


# Nudged Particle Filter with Optimal Resampling Applied to the Duffing Oscillator

Ryne Beeson 

*Mechanical and Aerospace Engineering  
Princeton University  
Princeton, New Jersey, USA  
ryne@princeton.edu*

Uwe Hanebeck 

*Intelligent Sensor-Actuator-Systems Laboratory (ISAS)  
Institute for Anthropomatics and Robotics  
Karlsruhe Institute of Technology (KIT)  
Karlsruhe, Germany  
uwe.hanebeck@kit.edu*

**Abstract**—Efficiently solving the continuous-time signal and discrete-time observation filtering problem for chaotic dynamical systems presents unique challenges in that the advected distribution between observations may encounter a separatrix structure that results in the prior distribution being far from the observation or the distribution may become split into multiple disjoint components. In an attempt to sense and overcome these dynamical issues, as well as approximate a non-Gaussian distribution, a nudged particle filtering approach has been introduced. In the nudged particle filter method a control term is added, but has the potential drawback of degenerating the weights of the particles. To counter this issue, we introduce an intermediate resampling approach based on the modified Cramér-von Mises distance. The new method is applied to a challenging scenario of the non-chaotic, unforced nonlinear Duffing oscillator, which possesses a separatrix structure. Our results show that it consistently outperforms the standard particle filter with resampling and original nudged particle filter.

## I. INTRODUCTION

The extension of particle filtering methods to complex systems, such as those encountered in the geosciences, has not yet been achieved. The core issue is the increased occurrence of particle degeneracy in not only high-dimensional systems, but also those that may be chaotic or turbulent, and may receive observations with a temporally sparse cadence. Classical studies on the degeneracy of the particle filter in high-dimensional systems include the work by Bengtsson et al. [1] and Synder et al. [2], where simply linear systems with additive Gaussian noise is used. The work of Synder et al. [3] has more recently looked at how the optimal proposal provides the best performance bounds to minimize degeneracy in the case of importance sampling particle filters.

With an aim toward generating samples from the optimal proposal using control theoretic techniques, Lingala et al. [4] introduced the nudged particle filter, which solves an optimal control problem to better advect particles toward the optimal proposal, given the future observation. We describe this approach in Section III and build on it to develop a new method described in Section IV. A potential drawback of the nudged method is that particles can degenerate during the advection to generate the prior due to excessive control. This was better understood in the work of Yeong et al. [5] and studied extensively on the Lorenz 1996 model in Beeson and Namachichivaya [6]. A simpler linear heuristic control

law with the same drawback was presented in van Leeuwen [7]. Although computationally more expensive than the simple control law of van Leeuwen [7], the nudged particle filter fully senses the nonlinear dynamics in the future forecast, enabling control that is sensitive to dynamical flow with separating behavior.

Another key advantage of the nudged particle filter for chaotic systems and those especially encountered in the geosciences, is the ability for the conditional distribution to remain near the dynamical attractor while generating samples approximating the optimal proposal. Methods that may move the distribution away from the attractor, such as Kalman approaches or recent flow-based particle methods, then require an advection time after the assimilation update to settle back onto the attractor and be physically realistic. The recent work of Zhou and Beeson [8] looked at the effectivity of projecting flow-based particle methods onto an attractor to maintain this property and Beeson [9] looked at the value of a hybrid approach using the nudged particle filter with a flow-based particle filter. Other approaches and ideas for effective particle filtering on high-dimensional systems have recently been surveyed by van Leeuwen et al. [10].

This paper contributes a new idea to the nudged particle filter by performing an intermediate resampling scheme using the modified Cramér-von Mises distance introduced by Hanebeck and Klumpp [11] and Eberhardt et al. [12]. The aim of this modification to the nudged particle filter is to remedy the degeneracy that may occur due to the control term. The original nudged particle filter maintained particle independence during advection between observations, whereas the intermediate resampling naturally results in correlation of the particles. Although this is a step back for proving possibly theoretical results, it enables a more continuous control action by the particles. We demonstrate this capability on a challenging low-dimensional stochastic nonlinear Duffing oscillator setup in Section V. The main result is that for a fixed number of particles, the new intermediate resampling nudged particle filter is able to consistently outperform both the standard and nudged particle filters to navigate the filtering distribution across separatrix boundaries and converge toward the true hidden signal.

We describe the dynamics of the Duffing oscillator test

problem in Section II, followed by the standard and nudged particle filters in Section III, the new importance resampling nudged particle filter in Section IV, the experimental results in Section V, and finally conclusions in Section VI.

## II. THE DUFFING OSCILLATOR

The unforced nonlinear Duffing oscillator is a one-degree of freedom Hamiltonian system, which we extend to the stochastic setting by driving it with a Brownian motion. The stochastic version is therefore given by the following stochastic differential equation (SDE),

$$d \begin{bmatrix} x_t \\ \dot{x}_t \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_t \\ \dot{x}_t \end{bmatrix} dt - \begin{bmatrix} 0 \\ x_t^3 \end{bmatrix} dt + \sigma dW_t, \quad (1)$$

where  $\sigma\sigma^* = 1\text{E-}3 \cdot \text{Id} \in \mathbb{R}^{2 \times 2}$ , with  $\text{Id}$  the identity matrix, and  $W$  is a two-dimensional standard Brownian motion (BM). The deterministic variant of (1) possesses three equilibrium points; one saddle equilibria at the origin, and two center equilibria at  $(-1, 0)$  and  $(1, 0)$ . The stable and unstable manifolds of the saddle equilibria generate homoclinic connections that separate the periodic motion about the center equilibria from each other, as well as from a periodic motion that exists in the remaining subset of the phase space. These realms of dynamical motion are shown in Fig. 1 and the separatrix structure from the saddle equilibria and its invariant manifolds will be exploited in the numerical experiments of Section V.

