A New Computational Method for Nonlinear Filtering

Hongjiang Qian,* George Yin,† Qing Zhang†

In numerous applications, the signals emitted from a dynamical system are often corrupted by noise. To restore the original signal, one must remove the noise. This process is termed filtering. When the underlying dynamics are nonlinear, the process is nonlinear filtering. Many nonlinear noise-removal filters typically examine the input signal and use certain stochastic models to make inference and estimate the most likely value for the original signal. Such a design is known as the filtering problem in estimation and control theory. Nonlinear filtering enjoys many applications in control of partially observed systems, target tracking, signal processing, robotics control, financial engineering, and environment monitoring.

This article reports a new method that addresses the long-standing issue of solving computational nonlinear filtering problems. These problems have persisted for over 60 years. Unlike the conventional method of approximating conditional means or distributions, this new method employs deep filtering with adaptive learning rates, which offers significant computational advantages. One key advantage of this approach is that it generates robust computational results, even when parameters change or noise perturbations occur. This breakthrough could potentially revive the field of computational nonlinear filtering.

Nonlinear Filtering. Consider a multi-dimensional stochastic system with state $X(t)$ at continuous time t, where the precise information of $X(t)$ is not known but only noise-corrupted observations on $X(t)$ denoted by $Y(t)$ is available. The dynamics of (state, observation)= $(X(t), Y(t))$ can be described by a pair of stochastic differential equations

$$\begin{aligned} \mathrm{d}X(t) &= g(X(t))\mathrm{d}t + S(X(t))\mathrm{d}W(t), \\ \mathrm{d}Y(t) &= h(X(t))\mathrm{d}t + H(t)\mathrm{d}V(t). \end{aligned} \tag{1}$$

In the above, $g(\cdot)$ and $h(\cdot)$ are suitable vector-valued functions, $S(\cdot)$ and $H(\cdot)$ are suitable matrix-valued functions of their arguments, and $W(t)$ and $V(t)$ are independent noise processes (standard Brownian motions). Nonlinear filtering focuses on estimating $X(t)$ based on the information of the observation $Y(t)$ up to time t. Extensive work has been done in the literature, which can be traced back to 1960s; see [1, 2, 3, 4, 5, 6], among others. In 1964, Kushner [2] derived the nonlinear filtering equation satisfied by a normalized conditional density (using the Itô calculus). The equation is now referred to as the Kushner equation. Subsequently, Duncan [4], Mortensen [5], and Zakai [6] independently derived filtering equations for un-normalized conditional densities. The unnormalized equation is referred to as the Duncan-Mortensen-Zakai equation (DMZ equation for short). Regardless the equations for normalized or unnormalized conditional densities, they are partial differential equations of infinite dimensions. A typical approach is to first find a solution to the DMZ equation, then show that it is indeed the conditional distribution under uniqueness of the solution of the differential equation. In spite of the celebrated results that settled the matters of nonlinear filtering theoretically, finding the solutions in closed-form

*Department of Mathematics, University of Connecticut, Storrs, CT 06269, hongjiang.qian@uconn.edu, gyin@uconn.edu

†Department of Mathematics, University of Georgia, Athens, GA 30602, qz@uga.edu

is virtually impossible because of the nonlinearity. Thus, much effort has been devoted to finding feasible computational methods ever since. The main difficulty is that the computational schemes based on conditional distributions suffer from curse of dimensionality. Various efforts on numerical methods for solving these equations have been made. For example, Lototsky et al. [7] developed a spectral approach with the advantage of separation of computations involving the observations and a system of parameters; see [9] for additional references. During the years, although significant progress has been made, finding feasible procedures for nonlinear filtering and constructing computable algorithms of nonlinear filtering is still an extremely challenging task. In contrast to the theoretical work and the various attempts on finding conditional means or conditional distributions and their approximations, we do not attempt to approximate the conditional distribution but rather convert the infinite-dimensional problem to a finite-dimensional parameter optimization problem using a machine learning approach. In lieu of numerically solving the stochastic partial differential equations, Monte Carlo samples are generated for the system states and observations. Then the problem is recast as finding a "function" based on the observed random samples. The "function" is parameterized by the weights of the neural network and estimated by stochastic approximation method [8].

Filtering Algorithms Based on Machine Learning Approach. In view of the rapid development in the field of machine learning, we reformulate the task as a stochastic optimization problem. We extend the traditional diffusion setup and consider the so-called switching diffusion model. The model is as (1), but the f, g, S, and H all depend on a continuous-time Markov chain $\alpha(t)$ taking values in a finite set, which is used to model hybrid uncertainty not covered by the continuous states. Next, we present our approximation procedure.

