Probabilistic Exponential Integrators — Appendix

A Proof of Proposition 1: Structure of the transition matrix

Proof of Proposition 1. The drift-matrix $A_{\text{IOUP}(d,q)}$ as given in Eq. (21) has block structure

$$A_{\text{IOUP}(d,q)} = \begin{bmatrix} A_{\text{IWP}(d,q-1)} & E_{q-1} \\ 0 & L \end{bmatrix}, \tag{31}$$

where $E_{q-1} \coloneqq \begin{bmatrix} 0 & \dots & 0 & I_d \end{bmatrix}^\top \in \mathbb{R}^{dq \times d}$. From Van Loan [47, Theorem 1], it follows

$$\Phi(h) = \begin{bmatrix} \exp\left(A_{\text{IWP}(d,q-1)}h\right) & \Phi_{12}(h) \\ 0 & \exp(Lh) \end{bmatrix}, \tag{32}$$

which is precisely Eq. (23). The same theorem also gives $\Phi_{12}(h)$ as

$$\Phi_{12}(h) = \int_0^h \exp(A_{\text{IWP}(d,q-1)}(h-\tau)) E_{q-1}^{(d-1)} \exp(L\tau) \, \mathrm{d}\tau. \tag{33}$$

Its ith $d \times d$ block is readily given by

$$\begin{aligned} (\Phi_{12}(h))_i &= \int_0^h E_i^\top \exp(A_{\text{IWP}(d,q-1)}(h-\tau)) E_{q-1} \exp(L\tau) \, \mathrm{d}\tau \\ &= \int_0^h \frac{(h-\tau)^{q-1-i}}{(q-1-i)!} \exp(L\tau) \, \mathrm{d}\tau \\ &= h^{q-i} \int_0^1 \frac{\tau^{q-1-i}}{(q-1-i)!} \exp(Lh(1-\tau)) \, \mathrm{d}\tau \\ &= h^{q-i} \varphi_{q-i}(Lh), \end{aligned} \tag{34}$$

where the second last equality used the change of variables $\tau = h(1-u)$, and the last line follows by definition. $\square$

B Proof of Proposition 2: Equivalence to a classic exponential integrator

We first briefly recapitulate the probabilistic exponential integrator setup for the case of the once integrated Ornstein–Uhlenbeck process, and then provide some auxiliary results. Then, we prove Proposition 2 in Appendix B.3.

B.1 The probabilistic exponential integrator with once-integrated Ornstein–Uhlenbeck prior

The integrated Ornstein–Uhlenbeck process prior with rate parameter L results in transition densities $Y(t+h) \mid Y(t) \sim \mathcal{N}\left(Y(t+h); \Phi(h)Y(t), Q(h)\right)$, with transition matrices (from Proposition 1)

$$\Phi(h) = \exp(Ah) = \begin{bmatrix} I & h\varphi_1(Lh) \\ 0 & \varphi_0(Lh) \end{bmatrix}, \tag{35}$$

$$Q(h) = \int_0^h \exp(A\tau) BB^\top \exp(A^\top \tau) \, \mathrm{d}\tau \tag{36}$$

$$= \int_0^h \begin{bmatrix} I & \tau\varphi_1(L\tau) \\ 0 & \varphi_0(L\tau) \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I & \tau\varphi_1(L\tau) \\ 0 & \varphi_0(L\tau) \end{bmatrix}^\top \mathrm{d}\tau \tag{37}$$

$$= \int_0^h \begin{bmatrix} \tau^2 \varphi_1(L\tau)\varphi_1(L\tau)^\top & \tau\varphi_1(L\tau)\varphi_0(L\tau)^\top \\ \tau\varphi_0(L\tau)\varphi_1(L\tau)^\top & \varphi_0(L\tau)\varphi_0(L\tau)^\top \end{bmatrix} \mathrm{d}\tau, \tag{38}$$

where we assume a unit diffusion $\sigma^2 = 1$. To simplify notation, we assume an equidistant time grid $\mathbb{T} = \{t_n\}_{n=0}^N$ with $t_n = n \cdot h$ for some step size h, and we denote the constant transition matrices simply by Φ and Q and write $Y_n = Y(t_n)$.

