Gibbs sampling for inference and reliability assessment in dynamic systems

M. Pozzi[1], D. Bernal[2]
(1) Carnegie Mellon University - USA, (2) Northestern University - USA

Abstract

We propose a numerical method to perform system identification and reliability analysis of a linear time-invariant dynamic system exposed to Gaussian noise. The method is based on the Gibbs Sampling approach, and it is able to generate a set of samples of models and state variable trajectories according to the posterior distribution of these quantities conditional to the observed measurements. Specifically, generation of samples of state trajectories is related to Kalman Filter algorithm. By post-processing the outcome of the sampling procedure, one can consistently estimate the probability of failure or damage events.
The method is described in details, and its performance is investigated and discussed on a simple mechanical problem.

1. Introduction

The task of system identification (SI) refers to the processing of data for calibrating an input-output map. This task can be accomplished by assuming a parametric form for the map, and estimating the corresponding parameters by fitting a dataset of measurements. In structural dynamics, the dataset can contain the recording of sensor readings and of forces applied to the structure (Yuen and Katafygiotis, 2002, Wu et al. 2007). Stochastic processes are usually assumed for modeling the measurement noise and the unknown contributions in the forces. For time-invariant linear systems, the input-output map can be modeled by a Markov chain of hidden state space variables, with transition and emission defined by a set of matrices, related to structural properties and position of sensors and actuators.
Usually, SI is performed by computing a single set of parameters, intended as a point estimate of the map, without assessing the uncertainty in this estimation (Ghahramani and Hinton, 1996). This practice is motivated by the need of performing SI at low computational cost and of using a single set of parameters in tasks as on-line structural control. However, the point estimate is not adequate for other purposes, as structural reliability, which involve the assessment of the probability of an event related to structural failure or damage. For example, suppose we want to estimate the probability that any structural property is in a range of values that we relate to damage: the point estimate of that property would produce just a binary indicator referring to the estimate being inside or outside the range, which is insufficient for the purpose of structural reliability.
In the Bayesian framework, to assess consistently the probability of damage, we need to compute the posterior probability density of the model parameters given the observations, and integrate that density on the failure domain. To do so, we need to develop a numerical approach able to perform a fully probabilistic inference of dynamic systems.
In this paper, we make use of the Probabilistic Graphical Model (PGM) framework to model a Linear Gaussian State Space Model (LG-SSM). Inference on model and state variables is performed by a Markov Chain Monte Carlo (MCMC) method adopting a Gibbs Sampling (GS) approach. Samples of state trajectory are generated by implementing a method based on the Kalman Filter (KF). Section 2 formulates the problem, Section 3 presents the method, Section 4 gives an example before drawing conclusions in Section 5.

2. Problem formulation

Consider a LG-SSM, i.e. a time invariant linear system exposed to known control forces, Gaussian stochastic forces, and monitored with instruments affected by Gaussian noise (Murphy, 2012). Model is defined by parameters $\boldsymbol{\theta}$, and prior knowledge on these parameters by probability density $p(\boldsymbol{\theta})$. Given a set of measures $\mathbf{Y}$, the task is to estimate probability of failure defined as $P_f = \mathbb{P}[g(\boldsymbol{\theta}, \mathbf{X}) < 0|\mathbf{Y}]$, where limit state function g can depend on hidden state variables $\mathbf{X}$ and model. In details, $\mathbf{x}_k$ and $\mathbf{y}_k$ indicate the state vector and measures at time t_k respectively. n measures are collected ($\mathbf{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$) with corresponding states ($\mathbf{X} = \{\mathbf{x}_0, \dots, \mathbf{x}_n\}$), and reliability assessment is perform offline, after all measures are collected.

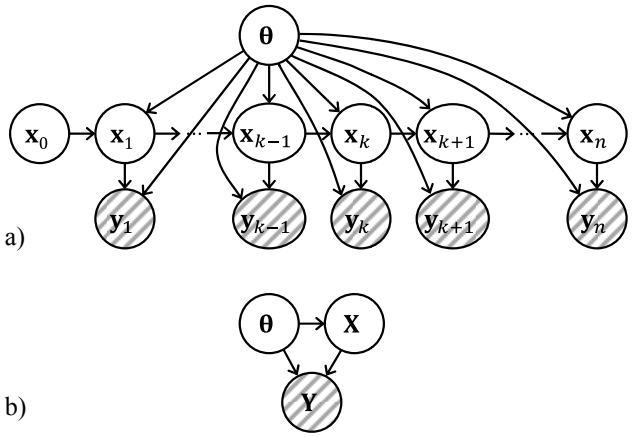

Fig.1 a) Extended and b) contracted PGM for LG-SSM.

2.1. Linear Gaussian models

A LG-SSM is modeled by the PGM reported in Figure 1. PGM in part (a) reports the extended graph, with complete structure of conditional independence among variables, while part (b) reports the contracted form. It is a partially observed graph: model parameters $\boldsymbol{\theta}$ and state variables $\mathbf{X}$ are unobserved, while $\mathbf{Y}$ is observed. The dynamic system is modeled as:

$$\forall k = 1, \dots, n \qquad \begin{cases} \mathbf{x}_k = \mathbf{A}(\boldsymbol{\theta})\mathbf{x}_{k-1} + \mathbf{B}(\boldsymbol{\theta})\mathbf{u}_{k-1} + \boldsymbol{\omega}_k \\ \mathbf{y}_k = \mathbf{C}(\boldsymbol{\theta})\mathbf{x}_k \quad + \mathbf{D}(\boldsymbol{\theta})\mathbf{u}_k \quad + \boldsymbol{\varepsilon}_k \end{cases} \tag{1a,b}$$

