A Novel Particle Filter for High-Dimensional Systems Using Penalized Perturbations

Yousef El-Laham⋆, Zahraa Krayem⋆, Jessica Maghakian†, Mónica Bugallo⋆
⋆Department of Electrical & Computer Engineering
†Department of Applied Mathematics & Statistics
Stony Brook University, Stony Brook (USA)
{yousef.ellaham, zahraa.krayem, jessica.maghakian, monica.bugallo}@stonybrook.edu

***Abstract*—In order to efficiently perform inference on high-dimensional nonlinear non-Gaussian state-space models using particle filtering, it is critical that particles are generated from the optimal proposal distribution. However, finding a closed-form to the optimal proposal proves to be difficult in practice, as many application problems do not satisfy the requirement of conjugate state and observation equations. In this paper, we overcome this challenge by designing a novel method that introduces conjugate artificial noise into the system and optimally perturbs the particles in a way that balances a bias-variance tradeoff. Our method is validated through extensive numerical simulations applied to a gene regulatory network problem, and results show better performance than that of state-of-the-art methods, especially in cases where the state noise is heavy-tailed.**

I. Introduction

State-space models (SSMs) are a popular class of mathematical models used to relate observed data to a set of hidden state variables. These models are used in a variety of fields, including epidemiology, genetics, finance, ecology and signal processing [1]–[4]. SSMs are characterized by: *(i)* a *state equation* that describes the dynamics of the latent states through some system transition function, and *(ii)* an *observation equation*, which relates the latent states to the observed data through a measurement function. A typical state estimation problem within the SSM framework is often presented in a Bayesian setting, where the goal is to compute the posterior probability distribution of the latent states given the set of available observations.

In the special case that the state and observation equations are linear and Gaussian, Kalman filtering (KF) provides the optimal closed-form solution to the state estimation problem [5]. Unfortunately in many applications nonlinearities can appear in both the transition and measurement functions. In such cases, approximate KF methods can be employed. For example, the extended KF (EKF) method utilizes a first order Taylor approximation to linearize the system [6]. For highly nonlinear systems, the unscented KF (UKF) technique achieves better performance than EKF by propagating a set of representative sample points that capture the mean and covariance of the posterior distribution at each time instant [7]. However, the performance of these methods heavily relies on the assumption that the system noise is Gaussian.

Particle filtering (PF) methods provide a flexible alternative approach to nonlinear and non-Gaussian SSMs. In these methods, the posterior density is approximated using a discrete random measure formed by a set of samples, called *particles*, and weights. The methodology is based on a Monte Carlo scheme called *sequential importance sampling*, that sequentially propagates a set of particles using a proposal density and then assigns weights to the particles proportionally to how well they represent the target posterior [8], [9]. For example, the famous bootstrap filter uses the state equation to propagate the particles. In high-dimensional state-spaces, the choice of proposal becomes critical and methods that do not use any information about the observations for propagation, such as the bootstrap filter, perform poorly. Ideally, a PF algorithm would use an optimal proposal distribution, which optimally incorporates information about the observations in the propagation step [10]. Unfortunately, a closed-form solution to the optimal proposal is only attainable for models with conjugate state and observation equations. As a result, the design of PF algorithms that can scale to high-dimensional state-spaces regardless of the distributions of the state and observation equations is a highly active area of research.

In this work, we propose a novel PF scheme that scales the state estimation of nonlinear and non-Gaussian SSMs to high-dimensional systems. To do so, we consider the approximated model introduced in [11] that biases the filtering problem through the addition of artificial noise. We derive the optimal artificial noise covariance matrix that minimizes the average Kullback-Leibler divergence (KLD) between a set of transformed particles and the observed data, while penalizing perturbations in instances when the PF is performing well. Finally, we apply the novel method to a gene regulatory network problem and conduct extensive simulations.

II. Prior Work

We consider a general state-space model of the form:

$$\begin{aligned} \mathbf{x}_t &= f(\mathbf{x}_{t-1}, \mathbf{u}_t), \quad \mathbf{u}_t \sim p(\mathbf{u}_t), \\ \mathbf{y}_t &= \mathbf{H}\mathbf{x}_t + \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}), \end{aligned} \tag{1}$$

This work was carried out thanks to the support of the National Science Foundation (NSF) under Award CCF-1617986. The authors thank the Research Computing and Cyberinfrastructure and the Institute for Advanced Computational Science at Stony Brook University for access to the high-performance SeaWulf computing system sponsored by NSF (# 1531492).

