Bayesian Interpolation and Parameter Estimation in a Dynamic Sinusoidal Model

Jesper Kjær Nielsen, Mads Græsbøll Christensen, *Member, IEEE*, A. Taylan Cemgil, *Member, IEEE*, Simon J. Godsill, *Member, IEEE*, and Søren Holdt Jensen, *Senior Member, IEEE*

Abstract—In this paper, we propose a method for restoring the missing or corrupted observations of nonstationary sinusoidal signals which are often encountered in music and speech applications. To model nonstationary signals, we use a time-varying sinusoidal model which is obtained by extending the static sinusoidal model into a dynamic sinusoidal model. In this model, the in-phase and quadrature components of the sinusoids are modeled as first-order Gauss–Markov processes. The inference scheme for the model parameters and missing observations is formulated in a Bayesian framework and is based on a Markov chain Monte Carlo method known as Gibbs sampler. We focus on the parameter estimation in the dynamic sinusoidal model since this constitutes the core of model-based interpolation. In the simulations, we first investigate the applicability of the model and then demonstrate the inference scheme by applying it to the restoration of lost audio packets on a packet-based network. The results show that the proposed method is a reasonable inference scheme for estimating unknown signal parameters and interpolating gaps consisting of missing/corrupted signal segments.

Index Terms—Bayesian signal processing, sinusoidal signal model, state space modeling.

I. INTRODUCTION

THE interpolation of missing, corrupted and future signal samples is an important task in several applications. For example, speech and audio signals are often transmitted over packet-based networks in which packets may be lost, delayed or corrupted. If the contents of neighboring packets are correlated, the erroneous packets can be approximately reconstructed by using suitable interpolation techniques. The simplest interpolation techniques employ signal repetition [1] and signal stretching [2]. More advanced interpolation techniques are based on filter bank methods such as GAPES and MAPES [3], [4] or based on signal models such as autoregressive models

[5], [6], hidden Markov models [7], and sinusoidal models [8]–[10]. An integral part of the techniques based on signal modeling is the estimation of the signal parameters. Given estimates of these parameters, signal samples are interpolated by simulating data from the model.

Within the applied speech and audio processing field, the sinusoidal signal model is one of the more popular parametric signal models because voiced speech and signals originating from several musical instruments can be accurately modeled as a sum of sinusoids [11]. In this paper, we initially consider the dampened sinusoidal signal model in its real form given by

$$x_n = \sum_{l=1}^{L} \rho_l^n \left[i_l \cos(\omega_l n) + q_l \sin(\omega_l n) \right] + w_n \qquad (1)$$

where the sampling indices $n = 1, \ldots, N$ label the uniform sampled data. In the model, $i_l, q_l, \omega_l \in [0, \pi]$, and $\rho_l > 0$ denote the undamped in-phase component, the undamped quadrature component, the (angular) frequency, and the damping coefficient of the lth sinusoid, respectively. The observed sample x_n at time index n is the sum of L such dampened sinusoids and a white Gaussian noise term w_n with variance σ_w^2. The model in (1) is also sometimes written in its polar form given by

$$x_n = \sum_{l=1}^{L} \rho_l^n \alpha_l \cos(\omega_l n - \varphi_l) + w_n \qquad (2)$$

where $\alpha_l = \sqrt{i_l^2 + q_l^2}$ and $\varphi_l = \arctan(q_l/i_l)$ are the undampened amplitude and phase of the lth sinusoid, respectively. In this paper, we refer to the models in (1) and (2) as static sinusoidal models. This naming convention is adopted in order to distinguish it from the dynamic sinusoidal model, which we introduce later.

The static sinusoidal model and its variations have been subject to extensive research for many years. This is primarily due to the large-scale applicability of the model, and because frequency parameters and damping coefficients enter the model in a nonlinear fashion. The latter complicates the estimation problem significantly and several methods for solving this problem have therefore been devised. Most of these estimators are aimed at estimating the frequency parameters. Well-known estimators comprise the Min–Norm method [12], nonlinear least squares estimators [13], [14], and the high-order Yule–Walker method [15]. Other well-known estimators are the subspace-based methods such as MUSIC [16], root-MUSIC [17], ESPRIT [18], and weighted subspace fitting [19]. A thorough review of most of these estimators is given in [20]. The theoretical foundation of these estimators is

based on classical statistics which is also known as frequentist or orthodox statistics. The other major approach to statistics is Bayesian statistics which offers some conceptual advantages to classical statistics (see, e.g., [21] and [22]). For instance, the Bayesian approach copes with nuisance parameters and signal interpolation in a highly standardised way. However, the history of Bayesian frequency estimators is much shorter because the Bayesian methods often struggle with practical problems such as the evaluation of high-dimensional and intractable integrals. In recent years, various developments in Markov chain Monto Carlo (MCMC) methods (see, e.g., [23]) have largely overcome these problems. Nevertheless, the methods still suffer from a high computational complexity.

Bayesian frequency estimation was first considered by Jaynes and Bretthorst in [24] and [25], respectively. In the pioneering work of the latter, the existence of analytical solutions to the Bayesian frequency estimation problem was demonstrated only in the case of a few sinusoids. Moreover, the general inference problem with multiple sinusoids was solved using suitable analytical approximations, under the assumptions that the sinusoids were well-separated, and enough data were available. This was not assumed in [26] and [27] in which the general frequency estimation problem was solved by use of an approximate MCMC technique which led to improved performance for closely spaced sinusoids. The performance was improved even further by Andrieu and Doucet in [28], where the case of unknown model orders was also considered and solved using reversible jump MCMC [29]. Recently, this work has been extended to the case of complex and dampened sinusoidal signals in [30]. In [31], Bayesian inference in the sinusoidal model was applied to the analysis of western tonal music.

In the static model in (2), the undampened amplitude α_l and the phase φ_l are assumed to be constant over a segment of N samples. Although this model is widely applicable, the model assumption violates the behavior of many real-world tonal signals. To better model these signals, the model in (2) has been modified in various ways. Typical modifications comprise amplitude and/or phase modulation [32], [33]. the representation of the amplitudes and/or phases as a linear combination of atoms from a suitable basis [34], and autoregressive (AR) frequency parameters [10]. In this paper, we use a dynamic sinusoidal model formulation in which the in-phase and quadrature components in (1) evolve as a first-order Gauss–Markov process. Within the field of econometrics, this class of dynamic models is referred to as stochastic cyclical models [35], [36]. Two slightly different stochastic cyclical models were given a fully Bayesian treatment using MCMC inference techniques in [37] and [38]. Independently, Cemgil *et al.* introduced a dynamic sinusoidal model for the application of polyphonic music transcription in [39]–[41]. In this model, the frequency parameters were discrete random variables, and significant attention was given to the problem of estimating note onset and offset. In the more recent papers [42], [43], Bayesian inference schemes for dynamic sinusoidal models were also considered. Like the proposed inference scheme by Cemgil *et al.*, they base their inference schemes on analytical approximations.

In this paper, we first analyze the dynamic model and discuss its interpretation. In this connection, we show that the in-phase and quadrature components of the dynamic sinusoidal model evolve as first-order Gauss–Markov processes. We also extend the cited work in the previous paragraph by developing an inference scheme for the dynamic sinusoidal model on basis of MCMC inference techniques. Moreover, we consider the more general case in which the frequency parameters are continuous random variables and some of the observations are missing. To achieve this, we develop a Gibbs sampler whose output can be used to form the histograms of the unknown parameters of interest. These histograms have the desirable property that they converge towards the probability distribution of these unknown parameters when the number of generated samples is increased, enabling us to extract statistical features for the model parameters and to perform the interpolation of the missing observations.

The paper is organized as follows. In Section II, we present and analyze the dynamic sinusoidal model. We set up the Bayesian framework for the model in Section III, and the proposed inference scheme based on a Gibbs sampler is derived in Section IV. Four simulations are performed in Section V illustrating the applicability of the model as well as the performance of the sampler and interpolator, and Section VI concludes this paper. The Appendix contains a list of the relevant probability distributions.

II. DYNAMIC SIGNAL MODEL

In the static sinusoidal model in (1), the undampened in-phase and quadrature components are constant throughout the segment of N samples. In the dynamic sinusoidal model, however, this restriction is no longer imposed. Similar to, e.g., [38], [39], we consider a dynamic sinusoidal model given by

$$y_n = \boldsymbol{b}^T \boldsymbol{s}_n + w_n \qquad \text{(observation equation)}$$
$$\boldsymbol{s}_{n+1} = \boldsymbol{A}\boldsymbol{s}_n + \boldsymbol{v}_n \qquad \text{(state equation)} \qquad (3)$$

where $\boldsymbol{s}_n$ is a state vector and $\boldsymbol{v}_n$ is a zero-mean Gaussian state noise[1] vector with covariance matrix

$$\boldsymbol{\Sigma_v} = \text{diag}\left(\sigma_{v,1}^2 \boldsymbol{I}_2, \ldots, \sigma_{v,l}^2 \boldsymbol{I}_2, \ldots, \sigma_{v,L}^2 \boldsymbol{I}_2\right). \qquad (4)$$

The state noise vectors are mutually independent and independent of the observation noise. Further, we have that

$$\boldsymbol{b} = [1 \quad 0 \quad \cdots \quad 1 \quad 0]^T \qquad (5)$$
$$\boldsymbol{A} = \text{diag}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_l, \ldots, \boldsymbol{A}_L) \qquad (6)$$
$$\boldsymbol{A}_l = \rho_l \begin{bmatrix} \cos \omega_l & \sin \omega_l \\ -\sin \omega_l & \cos \omega_l \end{bmatrix}. \qquad (7)$$

Notice that the state equation of (3) decouples into L independent state equations of the form

$$\boldsymbol{s}_{n+1,l} = \boldsymbol{A}_l \boldsymbol{s}_{n,l} + \boldsymbol{v}_{n,l} \qquad (8)$$

---

[1]In this paper, noise is not an unwanted component but a random process of interest. We use the term noise for w_n and $\boldsymbol{v}_n$ since this is common practice when working with the state space model in (3).

due to the block-diagonal structure of $\boldsymbol{A}$ and $\boldsymbol{\Sigma_v}$,

The dynamic model reduces to the static model if there is no state noise. For non-zero state noise, however, the dynamic model models the in-phase and quadrature components as first-order Gauss–Markov processes. In order to see this, we recursively insert the state equation into the observation equation and obtain

$$y_n = \boldsymbol{b}^T \boldsymbol{s}_n + w_n \tag{9}$$

$$= \boldsymbol{b}^T (\boldsymbol{A}\boldsymbol{s}_{n-1} + \boldsymbol{v}_{n-1}) + w_n \tag{10}$$

$$= \boldsymbol{b}^T \boldsymbol{A}^n \left(\boldsymbol{A}^{-1}\boldsymbol{s}_1 + \boldsymbol{A}^{-n} \sum_{k=1}^{n-1} \boldsymbol{A}^{k-1}\boldsymbol{v}_{n-k} \right) + w_n \tag{11}$$

$$= \sum_{l=1}^{L} [i_{n,l} \cos(\omega_l n) + q_{n,l} \sin(\omega_l n)] + w_n \tag{12}$$

where we have defined

$$\begin{bmatrix} i_{n,l} \\ q_{n,l} \end{bmatrix} \triangleq \rho_l^n \left(\boldsymbol{A}_l^{-1}\boldsymbol{s}_{1,l} + \boldsymbol{A}_l^{-n} \sum_{k=1}^{n-1} \boldsymbol{A}_l^{k-1}\boldsymbol{v}_{n-k,l} \right). \tag{13}$$

