)$ and let $\tilde{u}^k(t,x) = -\log(p_k(t,x))$, which means $\tilde{v}^k(t,x) = \tilde{u}^k(t,x) - \frac{1}{2}x^\top A(t)x$ with $t \in (t_{k-1}, t_k)$.

We can get that

$$\begin{aligned} \tilde{v}^0(0,x) &= -\log\sigma_0(x) - \frac{1}{2}x^\top A(0)x, \\ \tilde{v}^k(t_{k-1},x) &= \tilde{v}^{k-1}(t_{k-1},x) + \frac{1}{2}(z_k - h(x))^\top S^{-1}(z_k - h(x)) \\ &\quad - \frac{1}{2}x^\top (A(t_{k-1}^+) - A(t_{k-1}^-))x. \end{aligned} \tag{3.9}$$

By using (3.1.2), we will get

$$\begin{aligned} \frac{\partial \tilde{v}^k}{\partial t} &= \frac{1}{2}\sum_{i=1}^n \left(\frac{\partial^2 \tilde{v}^k}{\partial x_i^2} - \left(\frac{\partial \tilde{v}^k}{\partial x_i}\right)^2 \right) + (\nabla \tilde{v}^k)^\top A(t)x \\ &\quad - f\cdot\nabla\tilde{v}^k - \frac{1}{2}\mathrm{Tr}(A(t)) - \frac{1}{2}x^\top\left(\frac{d}{dt}A(t) + A^\top(t)A(t)\right)x - f\cdot A(t)x - \nabla f. \end{aligned} \tag{3.10}$$

Furthermore, the $f(t,x) = f_0(t) + f_1(t)x$. Fix $\tilde{v}^k$, and the function term of (3.10) by Theorem 2.3 is

$$\begin{aligned} &-b(t,x,\nabla\tilde{v}^k) = -\frac{1}{2}\|\nabla\tilde{v}^k\|^2 + (\nabla\tilde{v}^k)^\top A(t)x - f^\top\cdot\nabla\tilde{v}^k \\ &-\frac{1}{2}\mathrm{Tr}(A(t)) - \frac{1}{2}x^\top\left(\frac{d}{dt}A(t) + A^\top(t)A(t)\right)x - f^\top\cdot A(t)x - \nabla f. \end{aligned} \tag{3.11}$$

Taking the Hessian of $-b(t,x,\nabla\tilde{v}^k)$ yields

$$-\frac{d}{dt}A(t) - A^\top(t)A(t) - f_1(t)^\top A(t) - A(t)f_1(t). \tag{3.12}$$

By using (3.7), $-\frac{d}{dt}A(t) - A^\top(t)A(t) - f_1(t)^\top A(t) - A(t)f_1(t) = 0$ for any $t \in (t_{k-1}, t_k)$.

Since $\sigma_0(0,x)e^{\frac{1}{2}x^\top A(0)x}$ is log-concave, the initial condition for $\tilde{v}^0(0,x)$ is convex. So it implies that the function $\tilde{v}^k(t,x)$ is convex for any k with $t \in (t_{k-1}, t_k)$ if the initial $\tilde{v}^k(t_k,x)$ is convex. Furthermore, (3.9) and (3.8) show that if $\tilde{v}^k$ is convex at time $t = t_{k+1}$, then $\tilde{v}^{k+1}$ is convex at the same time. Now we have that $\tilde{v}^k(t,x)$ is convex for any k with $t \in [t_{k-1}, t_k]$, which means that the $\tilde{p}_k(t,x)e^{\frac{1}{2}x^\top A(t)x}$ log-concave $t \in (t_{k-1}, t_k)$. ∎

3.2. Examples of convex filtering systems. In this subsection, we want to point out the following fact for our proposed convex filter system. By using Theorem 3.2, it follows that

$$\text{Linear filtering} \subset \text{Convex filtering} \subset \text{Nonlinear filtering}.$$

3.2.1. The example with deterministic dynamical system. Now, it is important to find out whether the convex filtering is exactly linear. If so, convex filtering is a redundant concept. Next, we can give readers two vital examples of convex filtering.