with matrixes $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ function of model parameters $\boldsymbol{\theta}$, and with white noise assumption on unknown force and measurement noise: $\boldsymbol{\omega}_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\omega)$, $\boldsymbol{\varepsilon}_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$, indicating with $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ the Multivariate Normal (MVN) distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and with $\mathcal{N}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ that density computed in $\mathbf{x}$. In the following, for the easiness of notation we assume matrices $\mathbf{B}$ and $\mathbf{D}$ are irrelevant, as no control force $\mathbf{U} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ is applied to the structure. In this setting, a probabilistic formulation for initial condition, state evolution and observation, i.e. for all variables in the PGM of Fig.1(a), can be expressed as:

$$\forall k = 1, \dots, n \qquad \begin{cases} \mathbf{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \\ \mathbf{x}_{k+1}|\mathbf{x}_k, \boldsymbol{\theta} \sim \mathcal{N}(\mathbf{A}(\boldsymbol{\theta})\mathbf{x}_k, \boldsymbol{\Sigma}_\omega) \\ \mathbf{y}_k|\mathbf{x}_k, \boldsymbol{\theta} \sim \mathcal{N}(\mathbf{C}(\boldsymbol{\theta})\mathbf{x}_k, \boldsymbol{\Sigma}_\varepsilon) \end{cases} \tag{2a,b,c}$$

2.2. Normal model uncertainty

While the approach we investigate can be applied to a wider setting, in this paper we restrict our attention to a special case, assuming a normal distribution for the model uncertainty:

$$\begin{cases} \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta) \\ \mathbf{A}(\boldsymbol{\theta}) = \mathbf{A}_0 + \sum_{j=1}^{m} \mathbf{A}_j \theta_j \\ \mathbf{C}(\boldsymbol{\theta}) = \mathbf{C}_0 + \sum_{j=1}^{m} \mathbf{C}_j \theta_j \end{cases} \quad (3a,b,c)$$

where matrices $\{\mathbf{A}_0, \dots, \mathbf{A}_m\}$ and $\{\mathbf{C}_0, \dots, \mathbf{C}_m\}$ model the relationship between parameters $\boldsymbol{\theta} = [\theta_1 \quad \dots \quad \theta_m]^{\mathrm{T}}$ and system and observation matrices. In summary, the model is defined by m (possibly interdependent) parameters. As Eq.3b,c are linear, pair ($\mathbf{A}$,$\mathbf{C}$) is modeled by a MVN distribution. Numerically, this is particularly convenient, as the normal distribution is prior conjugate of itself, and this facilitates the implementation of the Gibbs approaches outlined in 3.5.

2.3. Limit state function

To define system reliability, we introduce limit state function $g(\boldsymbol{\theta}, \mathbf{X})$, with the assumption that failure event (F) occurs if g is negative. For the easiness of computation, here we assume a linear limit state function. For example, if failure is only a function of model, g can be expressed as by vector $\mathbf{v}$ and constant c by the following equation:

$$g = \mathbf{v}^{\mathrm{T}} \boldsymbol{\theta} + c \quad (4)$$

3. Numerical approaches for inference and reliability analysis

Posterior failure probability $P_{\mathrm{F}|\mathbf{Y}}$ can be computed by the classical reliability integral:

$$P_{\mathrm{F}|\mathbf{Y}} = \iint_{g(\boldsymbol{\theta},\mathbf{X})<0} p(\boldsymbol{\theta}, \mathbf{X}|\mathbf{Y}) d\mathbf{X} d\boldsymbol{\theta} \quad (5)$$

Even for a LG-SSM with normal prior on models, posterior distribution $p(\boldsymbol{\theta}, \mathbf{X}|\mathbf{Y})$ cannot be computed in close form. However, distributions $p(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Y})$ and $p(\mathbf{X}|\boldsymbol{\theta}, \mathbf{Y})$, are both MVN. This suggests that we could easily infer the model if we had observed the entire state trajectory and, vice versa, we could easily infer the entire state trajectory if we knew the model. This remark is the basis for a GS procedure: after having initialized model $\boldsymbol{\theta}$, we can sample $\mathbf{X}$, and we iterate sampling of model and of states. Following this MCMC procedure, we track a random walk in the joint domain of $(\boldsymbol{\theta}, \mathbf{X})$. Samples of model and states can intended as coming from posterior $p(\boldsymbol{\theta}, \mathbf{X}|\mathbf{Y})$, and can be used for evaluating the integral in Eq.(5).

3.1. EM algorithm

The GS iterative procedure we propose can be related to the Expectation-Maximization (EM) algorithm (see for example Barber, 2012), that we describe in this section.
In fully observed PGM, likelihood of observations can be easily expressed, so that model parameters can be identified by maximum likelihood estimation. For partially observed graph, as that in Fig. 1, the learning process is more complicated. EM algorithm is an iterative scheme to

identify maximum likelihood parameters in partially observed graphs. Each iteration is composed by two steps: after model has been initialized, in the Expectation step (E) inference is performed and expected value is assigned to each unobserved variable, while in the Maximization step (M) maximum likelihood parameters are identified as for fully observed graphs. E-step is then repeated using the updated model, and alternated with M-step until convergence. In the LG-SSM setting, EM algorithm requires select initial parameter set $\boldsymbol{\theta}_{(0)}$, and alternates the following steps for sufficiently high N:

$$\forall i = 1, \dots, N \qquad \begin{cases} \mathbf{X}_{(i)} \leftarrow \mathbb{E}\left(\mathbf{X}|\mathbf{Y}, \boldsymbol{\theta}_{(i-1)}\right) \\ \boldsymbol{\theta}_{(i)} \leftarrow \operatorname{argmax}_{\boldsymbol{\theta}} p\left(\mathbf{Y}, \mathbf{X}_{(i)}|\boldsymbol{\theta}\right) \end{cases} \tag{6a,b}$$