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

where $\mathbf{x}_t \in \mathbb{R}^{d_x}$ is a state vector, $\mathbf{y}_t \in \mathbb{R}^{d_y}$ is a vector of observations, $f(\cdot)$ is the system transition function, $\mathbf{u}_t$ and $\mathbf{v}_t$ are noise vectors, $p(\mathbf{u}_t)$ is the distribution of the noise vector $\mathbf{u}_t$, $\mathbf{H} \in \mathbb{R}^{d_y \times d_x}$ is the observation matrix, and t denotes a time index. Although in our setting the observation equation is linear and Gaussian, the transition equation may be nonlinear and non-Gaussian. The goal is to approximate $p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$, the posterior distribution of the states given the observations.

In [11], Wigren et al. introduce a bias to the model in (1) by approximating the SSM of interest as follows:

$$\begin{aligned} \tilde{\mathbf{x}}_t &= f(\mathbf{x}_{t-1}, \mathbf{u}_t), \quad & \mathbf{u}_t &\sim p(\mathbf{u}_t) \\ \mathbf{x}_t &= \tilde{\mathbf{x}}_t + \boldsymbol{\xi}_t, & \boldsymbol{\xi}_t &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}_t) \\ \mathbf{y}_t &= \mathbf{H}\mathbf{x}_t + \mathbf{v}_t, & \mathbf{v}_t &\sim \mathcal{N}(\mathbf{0}, \mathbf{R}). \end{aligned} \tag{2}$$

This formulation introduces an artificial noise process, whereby the latent state vector is perturbed using a Gaussian distribution with covariance matrix $\boldsymbol{\Lambda}_t$. We call the distribution $p(\mathbf{x}_t|\tilde{\mathbf{x}}_t) = \mathcal{N}(\mathbf{x}_t|\tilde{\mathbf{x}}_t, \boldsymbol{\Lambda}_t)$ the *perturbation distribution*. If $\boldsymbol{\Lambda}_t$ is non-zero and positive semidefinite, then this model introduces a bias. At each time instance t, a proposal distribution $q(\tilde{\mathbf{x}}_t, \mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_t)$ is used to propagate a set of weighted particles $\mathcal{X}_{t-1} = \{\mathbf{x}_{t-1}^{(n)}, w_{t-1}^{(n)}\}_{n=1}^N$, where $w_{t-1}^{(n)}$ is the weight of the nth particle stream $\mathbf{x}_{1:t-1}^{(n)}$. In general, the update rule for the weights of the nth particle stream is

$$w_t^{(n)} = w_{t-1}^{(n)} \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(n)}) p(\mathbf{x}_t^{(n)}|\tilde{\mathbf{x}}_t^{(n)}) p(\tilde{\mathbf{x}}_t^{(n)}|\mathbf{x}_{t-1}^{(n)})}{q(\tilde{\mathbf{x}}_t^{(n)}, \mathbf{x}_t^{(n)}|\mathbf{x}_{t-1}^{(n)}, \mathbf{y}_t)}, \tag{3}$$

where $\{\tilde{\mathbf{x}}_t^{(n)}, \mathbf{x}_t^{(n)}\} \sim q(\tilde{\mathbf{x}}_t^{(n)}, \mathbf{x}_t^{(n)}|\mathbf{x}_{t-1}^{(n)}, \mathbf{y}_t)$. If we select a proposal distribution that can be factored as $q(\tilde{\mathbf{x}}_t^{(n)}, \mathbf{x}_t^{(n)}|\mathbf{x}_{t-1}^{(n)}, \mathbf{y}_t) = q(\mathbf{x}_t^{(n)}|\tilde{\mathbf{x}}_t^{(n)}, \mathbf{y}_t) q(\tilde{\mathbf{x}}_t^{(n)}|\mathbf{x}_{t-1}^{(n)})$ and let $q(\tilde{\mathbf{x}}_t^{(n)}|\mathbf{x}_{t-1}^{(n)}) = p(\tilde{\mathbf{x}}_t^{(n)}|\mathbf{x}_{t-1}^{(n)})$, then

