Supplementary material for Rapid Bayesian identification of sparse nonlinear dynamics from scarce and noisy data

Lloyd Fung, Urban Fasel, Matthew Juniper

September 27, 2024

1 Formula for the library of variance

To calculate the variances of each column of the library of polynomial terms, we follow the method of Taylor expansion for the moments of functions of random variables [1].

Given a random variable X with Gaussian distribution $X \sim \mathcal{N}(\mu, \sigma_x^2)$, the moments of the function $f(X)$ can be calculated by the Taylor expansion of the function around the mean μ. In particular, the power function $f(X) = X^n$ has the following moments:

$$E[X^n] = \sum_{i=0}^{n} \frac{\sigma_x^i}{i!} \frac{n!}{(n-i)!} \mu^{n-i} a_i, \tag{1}$$

where the coefficients a_i are given by

$$a_i = \begin{cases} \prod_{j=1}^{i/2} (2j-1) & \text{if } i \text{ is even} \\ 0 & \text{if } i \text{ is odd} \end{cases} \tag{2}$$

and

$$\sigma_{x^n}^2 = \mathrm{Var}[X^n] = E[X^{2n}] - (E[X^n])^2. \tag{3}$$

To calculate the variance of the weighted sum or product of two independent random variables X_1 and X_2, we use the formula

$$\sigma_{ax_1+bx_2}^2 = \mathrm{Var}[aX_1 + bX_2] = a^2\mathrm{Var}[X_1] + b^2\mathrm{Var}[X_2] \tag{4}$$

and

$$\sigma_{x_1x_2}^2 = \mathrm{Var}[X_1X_2] = E[X_1^2]E[X_2^2] - (E[X_1X_2])^2. \tag{5}$$

2 Selection of hyperparameters values

Section 3 compares the learning performances of SINDy (STLS), Bayesian-SINDy and `SparseBayes` in four examples. To ensure a fair competition between algorithms, we tuned the hyperparameters of each algorithm to maximise their performance in each given example. The resulting hyperparameter values used in each example are summarised in the tables 1 and 2.

Table 1: Hyperparameters used in each example in §3.

	Prior Variance α^{-1} (B-SINDy)	Weak Formulation	Finite Difference
Van der Pol	10^2	$\phi = (t^2 - 1)^2$, 9 points	8th order
Cubic Oscillator	1^2	$\phi = (t^2 - 1)^4$, 6 points	
Lorenz	25^2		12th order
Lynx-Hare	10^2		8th order

For SINDy (STLS), the threshold parameter λ is tuned in increments of the first significant figure to maximise the overall success rate at each noise level in the synthetic examples and tuned to recreate the Lotka-Volterra equation with the least number of spurious terms in the Lynx-Hare example (table 2). For Bayesian-SINDy, the choice for the prior variance α^{-1} is chosen to be of similar order to the expected parameter value. The values of the prior variance represent prior knowledge of the order of magnitude of the parameter value. In practice, this rough knowledge of the parameters' order of magnitude can usually be obtained from simple scaling arguments or physical intuition. Meanwhile, the `SparseBayes` algorithm has no hyperparameter that requires tuning, as both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are optimised by the algorithm.

Table 2: Optimal values for thresholding parameter λ in SINDy (STLS) in each example in §3.

	Noise level σ_x	Threshold Parameter λ
Van der Pol	0.1	0.4
	0.2	0.4
Cubic Oscillator	0.005	0.04
	0.01	0.06
Lorenz	0.025	0.2
	0.05	0.3
	0.1	0.4
	0.2	0.5
Lynx-Hare	N/A	0.025

The choices between weak formulation and finite difference are made according to the sampling frequency of the data compared to the typical frequency of the data. As per the discussion in §2(c), the weak formulation is preferred if the sampling frequency is found to be suitably higher than the typical frequency of the dynamics. Note that in practice, the application of weak formulation or finite difference should be decided according to the signal in the data before the learning takes place. Here, the weak formulation is used in the Cubic Oscillator example because it does not significantly distort the signal. In other examples, the finite difference is chosen because crucial high-frequency signals are filtered out when the weak formulation is used.