Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Identification of a Duffing oscillator using particle Gibbs with ancestor sampling

To cite this article: T J Rogers *et al* 2019 *J. Phys.: Conf. Ser.* **1264** 012051

Identification of a Duffing oscillator using particle Gibbs with ancestor sampling

T J Rogers[1], T B Schön[2], A Lindholm[2], K Worden[1] and E J Cross[1]

[1]Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield, Mappin Street, S1 3JD, UK

[2]Uppsala University, Department of Information Technology, Box 337, SE-751 05 Uppsala, Sweden

E-mail: trogers3@sheffield.ac.uk, thomas.schon@it.uu.se, andreas.lindholm@it.uu.se, k.worden@sheffield.ac.uk, e.j.cross@sheffield.ac.uk

Abstract. The Duffing oscillator remains a key benchmark in nonlinear systems analysis and poses interesting challenges in nonlinear structural identification. The use of particle methods or sequential Monte Carlo (SMC) is becoming a more common approach for tackling these nonlinear dynamical systems, within structural dynamics and beyond. This paper demonstrates the use of a tailored SMC algorithm within a Markov Chain Monte Carlo (MCMC) scheme to allow inference over the latent states and parameters of the Duffing oscillator in a Bayesian manner. This approach to system identification offers a statistically more rigorous treatment of the problem than the common state-augmentation methods where the parameters of the model are included as additional latent states. It is shown how recent advances in particle MCMC methods, namely the particle Gibbs with ancestor sampling (PG-AS) algorithm is capable of performing efficient Bayesian inference, even in cases where little is known about the system parameters *a priori*. The advantage of this Bayesian approach is the quantification of uncertainty, not only in the system parameters but also in the states of the model (displacement and velocity) even in the presence of measurement noise.

1. Introduction

The Duffing equation presented 100 years ago by George Duffing [1] remains one of the most studied equations in nonlinear dynamics. Part of the reason for its continued popularity is that despite the simplicity of the differential equation it gives rise to many interesting phenomena, including chaos. It has been possible to create systems which show Duffing like behaviour both mechanically [2, 3] and electronically, most notably in the *silverbox* benchmark dataset [4, 5]. The *silverbox* data has been studied extensively within the nonlinear system identification community [6, 7, 8, 9, 10].

A comprehensive introduction to be behaviour of the Duffing oscillator can be found in [11] where many of its main characteristics are detailed. This paper presents a Bayesian approach to the identification of a nonlinear system that is known to be well described by the Duffing equation in the presence of measurement noise. This identification task was attempted in [12] where a low level of measurement noise was added and only the parameter estimation was attempted. The contribution of this paper is to simultaneously estimate the smoothing distributions of the states (displacement and velocity of the oscillator) alongside the parameters of the Duffing oscillator. This is achieved by use of a particle Gibbs scheme with ancestor sampling and particle rejuvenation. The use of particle rejuvenation [13, 14] improves the convergence of the Gibbs sampler with a low number of particles, since the model is degenerate — there is no process noise on the displacement state, i.e. the covariance matrix of the process noise is singular.

The layout of the paper is as follows, a brief introduction to SMC methods for state-space models, including the Conditional Particle Filter, is given in Section 2. The procedure for Particle Gibbs with Ancestor Sampling is shown in Section 3, the use of a Particle Rejuvenation scheme is also shown for handling degenerate models. The application of this methodology to the identification of a nonlinear dynamic system, namely the Duffing oscillator, is presented in Section 4 where the smoothing distributions of the states and the distributions of the model parameters are recovered. Finally the results are discussed alongside directions for further work in Section 5.

2. Sequential Monte Carlo

In a Monte Carlo approach, distributions are approximated by a number of point masses; SMC methods [15] are a subset of this which exploit an evolving relationship between the distributions. Originally SMC algorithms, namely the bootstrap particle filter [16], were developed for tackling nonlinear Bayesian filtering problems which appear in nonlinear state-space models (SSMs). A general nonlinear SSM can be considered where,

$$x_t \sim f_\theta\left(x_t \,|\, x_{t-1}, u_{t-1}\right) \tag{1a}$$

$$y_t \sim g_\theta\left(y_t \,|\, x_t, u_t\right) \tag{1b}$$

here x_t is a vector of some hidden (latent) states at time t, the evolution of which is governed by $f_\theta\left(x_t \,|\, x_{t-1}, u_{t-1}\right)$. u_t is a vector of 'control' inputs to the model at time t, in structural dynamics this would generally be the force input to the oscillator. These states are related to a vector of observed variables y_t through the probabilistic model defined by $g_\theta\left(y_t \,|\, x_t, u_t\right)$. In this formulation $f_\theta\left(x_t \,|\, x_{t-1}, u_{t-1}\right)$ is the transition density of the model and $g_\theta\left(y_t \,|\, x_t, u_t\right)$ the observation density of the model. Both of these distributions have their dependence on the unknown model parameters θ explicitly denoted for clarity. The first distribution of interest is the filtering distribution of an SSM, which is given by Bayes theorem,

$$p_\theta\left(x_{1:t} \,|\, y_{1:t}\right) = \frac{g_\theta\left(y_t \,|\, x_t, u_t\right) p_\theta\left(x_t \,|\, y_{1:t-1}\right)}{p_\theta\left(y_t \,|\, y_{1:t}\right)} \tag{2a}$$

$$p_\theta\left(x_t \,|\, y_{1:t-1}\right) = \int f_\theta\left(x_t \,|\, x_{t-1}, u_{t-1}\right) p_\theta\left(x_{t-1} \,|\, y_{1:t-1}\right) \mathrm{d}x_{t-1} \tag{2b}$$

By restricting the forms of $f_\theta\left(x_t \,|\, x_{t-1}, u_{t-1}\right)$ and $g_\theta\left(y_t \,|\, x_t, u_t\right)$ it is possible to obtain closed form solutions to these equations which are the well known Kalman filter formulation [17]. However, the restrictions in the Kalman filter model — linear dynamics and observation equations, and Gaussian noise — are too restrictive for many systems encountered; this includes all structural systems with nonlinearity.