$$w_t^{(n)} = w_{t-1}^{(n)} \frac{\mathcal{N}(\mathbf{y}_t|\mathbf{H}\mathbf{x}_t^{(n)}, \mathbf{R}) \mathcal{N}(\mathbf{x}_t^{(n)}|\tilde{\mathbf{x}}_t^{(n)}, \boldsymbol{\Lambda}_t)}{q(\mathbf{x}_t^{(n)}|\tilde{\mathbf{x}}_t^{(n)}, \mathbf{y}_t)}, \tag{4}$$

where $\tilde{\mathbf{x}}_t^{(n)} \sim p(\tilde{\mathbf{x}}_t^{(n)}|\mathbf{x}_{t-1}^{(n)})$ and $\mathbf{x}_t^{(n)} \sim q(\mathbf{x}_t^{(n)}|\tilde{\mathbf{x}}_t^{(n)}, \mathbf{y}_t)$. Since the numerator in (4) represents a Gaussian conjugate pair, $q(\mathbf{x}_t^{(n)}|\tilde{\mathbf{x}}_t^{(n)}, \mathbf{y}_t)$ can be chosen as the optimal proposal $\mathcal{N}(\mathbf{x}_t^{(n)}|\mathbf{m}_t^{(n)}, \mathbf{C}_t)$ where $\mathbf{m}_t^{(n)}$ and $\mathbf{C}_t$ are given by

$$\begin{aligned} \mathbf{m}_t^{(n)} &= \tilde{\mathbf{x}}_t^{(n)} + \boldsymbol{\Lambda}_t \mathbf{H}^\intercal (\mathbf{R} + \mathbf{H}\boldsymbol{\Lambda}_t \mathbf{H}^\intercal)^{-1} (\mathbf{y}_t - \mathbf{H}\tilde{\mathbf{x}}_t^{(n)}) \\ \mathbf{C}_t &= \boldsymbol{\Lambda}_t - \boldsymbol{\Lambda}_t \mathbf{H}^\intercal (\mathbf{R} + \mathbf{H}\boldsymbol{\Lambda}_t \mathbf{H}^\intercal)^{-1} \mathbf{H}\boldsymbol{\Lambda}_t. \end{aligned} \tag{5}$$

Furthermore, the conjugacy allows for the particle weights $w_t^{(n)} = w_{t-1}^{(n)} p(\mathbf{y}_t|\tilde{\mathbf{x}}_t^{(n)})$ to be obtained in closed-form

$$w_t^{(n)} = w_{t-1}^{(n)} \mathcal{N}(\mathbf{y}_t|\mathbf{H}\tilde{\mathbf{x}}_t^{(n)}, \mathbf{R} + \mathbf{H}\boldsymbol{\Lambda}_t \mathbf{H}^\intercal). \tag{6}$$

Unfortunately, there is no straightforward way to choose the perturbation covariance matrix $\boldsymbol{\Lambda}_t$. In [11] the form $\boldsymbol{\Lambda}_t = \epsilon^2 \mathbf{S}$ is assumed, and some heuristic suggestions on how to choose ϵ and $\mathbf{S}$ are discussed. However, no general procedure for choosing $\boldsymbol{\Lambda}_t$ is established that is robust regardless of the distributional assumptions on the state noises.

III. Proposed Methodology

In this work, we propose a robust particle perturbation strategy for the proposed model in (2). Specifically, we derive the optimal perturbation covariance matrix that minimizes the KLD between the distribution of the perturbed particles (projected to the space of the observations) and the observation distribution. For robustness, we consider an objective function that utilizes a regularization term which is calculated using the performance of the unperturbed particles. The regularization term guarantees that when the unperturbed particles perform well, the particle perturbations are correspondingly decreased.