(a) We discretize (1) by using stepsize $\delta > 0$ and the Euler-Maruyama method (see a classical treatment in Kloeden and Platen [10] for numerical solutions of diffusion equations) to obtain

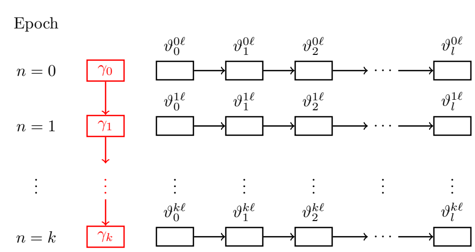
$$\begin{aligned} X_{n+1} &= X_n + \delta f(X_n, \alpha_n) + \sqrt{\delta}\, S(X_n, \alpha_n) W_n, \\ Y_{n+1} &= Y_n + \delta g(X_n, \alpha_n) + \sqrt{\delta}\, H(X_n, \alpha_n) V_n. \end{aligned} \tag{2}$$

(b) We set the number of training samples as N_{sample}, the training window size as n_0, and the total number of steps in the time horizon of the state and observation as N.

(c) We use Monte Carlo and (2) to generate data. For a sample point ω, use (2) to obtain (i) $\{Y_n(\omega), Y_{n+1}(\omega), \ldots, Y_{n+n_0-1}(\omega)\}$: input of the neural network with fixed sample point ω, $n = n_0, \ldots, N$, and (ii) $X_{n+n_0-1}(\omega)$: the corresponding state as the target.

(d) Then we use the least squares error between the target and calculated output as a loss function for the deep neural network training to generate a weight vector denoted by ϑ. Next, we use these weight vectors to another set of Monte Carlo samples of the actual dynamic system. We refer to the corresponding calculated DNN (Deep Neural Network)) output $\widetilde{X}_n$ as the deep filter.

We view estimates of X_n as a function of the observation $(Y_0, Y_1, \ldots, Y_n)$ and use the deep learning methods and the artificial neural network to find the function. In contrast to the existing machine learning literature, we propose a systematic approach. The strategy is while we are updating the parameter ϑ, we also adaptively update the learning rates. The updated learning rates are used in the next step parameter estimation for ϑ. We construct a pair of sequences of estimates for (ϑ, γ) in two loops. The estimate of ϑ requires more iterations, whereas the estimate of γ uses less frequent iterations.



The resulting algorithm is given in Algorithm 1. This algorithm can also be presented using a flow chat type diagram as below. As demonstrated in

Algorithm 1 Deep filtering with adaptive learning rates

function DF(ϑ_0^0, γ_0)
  for $n \leftarrow 0$ to Epoch **do**
    for $k \leftarrow 0$ to $\ell - 1$ **do**
      $\vartheta_{k+1}^{n\ell} = \vartheta_k^{n\ell} - \gamma_n \nabla_\vartheta J(\vartheta_k^{n\ell}, \xi_k^{n\ell})$
    $\vartheta_n^+ = \vartheta_\ell^{n\ell} - (\gamma_n + \Delta)\nabla_\vartheta J(\vartheta_\ell^{n\ell}, \xi_\ell^{n\ell})$
    $\vartheta_n^- = \vartheta_\ell^{n\ell} - (\gamma_n - \Delta)\nabla_\vartheta J(\vartheta_\ell^{n\ell}, \xi_\ell^{n\ell})$
    $\widehat{G}_n = (\chi(\vartheta_n^+, \zeta_n^+) - \chi(\vartheta_n^-, \zeta_n^-))/(2\Delta)$
    $\gamma_{n+1} = \gamma_n - \varepsilon \widehat{G}_n$

[9], under rather broad conditions, (a) γ_n converges to the optimal learning rate γ^*, (b) the approximation error bounds are obtained.

Numerical Examples. Here we briefly illustrate our approach by using a couple of examples. The details of the example, including various nonlinear functions used, the coefficients involved, and the numerical values can be found on GitHub at
https://github.com/hongjiang-qian/siam-news/blob/main/examples.pdf.
Note that we can show that the Euler-Maruyama approximations (X_n, α_n) and (Y_n, α_n), converge to the solution of the switching diffusions. Thus we will start with the discrete-time approximations rather than the original continuous-time systems.

Example 1 Consider a two-dimensional nonlinear system that involves sinusoidal nonlinearity and a Markov switching process. The state and observation variables x and y are both 2-dimensional. The computation results are given in Figure **??**. The performance in terms of relative errors is robust w.r.t. the initial learning rates.

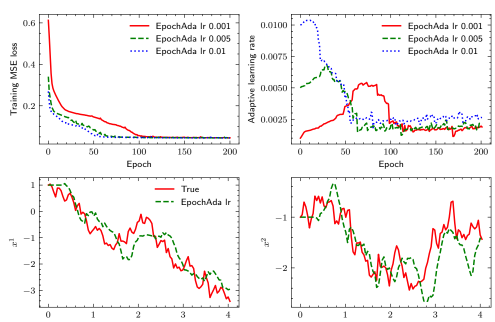


Figure 1: The training loss is presented on the top left and the path of adaptive learning rate is on the top right. The bottom figures plot the sample paths of out-of-sample state and the sample paths that of the deep filters.