The solution which allows the use of more flexible nonlinear models in SMC is to use sequential importance sampling to approximate the filtering distribution. The filtering density (2b) is then approximated by,

$$p_\theta\left(x_t \,|\, y_{1:t}\right) \approx \frac{g_\theta\left(y_t \,|\, x_t, u_t\right)}{p_\theta\left(y_t \,|\, y_{1:t-1}\right)} \sum_{i=1}^{N} w_{t-1}^{i} f_\theta\left(x_t \,|\, x_{t-1}, u_{t-1}\right) \tag{3}$$

where, throughout this paper, the superscript notation is used for indexing, for example w_t^i denotes the importance weight of the i^{th} particle at time t. The notation adopted here is to use indexing similar to the software package Matlab where a subscript $1:T$ indicates the section of the indices of a vector from 1 to T inclusively. For matrices the same notation is adopted but with commas separating dimensions and following column major order. This is again approximated using importance sampling such that the (unnormalised) importance weights of the filtering density are given by,

$$\tilde{w}_t^i = \frac{g_\theta \left(y_t \,|\, x_t, u_t\right) \sum_{i=1}^{N} w_{t-1}^i f_\theta \left(x_t \,|\, x_{t-1}, u_{t-1}\right)}{\sum_{j=1}^{N} \nu_{t-1}^j q_\theta \left(x_t^i \,|\, x_{t-1}^j, y_t\right)} \tag{4}$$

it remains for the proposal density $q_\theta \left(x_t^i \,|\, x_{t-1}^j, y_t\right)$ and proposal weights ν_{t-1}^j to be chosen. In the simplest application — a bootstrap particle filter — it is set such that,

$$\sum_{j=1}^{N} \nu_{t-1}^j q_\theta \left(x_t^i \,|\, x_{t-1}^j, y_t\right) = \sum_{i=1}^{N} w_{t-1}^i f_\theta \left(x_t \,|\, x_{t-1}, u_{t-1}\right) \tag{5}$$

i.e. the proposal weight is set to be the previous particle weight and the proposal density is chosen to be the transition density. In this case, the unnormalised importance weights are given by $\tilde{w}_t^i = g_\theta \left(y_t \,|\, x_t, u_t\right)$, as the proposal cancels out the other term in (4). The procedure for running a bootstrap particle filter is shown in Algorithm 1.

Algorithm 1 Bootstrap Particle Filter

```
Initialisation:
    $i \leftarrow \{1, \dots, N\}$                                  ▷ For $N$ particles
    $x_1^i \sim p_\theta(x_1)$
    $\tilde{w}_1^i = g_\theta\left(y_1 \,|\, x_1^i, u_1\right)$
    $w_1^i = \frac{\tilde{w}_1^i}{\sum_{j=1}^{N} \tilde{w}_1^j}$                        ▷ Normalisation
For $t = 2, \dots, T$:
    Resampling: $a_t^i \sim \mathcal{MN}(\mathbf{w}_t)$              ▷ Sample from Multinomial
    Propagation: $x_t^i \sim f_\theta\left(x_t^i \,|\, x_{t-1}^{a_t^i}, u_{t-1}\right)$
    Weighting:
        $\tilde{w}_t^i = g_\theta\left(y_t \,|\, x_t^i, u_t\right)$
        $w_t^i = \frac{\tilde{w}_t^i}{\sum_{j=1}^{N} \tilde{w}_t^j}$
```

It is useful to collapse some of the steps in the filter down notationally to form a general SMC algorithm where a system of particles x_t^i (for $i = 1, \dots, N$ particles) are propagated through time with their ancestors a_t^i (the particle index from which this particle transitioned) by a proposal kernel $M_{\theta_t}(a_t, x_t)$. These particles are then assessed through a weighting function $W_{\theta,t}(x_{1:t})$, which calculates the normalised weights of a set of particles x_t^i.

In Algorithm 2 the additional step of recording the paths of each particle has been included in line 7. Here, each ancestral path $x_{1:t}^i$ is updated by concatenating the current particle position x_t^i with the path of the ancestor particle $x_{1:t-1}^{a_t^i}$. Here, a_t^i is shorthand for the ancestor of particle i at time t since all the ancestors are recorded in the vector a_t at time t.. This records the trajectory of that particle through time when tracing back through its ancestors. Although only a bookkeeping step, this will be crucial when it comes to forming an effective particle

Algorithm 2 General Sequential Monte Carlo

```
Initialisation:
    $i \leftarrow \{1, \ldots, N\}$                          ▷ For $N$ particles
    $x_1^i \sim p_\theta(x_1)$
    $w_1^i = W_{\theta,t}(\mathbf{x}_1)$
For $t = 2, \ldots, T$:
    $\{a_t^i, x_t^i\} \sim M_{\theta_t}(a_t, x_t)$
    $x_{1:t}^i = \left\{x_{1:t-1}^{a_t^i}, x_t^i\right\}$
    $w_t^i = W_{\theta,t}(\mathbf{x_{1:t}})$
```

Gibbs (PG) algorithm. The ancestral paths of every particle $i = 1, \ldots, N$ for time $t = 1, \ldots, t$ are represented by the bold notation $\mathbf{x_{1:t}}$, likewise the weights of every particle at time t is represented by $\mathbf{w}$, the vector of weights.