A. Average Kullback-Leibler Minimization

Consider the random variables $\tilde{\mathbf{z}}_t = \mathbf{H}\tilde{\mathbf{x}}_t$ and $\mathbf{z}_t = \mathbf{H}\mathbf{x}_t$, which are just linearly transformed versions of $\tilde{\mathbf{x}}_t$ and $\mathbf{x}_t$, respectively. We define the *transformed perturbation distribution* of the nth particle as the distribution of $\mathbf{z}_t$ given $\tilde{\mathbf{z}}_t^{(n)}$

$$P_t^{(n)} = \mathcal{N}(\mathbf{z}_t|\tilde{\mathbf{z}}_t^{(n)}, \boldsymbol{\Phi}_t), \quad n = 1, \ldots, N, \tag{7}$$

where $\boldsymbol{\Phi}_t = \mathbf{H}\boldsymbol{\Lambda}_t\mathbf{H}^\intercal \in \mathbb{R}^{d_y \times d_y}$ is the covariance matrix in the transformed space. The observation distribution is $S_t = \mathcal{N}(\mathbf{y}_t|\mathbf{z}_t, \mathbf{R})$. To measure the distance between distributions, we utilize the KLD, denoted as $\mathcal{D}_{KL}(\cdot\|\cdot)$. For any two probability distributions P and Q, $\mathcal{D}_{KL}(P\|Q)$ is the expectation of the logarithmic difference of the probabilities. To minimize the distance between the perturbed particles and the observations, we construct an objective function that averages the KLD between each transformed perturbation distribution and the observation distribution

$$\boldsymbol{\Phi}_t^* = \underset{\boldsymbol{\Phi}_t \succeq 0}{\arg\min} \left(\frac{1}{N} \sum_{n=1}^N \mathcal{D}_{KL}(S_t \| P_t^{(n)}) \right), \tag{8}$$

where the constraint $\boldsymbol{\Phi}_t \succeq 0$ restricts $\boldsymbol{\Phi}_t$ to the set of positive semidefinite matrices. In the case that the underlying distributions are Gaussian, the KLD has a closed-form expression, which in turn yields the solution to (8) as

$$\boldsymbol{\Phi}_t^* = \mathbf{R} + \frac{1}{N} \sum_{n=1}^N (\tilde{\mathbf{z}}_t^{(n)} - \mathbf{z}_t)(\tilde{\mathbf{z}}_t^{(n)} - \mathbf{z}_t)^\intercal. \tag{9}$$

However, evaluating $\boldsymbol{\Phi}_t^*$ requires knowledge of $\mathbf{z}_t$, which is unknown. One solution is to replace $\mathbf{z}_t$ with an estimate $\hat{\mathbf{z}}_t$, such as the average of the unperturbed transformed particles:

$$\hat{\mathbf{z}}_t = \frac{1}{N} \sum_{n=1}^N \tilde{\mathbf{z}}_t^{(n)}. \tag{10}$$

The approximated optimal solution then becomes

$$\boldsymbol{\Phi}_t^* \approx \mathbf{R} + \hat{\boldsymbol{\Sigma}}_t, \tag{11}$$

where $\hat{\boldsymbol{\Sigma}}_t = \frac{1}{N} \sum_{n=1}^N (\tilde{\mathbf{z}}_t^{(n)} - \hat{\mathbf{z}}_t)(\tilde{\mathbf{z}}_t^{(n)} - \hat{\mathbf{z}}_t)^\intercal$ is an unweighted empirical estimator of the covariance matrix of the unperturbed states $\tilde{\mathbf{z}}_t$. Asymptotically, the approximation to $\boldsymbol{\Phi}_t^*$ can be seen as the sum of the transformed state and observation noise covariance matrices.

Given the solution in (11), we can compute the optimal perturbation covariance matrix as $\mathbf{\Lambda}_t^* = \mathbf{H}^{-1}\mathbf{\Phi}_t^*(\mathbf{H}^{-1})^\intercal$. In the case that $\mathbf{H}$ is not full rank and the inverse does not exist, we can instead use the Moore-Penrose inverse of $\mathbf{H}$.

The purpose of using perturbations is to drive the set of particles closer to the observations. However, in the case that the particles are already tracking the latent states well, or the observations are uninformative due to high noise, any perturbations will introduce an unnecessary bias in the estimation of the state. This lack of robustness for high observation noise is resolved in the following section by the introduction of a regularization term in the objective function.