Before getting to the actual proof, let us also briefly recapitulate the filtering formulas that are computed at each solver step. Given a Gaussian distribution $Y_n \sim \mathcal{N}(Y_n; \mu_n, \Sigma_n)$, the prediction step computes

$$\mu_{n+1}^- = \Phi \mu_n, \tag{39}$$

$$\Sigma_{n+1}^- = \Phi(h) \Sigma_n \Phi(h)^\top + Q(h). \tag{40}$$

Then, the combined linearization and correction step compute

$$\hat{z}_{n+1} = E_1 \mu_{n+1}^- - f(E_0 \mu_{n+1}^-), \tag{41}$$

$$S_{n+1} = H \Sigma_{n+1}^- H^\top, \tag{42}$$

$$K_{n+1} = \Sigma_{n+1}^- H^\top S_{n+1}^{-1}, \tag{43}$$

$$\mu_{n+1} = \mu_{n+1}^- - K_{n+1} \hat{z}_{n+1}, \tag{44}$$

$$\Sigma_{n+1} = \Sigma_{n+1}^- - K_{n+1} S_{n+1} K_{n+1}^\top, \tag{45}$$

with observation matrix $H = E_1 - L E_0 = [-L \quad I]$, since we perform the proposed `EKL` linearization.

B.2 Auxiliary results

In the following, we show some properties of the transition matrices and the covariances that will be needed in the proof of Proposition 2 later.

First, note that by defining $\varphi_0(z) = \exp z$, the φ-functions satisfy the following recurrence formula:

$$z \varphi_k(z) = \varphi_{k-1}(z) - \frac{1}{(k-1)!}. \tag{46}$$

See e.g. Hochbruck and Ostermann [14]. This property will be used throughout the remainder of the section.

Lemma B.1. *The transition matrices $\Phi(h), Q(h)$ of the once integrated Ornstein–Uhlenbeck process with rate parameter L satisfy*

$$H\Phi(h) = [-L \quad I], \tag{47}$$

$$Q(h) H^\top = \begin{bmatrix} h^2 \varphi_2(Lh) \\ h \varphi_1(Lh) \end{bmatrix}, \tag{48}$$

$$H Q(h) H^\top = hI, \tag{49}$$

Proof.

$$H\Phi(h) = (E_1 - L E_0) \begin{bmatrix} I & h\varphi_1(Lh) \\ 0 & \varphi_0(Lh) \end{bmatrix} = [0 \quad \varphi_0(Lh)] - L [I \quad h\varphi_1(Lh)] = [-L \quad I]. \tag{50}$$

$$\begin{aligned}
Q(h) H^\top &= \int_0^h \begin{bmatrix} \tau^2 \varphi_1(L\tau)\varphi_1(L\tau)^\top & \tau \varphi_1(L\tau)\varphi_0(L\tau)^\top \\ \tau \varphi_0(L\tau)\varphi_1(L\tau)^\top & \varphi_0(L\tau)\varphi_0(L\tau)^\top \end{bmatrix} H^\top \, \mathrm{d}\tau && (51) \\
&= \int_0^h \begin{bmatrix} \tau \varphi_1(L\tau)\varphi_0(L\tau)^\top - L\tau^2 \varphi_1(L\tau)\varphi_1(L\tau)^\top \\ \varphi_0(L\tau)\varphi_0(L\tau)^\top - L\tau \varphi_0(L\tau)\varphi_1(L\tau)^\top \end{bmatrix} \mathrm{d}\tau && (52) \\
&= \int_0^h \begin{bmatrix} \tau \varphi_1(L\tau) \left(\varphi_0(L\tau)^\top - L\tau \varphi_1(L\tau)^\top\right) \\ \varphi_0(L\tau) \left(\varphi_0(L\tau)^\top - L\tau \varphi_1(L\tau)^\top\right) \end{bmatrix} \mathrm{d}\tau && (53) \\
&= \int_0^h \begin{bmatrix} \tau \varphi_1(L\tau) \\ \varphi_0(L\tau) \end{bmatrix} \mathrm{d}\tau && (54) \\
&= \begin{bmatrix} h^2 \varphi_2(Lh) \\ h \varphi_1(Lh) \end{bmatrix} && (55)
\end{aligned}$$