Equation (12) is of the same form as (1) with one important difference: the in-phase and quadrature components are now time-varying which means that the amplitude and the phases of the polar form of (12) are also time-varying. We analyze the statistical behavior of the time-varying in-phase and quadrature components by introducing the stochastic process defined by $\boldsymbol{z}_{n,l} \triangleq [i_{n,l} \; q_{n,l}]^T$. First, we write $\boldsymbol{z}_{n,l}$ for $n = 1, \ldots, N$ in a recursive way given as

$$\boldsymbol{z}_{n+1,l} = \rho_l \boldsymbol{z}_{n,l} + \left(\rho_l^{-1}\boldsymbol{A}_l\right)^{-(n+1)} \boldsymbol{v}_{n,l} \tag{14}$$

with $\boldsymbol{z}_{1,l} = \rho_l \boldsymbol{A}_l^{-1}\boldsymbol{s}_{1,l}$. If we select a Gaussian distribution for the initial state, i.e., $\boldsymbol{s}_{1,l} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{s}_{1,l}}, \sigma_{\boldsymbol{s}_{1,l}}^2 \boldsymbol{I}_2)$, then $\boldsymbol{z}_{n,l}$ is a first-order Gauss–Markov process. Second, we notice that, the transformation $\tilde{\boldsymbol{v}}_{n,l} = (\rho_l^{-1}\boldsymbol{A}_l)^{-(n+1)} \boldsymbol{v}_{n,l}$ is an orthogonal transformation, and we therefore have that

$$p(\tilde{\boldsymbol{v}}_{n,l}) = p(\boldsymbol{v}_{n,l}) = \mathcal{N}\left(\boldsymbol{0}, \sigma_{v,l}^2 \boldsymbol{I}_2\right). \tag{15}$$

Thus, the statistical behavior of

$$\boldsymbol{z}_{n+1,l} = \rho_l \boldsymbol{z}_{n,l} + \tilde{\boldsymbol{v}}_{n,l} \tag{16}$$

is the same as that of (14). Therefore, $\boldsymbol{z}_{n,l}$ is a very simple first-order Gauss–Markov process evolving independently of the frequency parameter. Further, if we select the mean and variance of the initial state to be $\boldsymbol{\mu}_{\boldsymbol{s}_{1,l}} = \boldsymbol{0}$ and $\sigma_{\boldsymbol{s}_{1,l}}^2 = \sigma_{v,l}^2/(1-\rho_l^2)$, respectively, $\boldsymbol{z}_{n,l}$ is a stationary first-order Gauss–Markov process, i.e., a first-order autoregressive process (AR). Also, our model for the observations in (3) reduces to a simple AR(1) process if $\omega_l = 0$. In summary, the statistical behavior of the dynamic model in (3) is equivalent to that of the model given by[2]

$$\tilde{y}_n = \sum_{l=1}^{L} [\cos(\omega n) \quad \sin(\omega n)] \begin{bmatrix} \tilde{i}_{n,l} \\ \tilde{q}_{n,l} \end{bmatrix} + w_n$$

$$\begin{bmatrix} \tilde{i}_{n+1,l} \\ \tilde{q}_{n+1,l} \end{bmatrix} = \rho_l \begin{bmatrix} \tilde{i}_{n,l} \\ \tilde{q}_{n,l} \end{bmatrix} + \boldsymbol{v}_{n,l}. \tag{17}$$

---

[2]Here, we have introduced $\tilde{\cdot}$ meaning that, e.g., $\tilde{y}_n \neq y_n$ for the same noise realizations although they share the same statistical behavior.

in which the in-phase and quadrature components are explicitly evolving as a first-order Gauss–Markov process. In the model in (3), however, the frequencies have been separated from the time indices. This makes the inference problem for the frequencies more tractable.

We have shown that the in-phase and quadrature components are modeled as first order Gauss–Markov processes in the dynamic model. Unfortunately, it is not easy to make a statistical analysis of the time-varying amplitude and phase since the relationship between these and the in-phase and quadrature components are highly nonlinear. Instead, we make a simulation in Section V which give some insight into this.

III. PROBLEM FORMULATION

As stated in the introduction, we take a Bayesian approach to performing the interpolation and making inference about the unknown parameters of the dynamic sinusoidal model in (3). In the Bayesian approach, these variables are all random variables, and for the model in (3) they are all real and given by

$$
\begin{aligned}
\text{Observations:} \quad & \boldsymbol{y} = [y_1, y_2, \cdots, y_N]^T \\
\text{Latent variables:} \quad & \boldsymbol{S} = [\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_N] \\
\text{Model parameters:} \quad & \boldsymbol{\theta} = \{\boldsymbol{\omega}, \boldsymbol{\rho}, \sigma_v^2, \sigma_w^2\} \\
& \boldsymbol{\omega} = [\omega_1 \quad \omega_2 \quad \cdots \quad \omega_L]^T \\
& \boldsymbol{\rho} = [\rho_1 \quad \rho_2 \quad \cdots \quad \rho_L]^T \\
& \sigma_v^2 = [\sigma_{v,1}^2 \quad \sigma_{v,2}^2 \quad \cdots \quad \sigma_{v,L}^2]^T
\end{aligned}
$$

where $\boldsymbol{s}_n = [\boldsymbol{s}_{n,1}^T, \cdots, \boldsymbol{s}_{n,L}^T]^T$ consists of L two-dimensional state vectors pertaining to the L sinusoids. The evolution of these L two-dimensional state vectors is given by (8). We also assume that R of the elements in $\boldsymbol{y}$ are missing or corrupted, and that we know their indices $\mathcal{I} \subset \{1, \ldots, N\}$. Using this set of indices, we define the vectors $\boldsymbol{y}_{\mathrm{m}} \triangleq \boldsymbol{y}_{\mathcal{I}}$ and $\boldsymbol{y}_{\mathrm{o}} \triangleq \boldsymbol{y}_{\backslash\mathcal{I}}$ containing the R missing or corrupted observations and the $N - R$ valid observations, respectively. The notation $(\cdot)_{\backslash*}$ denotes "without element $*$".

A. Inference Aims

The primary aim is to perform the interpolation of the missing or corrupted samples, i.e., to reconstruct the elements of $\boldsymbol{y}_{\mathrm{m}}$ given the valid observations in $\boldsymbol{y}_{\mathrm{o}}$. In classical statistics, this interpolation task is often solved by using an EM-algorithm which iteratively maximizes the likelihood function $p(\boldsymbol{y}_{\mathrm{o}}|\boldsymbol{y}_{\mathrm{m}})$, whereas the MAP or MMSE estimate of the posterior distribution $p(\boldsymbol{y}_{\mathrm{m}}|\boldsymbol{y}_{\mathrm{o}})$ is often used in Bayesian statistics. For the purpose of interpolating music and speech, however, both of these methods tend to produce over-smoothed interpolants in the sense that they do not agree with the stochastic part of the valid observations [6], [44], [45]. In a Bayesian framework, a much more typical interpolant can be obtained by simply drawing a sample from the posterior distribution $p(\boldsymbol{y}_{\mathrm{m}}|\boldsymbol{y}_{\mathrm{o}})$.

B. Bayesian Inference

The posterior distribution for the missing samples given the valid samples is given by

$$p(\boldsymbol{y}_{\mathrm{m}}|\boldsymbol{y}_{\mathrm{o}}) = \int p(\boldsymbol{y}_{\mathrm{m}}, \boldsymbol{S}, \boldsymbol{\theta}|\boldsymbol{y}_{\mathrm{o}})d\boldsymbol{S}d\boldsymbol{\theta}. \qquad (18)$$

Unfortunately, we are not able to draw a sample directly from $p(\boldsymbol{y}_{\mathrm{m}}|\boldsymbol{y}_{\mathrm{o}})$ since we are not able to integrate the nuisance parameters $\boldsymbol{S}$ and $\boldsymbol{\theta}$ out analytically. However, we can obtain a sample from $p(\boldsymbol{y}_{\mathrm{m}}|\boldsymbol{y}_{\mathrm{o}})$ by taking a single sample from the joint posterior distribution $p(\boldsymbol{y}_{\mathrm{m}}, \boldsymbol{S}, \boldsymbol{\theta}|\boldsymbol{y}_{\mathrm{o}})$ and simply ignore the generated values for $\boldsymbol{s}$ and $\boldsymbol{\theta}$. The joint posterior distribution can be written as

$$p(\boldsymbol{y}_{\mathrm{m}}, \boldsymbol{S}, \boldsymbol{\theta}|\boldsymbol{y}_{\mathrm{o}}) = p(\boldsymbol{y}_{\mathrm{m}}|\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{y}_{\mathrm{o}})p(\boldsymbol{S}, \boldsymbol{\theta}|\boldsymbol{y}_{\mathrm{o}}) \qquad (19)$$

where $p(\boldsymbol{y}_{\mathrm{m}}|\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{y}_{\mathrm{o}})$ is known from the observation equation of (3). Thus, in order to generate a sample for $\boldsymbol{y}_{\mathrm{m}}$ the only problem left is computing $p(\boldsymbol{S}, \boldsymbol{\theta}|\boldsymbol{y}_{\mathrm{o}})$. By Bayes' theorem we may write it as

$$p(\boldsymbol{S}, \boldsymbol{\theta}|\boldsymbol{y}_{\mathrm{o}}) = \frac{p(\boldsymbol{y}_{\mathrm{o}}, \boldsymbol{S}_{\backslash 1}|\boldsymbol{s}_1, \boldsymbol{\theta})p(\boldsymbol{s}_1, \boldsymbol{\theta})}{p(\boldsymbol{y}_{\mathrm{o}})} \qquad (20)$$

where $p(\boldsymbol{y}_{\mathrm{o}}, \boldsymbol{S}_{\backslash 1}|\boldsymbol{s}_1, \boldsymbol{\theta})$, $p(\boldsymbol{s}_1, \boldsymbol{\theta})$ and $p(\boldsymbol{y}_{\mathrm{o}})$ are referred to as the likelihood, the prior and the model evidence, respectively. The likelihood can be factored as

$$p(\boldsymbol{y}_{\mathrm{o}}, \boldsymbol{S}_{\backslash 1}|\boldsymbol{s}_1, \boldsymbol{\theta}) = p(\boldsymbol{y}_{\mathrm{o}}|\boldsymbol{S}_{\backslash \mathcal{I}}, \boldsymbol{\theta}) \prod_{n=1}^{N-1} p(\boldsymbol{s}_{n+1}|\boldsymbol{s}_n, \boldsymbol{\theta}) \qquad (21)$$

which from (3) is seen to be a product of normal distributions. Since the state equation of (3) decouples into L independent state equations as in (8), we can factor the normal distribution $p(\boldsymbol{s}_{n+1}|\boldsymbol{s}_n, \boldsymbol{\theta})$ into L bivariate normal distributions given by

$$p(\boldsymbol{s}_{n+1}|\boldsymbol{s}_n, \boldsymbol{\theta}) = \prod_{l=1}^{L} p\left(\boldsymbol{s}_{n+1,l}|\boldsymbol{s}_{n,l}, \sigma_{v,l}^2, \omega_l, \rho_l\right). \qquad (22)$$

The form of the prior is considered in Section III-D. Implicit in the formulation of (20) is the model assumption which we consider as known (including its order L).[3] The model evidence in the denominator of (20) acts therefore as a mere scale factor since it is independent of $\boldsymbol{S}$ and $\boldsymbol{\theta}$. To reflect this, we simply write Bayes' theorem as

$$p(\boldsymbol{S}, \boldsymbol{\theta}|\boldsymbol{y}_{\mathrm{o}}) \propto p(\boldsymbol{y}_{\mathrm{o}}, \boldsymbol{S}_{\backslash 1}|\boldsymbol{s}_1, \boldsymbol{\theta})p(\boldsymbol{s}_1, \boldsymbol{\theta}) \qquad (23)$$

where $\propto$ denotes "proportional to."