3.2. Gibbs Sampling approach

GS is a widely used approach for performing inference in PGMs (Russell and Norvig, 2010, Mackay, 2003). In partially observed PGMs, after having initialized all unobserved variables to some values, each variable is updated, by sampling from the distribution conditional to its Markov blanket. GS can be applied to the learning process of LG-SSM, by an algorithm that is similar to EM (see Memarzadeh *et al.*, 2014). By considering Fig. 1(b), it is apparent that states has to be samples conditional to model and observations (similarly to the E-step) and models conditional to state and observation (similarly to the M-step). After having initialized model parameters to $\boldsymbol{\theta}_{(0)}$ and selected a burn-in period length $b \in \mathbb{N}$, GS is composed by two steps:

$$\forall i = 1, \dots, N + b \qquad \begin{cases} \mathbf{X}_{(i)} \sim p\left(\mathbf{X}|\mathbf{Y}, \boldsymbol{\theta}_{(i-1)}\right) \\ \boldsymbol{\theta}_{(i)} \sim p\left(\boldsymbol{\theta}|\mathbf{X}_{(i)}, \mathbf{Y}\right) \end{cases} \tag{7a,b}$$

Outcomes of the GS are set of joint samples $\left\{\boldsymbol{\theta}_{(i)}, \mathbf{X}_{(i)}\right\}_{N+b}^{i=1+b}$ that we can assume to be distributed according to joint posterior distribution $p(\boldsymbol{\theta}, \mathbf{X}|\mathbf{Y})$.

3.3. Kalman filter and sampler

To generate samples of space trajectories $\mathbf{X}$ when model $\boldsymbol{\theta}$ is assigned, we can adopt a Forward Filtering and Backward Sampling (FFBS) approach (Barber, 2012). A similar application to discrete states is presented by Memarzadeh *et al.* 2014. Specifically, the FF step can be performed by the KF algorithm. References to the KF are contained in many textbook, including Barber, 2012, Yuen, 2010. KF is consistent with LG-SSM and requires, as inputs, $\boldsymbol{\mu}_0$, $\boldsymbol{\Sigma}_0$, $\boldsymbol{\Sigma}_\omega$, $\boldsymbol{\Sigma}_\varepsilon$, $\mathbf{A}$, $\mathbf{C}$ and $\mathbf{Y}$. Posterior distribution at step k is expressed as:

$$p(\mathbf{x}_k|\mathbf{Y}_k, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{x}_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{8}$$

where $\mathbf{Y}_k = \{\mathbf{y}_1, \dots, \mathbf{y}_k\}$ the sub-set of measure up to step k, while $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are mean vector and covariance matrix of the updated distribution respectively. Note that, at the last step of the filtering process, probability of $\mathbf{x}_n$ is posterior respect to the entire dataset, as $\mathbf{Y}_n = \mathbf{Y}$. As it is easy to sample from the univariate standard normal distribution, we can easily get a sample from the MVN distribution $\mathcal{N}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ by transforming to the standard normal space (Murphy, 2012):

$$\mathbf{x} = \mathbf{L}\mathbf{u} + \boldsymbol{\mu} \tag{9}$$

where $\mathbf{u}$ is a vector of independent standard normal variables, and $\mathbf{L}$ is the lower triangular Cholesky decomposition of $\boldsymbol{\Sigma}$: we populate vector $\mathbf{u}$, and use Eq.7 for obtaining a sample in $\mathbf{x}$.
To get a sample of the entire state trajectory, we need to perform the BS process. To do so, we note that posterior probability $\mathbf{x}_k$ given $\mathbf{Y}_k$ and $\mathbf{x}_{k+1}$ can be expressed as:

$$p(\mathbf{x}_k|\mathbf{x}_{k+1}, \boldsymbol{\theta}, \mathbf{Y}_k) \propto p(\mathbf{x}_k, \mathbf{x}_{k+1}|\boldsymbol{\theta}, \mathbf{Y}_k) = p(\mathbf{x}_{k+1}|\mathbf{x}_k, \boldsymbol{\theta}) p(\mathbf{x}_k|\mathbf{Y}_k, \boldsymbol{\theta}) \tag{10}$$

Considering Fig.1(a), it is to be noted that x_k is conditional independent to all states after x_{k+1} and all observations after y_k, given model and x_k. Specifically, from Eq.2(b) we derive that:

$$p(\mathbf{x}_{k+1}|\mathbf{x}_k, \boldsymbol{\theta}) = \mathcal{N}(\mathbf{x}_{k+1}, \mathbf{A}\mathbf{x}_k, \boldsymbol{\Sigma}_\omega) \propto \mathcal{N}\left(\mathbf{x}_k, \mathbf{A}^{-1}\mathbf{x}_{k+1}, \boldsymbol{\Lambda}_g^{-1}\right) \tag{11}$$

with $\boldsymbol{\Lambda}_g = \mathbf{A}\boldsymbol{\Sigma}_\omega^{-1}\mathbf{A}^{\mathrm{T}}$ and conclude that

$$p(\mathbf{x}_k|\mathbf{x}_{k+1}, \boldsymbol{\theta}, \mathbf{Y}_k) = \mathcal{N}(x_k, \boldsymbol{\mu}_k', \boldsymbol{\Sigma}_k') \tag{12}$$