Example 1 (convex filtering with continuous observation). We can consider the following nonlinear filtering system:

$$\begin{cases} dx_t = 0, \quad x_0 \sim p_0(x), \\ dy_t = h(x_t)dt + dv_t, \end{cases} \tag{3.13}$$

where the x_t and y_t are the one-dimensional processes and the variance of dv_t is $\sigma_V^2 dt$. For the filtering problem the posterior densities are given as follows:

$$p(t, x|\mathcal{Y}_t) = c_t p_0(x) \times \exp\left(\frac{1}{\sigma_V^2}\left(h(x)y_t - \frac{t}{2}h(x)^2\right)\right), \tag{3.14}$$

where c_t is the normalizing factor of the distribution.

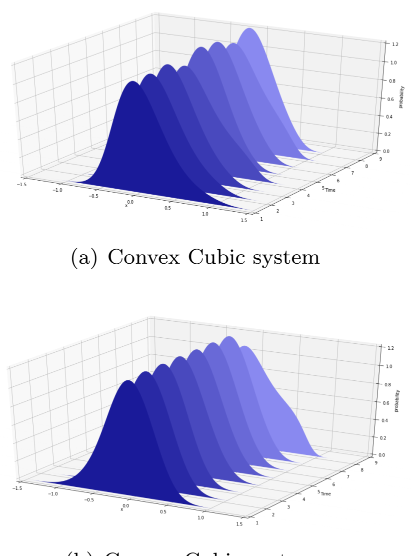
We shall assume that the $p_0(x)$ is a Gaussian distribution, $h(x)$ is the cubic function $h(x) = x^3$, and $\sigma_V = 1$. So, the posterior density in this case is $p(t, x|\mathcal{Y}_t) = c_t p_0(x) \times \exp(x^3 y_t - \frac{t}{2}x^6)$. It is easy to calculate that $U(x) = \frac{d^2}{dx^2}[-\log(p(t, x|\mathcal{Y}_t))] = \frac{1}{Var_0} - 6y_t x + 15x^4$. Suppose $U(x)$ achieves its minimal value at x_0. Then $\frac{d}{dx}U(x)|_{x_0} = 0$. So, $x_0 = (\frac{1}{10}y_0)^{\frac{1}{3}}$. It follows that $U(x) \geq 0$ if and only if

$$U(x_0) = \frac{1}{Var_0} - \frac{9}{2}y_t\left(\frac{y_t}{10}\right)^{\frac{1}{3}} \geq 0, \tag{3.15}$$

where Var_0 is the initial variance of $p_0(x)$. So if the y_t satisfy the condition (3.15), then the system is a convex filter.

Next, we will show a figure of the posterior distributions at different times ($t \in \{1, 2, 3, 4, 5, 6, 7, 8\}$) with $p_0(x) = \frac{1}{\sqrt{2\pi Var_0}}e^{-\frac{1}{Var_0}x^2}$ and Var_0 is 0.1. The continuous observations y_t in Figure 1(a) satisfy (3.15) so that the posterior distributions are log-concave. However, the observation y_t in Figure 1(b) in the last step $t = 8$ fails to satisfy this property, so that the last posterior distribution is not log-concave.

3.2.2. The case of FDEA systems. For the general FDEA system with maximum rank, we can construct many convex filters in FDEA for any dimension n. We shall start with the following lemma.



(a) Convex Cubic system

(b) Convex Cubic system

FIG. 1. *The x-axis denotes as space variable x and takes value in $[-1.5, 1.5]$, the y-axis denotes time and takes values in $\{1, 2, 3, 4, 5, 6, 7, 8\}$, and the z-axis denotes the value of the posterior distributions.*

THEOREM 3.12 (Theorem 12 in [17]). *Consider the following equation:*

$$-\Delta\Phi + |\nabla\Phi|^2 = \sum_{i=1}^{n} \lambda_i x_i^2 - c, \tag{3.16}$$

where $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $\lambda_i \geq 0$. Then we have the following:

- *(Existence) When $c < \sum_{i=1}^n \sqrt{\lambda_i}$, there is a family of convex solutions of* (3.16) *with $2n$ parameters such that $|\nabla\Phi|$ has at most linear growth at ∞, namely,*

$$|\nabla\Phi(x)| \leq C(1 + |x|), \tag{3.17}$$

for some constant C.