## III. STANDARD AND NUDGED PARTICLE FILTER

For an explanation of the standard particle filter (PF) with resampling, the nudged particle filter (nPF), and the new intermediate resampling nudged particle filter (IRnPF) to be introduced in Section IV, we consider the following general problem setup for a partially observable continuous-time signal, discrete-time observation system

$$\begin{aligned} dX_t &= f(X_t)dt + \sigma(X_t)dW_t, & X_0 &= x \in \mathbb{R}^m, \\ Y_{t_k} &= h(X_{t_k}) + \xi_{t_k}, & Y_0 &= 0 \in \mathbb{R}^d, \end{aligned} \quad (2)$$

where  $\xi_{t_k} \sim \mathcal{N}(0, \Sigma_y)$ . Standard assumptions on the independence of  $X_0 \perp (\xi_{t_k}) \perp W_t$  will be made. For brevity and to improve readability, we will often replace an element of the discrete-time set  $(t_k)$  with the corresponding index (e.g.,  $k$ ) and we use  $\mathcal{F}_k^Y$  to denote the  $\sigma$ -algebra of observations up to and including time  $t_k$ .

The filtering process follows the two recursive steps of: 1. generating the prior conditional distribution  $p_t(x|\mathcal{F}_{k-1}^Y)$  via the Chapman-Kolmogorov equation

$$p_t(x|\mathcal{F}_{k-1}^Y) = \int_{\mathbb{R}^m} p_{t|k-1}(x|z)p_k(dz|\mathcal{F}_{k-1}^Y), \quad t \in [t_{k-1}, t_k],$$

followed by; 2. the update of the posterior conditional distribution  $p_t(x|\mathcal{F}_k^Y)$  by Bayes' formula

$$p_t(x|\mathcal{F}_k^Y) \propto p_t(Y_k|x)p_t(x|\mathcal{F}_{k-1}^Y), \quad t = t_k.$$

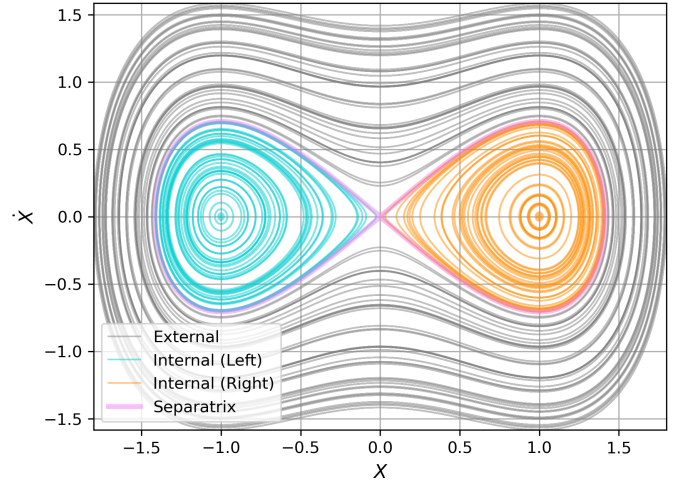


Fig. 1: Phase portrait for the deterministic unforced nonlinear Duffing oscillator, highlighting realms of dynamical significance.

### A. Standard Particle Filter with Resampling (PF)

The standard particle filter, introduced by Gordon et al. [13], with resampling solves the filtering problem as follows:

- (1) approximate the initial distribution by a sum of  $N$  equally weighted Dirac distributions that have support on the singletons  $(X_0^i \sim p_0)$  which are i.i.d.,

$$p_0^N(dx|Y_0) \equiv \sum_{i=1}^N w_0^i \delta_{X_0^i}(dx) = \frac{1}{N} \sum_{i=1}^N \delta_{X_0^i}(dx);$$

- (2) at time  $k-1$ , the prior is generated under the action of the Markov kernel  $P_X$  for the dynamics in (2),

$$p_k^N(dx|\mathcal{F}_{k-1}^Y) = p_{k-1}^N(dx|\mathcal{F}_{k-1}^Y)P_X = \sum_{i=1}^N w_{k-1}^i \delta_{X_k^i}(dx);$$

- (3) using the observation  $Y_k$ , the unnormalized weights are calculated as

$$\tilde{w}_k^i = w_{k-1}^i p_k(Y_k|X_k^i); \quad (3)$$

- (4) normalize the particle weights,

$$w_k^i = \tilde{w}_k^i / \sum_{i=1}^N \tilde{w}_k^i;$$

- (5) if the particle weights are degenerate, resample the particles and set all weights to be equal (i.e.,  $w_k^i = 1/N$ );
- (6) define the posterior at time  $k$  as

$$p_k^N(dx|\mathcal{F}_k^Y) \equiv \sum_{i=1}^N w_k^i \delta_{X_k^i}(dx),$$

and return to Step (2).

In this work, we use an approximation to the effective sample size, which was introduced in Bergman [14] and Liu and Chen [15], as the metric for Step (5) in deciding whether degeneracy

has occurred. Denoting the vector of normalized weights as  $w$ , the approximate effective sample size is defined as

$$N_{\text{eff}} \equiv 1/\langle w, w \rangle.$$

We use the threshold of  $N_{\text{eff}} < N/2$  to trigger resampling, and use the universal (also known as systematic) resampling approach of Kitagawa [16].

### B. Nudged Particle Filter (nPF)

As described in Section I, the nudged particle filter (nPF) aims to mitigate PF degeneracy in the chaotic and sparse temporal observation case by solving an optimal control problem that nudges particles independently toward the optimal proposal given the future observation. Hence, the nPF follows the same steps as the PF, but replaces the advection of particles in Step (2) of Section III-A with the following dynamical law

$$dX_t = f(X_t)dt + u_t dt + \sigma(X_t)dW_t, \quad t \in [t_k, t_{k+1}), \quad (4)$$

where the control term  $u$  is the minimizer of the finite-horizon objective function

$$\begin{aligned} J(u; k, x; Y_{k+1}) \\ = \mathbb{E} \left[ \int_{t_k}^{t_{k+1}} \frac{1}{2} \langle u_s, R_s^{-1} u_s \rangle ds + g(X_{k+1}^{k,x}, Y_{k+1}) \right]. \end{aligned}$$

In this objective function,  $R^{-1}$  takes values in the space of symmetric positive definite matrices and  $g$  is a terminal cost function that penalizes the realization  $X_{k+1}^{k,x}$  from being far from the observation  $Y_{k+1}$ . In this work, we use the negative log of the observation likelihood for  $g$ . The superscript for  $X_{k+1}^{k,x}$  indicates a realization of (4) starting at the state  $x \in \mathbb{R}^m$  at time  $t_k$  and advected to time  $t_{k+1}$ .