B. Regularized Objective Function

To tackle high observation noise, there needs to be a way of controlling the strength of the perturbations. Since the magnitude of the perturbations rely on the eigenvalues of $\mathbf{\Lambda}_t$, we augment the objective function in (11) with a regularizer that takes the eigenvalues of $\mathbf{\Phi}_t$ into consideration. The new objective function $\mathcal{L}$ with regularization parameter ρ_t is

$$\mathcal{L}(\mathbf{\Phi}_t, \rho_t) = \rho_t \log\left(\det \mathbf{\Phi}_t\right) + \frac{1}{N}\sum_{n=1}^{N} \mathcal{D}_{KL}(S_t \| P_t^{(n)}), \quad (12)$$

where $0 \leq \rho_t < \infty$. The term $\rho_t \log\left(\det \mathbf{\Phi}_t\right)$ penalizes solutions with large determinants, thus penalizing large eigenvalues. Minimizing $\mathcal{L}(\mathbf{\Phi}_t, \rho_t)$ with respect to $\mathbf{\Phi}_t$ yields the following approximated optimal solution:

$$\hat{\mathbf{\Phi}}_t^* = \underset{\mathbf{\Phi}_t \succeq 0}{\arg\min}\left(\mathcal{L}(\mathbf{\Phi}_t, \rho_t)\right) = \frac{1}{1+\rho_t}\left(\mathbf{R} + \hat{\mathbf{\Sigma}}_t\right). \quad (13)$$

Unlike the solution in (11), here the perturbation of the particles can be controlled by the parameter ρ_t. As $\rho_t \to \infty$, the solution in (13) approaches the null matrix, leading to no perturbation of the particles.

C. Choice of the Regularization Parameter

An important concern is the choice of the regularization parameter ρ_t. Large values of ρ_t are suitable for cases in which the unperturbed particles are tracking the state efficiently, and no perturbation of the particles is needed. On the other hand, if the unperturbed particles are tracking poorly, setting $\rho_t = 0$ improves the performance by incorporating more information about the observations.

To determine whether a particle filter is performing well, the effective sample size (ESS) of the unperturbed particles can be evaluated at each time instant. The ESS is a measure of the degeneracy of the particle weights, which is related to the performance of the PF scheme [12]. Consider the weight $r_t^{(n)}$ of the nth particle $\tilde{\mathbf{x}}_t^{(n)}$ under the case of no perturbation:

$$r_t^{(n)} = p(\mathbf{y}_t | \tilde{\mathbf{x}}_t^{(n)}) = \mathcal{N}(\mathbf{y}_t | \mathbf{H}\tilde{\mathbf{x}}_t^{(n)}, \mathbf{R}). \quad (14)$$

Given these weights, the ESS under no perturbation can be approximated as follows:

$$\widehat{\mathrm{ESS}}_t = \frac{\left(\sum_{n=1}^{N} r_t^{(n)}\right)^2}{\sum_{n=1}^{N}\left(r_t^{(n)}\right)^2}. \quad (15)$$

The approximation in (15) has the property that $1 \leq \widehat{\mathrm{ESS}}_t \leq N$. $\widehat{\mathrm{ESS}}_t = 1$ corresponds to the most extreme case of particle weight degeneracy and thus implies that the unperturbed particles are not tracking the state well. The case that $\widehat{\mathrm{ESS}}_t = N$ implies $r_t^{(1)} = r_t^{(2)} = \cdots = r_t^{(N)}$, mimicking the nature of true Monte Carlo sampling. We propose a strategy for selecting ρ_t based on the approximated ESS of the set of particles:

$$\rho_t = \widehat{\mathrm{ESS}}_t - 1, \quad (16)$$

When the particle weights are degenerate, $\rho_t = 0$, resulting in the largest perturbation of the particles.

D. Comments on Implementation

In high-dimensional scenarios, stable estimation of the empirical covariance matrix $\hat{\mathbf{\Sigma}}_t$ requires many particles. To stabilize the estimation, one can use a biased estimate of $\hat{\mathbf{\Sigma}}_t$ by restricting to isotropic or diagonal matrices. These biased estimates will have less variance than the full covariance matrix, and thus will stabilize $\hat{\mathbf{\Phi}}_t^*$.

