Companion Matrices for Differential Equations

We briefly describe companion matrices for turning n-order linear SDEs into first-order by representing the system as a linear operator on an augmented state variable.

Consider the 2-order system

$$a_0 x(t) + a_1 \frac{\mathrm{d}}{\mathrm{d}t} x(t) + a_2 \frac{\mathrm{d}^2}{\mathrm{d}t^2} x(t) = w(t)$$

Define a new variable, $z = \mathrm{d}x/\mathrm{d}t$, and substitute into the above equation:

$$a_0 x(t) + a_1 z(t) + a_2 \frac{\mathrm{d}}{\mathrm{d}t} z(t) = w(t)$$

This is now the 1-order system:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} x(t) &= z(t) \\ \frac{\mathrm{d}}{\mathrm{d}t} z(t) &= -\tilde{a}_0 x(t) - \tilde{a}_1 z(t) + a_2^{-1} w(t),\end{aligned}$$

where $\tilde{a}_0$ and $\tilde{a}_1$ are a_0/a_2 and a_1/a_2 respectively.

We can write this using a joint state, $\boldsymbol{x}(t) = \begin{bmatrix} x(t) \; z(t) \end{bmatrix}^\top$:

$$\frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{x}(t) = \begin{bmatrix} 0 & 1 \\ -\tilde{a}_0 & -\tilde{a}_1 \end{bmatrix} \boldsymbol{x}(t) + \begin{bmatrix} 0 \\ a_2^{-1} \end{bmatrix} w(t)$$

The companion matrix for linear n-order systems scales in a similar manner, with the final matrix consisting of a final row of scalars and off-diagonal band with value 1. The extension to non-linear systems is a straightforward extension here, simply replacing the matrix with a vector-valued function.

Unscented Transform

The unscented transform is a means for propagating a random variable, x through a non-linear functional, f, by optimally sampling about the mean and propagating each sample through f and combining the results as a weighted sum (Julier and Uhlmann, 1997). These so-called sigma points are defined as:

$$\chi_i = \begin{cases} \mathbb{E}[x] & i = 0 \\ \mathbb{E}[x] + \left[\sqrt{(n+\eta)\mathrm{cov}[x]}\right]_i & i = 1, \dots, n \\ \mathbb{E}[x] - \left[\sqrt{(n+\eta)\mathrm{cov}[x]}\right]_{n-i} & i = n+1, \dots, 2n \end{cases}.$$

Note that $[\cdot]_i$ indicates the i^{th} column of a matrix. n describes the dimension of the random variable x and η is a scaling parameter, defined such that $\eta = \alpha_\chi^2 (n + \kappa_\chi) - n$.

The unscented transform consists of transforming each sigma point, $\gamma_i = f(\chi_i)$ and constructing a weighted sum. The approximation of $y = f(x)$ is given by $y \sim \mathcal{N}(\mu, \Sigma)$, where

$$\begin{aligned}\mu &= \sum_{i=0}^{2n} \omega_i^{(m)} \gamma_i \\ \Sigma &= \sum_{i=0}^{2n} \omega_i^{(c)} \big(\mu - \gamma_i\big)\big(\mu - \gamma_i\big)^\top.\end{aligned}$$

The weights are defined by

$$\begin{aligned}\omega_0^{(m)} &= \eta (n + \eta)^{-1} \\ \omega_0^{(c)} &= \eta (n + \eta + 1 - \alpha_\chi^2 + \beta_\chi)^{-1} \\ \omega_i^{(m)} &= \omega_i^{(c)} = (2n + 2\eta)^{-1},\end{aligned}$$

where α, β, and κ are hyperparameters controlling the spread of sigma points. There are a number of reported settings for values of these hyperparameters, one such is $\alpha_\chi = 1$, $\beta_\chi = 0$, and $\kappa_\chi = n$ (Julier and Uhlmann, 1997). These are the values used in this paper.

Implementation

The implementation was written in TensorFlow, using TensorFlow probability. All experiments were optimised using Adam with a learning rate of 5e-3. For additional numerical stability, gradient clipping was performed based on the global norm.

The unscented filtering updates were implemented using the sigma-point dynamics as described in Sarkka (2007).

The number of flows for each experiment was set to $2d$, where d is the latent state dimension.