- *(Uniqueness) When $c = \sum_{i=1}^n \sqrt{\lambda_i}$, there is a quadratic polynomial, uniquely determined up to a constant, which satisfies* (3.16)*. Moreover, this is the unique solution up to a constant if either one of the following conditions holds:*
  - *$\lambda_i = 0$ for any i.*
  - *There is at least $n-2$ nonzero terms in λ_i.*
- *(Nonexistence). When $c > \sum_{i=1}^n \sqrt{\lambda_i}$, there is no smooth solution to* (3.16).

Example 2. Consider a potential function $\Phi : \mathbb{R}^n \to \mathbb{R}$ defined by

$$\Phi(x) = -\sum_{i=1}^{n} \log(\cosh(x_i)),$$

where $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. Let $f = -\nabla\Phi(x)$. Then, $|\nabla\Phi(x)|^2 - \Delta\Phi(x) = n$. So we can take $h(x) = cI_n x$ for $c > 1$. In particular, if the initial distribution is a Gaussian $\Pi_{i=1}^n \frac{1}{\sqrt{\pi}} e^{-x_i^2}$, the filter is convex by Theorem 3.2.

Remark 3.13. The Benes filter also can be constructed by considering $c = -n$ and $\lambda_i = 0$ in Theorem 3.12.

Through our detailed analysis, it is not difficult to see that the posterior density function of the convex filtering system is still unimodal distribution when the initial density is log-concave. This property can provide the basis for using MAP as a framework for a continuous filtering system.

As for the continuous-discrete setting, the posterior densities can be log-concave after a quadratic term correction in the case of a linear system, which provides a reasonable explanation for filtering algorithms of the MAP framework.

4. Algorithms for nonlinear filtering.

In this section, we set up a new framework for nonlinear filtering algorithms by using the log concavity of the posterior densities.

There is no real continuous observation data in practical applications. In this section, we will assume that the observations are only obtained at time steps

$$\{0 = \tau_0 < \cdots < \tau_N = T\}.$$

In a word, the MAP framework of filtering is to solve the following equation in order to estimate the states,

$$\hat{x}_t^{\mathrm{MAP}} = \arg\max_{x_t \in \mathbf{R}^n} p(t, x_t | \mathcal{Y}_{\tau_k}) \quad \text{for } t \in [\tau_k, \tau_{k+1}), \tag{4.1}$$

where $0 \leq k \leq N-1$, $\tau_{N+1} > T$, and the $p(t, \cdot|\mathcal{Y}_t)$ is the posterior density at time t.

In practical applications such as target tracking, we need to get the maximum point of the density, not the average effect of density. One common assumption frequently used in the numerical method is to assume the posterior densities are Gaussians, which means we apply the linearization method to the original system, and it is similar with the EKF and UKF methods. Setting [5, 33, 44]

$$\tilde{p}(t, x_t|\mathcal{Y}_{t_k}) = \frac{1}{\sqrt{(2\pi)^n \det(\hat{P}_{t|t_k})}} e^{-\frac{1}{2}(x-\hat{x}_{t|t_k})^\top \hat{P}_{t|t_k}^{-1}(x-\hat{x}_{t|t_k})}, \tag{4.2}$$

when $t \in (\tau_k, \tau_{k+1})$ for some $0 \le k \le N-1$, the density $\tilde{p}(t, x_t|\mathcal{Y}_{t_k})$ is determined by (2.1). So the estimate $\hat{x}_{t|\tau_k}$ is given by

$$\hat{x}_{t|\tau_k} = \hat{x}_{\tau_k|\tau_k} + \int_{\tau_k}^t f(s, \hat{x}_{s|\tau_k}) ds. \tag{4.3}$$

Now we only need to determine the update part of the filtering:

$$\tilde{p}(t, x_t|\mathcal{Y}_{\tau_k}) \propto \tilde{p}(t, x_t|\mathcal{Y}_{\tau_{k-1}})\tilde{p}(z_{\tau_k}|x_{\tau_k}). \tag{4.4}$$

It is easy to see that finding the maximal point of the posterior densities is equal to finding the minimum point of $-\log \tilde{p}(t, x_t|\mathcal{Y}_{\tau_k})$:

$$\begin{aligned} \hat{x}_{\tau_k|\tau_k}^{\mathrm{MAP}} = \arg\min_{x_t \in \mathbf{R}^n} & (x - \hat{x}_{t|\tau_{k-1}})^\top \hat{P}_{t|t_{k-1}}^{-1}(x - \hat{x}_{t|t_{k-1}}) \\ & + (z_{\tau_k} - H(\tau_k, x_t))^\top S^{-1}(\tau_k)(z_{\tau_k} - H(\tau_k, x_t)). \end{aligned} \tag{4.5}$$

IEKF and its variants can be obtained by using the iterative method to solve such optimization problems.