The joint posterior distribution encapsulates all knowledge about the states and model parameters by combining the prior knowledge with the information in the observed data through Bayes' theorem. Theoretically, it is also possible to derive posterior distributions, moments, probability intervals and other posterior characteristics for the individual variables by use of marginalization and various transformations. In practice, however, it is often either infeasible or impossible to compute these posterior characteristics, and we have to rely on numerical

[3]This assumption is quite common although not very realistic.

inference methods. The stochastic numerical methods offer various ways of generating samples from the posterior distribution. These samples are then used to form histograms which converge to the true posterior distributions for an increasing sample size. For an overview of some of the methods see, e.g., [22], [23], [46].

C. Markov Chain Monto Carlo Sampling

Markov chain Monto Carlo (MCMC) methods are currently a very popular class of stochastic sampling methods adopted by the Bayesian community in the late 1980s [47]. They work by selecting the transition kernel of an ergodic Markov chain such that the invariant distribution of the Markov chain is the desired posterior distribution which we wish to draw samples from. After an initial transient period in which the Markov chain converges, samples generated by the Markov chain are distributed according to the desired distribution. The two most well-known MCMC sampling schemes are the Metropolis–Hastings (MH) algorithm [48]–[50] and the Gibbs sampler [51]. In the MH algorithm, samples generated from the desired posterior distribution, say $p(\boldsymbol{x})$, which we know up to some normalizing constant Z with $p(\boldsymbol{x}) = \tilde{p}(\boldsymbol{x})/Z$, are generated by use of a user-defined proposal distribution $q(\boldsymbol{x}|\boldsymbol{x}^{[k]})$, where $\boldsymbol{x}^{[k]}$ is the current state of the Markov chain. By construction, $q(\boldsymbol{x}|\boldsymbol{x}^{[k]})$ is selected as a tradeoff between how similar it is to $p(\boldsymbol{x})$ and how easy it is to generate samples from. A candidate sample $\boldsymbol{x}' \sim q(\boldsymbol{x}|\boldsymbol{x}^{[k]})$ is accepted as the next state $\boldsymbol{x}^{[k+1]}$ with probability

$$\alpha\left(\boldsymbol{x}^{[k]}, \boldsymbol{x}'\right) = \min\left[1, \frac{\tilde{p}(\boldsymbol{x}')q\left(\boldsymbol{x}^{[k]}|\boldsymbol{x}'\right)}{\tilde{p}\left(\boldsymbol{x}^{[k]}\right)q\left(\boldsymbol{x}'|\boldsymbol{x}^{[k]}\right)}\right]. \qquad (24)$$

Otherwise, the current state of the Markov chain is retained. The Gibbs sampler is a special case of the MH-algorithm in which sampling from the multivariate distribution $p(\boldsymbol{x}) = p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K)$ is broken up into alternating sampling from the K lower dimensional conditional distribution $p(\boldsymbol{x}_k|\boldsymbol{x}_{\backslash k})$. Specifically, for the k's iteration, we sample for $k = 1, \ldots, K$ from

$$\boldsymbol{x}_k^{[k+1]} \sim p\left(\boldsymbol{x}_k|\boldsymbol{x}_1^{[k+1]}, \ldots, \boldsymbol{x}_{k-1}^{[k+1]}, \boldsymbol{x}_{k+1}^{[k]}, \ldots, \boldsymbol{x}_K^{[k]}\right). \qquad (25)$$

The generated samples from these conditional distributions are always accepted.

D. Prior Distributions

To complete the Bayesian setup, we need to specify prior distributions on the initial state as well as on the model parameters. In this paper, we assume that we have only vague prior information about the parameters whose joint prior distribution factor as

$$p(\boldsymbol{s}_1, \boldsymbol{\theta}) = p(\boldsymbol{s}_1)p(\boldsymbol{\omega})p(\boldsymbol{\rho})p\left(\sigma_v^2 v\right)p\left(\sigma_w^2\right)$$
$$= \left[\prod_{l=1}^{L} p(\boldsymbol{s}_{1,l})p(\omega_l)p(\rho_l)p\left(\sigma_{v,l}^2\right)\right]p\left(\sigma_w^2\right). \qquad (26)$$

For the joint distribution of the lth frequency parameter and damping coefficient, we use the Jeffreys' prior for the likelihood in (22), i.e., $p(\omega_l, \rho_l) = p(\omega_l)p(\rho_l) \propto \rho_l$ for $\omega_l \in [0, \pi]$ and $\rho_l > 0$. It is common to restrict the damping coefficient to be smaller than one since this ensures that the evolution of
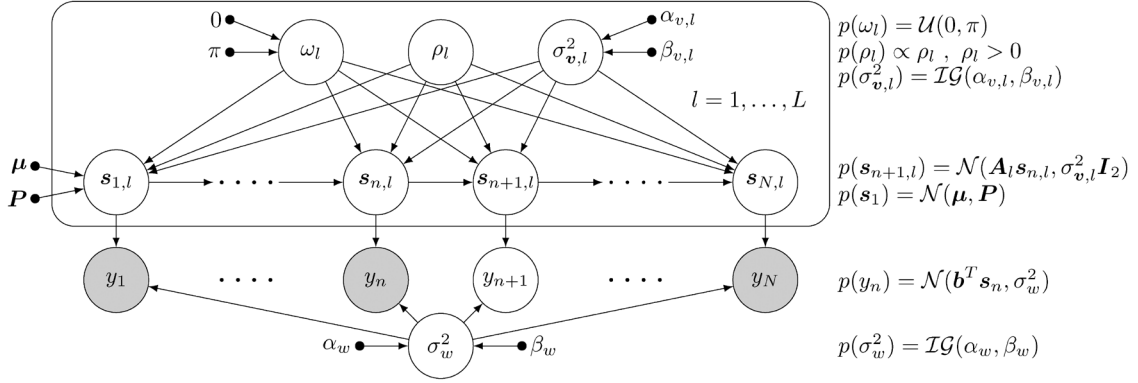
Fig. 1.   Directed graphical model for the parameter estimation and interpolation problem. Shaded nodes denote observed variables (notice that y_{n+1} is a missing observation), open circles denote latent variables, and smaller solid circles denote deterministic hyperparameters. The box denotes a plate, and it is labelled with $l = 1, \ldots, L$ indicating that there are L copies of the nodes inside.

the in-phase and quadrature components in (14) is stable. This yields a beta prior distribution with parameters 2 and 1 on the damping coefficient. In this paper, however, we do not impose this restriction since we wish to model nonstationary signal segments. The selected prior on the frequency parameters causes symmetry in the likelihood of the model parameters which leads to the problem of label switching [52]. More precisely, the joint posterior distribution is a mixture distribution of $L!$ similar distributions up to a permutation of labels [28]. For the interpolation of missing samples, which is the primary focus of this paper, this is not a problem. For making inference about the unknown parameters; however, the problem can be addressed by ensuring identifiability of the frequency parameters through a joint prior distribution on the frequency parameters given by

$$p(\boldsymbol{\omega}) \propto \mathbb{I}_{[0 \leq \omega_1 \leq \omega_2 \leq \cdots \leq \omega_L \leq \pi]}(\boldsymbol{\omega}) \qquad (27)$$

where $\mathbb{I}_{[A]}(\cdot)$ is the indicator function on the region A. Alternatively, the generated samples can also be postprocessed by applying various clustering techniques to the generated frequency parameters [52].

For the observation and state noise variances, we use inverse gamma distributions, i.e., $p(\sigma_w^2) = \mathcal{IG}(\alpha_w, \beta_w)$ and $p(\sigma_{\boldsymbol{v},l}^2) = \mathcal{IG}(\alpha_{v,l}, \beta_{v,l})$. These distributions can be made diffuse by choosing small values for the hyperparameters. They can also be used for preventing the noise variances from collapsing to zero which is often a necessary requirement in MCMC-based inference [5]. For the initial state distribution, we assume a normal distribution, i.e., $p(\boldsymbol{s}_1) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{P})$.

IV. DERIVATION OF INFERENCE SCHEME

The Bayesian model considered in the previous section is summarized in the directed graphical model in Fig. 1. The figure clearly reveals the assumptions, the conditional dependency between the variables, and the hierarchical structure to the setup also given by likelihood in (21) and (22), and the prior in (26). In our inference scheme for the variables of the model, we draw samples from the joint posterior distribution $p(\boldsymbol{S}, \boldsymbol{\theta}|\boldsymbol{y}_{\mathrm{o}})$ by means of a Gibbs sampler. As detailed in Section III-C,

we therefore have to group the variables into suitable blocks and derive conditional distributions for them. In this paper, we consider the following two conditional distributions given by

$$\text{States:} \qquad p(\boldsymbol{S}|\boldsymbol{\theta}, \boldsymbol{y}_{\mathrm{o}}) \qquad (28)$$

$$\text{Model parameters:} \qquad p(\boldsymbol{\theta}|\boldsymbol{S}, \boldsymbol{y}_{\mathrm{o}}). \qquad (29)$$

The selected grouping of variables in (28) and (29) leads to a set of conditional distributions which are fairly easy to sample from. Further, by sampling all model parameters in a single step, we increase the mixing properties of the sampler, i.e., we decrease the correlation of the generated samples leading to faster convergence of the underlying Markov chain. In the next section, we derive the particular form of these conditional distributions.

A. States

The conditional state distribution in (28) is a multivariate Gaussian distribution. However, the dimension of a sample from this distribution is $2LN \times 1$ which would render direct sampling from it infeasible for most applications. Instead, we use the simulation smoother for drawing samples from (28). The simulation smoother is an efficient sampling scheme using standard Kalman smoothing, and it is easily modified to handle the case of missing observations since this corresponds to skipping the update step of the build-in Kalman filter for these samples. The simulation smoother exists in several versions of which we use the version in [53] (see, e.g., [54]–[56] for other versions of the simulation smoother).

B. Model Parameters

Since the model parameter σ_w^2 of the observation equation and the L sets of model parameters $(\omega_l, \rho_l, \sigma_{\boldsymbol{v},l}^2)$ of the state equation are mutually independent conditioned on the states $\boldsymbol{S}$, we can factor (29) as

$$p(\boldsymbol{\theta}|\boldsymbol{S}, \boldsymbol{y}_{\mathrm{o}}) = \left[\prod_{l=1}^{L} p\left(\omega_l, \rho_l, \sigma_{v,l}^2|\boldsymbol{S}\right) \right] p\left(\sigma_w^2|\boldsymbol{S}, \boldsymbol{y}_{\mathrm{o}}\right). \qquad (30)$$

Thus, sampling from the conditional distribution in (29) can be done by sampling the $L + 1$ conditional distributions on the right-hand side of (30) independently.

