Examples in the SIAM News Article

We consider a nonlinear filtering problem for switching diffusions of the form

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

where we assume that $X(t) \in \mathbb{R}^d$ and $Y(t) \in \mathbb{R}^{d_1}$ are the state and observation vectors at time t, resp., $W(t)$ and $V(t)$ are d-dimensional and d_1-dimensional Brownian motions, $\alpha(t)$ is a continuous-time Markov chain taking values in a finite state space $\mathcal{M} = \{1,\dots,m_0\}$, $f : \mathbb{R}^d \times \mathcal{M} \mapsto \mathbb{R}^d$, $g : \mathbb{R}^{d_1} \times \mathcal{M} \mapsto \mathbb{R}^{d_1}$, $\Sigma : \mathbb{R}^d \times \mathcal{M} \mapsto \mathbb{R}^{d\times d}$, and $\Sigma_1 : \mathbb{R}^{d_1} \times \mathcal{M} \mapsto \mathbb{R}^{d_1\times d_1}$. We discretize (1) by using stepsize $\delta > 0$ and the Euler–Maruyama method, we obtain

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

Numerical Examples. The relative error of the state X_n and the deep filtering $\widetilde{X}_n$ is denoted by $\|X_n - \widetilde{X}_n\|$, which is defined as For a fixed ω, $X_n(\omega) = (X_n^l(\omega) : l \le d)$, and $\widetilde{X}_n(\omega) = (\widetilde{X}_n^l(\omega) : l \le d)$. The relative error is defined as:

$$
\|X_n - \widetilde{X}_n\| = \frac{\sum_{k=1}^{N_{\text{sample}}} \sum_{n=n_0}^{N} \sum_{l=1}^{d} |X_n^l(\omega_k) - \widetilde{X}_n^l(\omega_k)|}{\sum_{k=1}^{N_{\text{sample}}} \sum_{n=n_0}^{N} \sum_{l=1}^{d} (|X_n^l(\omega_k)| + |\widetilde{X}_n^l(\omega_k)|)}. \tag{3}
$$

Note that for the Euler-Maruyama approximation (X_n,α_n) and (Y_n,α_n), we can show that suitable interpolations of the sequences converge to the solution of the switching diffusions. Thus in what follows, 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, α_n a finite-state Markov chain taking values in $\{1,2\}$ whose transition matrix is

$$
I + \eta Q \ \text{ with } \ \eta = 0.04 \ \text{ and } \ Q = \begin{bmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix}.
$$

Consider system (2) with $X_n = (X_n^1, X_n^2)'$ (with z' denoting the transpose of z),

$$
\begin{aligned}
f(x,\alpha) &= \begin{bmatrix} \sin((0.3x^1 + 0.2x^2)\alpha) \\ \sin((0.2x^1 + 0.3x^2)\alpha) \end{bmatrix},\ S(x,\alpha) = \begin{bmatrix} 1 & -0.3 \\ 0 & 1 \end{bmatrix},\\
g(x,\alpha) &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x, \ \text{ and } \ H(x,\alpha) = \begin{bmatrix} 0.2 & 0.1 \\ 0 & 0.2 \end{bmatrix}.
\end{aligned}
$$

The initial state is $X_0 = (1,-1)'$. The relative errors of the deep filter with adaptive learning rates are $\|X_n - \widetilde{X}_n\| = 0.1128$ when $\rho_0 = 0.001$, $\|X_n - \widetilde{X}_n\| = 0.1126$ when $\rho_0 = 0.005$, and $\|X_n - \widetilde{X}_n\| = 0.1112$ when $\rho_0 = 0.01$. These relative errors display robustness of the deep filters w.r.t. the initial learning rates.