From our previous characterization of the log concavity of the posterior distribution, log concavity will be preserved as the system evolves in time, but not preserved in general by the update functions. We will assume a continuous update process by introducing an update parameter λ and a MAP flow, which is motivated by [15]; the maximum point at the later time should be perturbed from a maximum point at this time:

$$\tilde{p}_\lambda(t, x_t|\mathcal{Y}_{\tau_k}) \propto \tilde{p}(t, x_t|\mathcal{Y}_{\tau_{k-1}})\tilde{p}(z_{\tau_k}|x_{\tau_k})^\lambda \text{ with } \lambda \in [0, 1]. \tag{4.6}$$

Thus we can refine our main idea into the following optimization problem:

$$\hat{x}_{\tau_k|\tau_k}^{\lambda,\ \mathrm{MAP}} = \arg\min_{x \in \mathbf{R}^n} L_\lambda(x_t), \tag{4.7}$$

where

$$\begin{aligned} L_\lambda(x) = \arg\min_{x \in \mathbf{R}^n} & (x - \hat{x}_{t|\tau_{k-1}})^\top \hat{P}_{t|t_{k-1}}^{-1}(x - \hat{x}_{t|t_{k-1}}) \\ & + \lambda(z_{\tau_k} - H(\tau_k, x))^\top S^{-1}(\tau_k)(z_{\tau_k} - H(\tau_k, x)). \end{aligned} \tag{4.8}$$

5. Numerical experiments. In this section, we compare the relative performance of IEKF, UKF, and our proposed IOKF on a typical example that is well-known as BOT, and we discuss the advantages and effectiveness of the proposed IOKF.

In the experiments, for the purpose of comparing the performance of different methods, we introduce the mean of the squared error (MSE) based on 100 realizations, which is defined as follows:

$$\mathrm{MSE} := \frac{1}{100} \sum_{i=1}^{100} \frac{1}{K_2 + 1} \sum_{k=0}^{K_2} \|x_k^{(i)} - \hat{x}_k^{(i)}\|_2, \tag{5.1}$$

where $x_k^{(i)}$ is the real state at discrete time instant k in the ith experiment and $\hat{x}_k^{(i)}$ is the estimation of $x_k^{(i)}$, with $0 \le k \le K_2$, where $K_2 \in \mathbb{N}$ is total number of time steps.

$$\begin{cases} dx_t = F x_t dt + w_t, \\ z_{t_k} = h(x_{t_k}) + B_k, \end{cases} \tag{5.2}$$

where w_t is the standard Brownian motion of dimension 4 and B_k is the Gaussian $\mathcal{N}(0, I_2)$,

$$F = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

and $h(x)$ is the function $\mathbb{R}^4 \to \mathbb{R}^2$.

In Table 1, the iteration of IOKF is $2 * 3$, which means that 2 is a number of the Newton gradient descent steps and 3 is the split number of the optimization problem l in Algorithm 4.1. The algorithm of UKF is from [42] and with the $\alpha = 0.5$, $\beta = 2$, $\kappa = -0.5$ in the numerical experiments.

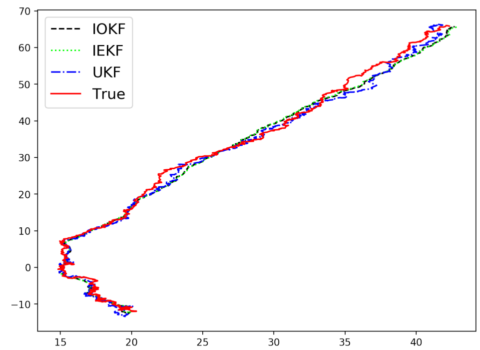
As we can see from Figure 2 and Table 1, the proposed algorithm, IOKF, outperforms IEKF when we fix the total number of iteration steps. Furthermore, the proposed algorithm IOKF has a similar performance to UKF, and UKF can be considered a standard benchmark. We ran our simulations on CPU clusters with an Intel Core i9-9880H (2.3 GHz/L3 16M) equipped with 16 GB memory.