The solution to this optimal control problem is given by the feedback control law solution

$$u(t, x) = -R_t \nabla_x V(t, x), \quad (5)$$

where  $\nabla_x V$  is the gradient of the value function that satisfies the Hamilton-Jacobi-Bellman equation

$$\begin{aligned} -\partial_t V(t, x) + H(t, x, \nabla_x V, \nabla_x^{\otimes 2} V) &= 0, \\ V(t_{k+1}, x) &= g(x, Y_{t_{k+1}}), \end{aligned}$$

and  $H$  is the control Hamiltonian

$$\begin{aligned} H(t, x, \nabla_x V, \nabla_x^{\otimes 2} V) = \\ \sup_{\nu} \left\{ -\langle f(x) + \nu, \nabla_x V \rangle - \frac{1}{2} \text{Tr} (R \nabla_x^{\otimes 2} V) - \frac{1}{2} \langle \nu, R^{-1} \nu \rangle \right\}. \end{aligned}$$

By using a log-transformation [17],  $V(t, x) = -\log \Phi(t, x)$ , the control solution takes the form

$$u(t, x) = \frac{1}{\Phi(t, x)} R_t \nabla_x \Phi(t, x).$$

Assuming the dispersion coefficient  $\sigma$ , is constant and independent of  $X_t$ , and then choosing  $R = \sigma \sigma^*$ , the evolution equations of  $\Phi$  and  $\nabla_x \Phi$  both become backward linear second

order parabolic partial differential equations. Therefore they have solutions given by the Feynman-Kac formulas

$$\Phi(t, x) = \mathbb{E} \left[ \exp(-g(\eta_{k+1}^{k,x}, Y_{k+1})) \right], \quad (6)$$

and

$$\begin{aligned} \nabla_x \Phi(t, x) &= -\mathbb{E} \left[ \exp(-g(\eta_{k+1}^{k,x}, Y_{k+1})) \right. \\ &\quad \left. \exp \left( \int_{t_k}^{t_{k+1}} \nabla_x f(\eta_s^{k,x}) ds \right) \nabla_x g(\eta_{k+1}^{k,x}, Y_{k+1}) \right], \quad (7) \end{aligned}$$

where  $\eta$  is a realization to an SDE with the same generator as the signal process in (2).

As alluded to in Section I, the drawback of the nPF is that the weights of the particles must be adjusted if control is performed. This is due to the fact that the controlled dynamics (4) are not the same as the original signal dynamics (2). The change in weight for a particle  $X^i$  is therefore given by a Radon-Nikodym derivative

$$\begin{aligned} \frac{d\mu^i}{d\hat{\mu}^i} &= \exp \left( - \int_{t_k}^{t_{k+1}} \langle v(s, X_s^i), dW_s \rangle \right. \\ &\quad \left. - \frac{1}{2} \int_{t_k}^{t_{k+1}} \langle v(s, X_s^i), v(s, X_s^i) \rangle ds \right), \quad (8) \end{aligned}$$

where  $v(s, X_s^i) = -\sigma^* \nabla_x V(s, X_s^i)$ , and the derivation follows from the results on Girsanov transformations. In contrast to (3), the unnormalized weights of the nPF are

$$\tilde{w}_k^i = \frac{d\mu^i}{d\hat{\mu}^i} w_{k-1}^i p_k(Y_k | X_k^i). \quad (9)$$

We direct the reader to Yeong et al. [5] for more details on the derivation and properties of the nPF.

The control solution for the nPF is continuous in time and requires the evaluation of the expectations in (6) and (7). An approximation must be made in both of these cases for numerical implementation. We therefore do the following in this work:

- (2a) we discretize the time interval  $[t_k, t_{k+1})$  into a uniform partition of  $M$  subintervals defined by the collection of times  $(t_{k(j)})_{j=1}^M$ ;
- (2b) at the start of each of these subintervals, a control solution satisfying (5) is computed using  $K$  realizations;
- (2c) the control solution is fixed for the control step interval  $[t_{k(j)}, t_{k(j+1)})$ , and the particle is advected under the dynamics of (4);
- (2d) the correction to the unnormalized weight is made according to (8);
- (2e) if the particle weight did not degenerate below some threshold, return to Step (2a) to solve for the next control step until the final time  $t_{k+1}$  is reached; otherwise if the particle did degenerate due to the control solution, advect the particle from  $t_{k(j)}$  until  $t_{k(j+1)}$  under the natural dynamics of (2) without control.

#### IV. INTERMEDIATE RESAMPLING NUGGED PARTICLE FILTER (IRNPF)

Each particle of the nPF calculates their control independently of all others and therefore the unnormalized weights during the intermediate time between observations can become degenerate relative to the original distribution's mean. A key contribution in this work is to introduce a deterministic optimal resampling after each control step, such that the advected distribution will have minimum particle weight variance throughout the advection (i.e.,  $w^i = 1/N$ ). The benefits of performing this resampling are at least twofold. On a local level, each particle has a weight that is reset to  $1/N$ , and therefore enables the ability to apply control in Step (2c) of the nPF algorithm without running into the degeneracy condition specified in Step (2e) during the next control step (i.e., any degeneration is isolated to a given control step). The new IRnPF is therefore able to continually apply control throughout the intermediate time to drive the particle toward the observation neighborhood based on the terminal constraint  $g$ . On a global level, the resampling enables the use of indirect information generated by each particle on the difficulty of the dynamics to be encountered, so that particles are resampled at locations with the expectation that they will have similar control efforts for the next control step, and hence a similar decrease in their unnormalized weights.

The resampling of particles is achieved by minimizing the modified Cramér-von Mises distance, which is defined in Definition IV.2. This distance was introduced by Hanebeck and Klumpp [11] and allows for the comparison of two probability distributions. The distance relies on the following definition for the localized cumulative distribution (LCD).