1) Frequency, Damping and State Noise Variance: The main difficulty of our Gibbs sampler is to draw samples from the joint conditional distribution of the frequency parameter, the damping coefficient, and the state noise variance given the states, i.e., $p(\omega_l, \rho_l, \sigma_{v,l}^2|\boldsymbol{S})$. To our knowledge, it is not possible to sample directly from $p(\omega_l, \rho_l, \sigma_{v,l}^2|\boldsymbol{S})$—although we come close in this paper. A Gibbs sampling scheme on the individual parameters is also not straight-forward since it suffers from poor mixing and since the lth damping coefficient conditioned on the lth frequency parameter and state noise variance has a nonstandard distribution. In order to improve mixing, we therefore propose sampling all parameters at once from $p(\omega_l, \rho_l, \sigma_{v,l}^2|\boldsymbol{S})$ by use of an MH-sampler previously discussed in Section III-C. For the proposed MH-sampler the candidate samples are easy to generate and the acceptance probability turns out to be very easy to evaluate.

Given the states, the posterior distribution for the lth set of model parameters of the state equation can be written as

$$p\left(\omega_l, \rho_l, \sigma_{v,l}^2|\boldsymbol{S}\right) \propto \left[\prod_{n=1}^{N-1} p\left(\boldsymbol{s}_{n+1,l}|\boldsymbol{s}_{n,l}, \sigma_{v,l}^2, \omega_l, \rho_l\right)\right]$$
$$\times p\left(\omega_l, \rho_l, \sigma_{v,l}^2\right) \quad (31)$$

where $p(\omega_l, \rho_l, \sigma_{v,l}^2)$ is the prior distribution which, as stated in Section III-D, factors into

$$p\left(\omega_l, \rho_l, \sigma_{v,l}^2\right) = p(\omega_l)p(\rho_l)p\left(\sigma_{v,l}^2\right).$$

The distribution for $p(\boldsymbol{s}_{n+1,l}|\boldsymbol{s}_{n,l}, \sigma_{v,l}^2, \omega_l, \rho_l)$ is a bivariate normal distribution and the product over n of $N-1$ of these can therefore be written as

$$\prod_{n=1}^{N-1} p\left(\boldsymbol{s}_{n+1,l}|\boldsymbol{s}_{n,l}, \sigma_{v,l}^2, \omega_l, \rho_l\right) \propto \sigma_{v,l}^{2}{}^{-(N-1)}$$
$$\times \exp\left\{\frac{-1}{2\sigma_{v,l}^2}\sum_{n=1}^{N-1}(\boldsymbol{s}_{n+1,l} - \boldsymbol{A}_l\boldsymbol{s}_{n,l})^T(\boldsymbol{s}_{n+1,l} - \boldsymbol{A}_l\boldsymbol{s}_{n,l})\right\}. \quad (32)$$

In order to write this distribution in a useful way in terms of the frequency parameter and the damping coefficient, we rewrite $\boldsymbol{A}_l\boldsymbol{s}_{n,l}$ into

$$\boldsymbol{A}_l\boldsymbol{s}_{n,l} = \begin{bmatrix}\boldsymbol{s}_{n,l} & \boldsymbol{s}_{n,l}^\perp\end{bmatrix}\boldsymbol{a}_l \quad (33)$$

where $\boldsymbol{s}_{n,l}^\perp$ is obtained by a 90° clockwise rotation of $\boldsymbol{s}_{n,l}$ and

$$\boldsymbol{a}_l \triangleq \rho_l[\cos\omega_l \quad \sin\omega_l]^T. \quad (34)$$

Inserting this into (32) and replacing the summation with an inner product yields

$$\prod_{n=1}^{N-1} p\left(\boldsymbol{s}_{n+1,l}|\boldsymbol{s}_{n,l}, \sigma_{v,l}^2, \omega_l, \rho_l\right) = \mathcal{N}\left(\boldsymbol{\varphi}_l; \boldsymbol{\Phi}_l\boldsymbol{a}_l, \sigma_{v,l}^2\boldsymbol{I}_{2(N-1)}\right)$$

where we have defined

$$\boldsymbol{\varphi}_l \triangleq \begin{bmatrix}\boldsymbol{s}_{2,l}^T & \boldsymbol{s}_{3,l}^T & \cdots & \boldsymbol{s}_{N,l}^T\end{bmatrix}^T \quad (35)$$

$$\boldsymbol{\phi}_l \triangleq \begin{bmatrix}\boldsymbol{s}_{1,l}^T & \boldsymbol{s}_{2,l}^T & \cdots & \boldsymbol{s}_{N-1,l}^T\end{bmatrix}^T \quad (36)$$

$$\tilde{\boldsymbol{\phi}}_l \triangleq \begin{bmatrix}\left(\boldsymbol{s}_{1,l}^\perp\right)^T & \left(\boldsymbol{s}_{2,l}^\perp\right)^T & \cdots & \left(\boldsymbol{s}_{N-1,l}^\perp\right)^T\end{bmatrix}^T \quad (37)$$

$$\boldsymbol{\Phi}_l \triangleq \begin{bmatrix}\boldsymbol{\phi}_l & \tilde{\boldsymbol{\phi}}_l\end{bmatrix}. \quad (38)$$

Now, by assuming a non-informative prior for $\boldsymbol{a}_l$ of the form

$$p\left(\boldsymbol{a}_l|\sigma_{v,l}^2\right) = \mathcal{N}\left(\boldsymbol{0}, \sigma_v^2\delta\boldsymbol{I}_2\right) \qquad \text{with } \delta \to \infty \quad (39)$$

and by using standard Bayesian inference for the linear model [21], we obtain after some algebra

$$p\left(\boldsymbol{a}_l, \sigma_{v,l}^2|\boldsymbol{S}\right) \propto p\left(\boldsymbol{S}|\boldsymbol{a}_l, \sigma_{v,l}^2\right)p\left(\boldsymbol{a}_l|\sigma_{v,l}^2\right)p\left(\sigma_{v,l}^2\right) \quad (40)$$
$$\propto \mathcal{NIG}\left(\boldsymbol{\mu}_{\boldsymbol{a},l}, \sigma_{\boldsymbol{a},l}^2\boldsymbol{I}_2, \alpha_{q,l}, \beta_{q,l}\right). \quad (41)$$

where the parameters of the normal-scaled inverse gamma distribution are defined by

$$\sigma_{\boldsymbol{a},l}^2 \triangleq \left(\boldsymbol{\phi}_l^T\boldsymbol{\phi}_l\right)^{-1} \quad (42)$$

$$\boldsymbol{\mu}_{\boldsymbol{a},l} \triangleq \sigma_{\boldsymbol{a},l}^2\boldsymbol{\Phi}_l^T\boldsymbol{\varphi}_l \quad (43)$$

$$\alpha_{\sigma_{v,l}^2} \triangleq \alpha_{v,l} + N - 1 \quad (44)$$

$$\beta_{\sigma_{v,l}^2} \triangleq \beta_{v,l} + \left(\boldsymbol{\varphi}_l^T\boldsymbol{\varphi}_l - \sigma_{\boldsymbol{a},l}^{-2}\boldsymbol{\mu}_{\boldsymbol{a},l}^T\boldsymbol{\mu}_{\boldsymbol{a},l}\right)/2. \quad (45)$$

The Jacobian determinant of the transformation from $\boldsymbol{a}_l$ to (ω_l, ρ_l) is given by

$$\begin{vmatrix}\dfrac{\partial\boldsymbol{a}_l}{\partial\rho_l} & \dfrac{\partial\boldsymbol{a}_l}{\partial\omega_l}\end{vmatrix} = \begin{vmatrix}\cos\omega_l & -\rho_l\sin\omega_l \\ \sin\omega_l & \rho_l\cos\omega_l\end{vmatrix} = \rho_l,$$

which is proportional to the prior distribution on the damping coefficient. Therefore, we may write (41) as

$$q\left(\omega_l, \rho_l, \sigma_{v,l}^2|\boldsymbol{S}\right) \propto p(\rho_l)p\left(\boldsymbol{a}_l, \sigma_{v,l}^2|\boldsymbol{S}\right) \quad (46)$$

with $\boldsymbol{a}_l$ replaced by the expression in (34). Thus, the distribution $q(\omega_l, \rho_l, \sigma_{v,l}^2|\boldsymbol{S})$ is nearly identical to the desired distribution $p(\omega_l, \rho_l, \sigma_{v,l}^2|\boldsymbol{S})$ in (31). The only difference between the two distributions is that the frequency parameter of $q(\omega_l, \rho_l, \sigma_{v,l}^2|\boldsymbol{S})$ is uniform on $[-\pi, \pi]$ whereas is it uniform on $[0, \pi]$ in $p(\omega_l, \rho_l, \sigma_{v,l}^2|\boldsymbol{S})$. In order to remedy for this, we use $q(\omega_l, \rho_l, \sigma_{vl}^2|\boldsymbol{S})$ as a proposal distribution in an MH-sampler. We draw a sample from this proposal by first sampling a set $(\boldsymbol{a}_l', \sigma_{v,l}^2{}')$ from the bivariate normal-scaled inverse gamma distribution in (41). Sampling from the bivariate normal-scaled inverse gamma distribution can be done in various ways. Here, we sample from its marginal densities given by

$$p\left(\sigma_{v,l}^2|\boldsymbol{S}\right) = \mathcal{IG}\left(\alpha_{\sigma_{v,l}^2}, \beta_{\sigma_{v,l}^2}\right) \quad (47)$$

$$p(\boldsymbol{a}_l|\boldsymbol{S}) = \text{St}\left(\boldsymbol{\mu}_{\boldsymbol{a},l}, \frac{\beta_{\sigma_{v,l}^2}}{\alpha_{\sigma_{v,l}^2}}\sigma_{\boldsymbol{a},l}^2\boldsymbol{I}_2, 2\alpha_{\sigma_{v,l}^2}\right). \quad (48)$$

This is done by sampling from [46]

$$\sigma_{v,l}^2{}' \sim \mathcal{IG}\left(\alpha_{\sigma_{v,l}^2}, \beta_{\sigma_{v,l}^2}\right) \quad (49)$$

$$\tau_l' \sim \mathcal{IG}\left(\alpha_{\sigma_{v,l}^2}, 1/2\right) \quad (50)$$

$$\boldsymbol{a}_l' = [a_{1,l}' \quad a_{2,l}']^T \sim \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{a},l}, 2\beta_{\sigma_{v,l}^2}\tau_l'\sigma_{\boldsymbol{a},l}^2\boldsymbol{I}_2\right). \quad (51)$$

Second, we transforms the generated sample $(\boldsymbol{a}'_l, \sigma^2_{v,l}{}')$ into $(\omega'_l, \rho'_l, \sigma^2_{\boldsymbol{v},l}{}')$ by use of the transformation

$$\omega'_l = \arctan\left(a'_{2,l}/a'_{1,l}\right) \tag{52}$$

$$\rho'_l = \sqrt{\boldsymbol{a}'^T_l \boldsymbol{a}'_l}. \tag{53}$$

Finally, the samples generated by this proposal distribution are accepted with probability

$$\alpha\left(\left(\omega_l, \rho_l, \sigma^2_{v,l}\right), \left(\omega'_l, \rho'_l, \sigma^2_{v,l}{}'\right)\right)$$

$$= \min\left[1, \frac{p\left(\boldsymbol{a}'_l, \sigma^2_{v,l}{}'|\boldsymbol{S}\right) p\left(\omega'_l\right) p\left(\rho'_l\right) q\left(\omega_l, \rho_l, \sigma^2_{v,l}|\boldsymbol{S}\right)}{p(\boldsymbol{a}_l, q_l|\boldsymbol{S}) p(\omega_l) p(\rho_l) q\left(\omega'_l, \rho'_l, \sigma^2_{v,l}{}'|\boldsymbol{S}\right)}\right]$$

$$= \min\left[1, \frac{p\left(\omega'_l\right)}{p\left(\omega_l\right)}\right]$$

$$= \mathbb{I}_{[0,\pi]}\left(\omega'_l\right) \tag{54}$$

If the sample is not accepted, the previous values $(\omega_l, \rho_l, \sigma^2_{\boldsymbol{v},l})$ are retained. In the case where we use the structured prior on the frequency parameters, the indicator function should be changed to $\mathbb{I}_{[\omega_{l-1}, \omega_{l+1}]}(\omega'_l)$.