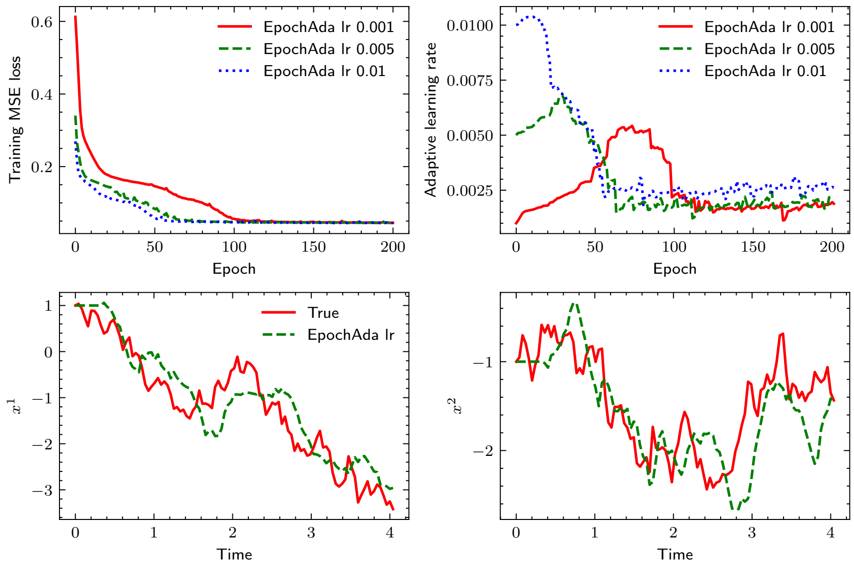

Figure 1: Example 1: the training loss is presented on the top left and the path of adaptive learning rate is on the top right. The bottom figures plot the sample paths of out-of-sample state and the sample paths that of the Deep filters with initial learning rate $\rho_0 = 0.002$.

Example 2 Denote state and observation variables and the corresponding sequences of state and observation vectors by $x = (x^i, i \le 6)'$, $y = (y^i, i \le 6)'$, $X_n = (X_n^i, i \le 6)'$, and $Y_n = (Y_n^i, i \le 6)'$, and W_n and V_n are the 6-dimensional vector-valued normal random variables with mean 0 and covariance matrix I_6 (6-dimensional identity). Consider system (2) with $f(x,\alpha) = \widetilde{F}x$, $S(x,\alpha) = \sigma(\alpha)I_6$, $g(x,\alpha) = h(x)$, and $H(x,\alpha) = \sigma_1(\alpha)I_6$, where

$$\widetilde{F} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & -\gamma_1^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -\gamma_2^2 & 0 \end{bmatrix}, \text{ and}$$

$$h(x) = [\sqrt{(x^1)^2 + (x^4)^2}, \tan^{-1}(x^4/x^1), x^2, x^3, x^5, x^6]'.$$

Note that the model is motivated by certain tracking problem. The first two components of $h(x)$ represent the distance and angle, respectively. The specific form of $\widetilde{F}$ indicates the matrix is sparse, which in fact, makes the filtering problem more difficult. Additional regularization and modification steps can be used to improve the performance but this is beyond the scope of the current article. For numerical experiments, take $\gamma_1 = 0.5, \gamma_2 = 0.9$ and initial value $X_0 = (1, 0.5, 1, -1, 0.5, 1)'$. The step size $\delta = 0.04$ and the Markov chain α has a finite state $\{1,2\}$ with generator $Q = \begin{bmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix}$. In addition, we take $\sigma(1) = 0.8, \sigma(2) = 1.0$, and $\sigma_1(1) = 0.2, \sigma_1(2) = 0.2$. The relative errors of the deep filter with adaptive learning rates are $\|X_n - \widetilde{X}_n\| = 0.1792$ when $\rho_0 = 0.001$, $\|X_n - \widetilde{X}_n\| = 0.1746$ when $\rho_0 = 0.005$, and $\|X_n - \widetilde{X}_n\| = 0.1747$ when $\rho_0 = 0.01$. Similar to the first example, these relative errors are robust with respect to the initial learning rate. 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 with initial learning rate $\rho_0 = 0.002$ are plotted in the rest of figures for each of the components.