**Definition IV.1** (Localized Cumulative Distribution [11]). Given a probability distribution  $\mu$ , a width parameter  $b \in \mathbb{R}_+^n$ , and kernel function  $\mathcal{K} : \mathbb{R}^n \times \mathbb{R}^n \rightarrow [0, 1]$ , the corresponding Localized Cumulative Distribution (LCD) is defined as

$$F_\mu(\cdot; b, \mathcal{K}) : \mathbb{R}^n \longrightarrow [0, \infty),$$

$$(m) \longmapsto \int_{\mathbb{R}^n} \mathcal{K}(x - m, b) \mu(dx).$$

We assume that the kernel function is a separable Gaussian kernel, dependent on a scalar width in each directional component,

$$\mathcal{K}_b(x) \equiv \mathcal{K}(x; b) = \exp(-\|x\|_2^2 / 2b^2), \quad (10)$$

which then gives the modified Cramér-von Mises distance defined in the following manner:

**Definition IV.2** (Modified Cramér-von Mises Distance [11]). Let  $F_\mu$  and  $F_\nu$  be the LCD for two probability distributions  $\mu$  and  $\nu$ , both with kernel  $\mathcal{K}$ , and  $w : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  a weighting function. Then their modified Cramér-von Mises distance is defined as

$$D(\mu, \nu) \equiv \int_{\mathbb{R}} w(b) \int_{\mathbb{R}^n} (F_\mu(m; b, \mathcal{K}) - F_\nu(m; b, \mathcal{K}))^2 dm db.$$

In this work, we make use of an asymptotic approximation for  $D(\mu, \nu)$  between two Dirac mixtures as  $b \rightarrow \infty$  and the weight function is restricted to  $(0, \bar{b}]$  and chosen according to  $w(b) = 1/b^{n-1}$  by Eberhardt et al. [12] (see also Hanebeck [18]) in the following theorem.

##### Theorem IV.1

With  $\bar{b} \gg 1$ ,  $w(b) = 1/b^{n-1}$ ,  $\mu, \nu$  Dirac mixtures with support  $(x_i), (y_i)$  and weights  $(w_x^i), (w_y^i)$  respectively, and  $\mathcal{K}$  given according to (10),  $D(\mu, \nu)$  is approximately,

$$D(\mu, \nu) \approx \frac{\pi^{n/2}}{8} \left( \tilde{D}(\mu) - 2\tilde{D}(\mu, \nu) + \tilde{D}(\nu) \right) + \frac{\pi^{n/2}}{4} C(\bar{b}) \bar{D}(\mu, \nu), \quad (11)$$

where

$$\tilde{D}(\mu, \nu) \equiv \sum_{i=1}^N \sum_{j=1}^K w_x^i w_y^j \text{xlog}(\|x_i - y_j\|_2^2)$$

$$\tilde{D}(\mu) \equiv \tilde{D}(\mu, \mu), \quad \tilde{D}(\nu) \equiv \tilde{D}(\nu, \nu),$$

$$\bar{D}(\mu, \nu) \equiv \|\mathbb{E}_\mu[x] - \mathbb{E}_\nu[x]\|_2^2,$$

$$C(\bar{b}) \equiv \log(4\bar{b}^2) - \Gamma,$$

where  $\Gamma \approx 0.5772$  is the Euler gamma constant and  $\text{xlog}(s) \equiv s \log(s)$ .

##### A. The IRnPF Method

The IRnPF method, an extension of the nPF in Section III-B is now described. In particular, following Step (2e) of the nPF, we have the additional steps:

- (2f) compute the normalized weights for the controlled Dirac mixture  $p_{k(j)}^N$  and generate a new mixture  $\tilde{p}_{k(j)}^K$  with  $\gamma K = N^1$ ,  $\gamma \in \mathbb{N}$ , evenly weighted samples of the normalized distribution  $p_{k(j)}^N$ ;
- (2g) translate the samples of  $\tilde{p}_{k(j)}^K$  to have the same mean as  $p_{k(j)}^N$ ;
- (2h) using the support locations of  $\tilde{p}_{k(j)}^K$  as optimization parameters, minimize the modified Cramér-von Mises distance between  $p_{k(j)}^N$  and  $\tilde{p}_{k(j)}^K$ ;
- (2i) using the optimized support locations, multiply each particle  $\gamma$ -times and normalize the weights by  $1/\gamma$  to produce the resampled distribution  $p_{k(j)}^N$ ; return to Step (2a) of the nPF.

Note that the translation given in Step (2g) implies that  $\bar{D} = 0$  in (11) for the initial iteration of the optimization. In this work we do not apply the intermediate resampling after the last control step, since the universal resampling algorithm will be applied after Bayes' update if needed.

#### V. NUMERICAL EXPERIMENTS

In the experiments to follow, we use the following discrete-time observation process,

$$Y_{t_k} = \begin{bmatrix} x_{t_k} \\ \dot{x}_{t_k} \end{bmatrix} + \xi_{t_k}, \quad \xi_{t_k} \sim \mathcal{N}(0, 1\text{E-}2 \cdot \text{Id}),$$

<sup>1</sup>For the IRnPF we specify  $\gamma, K$  first and then  $N$  before applying the method, so that this relation holds.



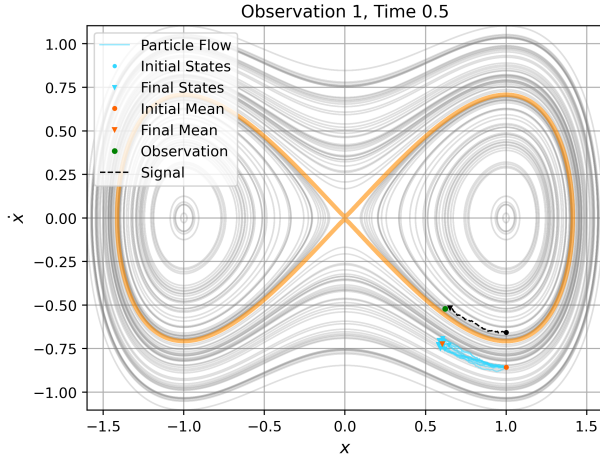


Fig. 2: Global view for a realization of the advected initial PF distribution with deterministic initial condition, separated from the true hidden signal state, over one observation interval.