To make use of SMC within an MCMC scheme such as PG, it is necessary to make a slight modification to Algorithm 2 which ensures it is a valid Markov kernel [18]. This will be referred to here as the conditional particle Filter (CPF). In the CPF one of the particle trajectories is held constant as a reference trajectory $x'_{1:t}$. By convention, this is usually the N^{th} particle in the particle system. At every time step this particle is propagated forward in time as usual, however, the value corresponding to x'_t is not updated in the resampling step. This has the effect of guiding each run of the SMC through the state-space [13].

3. Particle Gibbs with Ancestor Sampling

The PG algorithm can be thought of conceptually as a Gibbs sampler for an SSM, where samples are drawn iteratively from the CPF, for the state trajectories conditioned on the parameters, and then from the conditional distributions of the parameters given the states. For a more thorough introduction, along with proofs of PG as a valid Markov kernel, the reader is directed to Andrieu *et al.* [18]. A simple yet powerful modification to the PG algorithm was proposed by Lindsten *et al.* [13] which they termed Particle Gibbs with Ancestor Sampling (PG-AS). In this construction, rather than fixing the ancestors of the reference trajectory $a_{1:t}^N$ to be N at every time step, the ancestor for x_T^N is resampled at each time step t.

This change helps to tackle the path degeneracy problem encountered in PG, where all particles share a common ancestor if looking far enough back in time. In PG this leads to the state values close to $t = 1$ not being resampled very often — i.e. there is poor mixing in the Markov chain. This will lead to slow convergence of the model. To achieve better convergence rates the ancestors of the model are sampled such that,

$$\mathbb{P}\left(a_t^N = i\right) \propto \tilde{w}_{t-1|T}^i = w_{t-1}^i \, p_\theta\left(x'_t \,|\, x_{t-1}^i, y_{1:t-1}\right) \tag{6}$$

where $p_\theta\left(x'_t \,|\, x_{t-1}^i, y_{1:t-1}\right)$ is the likelihood of the reference particle given the dynamics of all the particles $i = 1, \ldots, N$ at the previous time step — note that this includes the previous point in the reference trajectory. Conceptually, the ancestor is sampled based on which is the most likely parent for the reference trajectory at time t. Applying Bayes rule, this is proportional to the prior for that particle w_{t-1}^i multiplied by the likelihood that the reference x'_t was drawn from the transition density for each possible ancestor $p_\theta\left(x'_t \,|\, x_{t-1}^i, y_{1:t-1}\right)$.

This methodology has been shown to be effective in a number of system identification tasks, e.g. [19, 20]; however, it has been shown that if the model is nearly degenerate or degenerate then the benefit of ancestor sampling is greatly diminished. Ancestor sampling relies on sampling from the *backward kernel* of the SSM,

$$p_\theta\left(x_t \,|\, x_{t+1}, y_{1:t}\right) \propto f\left(x_{t+1} \,|\, x_t\right) p_\theta\left(x_t \,|\, y_{1:t}\right) \tag{7}$$

In a degenerate model, all of the probability mass can be centred on only, the ancestor relating to the reference trajectory. If this is the case only this ancestor will be sampled, i.e. $a_t^N = N$ for every time step t, and the algorithm returns to the standard PG formulation without ancestor sampling. Since the model for a structural dynamic system is degenerate due to the absence of process noise on the displacement state.

Lindsten *et al.* [14] propose a solution to this based on a modification to the target distribution of the Gibbs sampler which they term *particle rejuvenation*. By also resampling a part of the reference trajectory with the ancestors at each time step the degeneracy in the model can be avoided as the reference is '*loosened up*'. To introduce PG-AS with particle rejuvenation it is necessary to develop some additional notation: $\tilde{x}'_{t:T}$ is the future reference trajectory and Ξ is some subset of the future reference trajectory $\Xi \in \tilde{x}'_{t:T}$.

To cope with degeneracy in the model, the Gibbs sampler is partially collapsed over a subset of future state variables Ξ, such that $\Xi = \{x_t, \ldots, x_{\kappa_t}\}$ with $\kappa_t = \min\{T, t+\ell-1\}$. Since the goal is to resample both the ancestor a_t^N and part of the future reference trajectory Ξ it is necessary to sample from the joint PDF of (a_t, Ξ_t) where,

$$p_\theta\left(a_t, \Xi_t\right) \propto w_{t-1}^{a_t} f_\theta\left(x'_{\kappa_t+1} \,|\, x_{\kappa_t}\right) \left\{ \prod_{s=t+1}^{\kappa_t} f_\theta\left(x_s \,|\, s_{s-1}\right) g\left(y_s \,|\, x_s\right) \right\} f\left(x_t \,|\, x_{t-1}^{a_t}\right) g\left(y_t \,|\, x_t\right) \tag{8}$$

In general it will not be possible to sample from this PDF in closed form, therefore, a Markov kernel is chosen which generates valid samples from this PDF. Lindsten *et al.* [14] indicate that the choice of this kernel will be problem dependent, but a sensible choice within an SMC framework is to employ an importance sampling approach. A conditional importance sampling scheme is established where the unnormalised importance weights are given by (8). If ℓ is chosen to be one then (8) simplifies to,

$$p\left(a_t, \Xi_t \,|\, \ell = 1\right) \propto w_{t-1}^{a_t} f_\theta\left(x'_{t+1} \,|\, x_t\right) f\left(x_t \,|\, x_{t-1}^{a_t}\right) g\left(y_t \,|\, x_t\right) \tag{9}$$

The complete procedure for PG-AS with particle rejuvenation is shown in Algorithm 3, with a Markov kernel K_t to sample the new ancestor and Ξ_t.