with $\boldsymbol{\Sigma}_k' = \left(\boldsymbol{\Sigma}_k^{-1} + \boldsymbol{\Lambda}_g\right)^{-1}$ and $\boldsymbol{\mu}_k' = \boldsymbol{\Sigma}_k'\left(\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k + \boldsymbol{\Lambda}_g x_{k+1}\right)$. Following Eqs.10-12, we can get samples of $\mathbf{x}_k$ for step k from $n-1$ down to 1, by using the sample previously generated for $\mathbf{x}_{k+1}$ in Eq.11. Sample from any normal distribution can be generated following the procedure outlined above.
The outcome of the FFBS procedure is a sample of the entire trajectory $\mathbf{X}$, generated according to distribution $p(\mathbf{X}|\mathbf{Y}, \boldsymbol{\theta})$, for any arbitrary model $\boldsymbol{\theta}$.

3.4. Model Likelihood

In this and the following sections, we describe how to compute distribution $p(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Y})$ and generate a sample of the model from it. We start computing likelihood function $p(\mathbf{X}, \mathbf{Y}|\boldsymbol{\theta})$.
Assuming marginal distribution p_ω and p_ε for unknown force and measurement noise, respectively, the conditional distribution of states and observations can be expressed as:

$$\begin{cases} p(\mathbf{X}|\boldsymbol{\theta}) & = \prod_{k=1}^{n} p_\omega(\mathbf{x}_k - \mathbf{A}(\boldsymbol{\theta})\mathbf{x}_{k-1}) \\ p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) & = \prod_{k=1}^{n} p_\varepsilon(\mathbf{y}_k - \mathbf{C}(\boldsymbol{\theta})\mathbf{x}_k) \end{cases} \tag{13a,b}$$

By chain rule, we can write:

$$p(\mathbf{X}, \mathbf{Y}|\boldsymbol{\theta}) = p(\mathbf{X}|\boldsymbol{\theta}) p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) \tag{14}$$

For LG-SSM, we can express the products in Eq.1 by using Eq.3, to get:

$$\begin{cases} \mathbf{A}(\boldsymbol{\theta})\mathbf{x}_k = \left[\mathbf{A}_0 + \sum_{j=1}^{m} \mathbf{A}_j\theta_j\right]\mathbf{x}_k = \mathbf{A}_0\mathbf{x}_k + \sum_{j=1}^{m} \mathbf{A}_j\mathbf{x}_k\theta_j = \mathbf{v}_{0,k}^a + \sum_{j=1}^{m} \mathbf{v}_{j,k}^a\theta_j \\ \mathbf{C}(\boldsymbol{\theta})\mathbf{x}_k = \left[\mathbf{C}_0 + \sum_{j=1}^{m} \mathbf{C}_j\theta_j\right]\mathbf{x}_k = \mathbf{C}_0\mathbf{x}_k + \sum_{j=1}^{m} \mathbf{C}_j\mathbf{x}_k\theta_j = \mathbf{v}_{0,k}^c + \sum_{j=1}^{m} \mathbf{v}_{j,k}^c\theta_j \end{cases} \tag{15a,b}$$

with $\mathbf{v}_{0,k}^a = \mathbf{A}_0\mathbf{x}_k$, $\mathbf{v}_{j,k}^a = \mathbf{A}_j\mathbf{x}_k$, $\mathbf{v}_{0,k}^c = \mathbf{C}_0\mathbf{x}_k$, and $\mathbf{v}_{j,k}^c = \mathbf{C}_j\mathbf{x}_k$. Previous Eq. can be re-written as:

$$\begin{cases} \mathbf{A}(\boldsymbol{\theta})\mathbf{x}_k = \mathbf{X}_k^a\boldsymbol{\theta} + \mathbf{v}_{0,k}^a \\ \mathbf{C}(\boldsymbol{\theta})\mathbf{x}_k = \mathbf{X}_k^c\boldsymbol{\theta} + \mathbf{v}_{0,k}^c \end{cases} \tag{16a,b}$$

with $\mathbf{X}_k^a = [\mathbf{v}_{1,k}^a \quad \dots \quad \mathbf{v}_{m,k}^a]$ and $\mathbf{X}_k^c = [\mathbf{v}_{1,k}^c \quad \dots \quad \mathbf{v}_{m,k}^c]$. Probability density function p_ω and p_ε can be express in normal canonical form, defining density form $\mathcal{N}'$ as:

$$\mathcal{N}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathcal{N}'(\mathbf{x}, \boldsymbol{\eta}, \boldsymbol{\Lambda}) \propto \exp\left[-\frac{1}{2}\mathbf{x}^{\mathrm{T}}\boldsymbol{\Lambda}\mathbf{x} + \mathbf{x}^{\mathrm{T}}\boldsymbol{\eta}\right] \tag{17}$$

where $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1}$ is the inverse of the covariance matrix and $\boldsymbol{\eta} = \boldsymbol{\Lambda}\boldsymbol{\mu}$. Eq. 13 can re-written as:

$$\begin{cases} p_\omega(\mathbf{x}_k - \mathbf{A}(\boldsymbol{\theta})\mathbf{x}_{k-1}) \propto \mathcal{N}'(\boldsymbol{\theta}; \boldsymbol{\eta}_k^a, \boldsymbol{\Lambda}_k^a) \\ p_\varepsilon(\mathbf{y}_k - \mathbf{C}(\boldsymbol{\theta})\mathbf{x}_k) \quad \propto \mathcal{N}'(\boldsymbol{\theta}; \boldsymbol{\eta}_k^c, \boldsymbol{\Lambda}_k^c) \end{cases} \tag{18a,b}$$