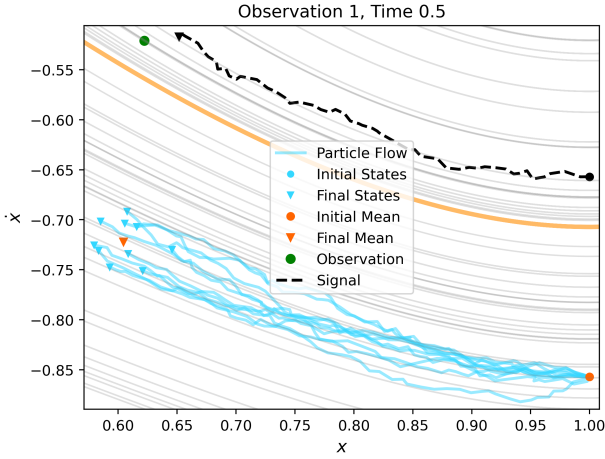


Fig. 3: Local view of Fig. 2.

with observations occurring every 0.5 time-units (TU) (i.e.,  $t_{k+1} - t_k = 0.5$  for all  $k$ ), and  $(\xi_{t_k}) \perp X_0 \perp (W_t)$ . Note that the nPF and IRnPF are not limited to the linear or fully observable case. In all simulations, we use a duration of 4.5 TU, and therefore 9 observations. All methods make use of the universal resampling approach if the effective sample size falls below 5 following the posterior update (Step (5) of the PF). An RK4Maruyama numerical integration scheme with a stepsize of  $1E-2$  is used for advection of particles.

Each particle filtering method uses 10 particles, and  $\gamma = 5$  for the IRnPF. Therefore  $N = 10$  for the PF and nPF, and for the IRnPF we have  $\gamma = 5, K = 10$  and hence  $N = 50$ . Note that the IRnPF will have a computational complexity on the same order as the nPF since the control is still only calculated for 10 particles for any of the control steps. The advection using 5 realizations for each control in the IRnPF enables a better approximation of the advected distribution, and hence improved resampling. The optimization at Step (2h) for the

IRnPF is accomplished using the `scipy.minimize` BFGS method with a (relative) tolerance of  $1E-3$ .

To demonstrate the qualitative behavior of the three methods, we choose deterministic initial conditions for the signal state at  $(x, \dot{x}) = (1, -0.657)$  and the initial filtering distribution a Dirac distribution at  $(x, \dot{x}) = (1, -0.857)$ . This choice results in the signal state being separated from the initial distribution by the deterministic stable manifold (separatrix) to the saddle equilibria. These initial conditions are shown in Fig. 2 and 3, which provides both a global and local (zoomed-in) view of the first observation interval for a PF simulation. The separatrix structure with invariant manifolds is shown in orange in these figures. The dispersion coefficient in (2) is such that the deterministic dynamics are not overpowered by the diffusion of the BM, and therefore forward in time, there is a high likelihood that the signal will remain in the *Internal (Right)* realm show in Fig. 1, the filtering distribution will remain in the *External* realm, and the two will diverge as they approach the saddle equilibria. The last observation period for this realization of the PF is shown in Fig. 4 and confirms that this is the typical behavior that occurs.

In Fig. 5 we show the final observation interval for the same simulation, but using the nPF. In particular, this simulation has the same initial conditions, the same signal path, and the same observations of the signal. The time between observations (0.5 TU) is partitioned into 50 equal control steps for the nPF, and  $K = 10$  realizations are used for each particle to calculate their control term in Step (2b). Comparing the final observation interval of Fig. 5 for the nPF to Fig. 4 for the PF is informative; it shows the nPF is lagging behind the PF in the natural dynamical flow. This implies that the nPF has been attempting to overcome the dynamical flow to move in the direction of the signal observation. This observation is further demonstrated in Fig. 7 by viewing the particle paths in the  $\dot{x}$  state over the simulation time, and coloring the particle paths according to their relative unnormalized weight during advection. Particles that are lightly shaded, tending toward white, is an indication that they have been actively using control to nudge themselves. As indicated in Step (2e) of the nPF algorithm, the particles are only allowed to nudge so much as to prevent possible full distributional degeneration, and hence control is only being applied for a fraction of the total simulation time.

Figure 6 shows the final observation interval of the IRnPF for this same case. In contrast to the nPF, the IRnPF is able to apply control more consistently throughout the simulation due to the intermediate resampling, and therefore is able to navigate the filtering distribution over the separatrix structure and approach the location and observations of the signal process. The improvement in the relative unnormalized weights of the particles is shown in Fig. 8. The final control steps of each observation interval shows some particles in light blue, which is expected because control is being used and as we stated after Step (2i) for the IRnPF algorithm, an intermediate resampling is not used on the final control step before an observation.

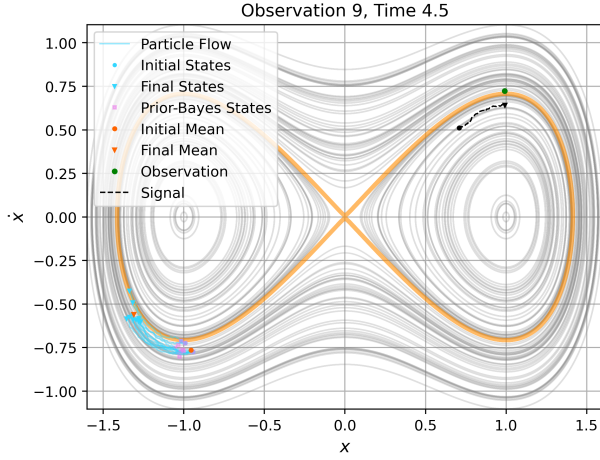


Fig. 4: Global view of the PF behavior during the last observation interval.