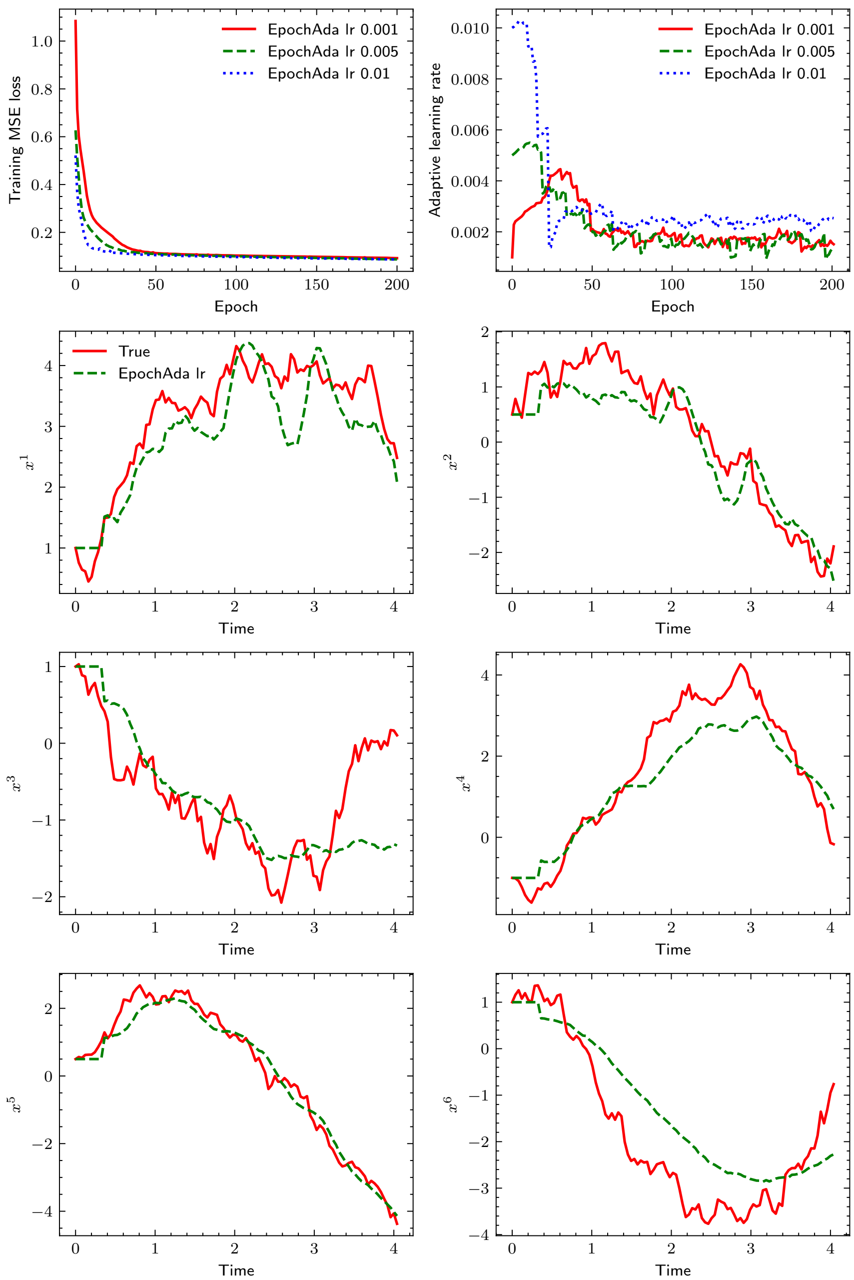

Figure 2: Figures for Example 2: the training loss graph is on the top left and the path of adaptive learning rate is plotted on the top right. The sample paths of out-of-sample state and that of the deep filters with initial learning rate $\rho_0 = 0.002$ are plotted in the rest of figures for each of the components.