with parameters $\boldsymbol{\Lambda}_k^a = [\mathbf{X}_{k-1}^a]^{\mathrm{T}}\boldsymbol{\Sigma}_\omega^{-1}\mathbf{X}_{k-1}^a$, $\boldsymbol{\eta}_k^a = [\mathbf{X}_{k-1}^a]^{\mathrm{T}}\boldsymbol{\Sigma}_\omega^{-1}(\mathbf{x}_k - \mathbf{v}_0^a)$, $\boldsymbol{\Lambda}_k^c = [\mathbf{X}_k^c]^{\mathrm{T}}\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{X}_k^c$ and $\boldsymbol{\eta}_k^c = [\mathbf{X}_k^c]^{\mathrm{T}}\boldsymbol{\Sigma}_\varepsilon^{-1}(\mathbf{x}_k - \mathbf{v}_0^a)$.

$$\begin{cases} p(\mathbf{X}|\boldsymbol{\theta}) \quad \propto \prod_{k=1}^n \mathcal{N}'(\boldsymbol{\theta}; \boldsymbol{\eta}_k^a, \boldsymbol{\Lambda}_k^a) = \mathcal{N}'(\boldsymbol{\theta}; \boldsymbol{\eta}_T^a, \boldsymbol{\Lambda}_T^a) \\ p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\theta}) \propto \prod_{k=1}^n \mathcal{N}'(\boldsymbol{\theta}; \boldsymbol{\eta}_k^a, \boldsymbol{\Lambda}_k^a) = \mathcal{N}'(\boldsymbol{\theta}; \boldsymbol{\eta}_T^a, \boldsymbol{\Lambda}_T^a) \end{cases} \tag{19a,b}$$

with $\boldsymbol{\eta}_T^a = \sum_{k=1}^n \boldsymbol{\eta}_k^a$, $\boldsymbol{\Lambda}_T^a = \sum_{k=1}^n \boldsymbol{\Lambda}_k^a$, $\boldsymbol{\eta}_T^c = \sum_{k=1}^n \boldsymbol{\eta}_k^c$ and $\boldsymbol{\Lambda}_T^c = \sum_{k=1}^n \boldsymbol{\Lambda}_k^c$. So joint probability likelihood can be expressed as:

$$p(\mathbf{X}, \mathbf{Y}|\boldsymbol{\theta}) \propto \mathcal{N}'(\boldsymbol{\theta}; \boldsymbol{\eta}_T, \boldsymbol{\Lambda}_T) \tag{20}$$

with $\boldsymbol{\eta}_T = \boldsymbol{\eta}_T^a + \boldsymbol{\eta}_T^c$ and $\boldsymbol{\Lambda}_T = \boldsymbol{\Lambda}_T^a + \boldsymbol{\Lambda}_T^c$.

3.5. Model posterior distribution and model sampling

Following previous section, we can derive the posterior distribution under the assumption that the prior distribution on $\boldsymbol{\theta}$ is normal. We can re-write Eq.3(a) in canonical form as:

$$p(\boldsymbol{\theta}) = \mathcal{N}'(\boldsymbol{\theta}; \boldsymbol{\eta}_\pi, \boldsymbol{\Lambda}_\pi) \tag{21}$$

with $\boldsymbol{\Lambda}_\pi = \boldsymbol{\Sigma}_\pi^{-1}$ and $\boldsymbol{\eta}_\pi = \boldsymbol{\Lambda}_\pi\boldsymbol{\mu}_\pi$, so that, by Bayes' formula, posterior distribution is:

$$p(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Y}) \propto p(\boldsymbol{\theta})p(\mathbf{X}, \mathbf{Y}|\boldsymbol{\theta}) = \mathcal{N}'(\boldsymbol{\theta}; \boldsymbol{\eta}_u, \boldsymbol{\Lambda}_u) \tag{22}$$

with $\boldsymbol{\eta}_u = \boldsymbol{\eta}_T + \boldsymbol{\eta}_\pi$ and $\boldsymbol{\Lambda}_u = \boldsymbol{\Lambda}_T + \boldsymbol{\Lambda}_\pi$. Posterior mean and covariance matrix can be derived by inverse relations: $\boldsymbol{\Sigma}_u = \boldsymbol{\Lambda}_u^{-1}$ and $\boldsymbol{\mu}_u = \boldsymbol{\Sigma}_u\boldsymbol{\eta}_u$. Again, to sample from a MVN distribution, we can use the procedure outlined above. Following the formulation of previous and current Section, we can derive model likelihood, model posterior and generate a consistent sample of the model.

3.6. Reliability analysis

Having samples of model $\boldsymbol{\theta}$ as an outcome of the GS procedure, we can use these for evaluating the integral in Eq. 5. However, we can hope to reduce the variance of the estimate by following a slightly different path. For the function defined in Eq.4, g is normally distributed if conditional to both $\mathbf{Y}$ and $\mathbf{X}$ as, in that case, $p(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Y})$ is normal (as shown in Eq.22) and so is any linear function of $\boldsymbol{\theta}$. The corresponding failure probability is:

$$P_{\mathrm{F}|\mathbf{X},\mathbf{Y}} = \mathbb{P}(\mathrm{F}|\mathbf{X}, \mathbf{Y}) = \Phi\left(-\mu_{g|\mathbf{X},\mathbf{Y}}/\sigma_{g|\mathbf{X},\mathbf{Y}}\right) \tag{23}$$