2) Observation Noise Variance: By Bayes' theorem, we can write $p(\sigma^2_w|\boldsymbol{S}, \boldsymbol{y}_\circ)$ as

$$p\left(\sigma^2_w|\boldsymbol{S}, \boldsymbol{y}_\circ\right) \propto p\left(\boldsymbol{y}_\circ|\boldsymbol{S}_{\backslash\mathcal{I}}, \sigma^2_w\right) p\left(\sigma^2_w\right) \tag{55}$$

where $p(\boldsymbol{y}_\circ|\boldsymbol{S}_{\backslash\mathcal{I}}, \sigma^2_w)$ is the likelihood of the observation equation in (3) and $p(\sigma^2_w)$ is the prior distribution for σ^2_w. Since $p(\boldsymbol{y}_\circ|\boldsymbol{S}_{\backslash\mathcal{I}}, \sigma^2_w) = \mathcal{N}(\boldsymbol{S}^T_{\backslash\mathcal{I}}\boldsymbol{b}, \sigma^2_w \boldsymbol{I}_{N-R})$ and $p(\sigma^2_w) = \mathcal{IG}(\alpha_w, \beta_w)$, the posterior distribution $p(\sigma^2_w|\boldsymbol{S}_{\backslash\mathcal{I}}, \boldsymbol{y}_\circ)$ is an inverse gamma distribution, $\mathcal{IG}(\sigma^2_w; \alpha_{\sigma^2_w}, \beta_{\sigma^2_w})$, with parameters

$$\alpha_{\sigma^2_w} = \alpha_w + N/2 \tag{56}$$

$$\beta_{\sigma^2_w} = \beta_w + \frac{1}{2}\left(\boldsymbol{y}_\circ - \boldsymbol{S}^T_{\backslash\mathcal{I}}\boldsymbol{b}\right)^T \left(\boldsymbol{y}_\circ - \boldsymbol{S}^T_{\backslash\mathcal{I}}\boldsymbol{b}\right). \tag{57}$$

C. Summary of Inference Scheme

Table I summarises our proposed Gibbs sampler for generating samples from $p(\boldsymbol{S}, \boldsymbol{\theta}|\boldsymbol{y}_\circ)$. The computational complexity of the algorithm is fairly high primarily due to the generation of the states by the simulation smoother. In our implementation with $N = 600$ observations and $L = 6$ sinusoids, it takes approximately 40 ms for generating a state sample $\boldsymbol{S}^{[k]}$. This corresponds to nearly 97% of the time consumption of one iteration of the Gibbs sampler. For the application of interpolation, we only need a single sample for the states and model parameters from the invariant distribution of the underlying Markov chain of the sampler. Once these have been generated, we may perform the interpolation by simulating from the observation equation of (3). Therefore, the computational complexity of the algorithm heavily depends on proper initialisation and the convergence speed of the chain.

TABLE I
SUMMARY OF PROPOSED GIBBS SAMPLER FOR
GENERATING SAMPLES FROM $p(\boldsymbol{S}, \boldsymbol{\theta}|\boldsymbol{y}_\circ)$

1) Select hyperparameters and initialise the Gibbs sampler.
2) Repeat for $k = 0, 1, 2, \ldots, K$
   a) $\boldsymbol{S}^{[k+1]} \sim p(\boldsymbol{S}|\boldsymbol{\theta}^{[k]}, \boldsymbol{y}_\circ)$ (simulation smoother)
   b) Repeat for $l = 1, 2, \ldots, L$
     i) $\sigma^2_{\boldsymbol{v},l}{}' \sim \mathcal{IG}(\alpha^{[k]}_{\sigma^2_{\boldsymbol{v},l}}, \beta^{[k]}_{\sigma^2_{\boldsymbol{v},l}})$
     ii) $\tau'_l \sim \mathcal{IG}(\alpha^{[k]}_{\sigma^2_{\boldsymbol{v},l}}, 1/2)$
     iii) $\boldsymbol{a}'_l \sim \mathcal{N}(\boldsymbol{\mu}^{[k]}_{\boldsymbol{a},l}, 2\beta^{[k]}_{\sigma^2_{\boldsymbol{v},l}} \tau'_l \sigma^2_{\boldsymbol{a},l}{}^{[k]} \boldsymbol{I}_2)$
     iv) $\omega'_l = \arctan(a'_{2,l}/a'_{1,l})$
     v) $\rho'_l = \sqrt{\boldsymbol{a}'^T_l \boldsymbol{a}'_l}$
     vi) if $\omega'_l > 0$
        • $(\omega^{[k+1]}_l, \rho^{[k+1]}_l, \sigma^2_{\boldsymbol{v},l}{}^{[k+1]}) = (\omega'_l, \rho'_l, \sigma^2_{\boldsymbol{v},l}{}')$
        else
        • $(\omega^{[k+1]}_l, \rho^{[k+1]}_l, \sigma^2_{\boldsymbol{v},l}{}^{[k+1]}) = (\omega^{[k]}_l, \rho^{[k]}_l, \sigma^2_{\boldsymbol{v},l}{}^{[k]})$
   c) $\sigma^{2\,[k+1]}_w \sim \mathcal{IG}(\alpha_{\sigma^2_w}, \beta^{[k]}_{\sigma^2_w})$

V. SIMULATIONS

In this section, we first demonstrate that the dynamic signal model is able to model signals with amplitude and frequency modulation. These phenomena are encountered in real-world signals. Second, we illustrate the proposed inference scheme on a synthetic signal and apply it to the application of reconstructing missing or corrupted audio packets on a packet-based network.[4] In our simulations, we use the following common setup for our Gibbs sampler. We use non-informative prior distributions with hyperparameters

$$[\boldsymbol{\mu} \quad \boldsymbol{P} \quad \alpha_i \quad \beta_i] = [\boldsymbol{0} \quad 10\boldsymbol{I}_2 \quad 0 \quad 10^{-5}].$$

where $i = \{v, w\}$. The small nonzero value for β_i is selected in order to prevent the noise variances from collapsing to zero. The Gibbs sampler is iterated 10 000 times and samples from the first 1000 iterations are discarded as burn-in samples. The initial values for the frequency and observation noise variance are found by using a simple matching pursuit algorithm [57]. The initial value for the damping coefficient and state noise variance are set to 1 and $\sigma^{2[0]}_w/10$, respectively. For the model order, we use $L = 1$ in the two examples with synthetic signals and $L = 6$ in the two examples with real-world signals.

A. Applicability of the Model

The static model in (2) is very useful for modeling the periodic parts of a signal. However, since the phase and frequency are modeled as constants and the amplitude with an exponentially decaying envelope, the static model is in general not able to capture common phenomena such as amplitude and frequency modulation [10]. As discussed in Section II, the dynamic model allows the in-phase and quadrature components to develop as a first order Gauss–Markov process. Thus, the

---

[4]The MATLAB code and audio samples used in the simulations can be obtained from http://kom.aau.dk/~jkn/publications/publications.php.
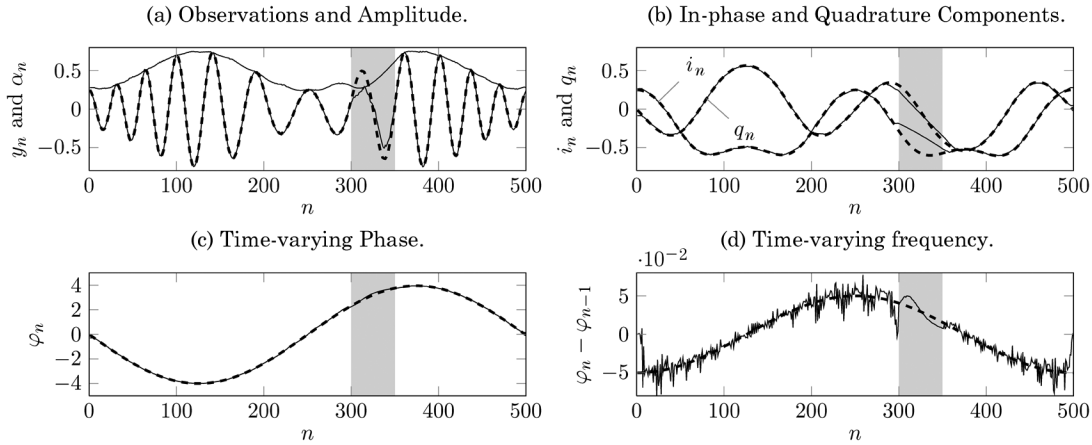
Fig. 2. Synthetic signal with both amplitude and phase modulation. The shaded area indicates a missing section of the signal. Plot (a) shows the signal (dashed line), a state generated by our Gibbs sampler, and the average amplitude. Plot (b) shows the true (dashed lines) and the average in-phase and quadrature components. Plot (c) and (d) show the true (dashed lines) and the average phase and time-varying part of the frequency. (a) Observations and apmplitude. (b) In-phase and quadrature components. (c) Time-varying phase. (d) Time-varying frequency.

model also allows the amplitude, the phase and hence the frequency to be time-varying. These are given by

$$\alpha_{n,l} = \rho_l^n \sqrt{i_{n,l}^2 + q_{n,l}^2} \tag{58}$$

$$\varphi_{n,l} = \arctan(q_{n,l}/i_{n,l}) \tag{59}$$

$$\omega_{n,l} = \omega_l + d\varphi_{t,l}/dt|_{t=nT} \tag{60}$$

where the time-varying frequency $\omega_{n,l}$ is a sum of the frequency ω_l from the dynamic model in (3) and the sampled derivative of the continuous-time phase $\varphi_{t,l}$.

In Fig. 2(a), we have shown a synthetic signal consisting of a single sinusoid with both sinusoidal amplitude and frequency modulation (dashed line). The signal consists of $N = 500$ samples and is given by

$$x_n = \alpha_n \cos(\theta_n) + w_n \tag{61}$$

$$\alpha_n = 0.5 + 0.25 \sin(4\pi n/N - \pi/2) \tag{62}$$

$$\theta_n = 0.15n - 0.05 \sum_{m=1}^{n} \sin(2\pi m/N - \pi/2) \tag{63}$$

where w_n is white Gaussian noise with variance 10^{-6}. The samples from index 300 to index 350 were removed and considered to be missing samples. We used the proposed inference scheme for analysing the signal x_n, and the full line on top of the dashed line in Fig. 2(a) shows a state vector generated by our Gibbs sampler. For all generated state samples, we also demodulated the states in order to obtain the samples for the in-phase and quadrature components. Based on these samples, we calculated the average amplitude $\alpha_{n,l}$, the average in-phase and quadrature components, the average phase $\varphi_{n,l}$ and the average derivative of the phase as $\varphi_{n,l} - \varphi_{n-1,l}$. The latter is an approximation to the derivative of the phase. These averages (full lines) are compared against their true values (dashed lines) in Fig. 2(a)–(d), respectively. Clearly, the model is able to capture both amplitude and frequency modulation. However, the figures also reveal a potential problem for the application of interpolating missing samples; In this example, the in-phase and quadrature components do not evolve as a typical Gauss–Markov process. There-

fore, we cannot expect the interpolation to be very successful since our interpolation scheme, on average, will reconstruct the missing samples in the in-phase and quadrature components with a straight line.