6. Conclusion.

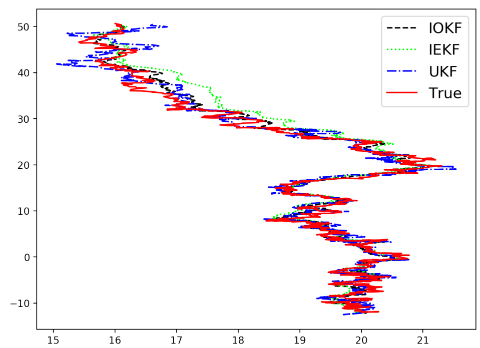
In this paper, we extend the Yau–Yau PDE method to the MAP framework. Convex filtering is important when we need to characterize the maximum point in the posterior densities. Furthermore, we give an iterative algorithm motivated by the structure of convex filtering. The effectiveness of the new algorithm reflects the theoretical value of convex filtering. And we hope there are still many efficient algorithms that can be designed by using the concept of convex filtering.

Algorithm 4.1 IOKF based on MAP framework.

1: **Input**: Functions f, h, H, $S(t)$, initial density $p(0,x)$, and let l, m be the number of iterations for the optimization problem and the optimization steps for a single problem, respectively.
2: Choose a sequence of numbers $0 = \lambda_0 < \cdots < \lambda_l = 1$ and calculate the initial estimation $\hat{x}_{0|0}^{\mathrm{MAP}}$. Set $k = 0$.
3: **While** $k \le N$ **Do:**
4: **For** $j = 1 : l$ **Do:**
5: Define functions $L_\lambda(x)$ by using the observation z_{t_k} as equation.
6: Solve $\hat{x}_{\tau_k|\tau_k}^{\lambda_j,\,\mathrm{MAP}} = \arg\min_{x \in \mathbf{R}^n} L_{\lambda_j}(x)$ with initial $\hat{x}_{\tau_k|\tau_k}^{\lambda_{j-1},\,\mathrm{MAP}}$ with Newton method for m time.
7: $\hat{x}_{\tau_k|\tau_k}^{MAP} = \hat{x}_{\tau_k|\tau_k}^{\lambda_l,\,\mathrm{MAP}}$
8: **Output:** $\hat{x}_{\tau_k|\tau_k}^{MAP}$
9: $k = k + 1$
10: **End**



(a) tracking



(b) tracking

FIG. 2. *Results of BOT. The x-axis denotes the first component in* x_t *and the y-axis denotes the second component in* x_t *for both* (a) *and* (b).

TABLE 1
MSE for 100 *realizations.*

Algorithm	MSE	Steps	Time (s)	Iteration
IOKF	0.6274	1000	0.50162	$2*3$
IEKF	0.68334	1000	0.52093	6
UKF	0.6295	1000	0.46794	–

Appendix.

Proof of Theorem 2.3. In Theorem 2.8 of [28], Korevaar considers parabolic equations in a slightly more general formulation of the form

$$\frac{\partial}{\partial t} u = Lu, \tag{6.1}$$

where L is a nonlinear operator of the form

$$Lu = \sum_{i,j=1}^{n} a_{ij}(\nabla u) \frac{\partial^2}{\partial x_i \partial x_j} - b(x, u, \nabla u) \tag{6.2}$$

such that $\frac{\partial b}{\partial u} \geq 0$. In the special case that $b(x, u, \nabla u)$ does not have a dependence on u, the $\frac{\partial b}{\partial u}$ condition is automatically satisfied. Under such assumptions, Korevaar considers a convexity function

$$\mathcal{C}(t,y_1,y_3,\mu) = u(t,\mu y_1 + (1-\mu)y_3) - \mu u(t,y_1) - (1-\mu)u(t,y_3),$$