with $\sigma^2_{g|\mathbf{X},\mathbf{Y}} = \mathbf{v}^{\mathrm{T}}\boldsymbol{\Sigma}_u\mathbf{v}$, $\boldsymbol{\mu}_{g|\mathbf{X},\mathbf{Y}} = \mathbf{v}^{\mathrm{T}}\boldsymbol{\mu}_u + c$ and Φ is the cumulative density function of the standard normal distribution. As parameters $\boldsymbol{\mu}_u$ and $\boldsymbol{\Sigma}_u$ can be computed for any sampled value of trajectory states $\mathbf{X}$, we can compute function $P_{\mathrm{F}|\mathbf{X},\mathbf{Y}}$ for any value in the set $\{\mathbf{X}_{(i)}\}_{N+b}^{i=b}$.
Following the classical Monte Carlo approach to reliability analysis, the corresponding marginal failure probability $P_{\mathrm{F}|\mathbf{Y}} = \mathbb{P}[\mathrm{F}|\mathbf{Y}]$ can be estimated as an arithmetical average:

$$P_{\mathrm{F}|\mathbf{Y}} \cong \hat{P}_{\mathrm{F}|\mathbf{Y}} = \frac{1}{N}\sum_{i=1+b}^{N+b} P_{\mathrm{F}|\mathbf{X}_{(i)},\mathbf{Y}} \tag{25}$$

Standard deviation of the estimator can be computed as:

$$\hat{\sigma} = \frac{1}{N^2}\sum_{i=1+b}^{N+b}\left(P_{\mathrm{F}|\mathbf{X}_{(i)},\mathbf{Y}} - \hat{P}_{\mathrm{F}|\mathbf{Y}}\right)^2 \tag{26}$$

So that 95% confidence interval can be represented as $CR_{95\%} = \left[\hat{P}_{\mathrm{F}|\mathbf{Y}} - 2\hat{\sigma}; \hat{P}_{\mathrm{F}|\mathbf{Y}} + 2\hat{\sigma}\right]$.
A summary of the entire approach to reliability analysis is reported in Fig. 2.

INPUT	
$\boldsymbol{\theta}_{(0)}, \boldsymbol{\mu}_{\pi}, \boldsymbol{\Sigma}_{\pi}, \{\mathbf{A}_0, \mathbf{C}_0\}_m^{j=1}$	prior on model
$\boldsymbol{\Sigma}_{\omega}, \boldsymbol{\Sigma}_{\varepsilon}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \mathbf{Y}$	noise, initial condition and data
$\mathbf{v}, c$	limit state function definition
N, b	burn-in length and number of samples
PROCESS	
for $i = 1:(N+b)$	
$\mathbf{X}_{(i)} \sim p(\mathbf{X}\|\mathbf{Y}, \boldsymbol{\theta}_{(i-1)})$	Kalman sampler
$P_{\mathrm{F}|\mathbf{X}_{(i)},\mathbf{Y}}$	Eq.18
$\boldsymbol{\theta}_{(i)} \sim p(\boldsymbol{\theta}\|\mathbf{X}_{(i)}, \mathbf{Y})$	model updating
OUTCOME	
$\hat{P}_{\mathrm{F}|\mathbf{Y}}, CR_{95\%}$	Eq.19-20

Fig.2 Pseudo-code for reliability analysis

4. Numerical investigation

4.1. Structural model

We test the method on a simple mechanical problem. Consider the 6 degrees of freedom (dof) system reported in Fig.3. Stiffness of each spring is unknown and vector $\boldsymbol{\theta}$ lists these values. For the simulation, we used stiffness values 42, 46, 40, 35, 38, 42 N/mm respectively. Sampling time is $\Delta t = 0.01$s and acquisition period is $T = 60$s. Masses m are known to be 1Ton each and damping forces d are 0.6Ns/mm. Random forces are spatially independent and so that, during Δt, the standard deviation of unknown force is σ_{ω} is 0.01mm for the first 6 components and $\sigma_{\omega} =$

0.1mm/s for the last 6. Measurement noise is also spatially independent, with standard deviation $\sigma_\varepsilon = 0.2\text{mm}$. 2 sensors are installed, measuring displacements on dof 3 and 6.
Prior distribution on stiffness values is defined by mean vector with equal component of $50 =$ N/mm. Marginal variances are $\sigma_\theta^2 = 125[\text{N/mm}]^2$ and correlation coefficient is 20% between each pair of stiffness values. This positive correlation models the assumption that stiffness values are supposed to be similar (but not identical).

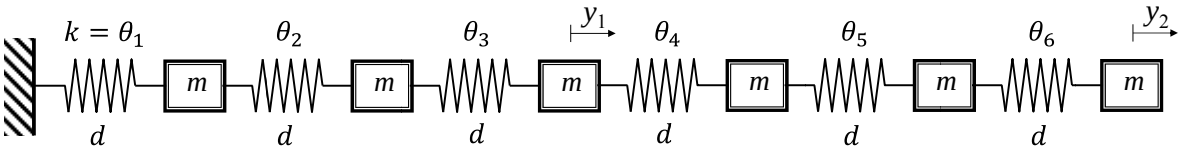

Fig.3 The mechanical system used for the numerical investigation

4.2. Model of structural dynamic

We can define constitutive matrix **E** arranging model parameters on the diagonal:

$$\mathbf{E} = \sum_{j=1}^{m} \theta_j \mathbf{F}_j \tag{27}$$

with $\mathbf{F}_j$ being a matrix of zeros except for unitary entry in position (j, j). Stiffness matrix **K**, can be expressed as:

$$\mathbf{K} = \mathbf{B}\,\mathbf{E}\,\mathbf{B}^{\mathrm{T}} = \sum_{j=1}^{m} \theta_j \left(\mathbf{B}\,\mathbf{F}_j\,\mathbf{B}^{\mathrm{T}}\right) \tag{28}$$

where **B** is the compatibility matrix. Continuous time system matrix can be expressed as:

$$\mathbf{A}_c = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{D} \end{bmatrix} \tag{29}$$

where **M** is the mass matrix and **D** the damping one. Discrete time matrix is, by using Eq. 28:

$$\mathbf{A} = \mathbf{I} + \mathbf{A}_c \Delta t = \begin{bmatrix} \mathbf{I} & \mathbf{I}\,\Delta t \\ -\mathbf{M}^{-1}\mathbf{K}\,\Delta t & \mathbf{I} - \mathbf{M}^{-1}\mathbf{D}\,\Delta t \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{I}\,\Delta t \\ -\sum_{j=1}^{m} \theta_j\, \mathbf{M}^{-1}\left(\mathbf{B}\,\mathbf{F}_i\,\mathbf{B}^{\mathrm{T}}\right)\Delta t & \mathbf{I} - \mathbf{M}^{-1}\mathbf{D}\,\Delta t \end{bmatrix} \tag{31}$$

and we identify matrixes$\{\mathbf{A}_0, \dots, \mathbf{A}_m\}$ as:

$$\mathbf{A}_0 = \begin{bmatrix} \mathbf{I} & \mathbf{I}\,\Delta t \\ \mathbf{0} & \mathbf{I} - \mathbf{M}^{-1}\mathbf{D}\,\Delta t \end{bmatrix} \qquad \mathbf{A}_i = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ -\mathbf{M}^{-1}\left(\mathbf{B}\,\mathbf{F}_i\,\mathbf{B}^{\mathbf{T}}\right)\Delta t & \mathbf{0} \end{bmatrix} \tag{32}$$

while we assume **C** does not depend on $\boldsymbol{\theta}$.

4.3. Outcomes of the simulations

Fig.4(a) reports some time histories related to dof 3: the thick red line shows the trajectory used for the simulation, the black one the simulated recordings used for the inference (as said before, recordings were also available for dof 6). Thin blues lines shows 50 realization of the corresponding component of the state trajectory, as resulting from the GS procedure. Fig.4(b) zooms on a part on

previous graph. Fig.4(c) shows the component of the random walk corresponding to θ_4, i.e. the stiffness of the fourth spring, in red, while initial blue line refers to the burn-in period, for which we use classical EM. Constant dotted upper line shows actual stiffness (35 N/mm). Fig.4(d) shows the empirical distribution of $p(\theta_4|\mathbf{Y})$.

4.4. Reliability analysis

We assume that a specific damage is defined by a very simple condition: θ_4 being below minimum allowable stiffness $\bar{\theta} = 34\text{N/mm}$, so limit state function g is defined by Eq.4 with $\mathbf{v}^{\mathrm{T}} = [0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0]$ and $c = -\bar{\theta}$. The lower dash-dotted line in Fig. 4(c) reports the value of $\bar{\theta}$. Fig.4(e) shows $CR_{95\%}$ for two alternative methods: blue lines refer to the method that directly uses the model samples (α), while black lines to the formulation in Section 3.6 (β). For the investigated case, the two approaches give very similar results.