It is clear from the algorithm that the methodology is very similar to that of the general SMC scheme presented in Algorithm 2. The value returned from the PG step (here the term is used to cover all methods that fall under this methodology) is a sample of the ancestral path for a particle, this is a sample of a path from the conditional smoothing distribution $p\left(x_{1:T} \,|\, y_{1:T}, \theta\right)$. To achieve this it is necessary to use a conditional particle filter to ensure validity of the Markov chain and the process of ancestor sampling and particle rejuvenation is used to help better mixing — i.e. more independent samples of the state trajectories are drawn.

PG-AS with particle rejuvenation allows efficient Gibbs sampling from the smoothing distribution of a nonlinear SSM. To perform inference; it is necessary to utilise this technique as part of a blocked Gibbs sampler where samples are drawn for the state trajectory conditioned on the parameters and then for the parameters conditioned on the state trajectory. The full procedure is shown in Algorithm 4, for notational convenience X' is used to represent the full sampled reference trajectory $x'_{1:T}$.

4. Identification of a Duffing Oscillator

This methodology is applied here to a Duffing oscillator to demonstrate its application to nonlinear system identification. There are four parameters to identify in the Duffing oscillator with displacement y forced by an input signal F,

Algorithm 3 Particle Gibbs with Ancestor Sampling and Particle Rejuvenation

```
**Initialisation**:
    Simulate $\Xi_1^\star \sim K_1(\Xi_1', \cdot)$
    Update $x'_{1:\kappa_t} \leftarrow \Xi_1^\star$
    Set $x_1^N \leftarrow x_1'$
    $x_1^i \sim p_\theta(x_1)$ for $i = 1, \dots, N-1$
    $w_1^i = W_{\theta,t}(\mathbf{x}_1)$ for $i = 1, \dots, N$
**For** $t = 2, \dots, T$:
    $\left\{a_t^i, x_t^i\right\} \sim M_{\theta_t}(a_t, x_t)$ for $i = 1, \dots, N-1$
    Simulate $\left(a_t^N, \Xi_t^\star\right) \sim K_1((N, \Xi_t'), \cdot)$
    Update $x'_{t:t+\kappa_t} \leftarrow \Xi_t^\star$
    $x_{1:t}^i = \left\{x_{1:t-1}^{a_t^i}, x_t^i\right\}$ for $i = 1, \dots, N$
    $w_t^i = W_{\theta,t}(\mathbf{x_{1:t}})$ for $i = 1, \dots, N$
Sample $k \sim \mathcal{MN}(\mathbf{w}_T)$
**return** $x'_{1:T} = x^k_{1:T}$
```

Algorithm 4 Blocked Gibbs Sampler for Inference in SSMs

```
Set $X_0'$ and $\theta_0$
Set $S$ as number of steps
**for** $s = 1, \dots, S$ **do**
    Sample $X_s'|\theta$ as in Algorithm 3
    Sample $\theta_s \sim p(\theta \mid X_s')$
**end for**
Discard first $s_b$ samples as burn-in
```

$$m\ddot{y} + c\dot{y} + ky + k_3 y^3 = F \tag{10}$$

these are the mass m, stiffness k, damping c, and cubic stiffness k_3. The system is simulated using a fifth order Runge-Kutta formulation [21], see Appendix A. One challenge in implementing a state-space approach to the identification of systems such as the Duffing oscillator is converting the continuous-time ordinary differential equation (ODE) into a discrete-time SSM for which the methods are developed. However, this can be solved in the same manner as the time-step integration used for simulation of nonlinear systems as the procedure is merely to produce a model for x_{t+1} given x_t where the state vector $x_t = [y\ \dot{y}]^\mathsf{T}$ as is common in application of numerical techniques to second-order ODEs. Therefore, if the same fifth-order Runge-Kutta scheme is used, the state transition density can be written down, assuming a Gaussian noise with unknown covariance across the states, as,

$$f_\theta(x_{t+1} \mid x_t) = \mathcal{N}(r(x_t),\ Q) \tag{11}$$

where $r(x_t)$ is the equation for the nonlinear propagation of the states, which is achieved with the same 5^{th} order Runge-Kutta scheme — although any time stepping method for an ODE could be used here provided it is stable. Considering the process noise covariance Q, for this model there is no process noise present so Q is set to be a diagonal matrix with small values on the diagonal to stabilise the model. This leads to a degenerate model which requires the use of the particle rejuvenation procedure in the PG-AS method [14]. Since the displacement of the

