Available online at www.sciencedirect.com
ScienceDirect
Journal of Magnetic Resonance 188 (2007) 367–379
JMR Journal of Magnetic Resonance
www.elsevier.com/locate/jmr

Time-domain Bayesian detection and estimation of noisy damped sinusoidal signals applied to NMR spectroscopy

Denis V. Rubtsov *, Julian L. Griffin

Department of Biochemistry, The Hopkins Building, Tennis Court Road, University of Cambridge, Cambridge CB2 1QW, UK

Received 19 June 2007; revised 11 August 2007
Available online 19 August 2007

Abstract

The problem of model detection and parameter estimation for noisy signals arises in different areas of science and engineering including audio processing, seismology, electrical engineering, and NMR spectroscopy. We have adopted the Bayesian modeling framework to jointly detect and estimate signal resonances. This considers a model of the time-domain complex free induction decay (FID) signal as a sum of exponentially damped sinusoidal components. The number of model components and component parameters are considered unknown random variables to be estimated. A Reversible Jump Markov Chain Monte Carlo technique is used to draw samples from the joint posterior distribution on the subspaces of different dimensions. The proposed algorithm has been tested on synthetic data, the ^{1}H NMR FID of a standard of L-glutamic acid and a blood plasma sample. The detection and estimation performance is compared with Akaike information criterion (AIC), minimum description length (MDL) and the matrix pencil method. The results show the Bayesian algorithm superior in performance especially in difficult cases of detecting low-amplitude and strongly overlapping resonances in noisy signals.

Keywords: Detection; Estimation; NMR spectroscopy; Bayesian inference; MCMC; Metabolomics

1. Introduction

The problem of detection and parameter estimation for noisy signals arises in different areas of science and engineering such as audio processing, seismology, electrical engineering and nuclear magnetic resonance spectroscopy (NMR). For the latter application the free induction decay (FID) detected in an NMR experiment can be considered as a sum of exponentially damped sinusoidal components, and this model has proven to be adequate in a number of applications, including the analysis of solution-state biofluid NMR spectroscopy. In this case model detection reduces to the estimation of the number of components. The detection and estimation problem for the damped sinusoidal model has received considerable attention over the past two decades, and this has increased since the advent of NMR spectroscopy based metabolomics [1,2], where it is often both laborious and mathematically challenging to identify all the components in a biological matrix such as urine, blood plasma or a tissue extract as detected by ^{1}H NMR spectroscopy. Ideally for this application one would like a method which is fully automated but still capable of resolving as many resonances as possible in a complex biological sample.

One class of algorithms, which includes popular methods such as linear prediction singular value decomposition (LP SVD), Hankel SVD (HSVD) [3], Prony method [4], and the matrix pencil [5], is based on the principle of linear prediction and space-state formalism. The algorithms require minimal user input and do not depend on starting values. However, their performance can be strongly affected by the choice of the model order which gives rise to a detection problem. The detection problem has been dealt with by a number of different approaches including the selection of a subset of 'significant' signal poles within

* Corresponding author. Fax: +44 (0) 1223 333345.
E-mail address: dvr22@cam.ac.uk (D.V. Rubtsov).

1090-7807/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2007.08.008

SVD framework, by statistical testing [6], Cadzow's technique for a minimum variance estimate [7], consecutive reconstructions with residual analysis [8] and by information theoretical criteria such as Akaike information criterion (AIC) [9] and minimum description length (MDL) [10]. In [11], for example, a matrix pencil technique is combined with AIC/MDL criteria in a joint detection–estimation procedure. A drawback of these techniques is that they do not offer much flexibility in terms of prior knowledge incorporation [12]. Recent and promising developments include filter diagonalization method and related high-resolution techniques rooted in quantum dynamics calculations [13,14].

Another type of approach, used by techniques such as VARPRO [15] and AMARES [16], is based on non-linear optimization. They seek to obtain maximum-likelihood parameter estimates of a chosen model, using local optimization procedures such as Levenberg–Marquardt [15], or global optimization procedures such as simulated annealing [17] or genetic algorithms [18]. Algorithms of this class are naturally very flexible and allow the use of a wide variety of prior knowledge. However, the detection problem has to be solved by external means before these methods can be applied. They also suffer from characteristic problems of non-linear optimization procedures applied to high-dimension spaces such as dependency on starting points and failure to converge to an acceptable solution in a reasonable amount of time. User input has been used to circumvent the starting value problem but it is not very reliable especially in low signal-to-noise ratio (SNR) cases. Dealing with nuisance peaks and baseline distortion can also pose a problem since it is difficult to define a model function for them [12].

To address the short comings of the above approaches Bayesian inference has been applied to the problem, particularly in a pioneering series of papers by Bretthorst [19–23], followed by Dou and Hodgson [24,25]. These techniques function effectively over various experimental setups. However, despite the considerable attention, the problem of detection and estimation remains generally unsolved for difficult cases when SNR is low and components' frequencies are closely spaced (co-resonant) [12].