defined for $t,y_1,y_3,\mu \in [0,T)\times\bar{\Omega}\times\bar{\Omega}\times[0,1]$. Korevaar concludes that if $\mathcal{C}$ is positive anywhere, it attains a positive maximum at a "boundary" point, that is, a point for which $t=0$ or a point for which at least one of $\{y_1,y_3,\mu y_1+(1-\mu)y_3\}$ lies on the boundary $\partial\Omega$. Therefore, if $u(t,x)$ fails to be convex for some t, then $\mathcal{C}>0$ for some (t,y_1,y_3,μ). Moreover, if Ω is strictly convex, then if $\mu y_1+(1-\mu)y_3\in\partial\Omega$, we must have $y_1=y_3$ or $\mu\in\{0,1\}$, and in this case, we have $\mathcal{C}=0$ (to avoid infinities here we can replace Ω with $\Omega_\delta=\{x : d(x,\partial\Omega)\geq\delta\}$ as in [28]). By the assumption that $u(0,x)$ is convex, we know the convexity function is nonpositive for $t=0$. So the only remaining case is if one of $y_1,y_3\in\partial\Omega$ and $\mu\in(0,1)$. But our assumptions on strong convexity and interior normal show that the hypotheses of Lemma 2.4 in [28] are satisfied, and the conclusion of that lemma rules out this last remaining case. Therefore, we conclude that $\mathcal{C}$ cannot be positive anywhere on the interior, and $u(t,x)$ must be convex for all t. ∎

Acknowledgment. Professor Stephen Shing-Toung Yau is grateful to the National Center for Theoretical Sciences (NCTS) for providing an excellent research environment while part of this research was done.

REFERENCES

[1] P. Albano, *Some properties of semiconcave functions with general modulus*, J. Math. Anal. Appl., 271 (2002), pp. 217–231, https://doi.org/10.1016/S0022-247X(02)00117-8.

[2] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, *A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking*, IEEE Trans. Signal Process., 50 (2002), pp. 174–188.

[3] B. Balaji, *Universal nonlinear filtering using Feynman path integrals* II*: The continuous-continuous model with additive noise*, PMC Phys. A, 3 (2009), pp. 1–28.

[4] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*, John Wiley & Sons, New York, 2004.

[5] B. Bell and F. Cathey, *The iterated Kalman filter update as a Gauss-Newton method*, IEEE Trans. Automat. Control, 38 (1993), pp. 294–297, https://doi.org/10.1109/9.250476.

[6] R. W. Brockett and J. M. C. Clark, *On the geometry of the conditional density equation*, in Analysis and Optimisation of Stochastic Systems, D. L. R. Jacobs, ed., Academic Press, New York, 1980, pp. 299–309.

[7] R. W. Brockett, *Nonlinear Systems and Nonlinear Estimation Theory*, in Stochastic Systems: The Mathematics of Filtering and Identification and Applications, Springer, New York, 1981, pp. 441–477.

[8] J. Chen and S. S. Yau, *Finite-dimensional filters with nonlinear drift* VII*: Mitter conjecture and structure of* η, SIAM J. Control Optim., 35 (1997), pp. 1116–1131.

[9] J. Chen, S. S. Yau, and C.-W. Leung, *Finite-dimensional filters with nonlinear drift* VIII*: Classification of finite-dimensional estimation algebras of maximal rank with state-space dimension* 4, SIAM J. Control Optim., 35 (1997), pp. 1132–1141.

[10] J. Chen and S. S.-T. Yau, *Finite-dimensional filters with nonlinear drift,* VI*: Linear structure of* Ω, Math. Control Signals Syst., 9 (1996), pp. 370–385.

[11] J. Chen, S. S.-T. Yau, and C.-W. Leung, *Finite-dimensional filters with nonlinear drift* IV*: Classification of finite-dimensional estimation algebras of maximal rank with state-space dimension* 3, SIAM J. Control Optim., 34 (1996), pp. 179–198.

[12] X. Chen, J. Shi, and S. S.-T. Yau, *Real-time solution of time-varying Yau filtering problems via direct method and Gaussian approximation*, IEEE Trans. Automat. Control, 64 (2019), pp. 1648–1654.