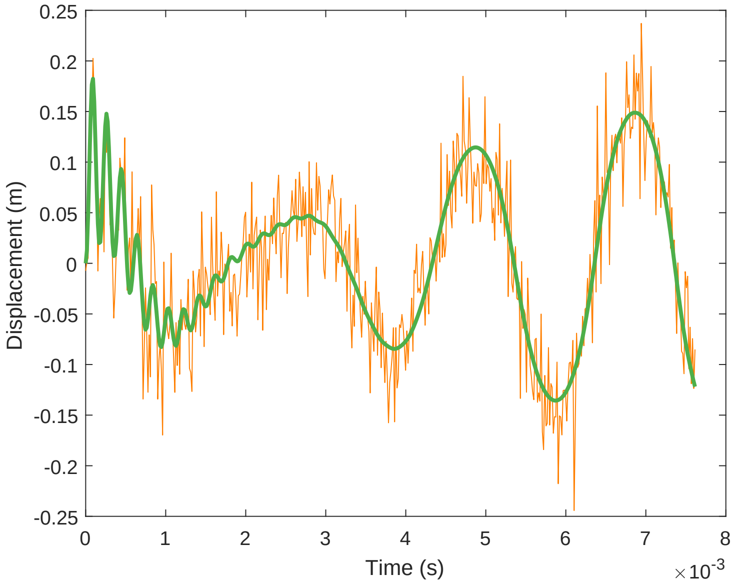


Figure 1. Figure showing the measured noisy signal in orange and the noise free true displacement in green.

oscillator is observed the observation density is a Gaussian with a linear mean which simply selects the first state as the observation. Therefore, the full SSM is given by,

$$x_t \sim \mathcal{N}\left(r(x_{t-1})\,,\ Q\right) \tag{12a}$$

$$y_t \sim \mathcal{N}\left(\begin{bmatrix}1 & 0\end{bmatrix} x_t\,,\ R\right) \tag{12b}$$

PG-AS with particle rejuvenation allows sampling of the state trajectories of the Duffing oscillator in order to estimate the smoothing distributions of the states — the displacement and velocity at each time step. However, it is still necessary to implement the distributions which allow sampling of the parameters of the model given the sampled state trajectories for each iteration of the Gibbs sampler. Once a state trajectory has been sampled the estimation of the parameters becomes a univariate Bayesian linear regression problem; since the quantities $\ddot{y}$, $\dot{y}$, y, and y^3 are now known. The system is approximated by a first order discretisation such that,

$$\dot{y}_{t+1} = \dot{y}_t + \Delta\left(\frac{1}{m}F_t - \frac{k}{m}y_t - \frac{c}{m}\dot{y}_t - \frac{k_3}{m}y_t^3\right) \tag{13a}$$

$$\dot{y}_{t+1} - \dot{y}_t = \Delta\left(\frac{1}{m}F_t - \frac{k}{m}y_t - \frac{c}{m}\dot{y}_t - \frac{k_3}{m}y_t^3\right) \tag{13b}$$

A design matrix can be constructed $X = \Delta\left[F_{1:T-1}, y_{1:T-1}, \dot{y}_{1:T-1}, y^3_{1:T-1}\right]$, and defining $\mathbf{y}' = \dot{y}_{2:T} - \dot{y}_{1:T-1}$ leads to the Bayesian linear regression problem,

$$\mathbf{y}' = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{14}$$

where $\boldsymbol{\varepsilon}$ is a vector in which each element $i = 1 : T-1$ is distributed $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(0\,,\ 1/\tau)$ and $\boldsymbol{\beta} = \left[\frac{1}{m}, \frac{k}{m}, \frac{c}{m}, \frac{k_3}{m}\right]^{\mathsf{T}}$. To perform Bayesian inference over the parameters conjugate priors are

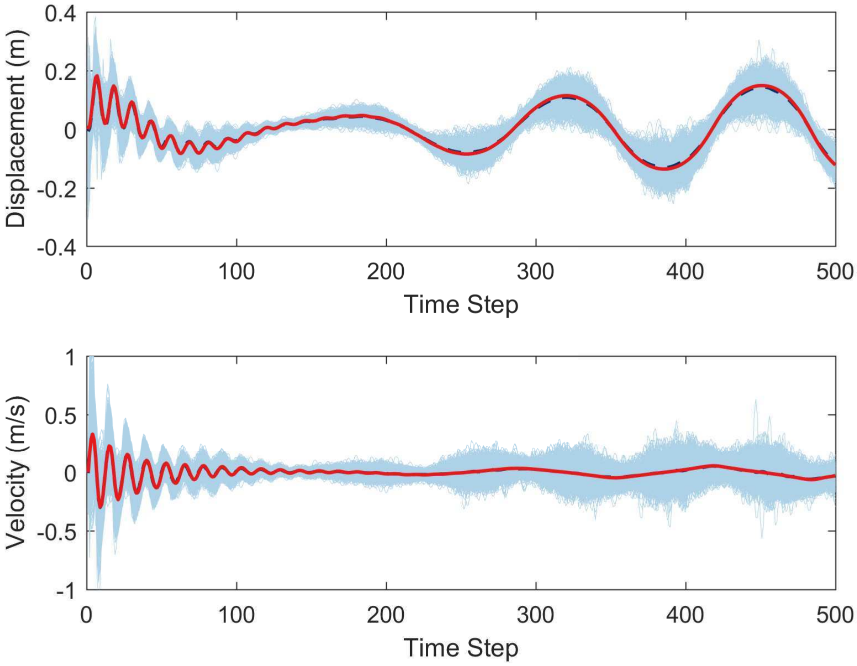


Figure 2. Plot showing sampled paths after the burn-in period in light blue. The mean of the sample paths is shown in dashed black and the true noise-free state trajectories are shown in red.