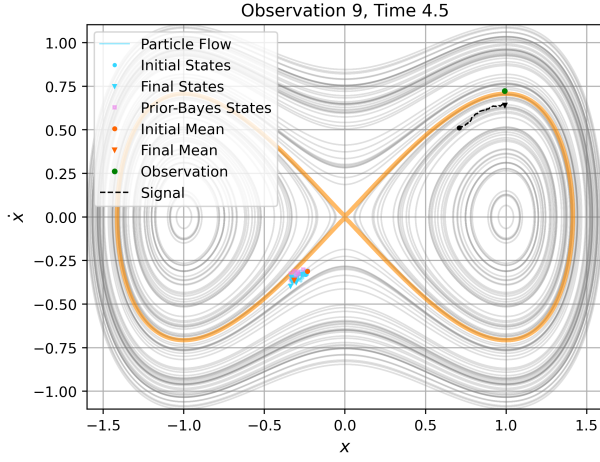


Fig. 5: Same as Fig. 4, but for the nPF.

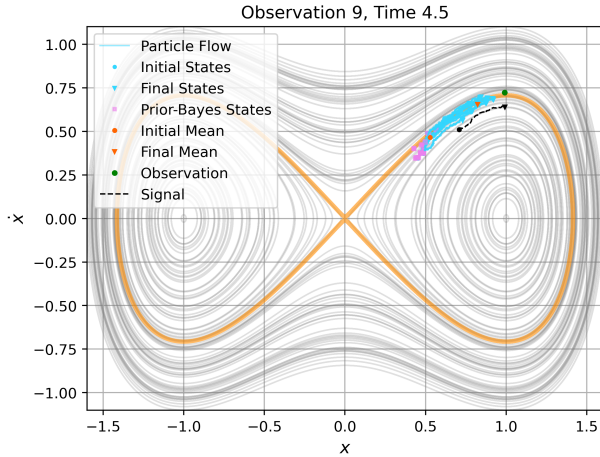


Fig. 6: Same as Fig. 4, but for the IRnPF.

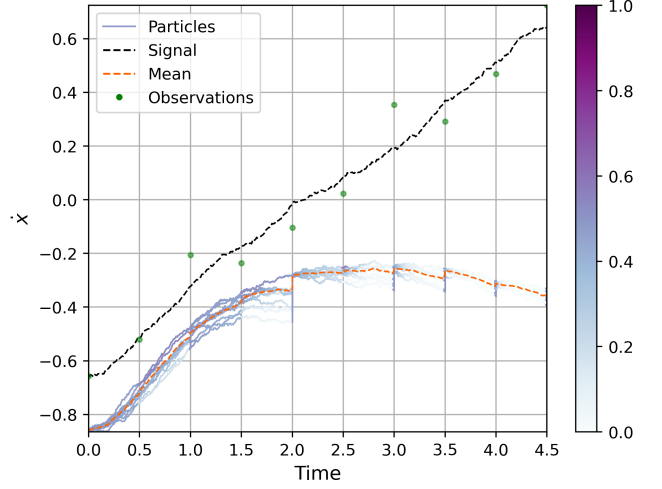


Fig. 7: The  $\dot{x}$  states versus time for particles of the nPF, with particle paths colored according to their relative unnormalized weights during advection. The figure illustrates the disadvantage of decreasing unnormalized weights during advection for the nPF; low relative weight (i.e., decreased relative to the mean) is shown in “white”, even relative weight in blue, and high relative weight in purple.

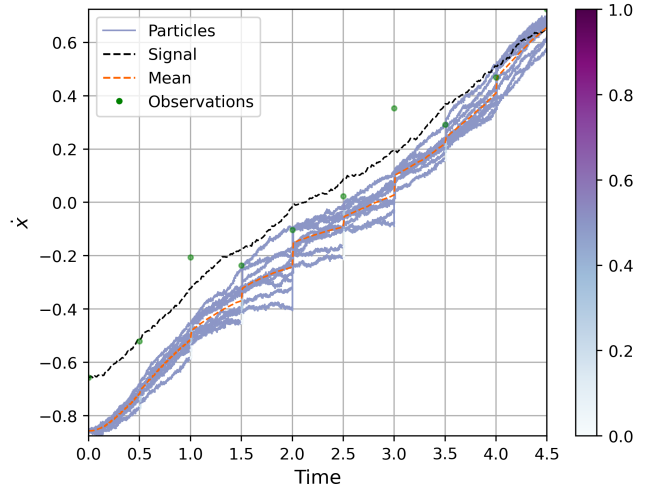


Fig. 8: Similar information as in Fig. 7, but for the IRnPF. One noticeable difference is that with the IRnPF, for each control solution,  $\gamma = 5$  advected realizations are produced. Hence the generation of the prior distribution has greater resolution at nearly the same computational cost.

TABLE I: Monte Carlo Analysis Results; Approximately Best Values for Each Criteria **Highlighted**.

Experiment Statistic	Filtering Method				
	PF <sub>1E1</sub>	PF <sub>1E2</sub>	PF <sub>1E3</sub>	nPF <sub>1E1</sub>	IRnPF <sub>1E1</sub> <sup><math>\gamma=10</math></sup>
Avg. RMSE	0.84	0.58	<b>0.42</b>	0.52	<b>0.40</b>
Avg. Min. RMSE	0.18	0.17	<b>0.16</b>	0.17	<b>0.16</b>
Avg. Max. RMSE	2.25	1.57	<b>0.91</b>	1.26	<b>0.87</b>
Avg. ( $N_{\text{eff}} / N$ )	0.34	0.25	0.26	0.40	<b>0.50</b>
Total Runtime (s)	4.3	19.5	178	814	1007

### A. Monte Carlo Results

A Monte Carlo (MC) analysis was performed to quantify and confirm the typical behavior of the PF, nPF, and IRnPF methods on the experimental setup just described. In the MC analysis, we use 20 different signal and observation history pairs, and for each pair we perform 20 different simulations. Therefore, a total of 400 simulations are produced for each method. To be clear, the benchmark of each method is on the same signal and observation pairs, so that direct comparison of performance is possible.