where we used $L\tau\varphi_1(L\tau) = \varphi_0(L\tau) - I$, and $\partial_\tau\left[\tau^k\varphi_k(L\tau)\right] = \tau^{k-1}\varphi_{k-1}(L\tau)$. It follows that

$$HQ(h)H^\top = H\begin{bmatrix} h^2\varphi_2(Lh) \\ h\varphi_1(Lh) \end{bmatrix} = h\left(\varphi_1(Lh) - Lh\varphi_2(Lh)\right) = hI, \tag{56}$$

where we used $L\tau\varphi_2(L\tau) = \varphi_1(L\tau) - I$. □

Lemma B.2. *The prediction covariance Σ^-_{n+1} satisfies*

$$\Sigma^-_{n+1}H^\top = Q(h)H^\top. \tag{57}$$

Proof. First, since the observation model is noiseless, the filtering covariance Σ_n satisfies

$$H\Sigma_n = \begin{bmatrix} 0 & 0 \end{bmatrix}. \tag{58}$$

This can be shown directly from the correction step formula:

$$\begin{aligned} H\Sigma_n &= H\Sigma^-_n - HK_nS_nK_n^\top & (59)\\ &= H\Sigma^-_n - H\left(\Sigma^-_n H^\top S_n^{-1}\right)S_nK_n^\top & (60)\\ &= H\Sigma^-_n - H\Sigma^-_n H^\top\left(H\Sigma^-_n H^\top\right)^{-1}S_nK_n^\top & (61)\\ &= H\Sigma^-_n - IS_nK_n^\top & (62)\\ &= H\Sigma^-_n - S_n\left(\Sigma^-_n H^\top S_n^{-1}\right)^\top & (63)\\ &= H\Sigma^-_n - S_nS_n^{-1}H\Sigma^-_n & (64)\\ &= \begin{bmatrix} 0 & 0 \end{bmatrix}. & (65) \end{aligned}$$

Next, since the observation matrix is $H = \begin{bmatrix} -L & I \end{bmatrix}$, the filtering covariance Σ_n is structured as

$$\Sigma_n = \begin{bmatrix} I \\ L \end{bmatrix}[\Sigma_n]_{00}\begin{bmatrix} I & L^\top \end{bmatrix}. \tag{66}$$

This can be shown directly from Eq. (58):

$$\begin{bmatrix} 0 & 0 \end{bmatrix} = H\Sigma = \begin{bmatrix} -L & I \end{bmatrix}\begin{bmatrix} \Sigma_{00} & \Sigma_{01} \\ \Sigma_{10} & \Sigma_{11} \end{bmatrix} = \begin{bmatrix} \Sigma_{10} - L\Sigma_{00} & \Sigma_{11} - L\Sigma_{01} \end{bmatrix}, \tag{67}$$

and thus

$$\begin{aligned} \Sigma_{10} &= L\Sigma_{00}, & (68)\\ \Sigma_{11} &= L\Sigma_{01} = L\Sigma_{10}^\top = L\Sigma_{00}L^\top. & (69) \end{aligned}$$

It follows

$$\Sigma = \begin{bmatrix} \Sigma_{00} & L\Sigma_{00} \\ \Sigma_{00}L^\top & L\Sigma_{00}L^\top \end{bmatrix} = \begin{bmatrix} I \\ L \end{bmatrix}\Sigma_{00}\begin{bmatrix} I & L^\top \end{bmatrix}. \tag{70}$$

Finally, together with Lemma B.1 we can derive the result:

$$\begin{aligned} \Sigma^-_{n+1}H^\top &= \Phi(h)\Sigma_n\Phi(h)^\top H^\top + Q(h)H^\top & (71)\\ &= \Phi(h)\begin{bmatrix} I \\ L \end{bmatrix}\bar{\Sigma}_n\begin{bmatrix} I & L^\top \end{bmatrix}\begin{bmatrix} -L^\top \\ I \end{bmatrix} + Q(h)H^\top & (72)\\ &= \Phi(h)\begin{bmatrix} I \\ L \end{bmatrix}\bar{\Sigma}_n \cdot 0 + Q(h)H^\top & (73)\\ &= Q(h)H^\top. & (74) \end{aligned}$$

□

B.3 Proof of Proposition 2

With these results, we can now prove Proposition 2.