B. Synthetic Signal

We consider a simple synthetic signal generated by the static sinusoidal model. We do this in order to illustrate some of the features of the proposed Bayesian inference scheme and interpolator. Specifically, we generated $N = 500$ observations from the static model in (1) with a single sinusoidal component with parameters

$$\begin{bmatrix} \alpha & \beta & \omega & \rho & \sigma_w^2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0.2 & 0.997 & 0.01 \end{bmatrix}.$$

We also removed 50% of the observations distributed over three sections as illustrated in Fig. 3(e). Fig. 3 shows the results of running the Gibbs sampler. In Fig. 3(a)–3(d), the traces of the 10 000 generated samples for the model parameters are shown, and Fig. 3(e) shows the results of interpolating the sections of missing observations. The underlying Markov chain seems to have converged to its invariant distribution after approximately 500 samples. The histograms in the margin of the first four plots are based on the last 9000 generated samples. They are an approximation to the marginal distribution for the individual model parameters, and they converge to it for an increasing number of iterations of the Gibbs sampler. As previously discussed in Section IV, the histograms can be used for summarizing various posterior features such as point and interval estimates. For example, computing their means yields the estimates $\hat{\omega} = 0.1999$, $\hat{\rho} = 0.997 \cdot 10^{-3}$, $\hat{q} = 7.044 \cdot 10^{-6}$, and $\hat{\sigma}_w^2 = 9.606 \cdot 10^{-3}$.

In Fig. 3(e), the three interpolation sections are shown with a shaded background. In all three simulation sections, we have shown the 95% credible interval for the missing observations.[5] In the last interpolation section, we have used the mean estimate of the interpolated samples whereas the interpolation

---

[5]The credible intervals were computed by assuming that the missing observations were normally distributed. More precise, but also more complex, methods for estimating the credible interval can be found in, e.g., [58].

(a) Trace of samples for the frequency

(b) Trace of samples for the damping coefficient

(c) Trace of samples for the state noise variance

(d) Trace of samples for the observation noise variance

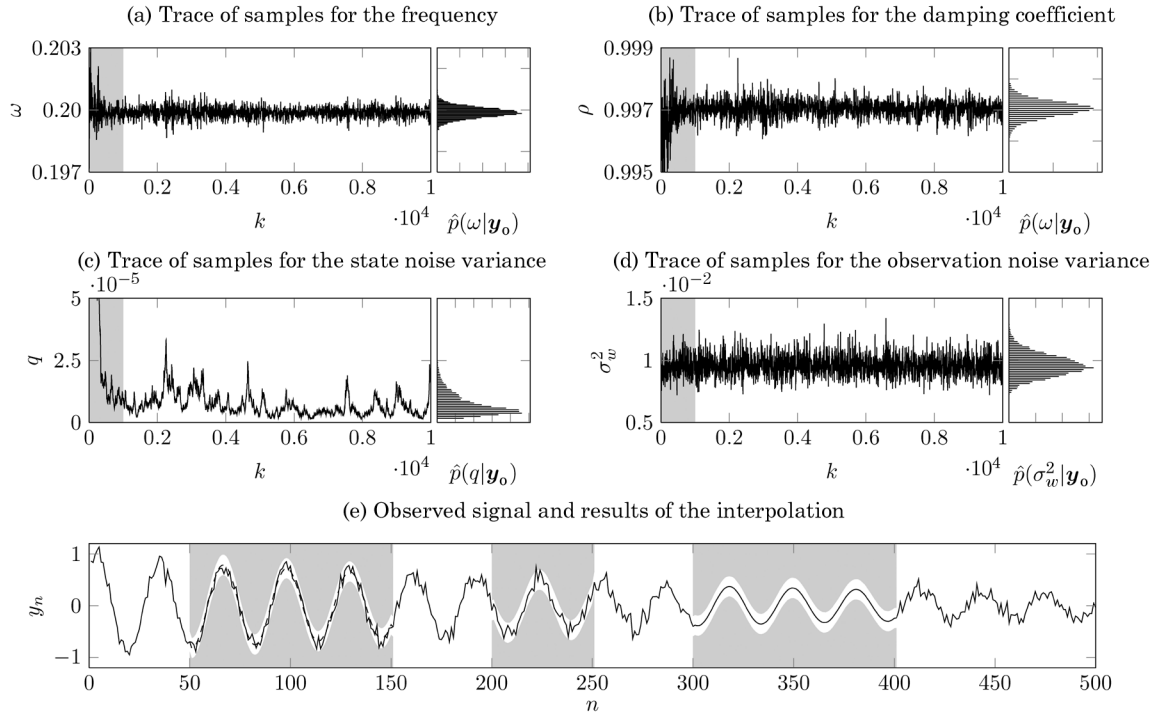(e) Observed signal and results of the interpolation

Fig. 3. Traces of 10 000 generated samples for the (a) frequency, (b) damping coefficient, (c) state noise variance, and (d) observation noise variance. The histograms in the right margin of plots (a)–(d) are computed based on the last 9000 samples. Only 50% of the signal is observed and the three missing sections are indicated by a shaded background. In the three interpolation sections, the 95% credible interval for the posterior distribution $\hat{p}(\boldsymbol{y}_{\mathrm{m}}|\boldsymbol{y}_{\mathrm{o}})$ of the interpolated samples are shown along with interpolated samples based on the mean of $\hat{p}(\boldsymbol{y}_{\mathrm{m}}|\boldsymbol{y}_{\mathrm{o}})$ (last section), a sample from $\hat{p}(\boldsymbol{y}_{\mathrm{m}}|\boldsymbol{y}_{\mathrm{o}})$ (middle section) and both (first section). (a) Trace of samples for the frequency. (b) Trace of samples for the damping coefficient. (c) Trace of samples for the state noise varience. (d) Trace of samples for observation noise variance. (e) Observed signal and results of the interpolation.

(a) Traces of samples for the frequencies

(b) Spectrogram

(c) Periodogram

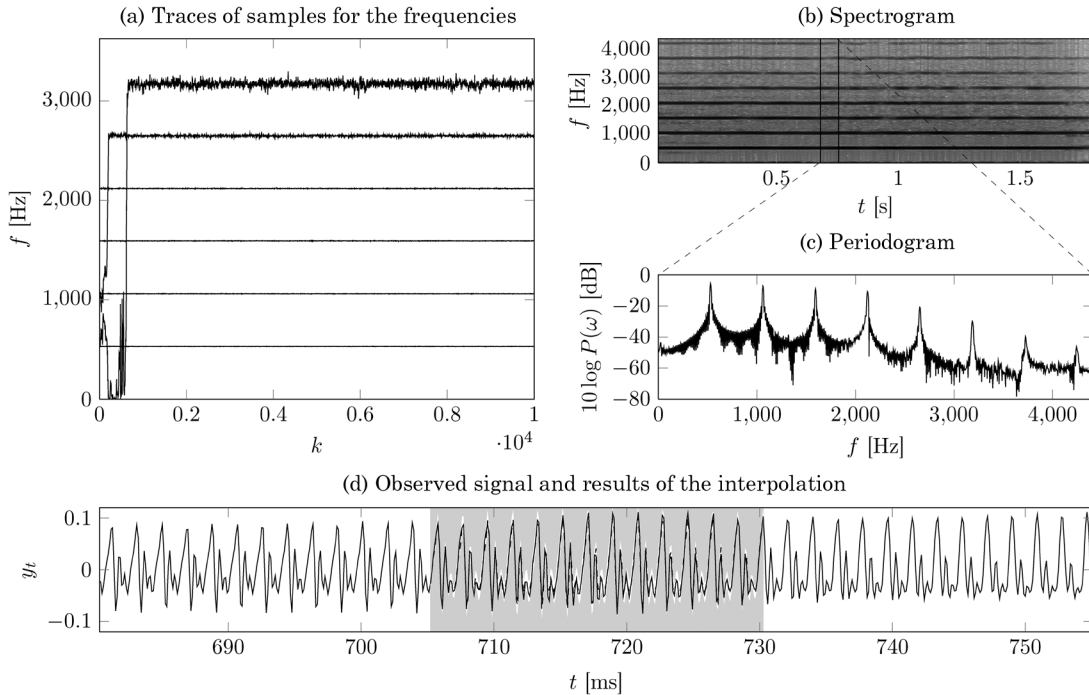(d) Observed signal and results of the interpolation

Fig. 4. Plot (a) shows the six traces for the frequencies each consisting of 10 000 samples. Plot (b) shows the spectrogram for the complete trumpet signal whereas plot (c) shows the periodogram for the section indicated in plot (b). The time series corresponding to this section is shown in plot (d) with the middle section of 25 ms audio missing. The plot also shows the result of the interpolation in terms of 95% probability interval, a sample for the posterior distribution $\hat{p}(\boldsymbol{y}_{\mathrm{m}}|\boldsymbol{y}_{\mathrm{o}})$ and the true missing observations (dotted). (a) Traces of samples for the frequencies. (b) Spectrogram. (c) Periodogram. (d) Observed signal and results of the interpolation.

in the middle section is a random sample from the posterior distribution. Both methods are shown in the first interpolation section. Clearly, sampling from the posterior distribution yields a much more typical sample than using the mean estimate. The latter has higher probability, but it does not model the noise.
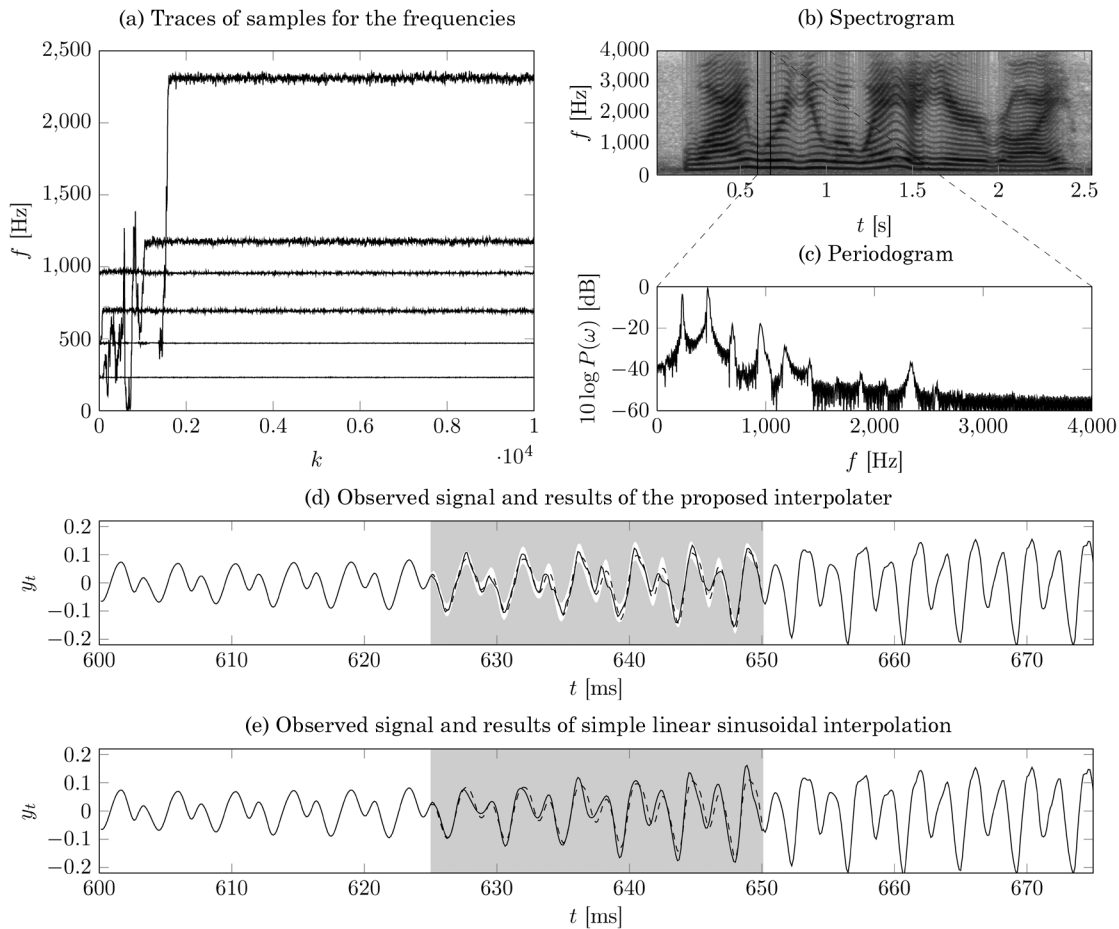
Fig. 5. Plot (a) shows the six traces for the frequencies each consisting of $10\,000$ samples. Plot (b) shows the spectrogram for the complete speech signal whereas plot (c) shows the periodogram for the section indicated in plot (b). The time series corresponding to this section is shown in plot (d) with the middle section of 25 ms audio missing. The plot also shows the result of the interpolation in terms of 95% probability interval, a sample for the posterior distribution $\hat{p}(\boldsymbol{y}_{\mathrm{m}}|\boldsymbol{y}_{\mathrm{o}})$ and the true missing observations (dotted). For comparison, the missing packet was also interpolated in plot (e) by use of a simpler interpolator. (a) Traces of samples for the frequencies. (b) Spectrogram. (c) Periodogram. (d) Observed signal and results of the proposed interpolater. (e) Observed signal and results of simple linear sinusoidal interpolation.