[13] W.-L. Chiou and S. S.-T. Yau, *Finite-dimensional filters with nonlinear drift.* II*: Brockett's problem on classification of finite-dimensional estimation algebras*, SIAM J. Control Optim., 32 (1994), pp. 297–310.

[14] J. D. Cole, *On a quasi-linear parabolic equation occurring in aerodynamics*, Quart. Appl. Math., 9 (1951), pp. 225–236.

[15] F. Daum and J. Huang, *Particle flow for nonlinear filters with log-homotopy*, in Signal and Data Processing of Small Targets 2008, Proc. SPIE 6969, 2008, 696918.

[16] M. Davis, *On a multiplicative functional transformation arising in nonlinear filtering theory*, Z. Wahrscheinlichkeitstheorie verwandte Gebiete, 54 (1980), pp. 125–139.

[17] R.-T. Dong, L.-F. Tam, W. S. Wong, and S. S.-T. Yau, *Structure and classification theorems of finite-dimensional exact estimation algebras*, SIAM J. Control Optim., 29 (1991), pp. 866–877.

[18] W. Dong, X. Chen, and S. S.-T. Yau, *The novel classes of finite dimensional filters with non-maximal rank estimation algebra on state dimension four and rank of one*, Internat. J. Control, 94 (2019), pp. 1–20.

[19] W. Dong, X. Luo, and S. S.-T. Yau, *Solving nonlinear filtering problems in real time by Legendre Galerkin spectral method*, IEEE Trans. Automat. Control, 66 (2021), pp. 1559–1572.

[20] T. E. Duncan, *Probability Densities for Diffusion Processes with Applications to Nonlinear Filtering System,* Ph.D. dissertation, Stanford University, 1967.

[21] S. Godsill, *Particle filtering: The first* 25 *years and beyond*, in Proceedings of the International Conference on Acoustics, Speech and Signal Processing, IEEE, 2019, pp. 7760–7764.

[22] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, *Novel approach to nonlinear/non-Gaussian Bayesian state estimation*, IEE Proc. F Radar Signal Proc., 140 (2002), pp. 107–113.

[23] A. C. Harvey, *A unified view of statistical forecasting procedures*, J. Forecasting, 3 (1984), pp. 245–275.

[24] E. Hopf, *The Partial Differential Equation* $u_t + uu_x = \mu_{xx}$, Technical report, Indiana University at Blomington, 1950.

[25] G.-Q. Hu and S.-T. Yau, *Finite-dimensional filters with nonlinear drift.* XV. *New direct method for construction of universal finite-dimensional filter*, IEEE Trans. Aeros. Electron. Syst., 38 (2002), pp. 50–57.

[26] X.-p. Jiao and S. S.-T. Yau, *New classes of finite dimensional filters with nonmaximal rank estimation algebra on state dimension n and linear rank n*-2, SIAM J. Control Optim., 58 (2020), pp. 3413–3427.

[27] R. E. Kalman, *A new approach to linear filtering and prediction problems*, Trans. ASME J. Basic Eng., 82 (1960), pp. 35–45.

[28] N. J. Korevaar, *Convex solutions to nonlinear elliptic and parabolic boundary value problems*, Indiana Univ. Math. J., 32 (1983), pp. 603–614.

[29] J. Kotecha and P. Djuric, *Gaussian particle filtering*, IEEE Trans. Signal Process., 51 (2003), pp. 2592–2601, https://doi.org/10.1109/TSP.2003.816758.

[30] C. Kuptametee and N. Aunsri, *A review of resampling techniques in particle filtering framework*, Measurement (2022), 110836.

[31] J. Leitão and F. D. Nunes, *A nonlinear filtering approach to carrier tracking and symbol detection in digital phase modulation*, in Proceedings of ICC/SUPERCOMM'94, IEEE, 1994, pp. 386–390.

[32] X. R. Li, *Recursibility and optimal linear estimation and filtering*, in Proceedings of the 43rd IEEE Conference on Decision and Control, Vol. 2, IEEE, 2004, pp. 1761–1766.

[33] Y. Li, Y. Cheng, X. Li, H. Wang, X. Hua, and Y. Qin, *Bayesian nonlinear filtering via information geometric optimization*, Entropy, 19 (2017), 655.

[34] L. Ljung, *Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems*, IEEE Trans. Automat. Control, 24 (1979), pp. 36–50.