Proof of Proposition 2. We prove the proposition by induction, showing that the filtering means are all of the form

$$\mu_n := \begin{bmatrix} y_n \\ Ly_n + N(\tilde{y}_n) \end{bmatrix}, \tag{75}$$

where $y_n, \tilde{y}_n$ are defined as

$$\begin{aligned} \tilde{y}_0 &:= y_0, &(76)\\ \tilde{y}_{n+1} &:= \varphi_0(Lh)y_n + h\varphi_1(Lh)N(\tilde{y}_n), &(77)\\ y_{n+1} &:= \varphi_0(Lh)y_n + h\varphi_1(Lh)N(\tilde{y}_n) - h\varphi_2(Lh)\left(N(\tilde{y}_n) - N(\tilde{y}_{n+1})\right). &(78) \end{aligned}$$

This result includes the statement of Proposition 2.

Base case $n = 0$ The initial distribution of the probabilistic solver is chosen as

$$\mu_0 = \begin{bmatrix} y_0 \\ Ly_0 + N(\tilde{y}_0) \end{bmatrix}, \Sigma_0 = 0. \tag{79}$$

This proves the base case $n = 0$.

Induction step $n \to n+1$ Now, let

$$\mu_n = \begin{bmatrix} y_n \\ Ly_n + N(\tilde{y}_n) \end{bmatrix} \tag{80}$$

be the filtering mean at step n and Σ_n be the filtering covariance. The prediction mean is of the form

$$\mu_{n+1}^- = \Phi(h)\mu_n = \begin{bmatrix} y_n + h\varphi_1(Lh)(Ly_n + N(\tilde{y}_n)) \\ \varphi_0(Lh)(Ly_n + N(\tilde{y}_n)) \end{bmatrix} = \begin{bmatrix} \varphi_0(Lh)y_n + h\varphi_1(Lh)N(\tilde{y}_n) \\ \varphi_0(Lh)(Ly_n + N(\tilde{y}_n)) \end{bmatrix}. \tag{81}$$

The residual $\hat{z}_{n+1}$ is then of the form

$$\begin{aligned} \hat{z}_{n+1} &= E_1\mu_{n+1}^- - f(E_0\mu_{n+1}^-) &(82)\\ &= \varphi_0(Lh)(Ly_n + N(\tilde{y}_n)) - f\left(\varphi_0(Lh)y_n + h\varphi_1(Lh)N(\tilde{y}_n)\right) &(83)\\ &= \varphi_0(Lh)(Ly_n + N(\tilde{y}_n)) - L\left(\varphi_0(Lh)y_n + h\varphi_1(Lh)N(\tilde{y}_n)\right) - N\left(\tilde{y}_{n+1}\right) &(84)\\ &= \varphi_0(Lh)Ly_n + \varphi_0(Lh)N(\tilde{y}_n) - L\varphi_0(Lh)y_n - Lh\varphi_1(Lh)N(\tilde{y}_n) - N\left(\tilde{y}_{n+1}\right) &(85)\\ &= \left(\varphi_0(Lh) - Lh\varphi_1(Lh)\right)N(\tilde{y}_n) - N\left(\tilde{y}_{n+1}\right) &(86)\\ &= N(\tilde{y}_n) - N\left(\tilde{y}_{n+1}\right), &(87)\\ & &(88) \end{aligned}$$

where we used properties of the φ-functions, namely $Lh\varphi_1(Lh) = \varphi_0(Lh)$ and the commutativity $\varphi_0(Lh)L = L\varphi_0(Lh)$. With Lemma B.2, the residual covariance S_{n+1} and Kalman gain K_{n+1} are then of the form

$$\begin{aligned} S_{n+1} &= H\Sigma_{n+1}^- H^\top = HQ(h)H^\top = hI, &(89)\\ K_{n+1} &= \Sigma_{n+1}^- H^\top S_{n+1}^{-1} = Q(h)H^\top\,(hI)^{-1} = \begin{bmatrix} h\varphi_2(Lh) \\ \varphi_1(Lh) \end{bmatrix}. &(90) \end{aligned}$$