IV. Simulations

In order to validate the novel method, we run numerical simulations on synthetic data generated from a SSM describing the time evolution of gene expressions [13]. We considered a network of d_x genes, where $\mathbf{x}_t = [x_{1,t}, \ldots, x_{d_x,t}]^\intercal \in \mathbb{R}^{d_x}$ denotes the vector of gene expressions at time instant t. The state equation of the SSM is given as follows:

$$\mathbf{x}_t = \mathbf{A}\mathbf{g}(\mathbf{x}_{t-1}) + \mathbf{u}_t, \quad \mathbf{u}_t \sim p(\mathbf{u}_t), \quad (17)$$

where $\mathbf{A} \in \mathbb{R}^{d_x \times d_x}$ is a sparse coefficient matrix and $\mathbf{g}(\mathbf{x}_{t-1}) = [g(x_{1,t-1}), \ldots, g(x_{d_x,t-1})]^\intercal \in \mathbb{R}^{d_x}$ is a vector of nonlinear transformations applied to each gene expression as determined by the following expression,

$$g(x_{i,t-1}) = \frac{1}{1+\mathrm{e}^{-x_{i,t-1}}}, \quad (18)$$

for $i = 1, \ldots, d_x$. The observation model is given by

$$\mathbf{y}_t = \mathbf{x}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \sigma_r^2 \mathbb{I}_{d_x}). \quad (19)$$

Given a set of observations $\mathbf{y}_1, \ldots, \mathbf{y}_T$, our goal was to estimate $\mathbf{x}_t$ for $t = 1, \ldots, T$. We considered two scenarios for the experiment: *(i)* additive Gaussian state noise, and *(ii)* additive heavy-tailed non-Gaussian state noise.

We note that all simulated datasets were based on the same coefficient matrix $\mathbf{A}$, which was randomly generated with 10% sparsity and $d_x = 25$ dimensions. The length of the time-series simulated was $T = 150$. The results were averaged over 5000 runs where $N = 50$ particles were used for each of the PF methods. Our performance metric was the cumulative sum of the mean squared error (MSE) over time. For each of the numerical simulations, we added a burn-in period of 50 time steps to remove dependence of the MSE on the initialization of the algorithm.

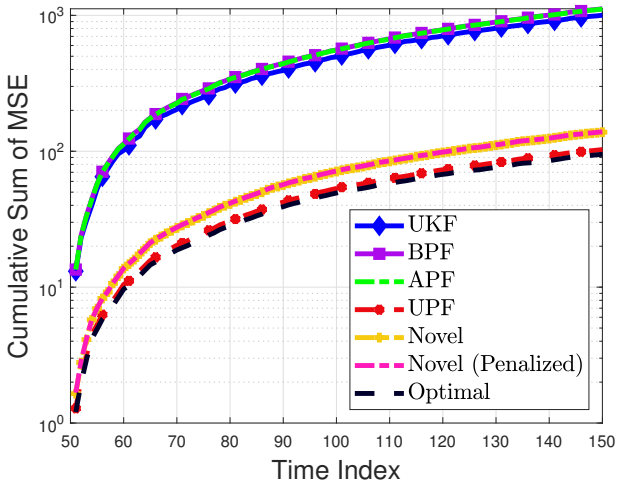

Fig. 1: Gaussian noise ($\sigma_q^2 = 10$ and $\sigma_r^2 = 1$).

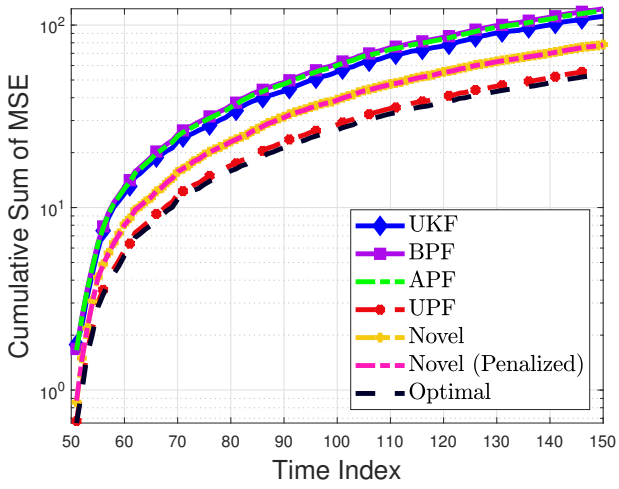

Fig. 2: Gaussian noise ($\sigma_q^2 = 1$ and $\sigma_r^2 = 1$).