The main results are summarized in Table I. The *Total Runtime* is for all 400 simulations. We indicate the number of particles used in the PF method with the subscript (e.g., 1E1) and the number of nudged particles for the nPF and IRnPF similarly. The superscript for the IRnPF method indicates the multiplicative number of particles advected for each control (see Section IV-A). The root mean square error (RMSE) and the Min./Max. RSME for a given simulation realization  $\omega$  are defined as,

$$\text{RMSE}(X_t, p_t^N)(\omega) \equiv \frac{1}{4.5} \int_0^{4.5} \|X_s(\omega) - \mathbb{E}_{p_s^N(\omega)}[x]\|_2 ds,$$

$$\text{Min. RMSE}(X_t, p_t^N)(\omega) \equiv \min_{s \in [0, 4.5]} \|X_s(\omega) - \mathbb{E}_{p_s^N(\omega)}[x]\|_2,$$

where  $X_s$  is the true signal path for a given realization and Max. RMSE is defined in an analogous way. The Avg. of each of these metrics, as reported in Table I, is the average over the 400 simulations. Histograms<sup>2</sup> of the RMSE data in Table I is shown in Figs. 9, 10, and 11 to provide a greater understanding of the statistics of the MC runs.

The behavior and performance of the three methods for the example realization described in Section V and shown in Figs. 4 through 8 is apparent in the data of Table I. The lower Avg. Max. RMSE for the nPF<sub>1E1</sub> versus the PF<sub>1E1</sub> shows that on average the nPF is able to counter the dynamical flow sufficiently long to loiter near the stable manifold before slowly being forced along the unstable direction as shown in Fig. 5. Given that the initial RMSE is 0.2, the Avg. Min. RMSE for the PF<sub>1E1</sub> case shows that the PF with few particles is largely unable to ever get closer to the signal. Increasing the number of particles to the PF<sub>1E2</sub> case enables the standard PF to produce RMSE metrics nearer to that of the nPF<sub>1E1</sub>, but the nPF still has superior effective sample size. Of course, the main disadvantage of the nPF<sub>1E1</sub> is the computation complexity due to calculating the control. Increasing the number of particles further in the standard approach, the PF<sub>1E3</sub> case is able to outperform the nPF<sub>1E1</sub> in all RMSE criteria with a saving of about 4.5 times the computational effort.

The new approach proposed in this paper, the IRnPF, is competitive with the PF<sub>1E3</sub> case, even if only 10 particles are used. The IRnPF method retains superior effective sample size and has approximately the same computational complexity as the nPF. The nudging parameters for the nPF variants (nPF

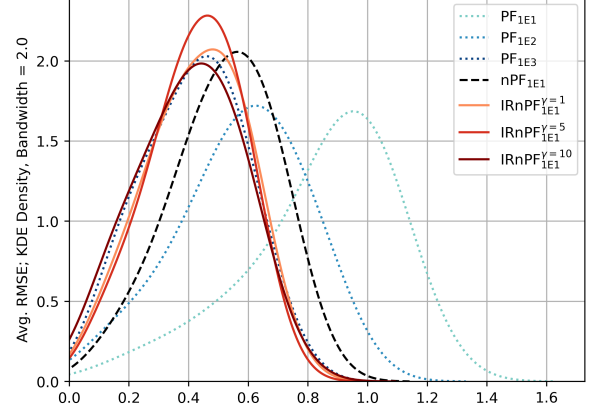


Fig. 9: Histogram for Avg. RMSE data shown in Table I.

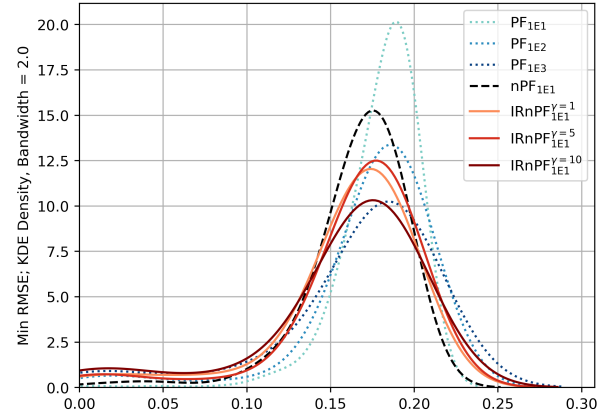


Fig. 10: Histogram for Min. RMSE data shown in Table I.

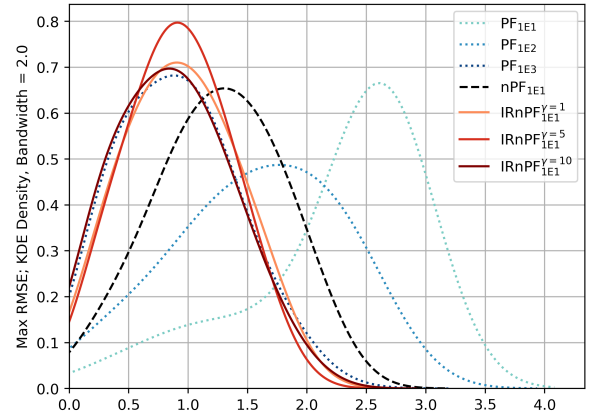


Fig. 11: Histogram for Max. RMSE data shown in Table I.

<sup>2</sup>Kernel Density Estimates (KDE)



and IRnPF) have not been optimized in these experiments and could potentially be more competitive. It is known that the number of particles for the standard PF scales exponentially with the dimension of the problem. Therefore there may be a crossover where the nPF and IRnPF methods are more advantageous in terms of RMSE, effective sample size, and computational complexity criteria, as the dimensionality of the problem with similar dynamical features is increased.

## VI. CONCLUSIONS

This paper has built on the optimal control approach of the nPF developed in Lingala et al. [4], Yeong et al. [5], and Beeson and Namachchivaya [6] for increasingly chaotic systems, by adding an intermediate optimal deterministic resampling approach using the modified Cramér-von Mises distance developed by Hanebeck and Klumpp [11] and Eberhardt et al. [12], to mitigate the isolated intermediate particle degeneracy condition of the nudged method. The result is a method that can more continuously apply control to adapt to difficult dynamical regimes such as the separating flow problem of Section V for the stochastic Duffing oscillator with erroneous deterministic initial condition. The new intermediate resampling nudged particle filter is consistently able to overcome the separating flow with only a few particles.