This gives the updated mean

$$\begin{aligned} \mu_{n+1} &= \mu_{n+1}^- - K_{n+1}\hat{z}_{n+1} &(91)\\ &= \begin{bmatrix} \varphi_0(Lh)y_n + h\varphi_1(Lh)N(\tilde{y}_n) \\ \varphi_0(Lh)(Ly_n + N(\tilde{y}_n)) \end{bmatrix} - \begin{bmatrix} h\varphi_2(Lh) \\ \varphi_1(Lh) \end{bmatrix}\left(N(\tilde{y}_n) - N(\tilde{y}_{n+1})\right) &(92)\\ &= \begin{bmatrix} \varphi_0(Lh)y_n + h\varphi_1(Lh)N(\tilde{y}_n) - h\varphi_2(Lh)\left(N(\tilde{y}_n) - N(\tilde{y}_{n+1})\right) \\ \varphi_0(Lh)(Ly_n + N(\tilde{y}_n)) - \varphi_1(Lh)\left(N(\tilde{y}_n) - N(\tilde{y}_{n+1})\right) \end{bmatrix}. &(93) \end{aligned}$$

This proves the first half of the mean recursion:

$$E_0 \mu_{n+1} = \varphi_0(Lh) y_n + h\varphi_1(Lh) N(\tilde{y}_n) - h\varphi_2(Lh) \left(N(\tilde{y}_n) - N(\tilde{y}_{n+1})\right) = y_{n+1}. \tag{94}$$

It is left to show that

$$E_1 \mu_{n+1} = L y_{n+1} - N(\tilde{y}_{n+1}). \tag{95}$$

Starting from the right-hand side, we have

$$\begin{aligned}
& L y_{n+1} + N(\tilde{y}_{n+1}) && (96)\\
&= L\left(\varphi_0(Lh) y_n + h\varphi_1(Lh) N(\tilde{y}_n) - h\varphi_2(Lh)\left(N(\tilde{y}_n) - N(\tilde{y}_{n+1})\right)\right) + N(\tilde{y}_{n+1}) && (97)\\
&= \varphi_0(Lh) L y_n + Lh\varphi_1(Lh) N(\tilde{y}_n) - Lh\varphi_2(Lh)\left(N(\tilde{y}_n) - N(\tilde{y}_{n+1})\right) N(\tilde{y}_{n+1}) && (98)\\
&= \varphi_0(Lh) L y_n + (\varphi_0(Lh) - I) N(\tilde{y}_n) - (\varphi_1(Lh) - I)\left(N(\tilde{y}_n) - N(\tilde{y}_{n+1})\right) N(\tilde{y}_{n+1}) && (99)\\
&= \varphi_0(Lh)(L y_n + N(\tilde{y}_n)) - \varphi_1(Lh)(N(\tilde{y}_n) - N(\tilde{y}_{n+1})) && (100)\\
&= E_1 \mu_{n+1}. && (101)
\end{aligned}$$

This concludes the proof of the mean recursion and thus shows the equivalence of the two recursions. □

C Proof of Proposition 3: L-stability

We first provide definitions of L-stability and A-stability, following [26, Section 8.6].

Definition 1 (L-stability)**.** *A one-step method is said to be L-stable if it is A-stable and, in addition, when applied to the scalar test-equation $\dot{y}(t) = \lambda y(t)$, $\lambda \in \mathbb{C}$ a complex constant with $\mathrm{Re}(\lambda) < 0$, it yields $y_{n+1} = R(h\lambda) y_n$, and $R(h\lambda) \to 0$ as $\mathrm{Re}(h\lambda) \to -\infty$.*

Definition 2 (A-stability)**.** *A one-step method is said to be A-stable if its region of absolute stability contains the whole of the left complex half-plane. That is, when applied to the scalar test-equation $\dot{y}(t) = \lambda y(t)$ with $\lambda \in \mathbb{C}$ a complex constant with $\mathrm{Re}(\lambda) < 0$, the method yields $y_{n+1} = R(h\lambda) y_n$, and $\{z \in \mathbb{C} : \mathrm{Re}(z) < 0\} \subset \{z \in \mathbb{C} : R(z) < 1\}$.*