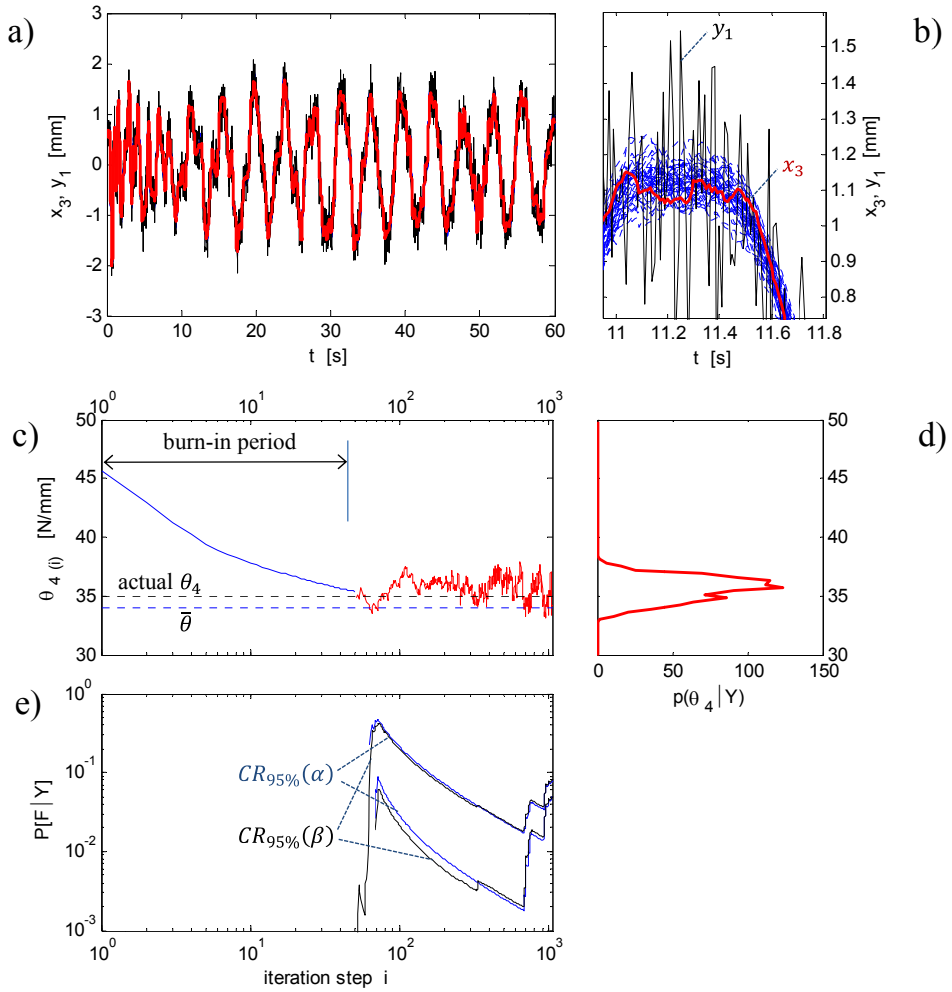

Fig.4 Results of the numerical investigation: a) signal and sampled trajectories, b) zoom in a region of the previous graph, c) sample of stiffness values for the fourth spring, d) corresponding empirical distribution, e) confidence bound of estimate of failure probability.

5. Conclusions

We have proposed a method based on Gibbs Sampling to perform offline structural System Identification and reliability analysis. For easiness of computation, we have assumed a normal prior on model parameters and a linear relationship between parameters and response. These assumptions can be removed, but at a higher computational cost: while likelihood function $p(\mathbf{X},\mathbf{Y}|\mathbf{A},\mathbf{C})$ is always normal, function $p(\mathbf{X},\mathbf{Y}|\boldsymbol{\theta})$ is not, if relation between $(\mathbf{A},\mathbf{C})$ and $\boldsymbol{\theta}$ is not linear. Furthermore, even when $p(\mathbf{X},\mathbf{Y}|\boldsymbol{\theta})$ is normal, posterior distribution $p(\boldsymbol{\theta}|\mathbf{X},\mathbf{Y})$ is generally not, if prior $p(\boldsymbol{\theta})$ is not. For those cases, a Metropolis-Hasting approach has to be used for sampling models (see Memarzadeh *et al.* 2014b, for an application to discrete states).
The method offers a consistent approach for the challenging task of representing posterior distribution $p(\boldsymbol{\theta},\mathbf{X}|\mathbf{Y})$, that is key for probabilistic reliability analysis. Other methods can be used for the same task: for example, it is well known that the KF algorithm can also be used for evaluating likelihood $p(\mathbf{Y}|\boldsymbol{\theta})$, and thus can be used in a Monte Carlo approach when models are generated according to the prior distribution $p(\boldsymbol{\theta})$. The benefit of the MCMC approach is that samples are generated according to the posterior distribution, which is much needed when posterior distribution is concentrated in a small region of the parameter domain. Both approaches (that one we propose and that based on KF) can benefit from the use of an importance sampling approach that considers posterior distribution and failure domain for the reliability assessment. Those comparisons and possible extensions, and a more accurate study of the performance of the algorithm in other settings, are future work for this research.

Acknowledgements
The first author acknowledges the Pennsylvania Infrastructure Technology Alliance, a partnership of Carnegie Mellon, Lehigh University and the Commonwealth of Pennsylvania's Department of Community and Economic Development (DCED), via grant PITA YR16 31571.1.9.1042204.

REFERENCES

BARBER, B. BAYESIAN REASONING AND MACHINE LEARNING. CAMBRIDGE UP, 2012.

GHAHRAMANI Z., HINTON G.. PARAMETER ESTIMATION FOR LINEAR DYNAMICAL SYSTEMS. TECHNICAL REPORT CRG-TR-96-2, DEPT. COMP. SCI., UNIV. TORONTO, 1996.

MACKAY, D. INFORMATION THEORY. INFERENCE AND LEARNING ALGORITHMS. CAMBRIDGE UNIVERSITY PRESS, 2003.

MEMARZADEH, M., POZZI, M., AND ZICO KOLTER, J. "OPTIMAL PLANNING AND LEARNING IN UNCERTAIN ENVIRONMENTS FOR THE MANAGEMENT OF WIND FARMS." J. COMPUT. CIV. ENG., 10.1061/(ASCE)CP.1943-5487.0000390 , 04014076. (2014).

MEMARZADEH, M., POZZI, M. AND KOLTER, J.Z., "MULTIPLE POMDP FORMULATIONS FOR SYSTEMS OF DEGRADING COMPONENTS", THE SIXTH WORLD CONFERENCE ON STR. CONTROL AND MONITORING, BARCELONA, 15-17 JULY 2014B.

MURPHY, K. MACHINE LEARNING, A PROBABILISTIC PERSPECTIVE. MIT PRESS, 2012.

RUSSELL, S. AND P. NORVIG. ARTIFICIAL INTELLIGENCE: A MODERN APPROACH. PEARSON EDUCATION, 2010.

WU M., SMYTH, A.W. "APPLICATION OF THE UNSCENTED KALMAN FILTER FOR REAL-TIME NONLINEAR STRUCTURAL SYSTEM IDENTIFICATION". STRUCTURAL CONTROL AND HEALTH MONITORING, WILEY ONLINE LIBRARY. 2007

YUEN, K.V., KATAFYGIOTIS, L.S., "BAYESIAN MODAL UPDATING USING COMPLETE INPUT AND INCOMPLETE RESPONSE NOISY MEASUREMENTS," JOURNAL OF ENGINEERING MECHANICS, VOL. 128, NO. 3, PP. 340-350, 2002.

YUEN, K-V. BAYESIAN METHODS FOR STRUCTURAL DYNAMICS AND CIVIL ENGINEERING. WILEY, 2010.