Example 2 Motivated by a target tracking problem, consider a 6-dimensional dynamical system of a two-dimensional moving particle. In the observation equation, the first two components represent the distance and angle, respectively. Numerical results are displayed in Figure **??**.

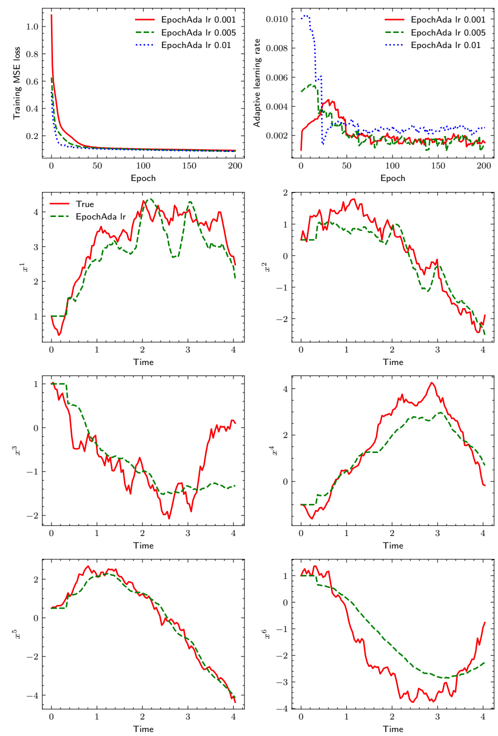


Figure 2: The training loss graph is on the top left and the path of adaptive learning rate is plotted on the top right. The sample paths of out-of-sample state and that of the deep filters are plotted in the rest of figures for each of the components.

Summary. This article presented a recent development on a new computational method to solve challenging nonlinear filtering problems. Let us make a few observations.

(i) In general, filtering involving random switching is a difficult problem. There is no easy way out not to mention any ready-made method from the computational tool box. Our deep filtering algorithm offers an effective procedure.

(ii) Although the algorithm targets nonlinear filtering, it is naturally applicable to linear systems. We have carried out numerical experiments for linear systems as well and compared them with the Kalman filters that are known to be optimal. We noted that the Kalman filters depend heavily on the nominal models used, whereas such dependence is less pronounced with deep filtering.

(iii) In nonlinear filtering, treating functions involving sinusoidal signals are more difficult and often the computation breaks down. Deep filters can treat such signals with no substantial difficulties.

(iv) Sparsity in the model can render undesirable output. In such a case, a recommendation is to use a deep filter together with some regularization procedures.

(v) In switching diffusions, if the Markov chain is also part of the state to be restored, a deep filter can be used to device a computational procedure. So far, the computational can be carried out, but the computational accuracy needs to be further improved.

Finally, the computational filtering method presented here, based on a deep learning approach, opens up new avenues for solving numerous problems in which computing and estimating unknown signals of systems are needed. What reported in this article uses a straight forward implementation. To improve the performance, to modify, and to improve the basic algorithms will be much welcomed.

References

[1] R. Stratonovich, Conditional markov processes, *Theor. Probab. Appl.* **5** (1960), 156–178.

[2] H. Kushner, On the differential equations satisfied by conditional probability densities of markov processes, with applications, *J. SIAM Control Ser. A* **2** (1964), 106–119.

[3] A. Shiryaev, On stochastic equations in the theory of conditional markov processes, *Theor. Probab. Appl.* **11** (1966), 179–184.

[4] T. Duncan, Probability Densities for Diffusion Processes with Applications to Nonlinear, Filtering Theory and Detection Theory, Ph.D. dissertation, Stanford Univ., 1967.

[5] R. Mortensen, Maximum-likelihood recursive nonlinear filtering. *J. Optim. Theory Appl.* **2** (1968), 386–394.

[6] M. Zakai, On the optimal filtering of diffusion processes. *Z. Wahrsch. Verw. Gebiete.* **11** (1969), 230–243.

[7] S. Lototsky, R. Mikulevicius, and B. Rozovskii, Nonlinear filtering revisited: A spectral approach. *SIAM J. Control Optim.* **35** (1997), 435–461.

[8] H. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, Springer, New York, 2nd ed., 2003.

[9] H. Qian, G. Yin, and Q. Zhang, Deep filtering with adaptive learning rates. *IEEE Trans. Automat. Control* (2022), 15 pp., DOI: 10.1109/TAC.2022.3183147.

[10] P. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, Berlin, 1992.