[35] X. Luo and S. S.-T. Yau, *Complete real time solution of the general nonlinear filtering problem without memory*, IEEE Trans. Automat. Control, 58 (2013), pp. 2563–2578.

[36] S. K. Mitter, *On the analogy between mathematical problems of nonlinear filtering and quantum physics*, Ric. Automat., 10 (1979), pp. 163–216.

[37] R. E. Mortensen, *Optimal Control of Continuous Time Stochastic Systems*, Ph.D. dissertation, University of California, Berkeley, 1966.

[38] F. D. Nunes and J. M. Leitao, *A nonlinear filtering approach to estimation and detection in mobile communications*, IEEE J. Sel. Area Commun., 16 (1998), pp. 1649–1659.

[39] H. Ogawa and E. Oja, *Projection filter, Wiener filter, and Karhunen-Loeve subspaces in digital image restoration*, J. Math. Anal. Appl., 114 (1986), pp. 37–51.

[40] S. Pathiraja, S. Reich, and W. Stannat, *McKean-Vlasov SDEs in nonlinear filtering*, SIAM J. Control Optim., 59 (2021), pp. 4188–4215.

[41] H. Poor, *On robust Wiener filtering*, IEEE Trans. Automat. Control, 25 (1980), pp. 531–536.

[42] S. Sarkka, *On unscented Kalman filtering for state estimation of continuous-time nonlinear systems*, IEEE Trans. Automat. Control, 52 (2007), pp. 1631–1641, https://doi.org/10.1109/TAC.2007.904453.

[43] J. Shi, X. Chen, W. Dong, and S. S.-T. Yau, *New classes of finite dimensional filters with non-maximal rank*, IEEE Control Systems Lett., 1 (2017), pp. 233–237.

[44] J. Shi, Z. Yang, and S.-T. Yau, *Direct method for Yau filtering system with nonlinear observations*, Internat. J. Control, 91 (2018), pp. 678–687.

[45] J. Shi and S. S.-T. Yau, *Finite dimensional estimation algebras with state dimension* 3 *and rank* 2, I*: Linear structure of Wong matrix*, SIAM J. Control Optim., 55 (2017), pp. 4227–4246.

[46] J. Shi and S. S.-T. Yau, *Finite dimensional estimation algebras with state dimension* 3 *and rank* 2*, Mitter conjecture*, Internat. J. Control, 93 (2020), pp. 2177–2186.

[47] E. A. Wan and R. Van Der Merwe, *The unscented Kalman filter for nonlinear estimation*, in Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium, IEEE, 2000, pp. 153–158.

[48] Z. Wang, X. Luo, S. S.-T. Yau, and Z. Zhang, *Proper orthogonal decomposition method to nonlinear filtering problems in medium-high dimension*, IEEE Trans. Automat. Control, 65 (2020), pp. 1613–1624.

[49] X. Wu and S. S.-T. Yau, *Classification of estimation algebras with state dimension* 2, SIAM J. Control Optim., 45 (2006), pp. 1039–1073.

[50] S.-T. Yau and G.-G. Hu, *Finite-dimensional filters with nonlinear drift* X*: Explicit solution of DMZ equation*, IEEE Trans. Automat. Control, 46 (2001), pp. 142–148.

[51] S. S.-T. Yau and G. Q. Hu, *Classification of finite-dimensional estimation algebras of maximal rank with arbitrary state-space dimension and Mitter conjecture*, Internat. J. Control, 78 (2005), pp. 689–705.

[52] S.-T. Yau and S. S.-T. Yau, *Finite-dimensional filters with nonlinear drift.* III*: Duncan-Mortensen-Zakai equation with arbitrary initial condition for the linear filtering system and the Benes filtering system*, IEEE Trans. Aero. Electron. Syst., 33 (1997), pp. 1277–1294, https://doi.org/10.1109/7.625127.

[53] S.-T. Yau and S. S.-T. Yau, *Real time solution of the nonlinear filtering problem without memory* II, SIAM J. Control Optim., 47 (2008), pp. 163–195.

[54] M. Zakai, *On the optimal filtering of diffusion processes*, Probab. Theory Related Fields, 11 (1969), pp. 230–243.