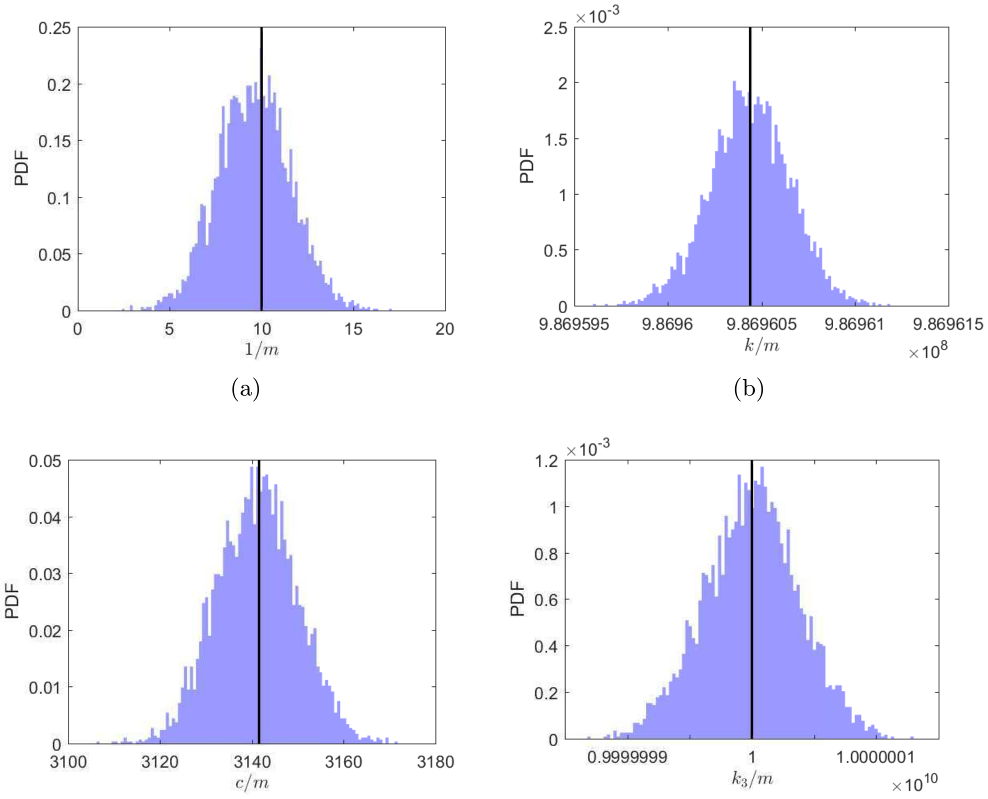
chosen which are Gaussian for each element of $\boldsymbol{\beta}$ and a Gamma distribution over τ. These are defined as,

$$\boldsymbol{\beta}_p \sim \mathcal{N}\left(\mu_p\,,\; 1/\tau_p\right) \qquad \tau \sim \mathcal{G}a\left(a, b\right) \tag{15}$$

where p indexes which element of $\boldsymbol{\beta}$ is being considered and $\mathcal{G}a$ represents the Gamma distribution. At each iteration of the Gibbs sampler, the sampled state trajectory X' is used to assemble the design matrix such that $X = \Delta\left[F_{1:T-1}, X'_{1:T-1,1}, X'_{1:T-1,2}, \left(X'_{1:T-1,1}\right)^3\right]$, sampling from the conditional distributions of the parameters is achieved by the standard results for Bayesian linear regression [22, 23].

Test data were simulated for the Duffing oscillator using a sampling frequency of 2^{16} Hz and generated as a time series of 500 points. The system was forced with an odd multisine signal [4] from 450 Hz to 500 Hz with a frequency resolution of 5 Hz. The system was defined to have a mass of 0.1 kg, linear natural frequency of 500 Hz, and damping ratio of 0.05. Additionally k_3 was set to be equal to 1×10^9 N m^{-1} and measurement noise equal to 50% RMS of the signal was added artificially after the system was simulated. The simulated measured signal is shown in Figure 1. This defines the *ground-truth* parameters $\theta^\star = [m^\star, k^\star, c^\star, k_3^\star]$. The prior distributions in the model are set such that $\boldsymbol{\mu}$, the prior means, is equal to $\theta^\star$ perturbed by a Gaussian distributed random number with variance equal to half of the true value. The prior variances for the parameters are set to be equal to twice the true values of the parameters. The hyperparameters of the Gamma prior for τ are set such that $a = 1$, $b = 500$ when the distribution is parameterised in terms of its shape and rate[1].

[1] Code is available to reproduce these results at https://github.com/TimothyRogers/rasd2019



(a) (b)

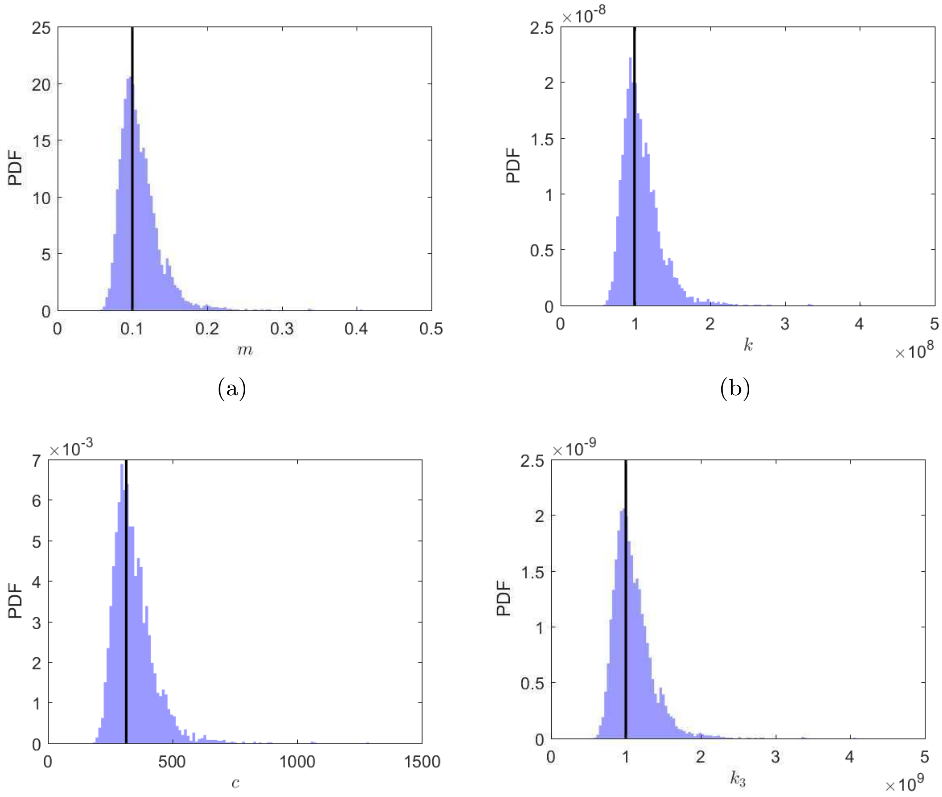
(c) (d)