C. Music Signal

In the third simulation, we considered a segment of observations from a downsampled trumpet signal whose spectrogram can be seen in Fig. 4(b). The considered snapshot corresponds to 75 ms of audio and is shown in Fig. 4(d). The periodogram of the $N = 660$ observations in the snapshot is shown in Fig. 4(c). Prior to running the Gibbs sampler, we removed the middle section thus emulating a lost audio packet of 25 ms on a packet-based network. In Fig. 4(a), we have shown the six traces of samples for the frequencies. We see that the sampler reached a stationary point after approximately 500 iterations after which samples for the dominating six frequency components were generated. The results of the interpolation are shown in Fig. 4(d). It is observed, that the 95% credible interval was very tight and that the generated sample from the posterior distribution for the missing observations therefore almost coincided with the true missing observations. An informal listening test also confirmed that the music segment had been restored with almost no perceptual loss.

D. Speech Signal

In the fourth and final simulation, we considered a more challenging segment of observations originating from a speech

signal; where the frequency spectrum and amplitudes of the trumpet signal in the previous simulation were approximately constant, the snapshot shown in Fig. 5(d) is clearly nonstationary. Additionally, as can be seen from the spectrogram in Fig. 5(b), some of the frequencies are nonconstant in the snapshot. The speech signal originates from a female voice uttering "*Why were you away a year, Roy?*," and it was downsampled to 8000 Hz. As in the previous simulation, we removed the middle section of 25 ms prior to running the Gibbs sampler. The traces of samples for the frequencies are shown in Fig. 5(a). The sampling scheme seemed to have reached a stationary point after approximately 1500 iterations. The interpolated samples in Fig. 5(d) follows the same increasing trend as the true signal. Compared against the interpolation of the trumpet signal, the 95% confidence interval is wider reflecting the more complex structure of the signal. Despite this, an informal listening test revealed that the music segment had been restored with only little perceptual loss. For comparison, we have also performed the interpolation of the missing packet by use of a simpler interpolator based on [9]. In this interpolation scheme, the amplitudes and frequencies are estimated on both sides of the missing packet which is recovered by linearly interpolating these amplitudes and frequencies. The result of this inter-

polation is shown in Fig. 5(e). In order to compare the two methods, we have measured the reconstruction signal-to-noise ratio (SNR) for both methods. For the simple linear sinusoidal interpolater, the SNR was 7.7 dB whereas a sample from the posterior distribution $p(\boldsymbol{y}_{\mathrm{m}}|\boldsymbol{y}_{\mathrm{o}})$ resulted in an SNR of 10.8 dB. If we instead used the posterior mean as an interpolant, the SNR was 15.8 dB. It should be noted, however, that SNR cannot be used as an objective measure for the reconstruction performance since the human auditory system does not perceive sound degradation in the two norm.

VI. CONCLUSION

In this paper, we have presented a Bayesian interpolation and parameter estimation inference scheme based on a dynamic signal model hypothesis for the observed segment of data. The dynamic model enables modeling of real-world signals with nonstationary, but smooth evolution since the in-phase and quadrature components were modeled as first-order Gauss–Markov processes. The proposed inference scheme for the dynamic model was developed in a Bayesian framework and comprised two stages. In the first stage, a two state Gibbs sampler alternated between sampling from the conditional distribution for the hidden states given the model parameters and sampling from the conditional distribution for the model parameters given the hidden states. In the second stage, a single draw from the posterior distribution for the missing observations given the last sample for the hidden states and model parameters was obtained. This sample was used for replacing the missing sample with a typical interpolant for the underlying process.

In the simulations, we demonstrated that the inference scheme can be used for generating histograms for the unknown parameters from which, e.g., point and interval estimates can be derived. We also demonstrated the applicability of the proposed inference scheme to audio restoration. For a simple segment from a trumpet signal and a more complex segment from a speech signal, we recovered a 25-ms packet by use of the two neighboring packets. Informal listening tests revealed that the restoration procedure restored the audio signal segments with a slight perceptual loss.

APPENDIX
PROBABILITY DISTRIBUTIONS

In the following list, τ is a scalar positive random variable and $\boldsymbol{x}$ is an N-dimensional random vector.

Exponential Distribution: The exponential distribution with rate parameter λ has the probability distribution

$$p(\tau|\lambda) = \lambda \exp\{-\lambda\tau\}$$

and is denoted by $\mathrm{Exp}(\tau;\lambda)$.

Inverse Gamma Distribution: The inverse gamma distribution with shape parameter α and scale parameter β has the probability distribution

$$p(\tau|\alpha,\beta) = [\beta^{\alpha}/\Gamma(\alpha)]\, \tau^{-(\alpha+1)} \exp\{-\beta/\tau\}$$

and is denoted by $\mathcal{IG}(\tau;\alpha,\beta)$.

Multivariate Normal Distribution: The multivariate normal distribution with the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ has the probability distribution

$$p(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \left[(2\pi)^N|\boldsymbol{\Sigma}|\right]^{-1/2} \exp\left\{\frac{-1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

and is denoted by $\mathcal{N}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma})$.

Multivariate Normal-Scaled Inverse Gamma Distribution: The multivariate normal-scaled inverse gamma distribution with the location vector $\boldsymbol{\mu}$, covariance scale matrix $\boldsymbol{C}$, shape parameter α, and scale parameter β has the probability distribution

$$p(\boldsymbol{x},\tau|\boldsymbol{\mu},\boldsymbol{C},\alpha,\beta) = \mathcal{N}(\boldsymbol{x};\boldsymbol{\mu},\tau\boldsymbol{C})\mathcal{IG}(\tau;\alpha,\beta)$$

and is denoted by $\mathcal{NIG}(\boldsymbol{x},\tau;\boldsymbol{\mu},\boldsymbol{C},\alpha,\beta)$.

Multivariate Student's t-Distribution: The multivariate student's t-distribution with the mean vector $\boldsymbol{\mu}$, covariance matrix $\boldsymbol{\Sigma}$, and ν degrees of freedom has the probability distribution

$$p(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) = \frac{\Gamma(N/2+\nu/2)}{\Gamma(\nu/2)\sqrt{(\pi\nu)^N|\boldsymbol{\Sigma}|}} \left[1+\frac{\Delta^2}{\nu}\right]^{-\frac{N+\nu}{2}}$$
$$\Delta^2 = (\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$$

and is denoted by $\mathrm{St}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$.

Uniform Distribution: For $N=1$, the uniform distribution with lower and upper boundary parameters a and b has the probability distribution

$$p(x|a,b) = \begin{cases} (b-a)^{-1}, & \text{for } a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

and is denoted by $\mathcal{U}(x;a,b)$.

REFERENCES

[1] D. Goodman, G. Lockhart, O. Wasem, and W.-C. Wong, "Waveform substitution techniques for recovering missing speech segments in packet voice communications," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 6, pp. 1440–1448, Dec. 1986.

[2] Y. Liang, N. Farber, and B. Girod, "Adaptive playout scheduling and loss concealment for voice communication over IP networks," *IEEE Trans. Multimedia*, vol. 5, no. 4, pp. 532–543, Dec. 2003.

[3] Y. Wang, J. Li, and P. Stoica, *Spectral Analysis of Signals: The Missing Data Case*. San Rafael, CA: Morgan & Claypool, Jun. 2005.

[4] H. Ofir, D. Malah, and I. Cohen, "Audio packet loss concealment in a combined MDCT-MDST domain," *IEEE Signal Process. Lett.*, vol. 14, no. 12, pp. 1032–1035, Dec. 2007.

[5] J. Rajan, P. Rayner, and S. Godsill, "Bayesian approach to parameter estimation and interpolation of time-varying autoregressive processes using the Gibbs sampler," *IEE Proc. Vis. Image Signal Process.*, vol. 144, no. 4, pp. 249–256, Aug. 1997.

[6] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration*. London, U.K.: Springer-Verlag, 1998.

[7] C. A. Rødbro, M. N. Murthi, S. V. Andersen, and S. H. Jensen, "Hidden Markov model-based packet loss concealment for voice over IP," *IEEE Trans. Audio, Speech, Lang., Process.*, vol. 14, no. 5, pp. 1609–1623, Sep. 2006.

[8] J. Lindblom, "A sinusoidal voice over packet coder tailored for the frame-erasure channel," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 787–798, Sep. 2005.

[9] C. A. Rødbro, M. G. Christensen, S. V. Andersen, and S. H. Jensen, "Compressed domain packet loss concealment of sinusoidally coded speech," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2003, vol. 1, pp. 104–107.

[10] M. Lagrange, S. Marchand, and J.-B. Rault, "Enhancing the tracking of partials for the sinusoidal modeling of polyphonic sounds," *IEEE Trans. Audio, Speech, Lang., Process.*, vol. 15, no. 5, pp. 1625–1634, Jul. 2007.

[11] M. M. Goodwin, *Adaptive Signal Models: Theory, Algorithms and Audio Applications*. New York: Springer, Oct. 1998.

[12] R. Kumaresan and D. Tufts, "Estimating the angles of arrival of multiple plane waves," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-19, no. 1, pp. 134–139, Jan. 1983.

[13] P. Stoica, R. L. Moses, B. Friedlander, and T. Söderström, "Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 3, pp. 378–392, Mar. 1989.