Proof of Proposition 3. Both L-stability and A-stability directly follow from Remark 1: Since the probabilistic exponential integrator solves linear ODEs exactly its stability function is the exponential function, i.e. $R(z) = \exp(z)$. A-stability and L-stability then follow: Since $\mathbb{C}^- \subset \{z : |R(z)| \leq 1\}$ holds the method is A-stable. And since $|R(z)| \to 0$ as $\mathrm{Re}(z) \to -\infty$ the method is L-stable. □

D Experiment details

D.1 Burger's equation

Burger's equation is a semi-linear partial differential equation (PDE) of the form

$$\partial_t u(x,t) = -u(x,t)\partial_x u(x,t) + D\partial_x^2 u(x,t), \qquad x \in \Omega, \quad t \in [0,T], \tag{102}$$

with diffusion coefficient $D \in \mathbb{R}_+$. We discretize the spatial domain Ω on a finite grid and approximate the spatial derivatives with finite differences to obtain a semi-linear ODE of the form

$$\dot{y}(t) = D \cdot L \cdot y(t) + F(y(t)), \qquad t \in [0,T], \tag{103}$$

with N-dimensional $y(t) \in \mathbb{R}^N$, $L \in \mathbb{R}^{N \times N}$ the finite difference approximation of the Laplace operator ∂_x^2, and a non-linear part F.

More specifically, we consider a domain $\Omega = (0,1)$, which we discretize with a grid of $N = 250$ equidistant locations, thus we have $\Delta x = 1/N$. We consider zero-Dirichlet boundary conditions, that is, $u(0,t) = u(1,t) = 0$. The discrete Laplacian is then

$$[L]_{ij} = \frac{1}{\Delta x^2} \cdot \begin{cases} -2 & \text{if } i = j, \\ 1 & \text{if } i = j \pm 1, \\ 0 & \text{otherwise.} \end{cases} \tag{104}$$

The non-linear part of the discretized Burger's equation results from another finite-difference approximation of the term $u \cdot \partial_x u$, and is chosen as

$$[F(y)]_i = \frac{1}{4\Delta x} \begin{cases} y_2^2 & \text{if } i = 1, \\ y_{d-1}^2 & \text{if } i = d, \\ y_{i+1}^2 - y_{i-1}^2 & \text{else.} \end{cases} \tag{105}$$

The initial condition is chosen as

$$u(x, 0) = \sin(3\pi x)^3 (1 - x)^{3/2}. \tag{106}$$

We consider an integration time-span $t \in [0, 1]$, and choose a diffusion coefficient $D = 0.075$.

D.2 Reaction-diffusion model

The reaction-diffusion model presented in the paper, with logistic reaction term, has been used to describe the growth and spread of biological populations [21]. It is given by a semi-linear PDE

$$\partial_t u(x, t) = D \partial_x^2 u(x, t) + R(u(x, t)), \qquad x \in \Omega, \quad t \in [0, T], \tag{107}$$

where $D \in \mathbb{R}_+$ is the diffusion coefficient and $R(u) = u(1 - u)$ is a logistic reaction term. We discretize the spatial domain Ω on a finite grid and approximate the spatial derivatives with finite differences, and obtain a semi-linear ODE of the form

$$\dot{y}(t) = D \cdot L \cdot y(t) + R(y(t)), \qquad t \in [0, T], \tag{108}$$

with N-dimensional $y(t) \in \mathbb{R}^N$, $L \in \mathbb{R}^{N \times N}$ the finite difference approximation of the Laplace operator, and the reaction term R is as before but applied element-wise.

We again consider a domain $\Omega = (0, 1)$, which we discretize on a grid of $N = 100$ points. This time we consider zero-Neumann conditions, that is, $\partial_x u(0, t) = \partial_x u(1, t) = 0$. Including these directly into the finite-difference discretization, the discrete Laplacian is then

$$[L]_{ij} = \frac{1}{\Delta x^2} \cdot \begin{cases} -1 & \text{if } i = j = 1 \text{ or } i = j = d, \\ -2 & \text{if } i = j, \\ 1 & \text{if } i = j \pm 1, \\ 0 & \text{otherwise.} \end{cases} \tag{109}$$

The initial condition is chosen as

$$u(x, 0) = \frac{1}{1 + e^{30x - 10}}. \tag{110}$$

The discrete ODE is then solved on a time-span $t \in [0, 2]$, and we choose a diffusion coefficient $D = 0.25$.