Figure 3. Figure showing the CDF and PDF of the samples of the individual system parameters $1/m, k/m, c/m, k_3/m$ from the Gibbs sampler after burn-in using the PGAS algorithm with particle rejuvenation. The known ground truth is shown in black.

The Gibbs sampler was run for a total of 5000 iterations with a burn-in of 250 iterations. Figure 2 shows, each of the sampled paths X' for both the displacement and velocity states after the burn-in period with the true trajectories superimposed. It can be seen that the true noise free trajectories lie within the smoothing distributions and taking the means of the samples gives a normalised mean squared error of 0.24 in the displacement and 0.45 in the velocity state, when compared to the noise free state trajectories. This shows the ability of the PG-AS scheme with particle rejuvenation to accurately recover the smoothing distributions over the states.

The normalised mean squared error metric is defined as $\mathrm{NMSE} = \frac{100}{N\sigma_{\mathbf{y}}^2}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2$ for N test points, where $\sigma_{\mathbf{y}}$ is the variance of the measured data, $\hat{y}_i$ is the predicted points, and y_i is the measured data point for $i = 1, \dots, N$. This will be equal to 0 for a perfect prediction and a value of 100 is equivalent to predicting as well as taking the mean of the measured data.

The parameter distributions estimated in the Gibbs sampler, which are transforms of the individual system parameters, are shown in Figure 3. The known true parameters are shown by the vertical green line in the plots. It can be seen that all of the parameters are well estimated by the Gibbs sampler. By taking the means of the distributions, the accuracy of the *maximum a posteriori* estimate from the Gibbs sampler can be assessed. For the first parameter $1/m$ the percentage error is -4.20% of the true parameter value, for the c/m coefficient the percentage error is -0.015%, and for the other two parameters k/m and k_3/m the percentage error is less that 1×10^{-6}%. The error in the first parameter value is linked to the approximation error in



(a)

(b)

(c)

(d)

Figure 4. Figure showing the CDF and PDF of the samples of the transformed system parameters m, k, c, k_3 from the Gibbs sampler after burn-in using the PGAS algorithm with particle rejuvenation. The known ground truth is shown in black.

the first-order discretisation used in Gibbs sampling the parameters. Although not shown here due to space constraints, the error in this parameter reduces as the sample rate of the data is increased.

It is then possible to recover distributions over the system parameters themselves by transforming the variables which were used in the Gibbs sampler. The results of this procedure is shown in Figure 4. The transformation is seen to skew the distributions but the modes of these are well aligned with the known ground truth for the parameters. Taking *maximum a posteriori* estimates of each of the parameters the percentage error is -3.39% for all parameters, this is due to the influence of the misidentification of the $1/m$ parameter in the transformed space due to the first order discretisation.

5. Discussion

This paper has presented the use of Particle Gibbs with Ancestor Sampling and Particle Rejuvenation as a viable technique for Bayesian estimation of nonlinear dynamical systems encountered in structural dynamics. It has been shown that the methodology is capable of recovering the smoothing distributions of the states and the distributions of the model parameters with low error in relation to the *maximum a posteriori* estimates and that the known true values and trajectories are within the probability mass of the estimated distributions. This has been demonstrated on a case where observations are only made of the displacement and the measurement noise is very high (50% RMS).

The method is seen to be sensitive to the sample frequency of the signal due to the first order approximation in the Gibbs sampler for the parameters. This warrants further investigation, including the possibility of adopting a *Metropolis within Gibbs* scheme to allow a higher-order discretisation. It has also been identified that the damping parameter is sensitive to the amount of process noise estimated by the Gibbs sampler. For a simulated system of the Duffing oscillator this can cause problems since there is no process noise, however, it is expected that on measured data from a physical system this will not be a limitation. It is likely that, for most systems of interest, it will not be possible to exactly write down the differential equations of motion. Alongside the occurrence of actual process noise, this process noise can help to compensate for model form errors when describing the system. Of interest for further research is to move to a nonparametric representation of the nonlinear system which would avoid issues with needing to specify the model form *a priori*. While it has been shown that this method is effective on a simulated dataset, future research is needed to prove the effectiveness of this approach on measured data. However, the results shown in this paper would motivate continued investigation of Sequential Monte Carlo approaches to nonlinear system identification in structural dynamics and a key contribution is demonstrating the use of the Particle Rejuvenation scheme to handle degeneracy in the models encountered in structural dynamics.