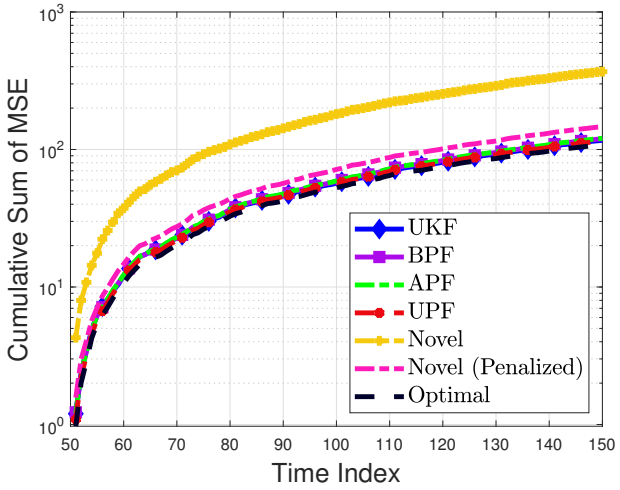

Fig. 3: Gaussian noise ($\sigma_q^2 = 1$ and $\sigma_r^2 = 10$).

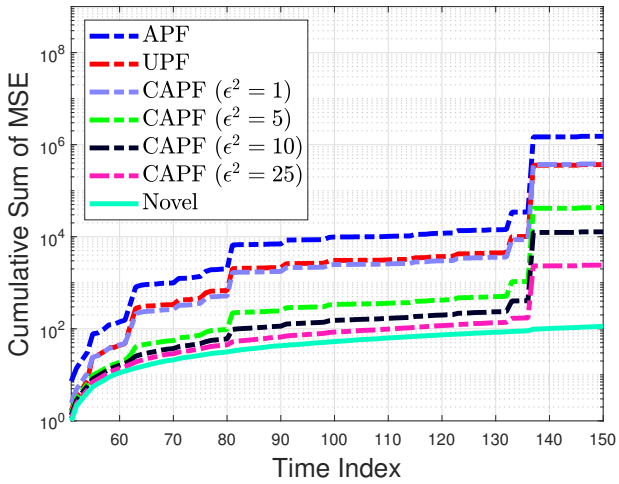

Fig. 4: Student-t noise ($\nu = 1$).

A. Additive Gaussian State Noise

We considered the state noise to be additive and Gaussian, $\mathbf{u}_t \sim \mathcal{N}(\mathbf{0}, \sigma_q^2 \mathbb{I}_{d_x})$, and performed a comparative study by varying the levels of state and observation noises. In particular, we tested three cases: *(a)* high state noise ($\sigma_q^2 = 10$ and $\sigma_r^2 = 1$), *(b)* moderate noise ($\sigma_q^2 = 1$ and $\sigma_r^2 = 1$), and *(c)* high observation noise ($\sigma_q^2 = 1$ and $\sigma_r^2 = 10$). For each case, we compared the performance of the novel method with the following state-of-the-art methodologies: UKF, bootstrap PF (BPF), auxiliary PF (APF) [14], and unscented PF (UPF) [15]. We also compared each of the methods to PF with the optimal proposal, which was attainable for this model.

Figure 1 shows the evolution of the cumulative MSE in the high state noise scenario. In this case, the novel method attains performance close to UPF and the optimal, while the other methodologies accumulated a large error. In the case of moderate noises, shown in Fig. 2, the novel methods still perform close to UPF and the optimal and significantly outperform the other methodologies. The results for the case of high observation noise are shown in Fig. 3. In this scenario, the observations are less informative of the ground truth, so methods like UKF, BPF, and APF were able to attain performance close to that of UPF and the optimal method. These results emphasize why the novel methodology without a regularization term performs poorly and a penalty on the perturbations is necessary to attain good performance.

B. Additive Heavy-Tailed State Noise

Finally, we considered a heavy-tailed, additive multivariate student-t state noise, $\mathbf{u}_t \sim t_\nu(\mathbf{0}, \mathbb{I}_{d_x})$, where ν corresponds to the degrees of freedom. We simulated synthetic data under the setting that $\nu = 1$ (corresponds to Cauchy noise) and $\mathbf{R} = \mathbb{I}_{d_x}$. We compared the performance of the novel method with APF, UPF and the original conjugate artificial PF method with arbitrary perturbations (CAPF) [11]. For the original CAPF we assumed $\mathbf{\Lambda}_t = \epsilon^2 \mathbb{I}_{d_x}$ and tested different settings for ϵ^2: *(a)* $\epsilon^2 = 1$, *(b)* $\epsilon^2 = 5$, *(c)* $\epsilon^2 = 10$, and *(d)* $\epsilon^2 = 25$.