[14] J. Li and P. Stoica, "Efficient mixed-spectrum estimation with applications to target feature extraction," in *Rec. Asilomar Conf. Signals, Syst., Comput.*, Oct. 1995, vol. 1, pp. 428–432.

[15] J. Cadzow, "Spectral estimation: An overdetermined rational model equation approach," *Proc. IEEE*, vol. 70, no. 9, pp. 907–939, Sep. 1982.

[16] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.

[17] A. Barabell, "Improving the resolution performance of eigenstructure-based direction-finding algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 1983, vol. 8, pp. 336–339.

[18] A. Paulraj, R. Roy, and T. Kailath, "Estimation of signal parameters via rotational invariance techniques- ESPRIT," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Nov. 1985, pp. 83–89.

[19] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Process.*, vol. 39, no. 5, pp. 1110–1121, May 1991.

[20] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ: Prentice-Hall, 2005.

[21] J. M. Bernardo and A. Smith, *Bayesian Theory*, 1st ed. New York: Wiley, 1994.

[22] D. J. C. MacKay, *Information Theory, Inference & Learning Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, Jun. 2002.

[23] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. Secaucus, NJ: Springer-Verlag, 2005.

[24] E. T. Jaynes, , C. R. Smith and G. J. Erickson, Eds., "Bayesian spectrum and chirp analysis," in *Maximum Entropy and Bayesian Spectral Analysis and Estimation Problems*. Dordrecht, The Netherlands: D. Reidel, 1987, pp. 1–37.

[25] G. L. Bretthorst, *Bayesian Spectrum Analysis and Parameter Estimation*. Berlin, Heidelberg, Germany: Springer-Verlag, 1988.

[26] L. Dou and R. J. W. Hodgson, "Bayesian inference and Gibbs sampling in spectral analysis and parameter estimation I," *Inverse Probl.*, vol. 11, no. 5, pp. 1069–1085, 1995.

[27] L. Dou and R. J. W. Hodgson, "Bayesian inference and Gibbs sampling in spectral analysis and parameter estimation II," *Inverse Probl.*, vol. 12, no. 2, pp. 121–137, 1996.

[28] C. Andrieu and A. Doucet, "Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2667–2676, Oct. 1999.

[29] P. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrica*, vol. 82, no. 4, pp. 711–732, Dec. 1995.

[30] D. V. Rubtsov and J. L. Griffin, "Time-domain Bayesian detection and estimation of noisy damped sinusoidal signals applied to NMR spectroscopy," *J. Magn. Resonance*, vol. 188, no. 2, pp. 367–379, Aug. 2007.

[31] M. Davy, S. J. Godsill, and J. Idier, "Bayesian analysis of polyphonic western tonal music," *J Acoust. Soc. Amer.*, vol. 119, no. 4, pp. 2498–2517, 2006.

[32] M. G. Christensen, S. V. Andersen, and S. H. Jensen, "Amplitude modulated sinusoidal models for audio modeling and coding," in *Knowledge-Based Intelligent Information and Engineering Systems*. Berlin, Germany: Springer-Verlag, Oct. 2003, vol. 2773, pp. 1334–1342.

[33] M. G. Christensen and S. van de Par, "Efficient parametric coding of transients," *IEEE Trans. Audio, Speech, Lang., Process.*, vol. 14, no. 4, pp. 1340–1351, Jul. 2006.

[34] S. J. Godsill and M. Davy, "Bayesian harmonic models for musical pitch estimation and analysis," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2002, vol. 2, pp. 1769–1772.

[35] A. C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge, U.K.: Cambridge Univ. Press, 1989.

[36] A. C. Harvey and A. Jaeger, "Detrending, stylized facts and the business cycle," *J. Appl. Econ.*, vol. 8, no. 3, pp. 231–247, Jul. 1993.

[37] A. Harvey, T. Trimbur, and H. van Dijk, Cyclical components in economic time series Erasmus Univ. Rotterdam, Econometric Inst., Econometric Inst. Rep. EI 2002-20, Nov. 2002.

[38] R. Kleijn and H. K. van Dijk, "Bayes model averaging of cyclical decompositions in economic time series," *J. Appl. Econ.*, vol. 21, no. 2, pp. 191–212, Mar. 2006.

[39] A. T. Cemgil and S. J. Godsill, "Probabilistic phase vocoder and its application to interpolation of missing values in audio signals," in *Proc. Eur. Signal Process. Conf.*, 2005.

[40] A. T. Cemgil and S. J. Godsill, "Efficient variational inference for the dynamic harmonic model," in *Proc. IEEE Workshop Appl. Signal Process. Aud. Acoust.*, Oct. 2005, pp. 271–274.

[41] A. T. Cemgil, H. J. Kappen, and D. Barber, "A generative model for music transcription," *IEEE Trans. Audio, Speech, Lang., Process.*, vol. 14, no. 2, pp. 679–694, Mar. 2006.

[42] C. Dubois and M. Davy, "Joint detection and tracking of time-varying harmonic components: A flexible Bayesian approach," *IEEE Trans. Audio, Speech, Lang., Process.*, vol. 15, no. 4, pp. 1283–1295, May 2007.

[43] E. Vincent and M. D. Plumbley, "Efficient Bayesian inference for harmonic models via adaptive posterior factorization," *Neurocomput.*, vol. 72, no. 1–3, pp. 79–87, 2008.

[44] J. O Ruanaidh and W. Fitzgerald, "Interpolation of missing samples for audio restoration," *Electron. Lett.*, vol. 30, no. 8, pp. 622–623, Apr. 1994.

[45] J. O Ruanaidh and W. Fitzgerald, *Numerical Bayesian Methods Applied to Signal Processing*, 1st ed. New York: Springer-Verlag, Feb. 1996.

[46] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, 2nd ed. London, U.K.: Chapman & Hall/CRC, Jul. 2003.

[47] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, Aug. 2006.

[48] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, no. 6, pp. 1087–1092, Mar. 1953.

[49] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrica*, vol. 57, no. 1, pp. 97–109, Apr. 1970.

[50] S. Chib and E. Greenberg, "Understanding the Metropolis-Hastings algorithm," *Amer. Stat.*, vol. 49, no. 4, pp. 327–335, Nov. 1995.

[51] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.

[52] M. Stephens, "Dealing with label switching in mixture models," *J. Roy. Stat. Soc., Series B*, vol. 62, no. 4, pp. 795–809, 2000.

[53] J. Durbin and S. J. Koopman, "A simple and efficient simulation smoother for state space time series analysis," *Biometrica*, vol. 89, no. 3, pp. 603–616, Aug. 1, 2002.

[54] C. K. Carter and R. Kohn, "On Gibbs sampling for state space models," *Biometrica*, vol. 81, no. 3, pp. 541–553, Sep. 1994.

[55] P. De Jong and N. Shephard, "The simulation smoother for time series models," *Biometrica*, vol. 82, no. 2, pp. 339–350, Jun. 1995.

[56] J. Durbin and S. Koopman, *Time Series Analysis by State Space Methods*. New York: Oxford Univ. Press, 2001.

[57] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[58] L. E. Eberly and G. Casella, "Estimating Bayesian credible intervals," *J. Stat. Planning Inference*, vol. 112, no. 1–2, pp. 115–132, Mar. 2003.

Jesper Kjær Nielsen was born in Struer, Denmark, in 1982. He received the B.Sc. and M.Sc. (*cum laude*) degrees in electrical engineering from Aalborg University, Aalborg, Denmark, in 2007 and 2009, respectively. He is currently pursuing the Ph.D. degree in the Department of Electronic Systems, Aalborg University.

His research interests include spectral estimation, (sinusoidal) parameter estimation, as well as statistical and Bayesian methods for signal processing.

Mads Græsbøll Christensen (S'00–M'05) was born in Copenhagen, Denmark, in March 1977. He received the M.Sc. and Ph.D. degrees from Aalborg University, Aalborg, Denmark, in 2002 and 2005, respectively.

He was formerly with the Department of Electronic Systems, Aalborg University, and is currently an Associate Professor in the Department of Architecture, Design, and Media Technology. He has been a Visiting Researcher at Philips Research Labs, Ecole Nationale Supérieure des Télécommunications (ENST), University of California, Santa Barbara (UCSB), and Columbia University, New York. He has published more than 75 papers in peer-reviewed conference proceedings and journals and is coauthor (with A. Jakobsson) of the book *Multi-Pitch Estimation* (Morgan & Claypool, 2009). His research interests include digital signal processing theory and methods with application to speech and audio, in particular parametric analysis, modeling, and coding.

Dr. Christensen has received several awards, namely an IEEE International Conference on Acoustics, Speech, and Signal Processing Student Paper Contest Award, the Spar Nord Foundation's Research Prize awarded annually for an excellent Ph.D. dissertation, and a Danish Independent Research Councils Young Researcher's Award. He is an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS.

A. Taylan Cemgil (M'04) received the B.Sc. and M.Sc. degrees in computer engineering from Boğaziçi University, Istanbul, Turkey, and the Ph.D. degree from Radboud University, Nijmegen, The Netherlands, with a dissertation on Bayesian music transcription.

He was a Postdoctoral Researcher at the University of Amsterdam and as a Research Associate at the Signal Processing and Communications Laboratory, University of Cambridge, Cambridge, U.K. He is currently an Assistant Professor at Boğaziçi University, where he cultivates his interests in machine learning methods, stochastic processes, and statistical signal processing. His research is focused towards developing computational techniques for audio, music, and multimedia processing.

Simon J. Godsill (M'95) is a Professor of Statistical Signal Processing in the Engineering Department, Cambridge University, Cambridge, U.K. He has research interests in Bayesian and statistical methods for signal processing, Monte Carlo algorithms for Bayesian problems, modeling and enhancement of audio and musical signals, tracking, and high-frequency financial data. He has published extensively in journals, books, and conferences. He is currently co-organizing a year-long program on sequential Monte Carlo Methods at the SAMSI Institute in North Carolina.

Prof. Godsill was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the journal *Bayesian Analysis*, and is a member of IEEE Signal Processing Theory and Methods Committee. He has coedited in 2002 a special issue of the IEEE TRANSACTIONS ON SIGNAL PROCESSING on Monte Carlo Methods in Signal Processing and organized many conference sessions on related themes.

Søren Holdt Jensen (S'87–M'88–SM'00) received the M.Sc. degree in electrical engineering from Aalborg University, Aalborg, Denmark, in 1988, and the Ph.D. degree in signal processing from the Technical University of Denmark, Lyngby, in 1995.

Before joining the Department of Electronic Systems, Aalborg University, he was with the Telecommunications Laboratory of Telecom Denmark, Ltd, Copenhagen, Denmark, the Electronics Institute of the Technical University of Denmark, the Scientific Computing Group of Danish Computing Center for Research and Education (UNI•C), Lyngby, the Electrical Engineering Department of Katholieke Universiteit Leuven, Leuven, Belgium, and the Center for PersonKommunikation (CPK), Aalborg University. He is Full Professor and is currently heading a research team working in the area of numerical algorithms and signal processing for speech and audio processing, image and video processing, multimedia technologies, and digital communications.

Prof. Jensen was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and *Elsevier Signal Processing*, and is currently Member of the Editorial Board of the *EURASIP Journal on Advances in Signal Processing*. He is a recipient of an European Community Marie Curie Fellowship, former Chairman of the IEEE Denmark Section, and Founder and Chairman of the IEEE Denmark Section's Signal Processing Chapter. In January 2011, he was appointed as member of the Danish Council for Independent Research—Technology and Production Sciences by the Danish Minister for Science, Technology and Innovation.