Ongoing improvements by the authors in tuning the nudged component and intermediate resampling may result in the method outperforming the PF, not only in estimation quality, but also in computational runtime, for similarly complex dynamical problems in higher dimensions. Improvements in the nudging optimal control theory are also being pursued.

One theoretical drawback of including the intermediate resampling is the potential loss of independence of the particles; this independence is also potentially lost after Bayes' update if resampling occurs due to the effective sample size decreasing below a user defined threshold.

## REFERENCES

- [1] T. Bengtsson, P. Bickel, and B. Li, "Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems," English, in *Probability and statistics: Essays in honor of David A. Freedman*. Beachwood, OH: IMS, Institute of Mathematical Statistics, 2008, pp. 316–315, ISBN: 978-0-940600-74-4. DOI: [10.1214/193940307000000518](https://doi.org/10.1214/193940307000000518).
- [2] C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson, "Obstacles to high-dimensional particle filtering," *Monthly Weather Review*, vol. 136, no. 12, pp. 4629–4640, 2008. DOI: [10.1175/2008MWR2529.1](https://doi.org/10.1175/2008MWR2529.1).
- [3] C. Snyder, T. Bengtsson, and M. Morzfeld, "Performance bounds for particle filters using the optimal proposal," *Monthly Weather Review*, vol. 143, no. 11, pp. 4750–4761, 2015. DOI: [10.1175/MWR-D-15-0144.1](https://doi.org/10.1175/MWR-D-15-0144.1).
- [4] N. Lingala, N. Perkowski, H. C. Yeong, N. S. Namachchivaya, and Z. Rapti, "Optimal nudging in particle filters," *Probabilistic Engineering Mechanics*, vol. 37, pp. 160–169, Jul. 2014. DOI: [10.1016/j.probengmech.2013.08.007](https://doi.org/10.1016/j.probengmech.2013.08.007).
- [5] H. C. Yeong, R. T. Beeson, N. S. Namachchivaya, and N. Perkowski, "Particle filters with nudging in multiscale chaotic systems: With application to the lorenz '96 atmospheric model," *Journal of Nonlinear Science*, vol. 30, no. 4, pp. 1519–1552, 2020. DOI: [10.1007/s00332-020-09616-x](https://doi.org/10.1007/s00332-020-09616-x).

- [6] R. Beeson and N. Sri Namachchivaya, "Particle filtering for chaotic dynamical systems using future right-singular vectors," *Nonlinear Dynamics*, vol. 102, pp. 679–696, 2020. DOI: [10.1007/s11071-020-05727-y](https://doi.org/10.1007/s11071-020-05727-y).
- [7] P. J. van Leeuwen, "Nonlinear data assimilation in geosciences: an extremely efficient particle filter," *Quart. J. Royal Meteor. Soc.*, vol. 136, pp. 1991–1999, Dec. 2010. DOI: [10.1002/qj.699](https://doi.org/10.1002/qj.699).
- [8] B. Zhou and R. Beeson, "Projected feedback particle filtering for chaotic dynamical systems using lyapunov vectors," in *2024 27th International Conference on Information Fusion (FUSION)*, 2024, pp. 1–8. DOI: [10.23919/FUSION59988.2024.10706376](https://doi.org/10.23919/FUSION59988.2024.10706376).
- [9] R. Beeson, "A hybrid controlled particle filter for chaotic systems with sparse observations," in *11th European Nonlinear Dynamics Conference 2024*, Delft, Netherlands, Jul. 2024.
- [10] P. J. van Leeuwen, H. R. Künsch, L. Nerger, R. Potthast, and S. Reich, "Particle filters for high-dimensional geoscience applications: A review," *Quarterly Journal of the Royal Meteorological Society*, vol. 145, no. 723, pp. 2335–2365, 2019. DOI: [10.1002/qj.3551](https://doi.org/10.1002/qj.3551).
- [11] U. D. Hanebeck and V. Klumpp, "Localized cumulative distributions and a multivariate generalization of the cramer-von mises distance," in *2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2008, pp. 33–39. DOI: [10.1109/MFI.2008.4648104](https://doi.org/10.1109/MFI.2008.4648104).
- [12] H. Eberhardt, V. Klumpp, and U. D. Hanebeck, "Optimal dirac approximation by exploiting independencies," in *Proceedings of the 2010 American Control Conference*, 2010, pp. 1392–1398. DOI: [10.1109/ACC.2010.5530503](https://doi.org/10.1109/ACC.2010.5530503).
- [13] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings F - Radar and Signal Processing*, vol. 140, no. 2, pp. 107–113, 1993. DOI: [10.1049/ip-f-2.1993.0015](https://doi.org/10.1049/ip-f-2.1993.0015).
- [14] N. Bergman, "Recursive bayesian estimation: Navigation and tracking applications," Ph.D. dissertation, Linköping University, Linköping, Sweden, 1999. [Online]. Available: <https://www.rt.isy.liu.se/research/reports/Ph.D.Thesis/PhD579.pdf>.
- [15] J. S. Liu and R. Chen, "Sequential monte carlo methods for dynamical systems," *Journal of the American Statistical Association*, vol. 93, no. 433, pp. 1032–1044, 1998, ISSN: 01621459, 1537274X. DOI: [10.2307/2669847](https://doi.org/10.2307/2669847).
- [16] G. Kitagawa, "Monte carlo filter and smoother for non-gaussian nonlinear state space models," *Journal of Computational and Graphical Statistics*, vol. 5, no. 1, pp. 1–25, 1996, ISSN: 10618600. [Online]. Available: <http://www.jstor.org/stable/1390750>.
- [17] W. H. Fleming, "Logarithmic transformations and stochastic control," in *Advances in Filtering and Optimal Stochastic Control*, W. H. Fleming and L. G. Gorostiza, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 1982, pp. 131–141, ISBN: 978-3-540-39517-1. DOI: [10.1007/BFb0004532](https://doi.org/10.1007/BFb0004532).
- [18] U. D. Hanebeck, *at - Automatisierungstechnik*, vol. 63, no. 4, pp. 265–278, 2015. DOI: [doi:10.1515/auto-2015-0005](https://doi.org/10.1515/auto-2015-0005).