Figure 4 shows the cumulative sum of the MSE averaged over 1000 simulations. The results indicate that the novel method outperforms APF, UPF and CAPF by several orders of magnitude, which is consistent with the theory explained in previous sections. Recall that in the case of Gaussian noise, UPF was able to attain close to optimal performance for all cases. Now, with heavy-tailed noise, the performance of UPF is much worse. The sudden accumulations of MSE for APF, UPF and CAPF can be explained by the methodologies' lack of robustness to tail events. Note that with the increase of ϵ^2, the performance of CAPF improves, but it never exceeds the performance of the novel method. This highlights the main advantage of the novel method, which is that it does not require specification of artificial noise parameters, making it more suitable for arbitrary problems.

V. Conclusions

In this work, we proposed a novel PF scheme which uses penalized perturbations to improve performance for high-dimensional systems. Simulation results indicated that the proposed methodology outperforms state-of-the-art filtering methodologies in high-dimensional scenarios for additive Gaussian and additive student-t state noise.

References

[1] M. Fasiolo, N. Pya, S. N. Wood, et al., "A comparison of inferential methods for highly nonlinear state space models in ecology and epidemiology," *Statistical Science*, vol. 31, no. 1, pp. 96–118, 2016.

[2] C. Rangel, J. Angus, Z. Ghahramani, M. Lioumi, E. Sotheran, A. Gaiba, D. L. Wild, and F. Falciani, "Modeling t-cell activation using gene expression profiling and state-space models," *Bioinformatics*, vol. 20, no. 9, pp. 1361–1372, 2004.

[3] J. R. Stroud, N. G. Polson, and P. Müller, "Practical filtering for stochastic volatility models," *State Space and Unobserved Components Models*, pp. 236–247, 2004.

[4] W. H. Aeberhard, J. Mills Flemming, and A. Nielsen, "Review of state-space models for fisheries science," *Annual Review of Statistics and Its Application*, vol. 5, pp. 215–235, 2018.

[5] A. H. Jazwinski, *Stochastic processes and filtering theory*, Courier Corporation, 2007.

[6] L. Ljung, "Asymptotic behavior of the extended kalman filter as a parameter estimator for linear systems," *IEEE Transactions on Automatic Control*, vol. 24, no. 1, pp. 36–50, 1979.

[7] 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 (Cat. No. 00EX373)*. Ieee, 2000, pp. 153–158.

[8] A. Doucet, N. de Freitas, and N. Gordon, "Sequential monte carlo methods in practice," .

[9] P. M. Djuric, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Miguez, "Particle filtering," *IEEE signal processing magazine*, vol. 20, no. 5, pp. 19–38, 2003.

[10] C. Snyder, "Particle filters, the optimal proposal and high-dimensional systems," in *Proceedings of the ECMWF Seminar on Data Assimilation for atmosphere and ocean*, 2011, pp. 1–10.

[11] A. Wigren, L. Murray, and F. Lindsten, "Improving the particle filter for high-dimensional problems using artificial process noise," *arXiv preprint arXiv:1801.07000*, 2018.

[12] L. Martino, V. Elvira, and F. Louzada, "Effective sample size for importance sampling based on discrepancy measures," *Signal Processing*, vol. 131, pp. 386–401, 2017.

[13] B. Jia and X. Wang, "Gene regulatory network inference by point-based gaussian approximation filters incorporating the prior information," *EURASIP Journal on Bioinformatics and Systems Biology*, vol. 2013, no. 1, pp. 16, 2013.

[14] M. K. Pitt and N. Shephard, "Filtering via simulation: Auxiliary particle filters," *Journal of the American statistical association*, vol. 94, no. 446, pp. 590–599, 1999.

[15] R. Van Der Merwe, A. Doucet, N. De Freitas, and E. A. Wan, "The unscented particle filter," in *Advances in neural information processing systems*, 2001, pp. 584–590.