Acknowledgments

The authors from Sheffield would like to gratefully acknowledge funding from the EPSRC through grants EP/S001565/1 for Rogers and Cross and EP/R003645/1 for Worden. Schön and Lindholm were financially supported by the Swedish Foundation for Strategic Research (SSF) via the project *ASSEMBLE* (contract number: RIT15-0012) and by the project *NewLEADS - New Directions in Learning Dynamical Systems* (contract number: 621-2016-06079), funded by the Swedish Research Council.

[1] Duffing G 1918 *Erzwungene Schwingungen bei veränderlicher Eigenfrequenz und ihre technische Bedeutung* 41-42 (F. Vieweg & sohn)

[2] Virgin L 2001 *Introduction to Experimental Nonlinear Dynamics* (IOP Publishing)

[3] Moon F and Holmes P J 1979 *Journal of Sound and Vibration* **65** 275–296

[4] Pintelon R and Schoukens J 2012 *System Identification: a Frequency Domain Approach* (John Wiley & Sons)

[5] Wigren T and Schoukens J 2013 *Proceedings of the Control Conference (ECC), 2013 European* 2933–2938

[6] Ljung L, Zhang Q, Lindskog P and Juditsky A 2004 *Proceedings of the IFAC Symposium on Nonlinear Control Systems, Stuttgart, Germany*

[7] Verdult V 2004 *IFAC Proceedings Volumes* **37** 393–398

[8] Espinoza M, Pelckmans K, Hoegaerts L, Suykens J A and De Moor B 2004 *Proceedings of the 6th IFAC Symposium on Nonlinear Control Systems (NOLCOS)*

[9] Wills A and Ninness B 2009 *Proceedings of the 15th IFAC Symposium on System Identification (SYSID 2009)* 1104–1109

[10] Marconato A, Sjöberg J, Suykens J and Schoukens J 2012 *IFAC Proceedings of the 16th IFAC Symposium on System Identification* **16** 632–637

[11] Kovacic I and Brennan M J 2011 *The Duffing Equation: Nonlinear Oscillators and Their Behaviour* (John Wiley & Sons)

[12] Worden K and Hensman J 2012 *Mechanical Systems and Signal Processing* **32** 153–169

[13] Lindsten F, Jordan M I and Schön T B 2014 *The Journal of Machine Learning Research* **15** 2145–2184

[14] Lindsten F, Bunch P, Singh S S and Schön T B 2015 *arXiv preprint arXiv:1505.06356*

[15] Doucet A and Johansen A M 2012 *Handbook of nonlinear filtering* **12** 3

[16] Gordon N J, Salmond D J and Smith A F 1993 *IEE Proceedings F (Radar and Signal Processing)* **140** 107–113

[17] Kalman R E 1960 *Journal of basic Engineering* **82** 35–45

[18] Andrieu C, Doucet A and Holenstein R 2010 *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72** 269–342

[19] Marcos M, Calafat F M, Berihuete Á and Dangendorf S 2015 *Journal of Geophysical Research: Oceans* **120** 8115–8134

[20] Svensson A and Schön T B 2017 *Automatica* **80** 189–199
[21] Butcher J 1995 *BIT Numerical Mathematics* **35** 202–209
[22] Gelman A, Carlin J B, Rubin D B, Vehtari A, Dunson D B and Stern H S 2013 *Bayesian Data Analysis, Third Edition* (CRC Press)
[23] Murphy K P 2007 *Technical Report*

Appendix A. Details of 5th Order Runge-Kutta Scheme

The numerical integration scheme is a 5th order Runge-Kutta method [21], whose details are shown here for completeness.

$$x_{t+1} = x_t + \frac{\Delta}{90}\left(7a_1 + 32a_3 + 12a_4 + 32a_5 + 7a_6\right) \tag{A.1a}$$

$$a_1 = \mathbf{f}\left(t, x_t\right) \tag{A.1b}$$

$$a_2 = \mathbf{f}\left(t + 0.25\Delta, x_t + 0.25\Delta a_1\right) \tag{A.1c}$$

$$a_3 = \mathbf{f}\left(t + 0.125\Delta, x_t + 0.125\Delta a_1 + 0.125\Delta a_2\right) \tag{A.1d}$$

$$a_4 = \mathbf{f}\left(t + 0.5\Delta, x_t - 0.5\Delta a_2 + \Delta a_3\right) \tag{A.1e}$$

$$a_5 = \mathbf{f}\left(t + 0.75\Delta, x_t + 0.1875\Delta a_1 + 0.5625\Delta a_4\right) \tag{A.1f}$$

$$a_6 = \mathbf{f}\left(t + \frac{6}{7}\Delta, x_t - \frac{3}{7}\Delta a_1 + \frac{2}{7}\Delta a_2 + \frac{12}{7}\Delta a_3 - \frac{12}{7}\Delta a_4 + \frac{8}{7}\Delta a_5\right) \tag{A.1g}$$

Where Δ is the time step and $\mathbf{f}(\cdot)$ is the vector Markov form of the ODE which for the Duffing oscillator (10) is given by,

$$\mathbf{f}\left(t, x_t\right) = \begin{bmatrix} x_{t,2} \\ \frac{F(t)}{m} - \frac{c}{m}x_{t,1} - \frac{k}{m}x_{t,1} - \frac{k_3}{m}x_{t,1}^3 \end{bmatrix} \tag{A.2}$$

The notation $x_{t,1}$ denotes the value of the first state in x_t this corresponds to the displacement of the oscillator. Equivalently, $x_{t,2}$ corresponds to the velocity value.