Green [26] has proposed a technique called Reversible Jump Markov Chain Monte Carlo which employs a full Bayesian inference to joint detection–estimation problems. The technique has been successfully applied to various problems including mixture estimation [27]. Furthermore, Andrieu and Doucet [28] applied Reversible Jump MCMC successfully to the problem of detection–estimation of real-valued sinusoids in a noisy signal. The main advantage of the approach is that it requires minimal user intervention while increasing sensitivity and resolution of the analysis. The Bayesian framework allows great flexibility in using prior information on the model and its parameters. At the same time it offers rigorous treatment of the available information based on the probability theory.

In this work, we offer a joint detection–estimation algorithm for the complex-valued case with exponential damping following the Reversible Jump approach developed by Andrieu and Doucet [28]. Specifically, this is applied to the problem of estimating the number of components in a FID generated during a NMR experiment. The model and posterior density function are described in Section 2. Section 3 gives a formulation of the Reversible Jump MCMC algorithm. In Section 4 a numerical study of the algorithm's performance is presented, including a Monte Carlo study of detection and estimation properties using simulated dataset. The comparison with combined MDL/AIC and matrix pencil technique is given which shows the algorithm's superior performance especially in low SNR cases. Next, the algorithm is used to process a ^{1}H NMR FID of a standard sample of glutamic acid and a human blood plasma sample, representing a challenging biological mixture typically analyzed using NMR spectroscopy based metabolomics.

2. The data model and posterior distribution function

2.1. The data model

Let $\mathbf{y} = [y_t]$, $t = 0, \ldots, N-1$, represent an observed complex data sequence of length N. The model with Lorentzian decay can be expressed as

$$y(t) = \sum_{i=1}^{k} a_i z_i(t) + \varepsilon_k(t) \tag{1}$$

where $z_i(t) = \mathrm{e}^{j2\pi\omega_i t\Delta}\mathrm{e}^{-\alpha_i t\Delta}$, $a_i = A_i\mathrm{e}^{j2\pi\varphi_i}$, $\omega_i \in [0,1)$—angular frequency, $\alpha_i \in \mathbb{R}^+$—damping factor, a_i—complex amplitude, $A_i \in \mathbb{R}^+$—magnitude, $\varphi_i \in [0,1)$—angular phase shift, $i = 1, \ldots, k, k < k_{\max}$—number of sinusoidal components with $k_{\max} < N$. $\varepsilon_k(t)$ represents added complex-valued noise and assumed to be independently identically distributed (i.i.d.) zero mean white circular Gaussian with variance σ_k^2 for both complex and real part. Δ is a sampling interval. We let Δ to be 1 without loss of generality.

Such a model can arise in several contexts, including the FID of an NMR experiment obtained by quadrature detection.

In vector-matrix form the model in Eq. (1) can be written as

$$\mathbf{y} = \mathbf{D}_k\mathbf{a}_k + \boldsymbol{\varepsilon}_k \tag{2}$$

where $\mathbf{y} = [y_1, \ldots, y_N]^{\mathrm{T}}$,

$$\mathbf{D}_k = \begin{bmatrix} z_1^0 & z_2^0 & \cdots & z_k^0 \\ z_1^1 & z_2^1 & \cdots & z_k^1 \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{N-1} & z_2^{N-1} & \cdots & z_k^{N-1} \end{bmatrix}$$

is $N \times k$ complex matrix, $\mathbf{a}_k = [a_1, \ldots, a_k]^{\mathrm{T}}$ is a vector of complex amplitudes and $\boldsymbol{\varepsilon}_k = [\varepsilon_1, \ldots, \varepsilon_N]^{\mathrm{T}}$ is a noise sequence.

Our objective in this paper is to jointly estimate k and a vector of parameters $\boldsymbol{\theta}_k = (\mathbf{z}_k, \mathbf{a}_k, \sigma_k^2)^{\mathrm{T}}$.

2.2. Prior distributions and likelihood

Following the model, Eq. (2), the likelihood function is

$$p(\mathbf{y} \mid k, \boldsymbol{\theta}_k) = \frac{1}{(2\pi\sigma_k^2)^N} \mathrm{e}^{\left(-(\mathbf{y}-\mathbf{D}_k\mathbf{a}_k)^H(\mathbf{y}-\mathbf{D}_k\mathbf{a}_k)/2\sigma_k^2\right)} \tag{3}$$

where H indicates complex conjugation.

The prior probability distribution for k is a truncated Poisson distribution with $k < k_{\max}$:

$$k \sim \frac{\Lambda^k}{k!}\mathrm{e}^{-\Lambda}$$

with Λ being the "expected" number of components. The prior distribution for σ_k^2 is an independent inverse gamma distribution: $\sigma_k^2 \sim \mathrm{IG}(\upsilon_0/2, \gamma_0/2)$. We assume that $\mathbf{a}_k$ is independently and normally distributed: $\mathbf{a}_k \sim N(\mathbf{0}, \sigma_k^2\boldsymbol{\Sigma}_k)$ with $\boldsymbol{\Sigma}_k^{-1} = \frac{1}{\delta^2}\mathbf{D}_k^H\mathbf{D}_k$ where weighting parameter δ^2 can be interpreted as the expected SNR [28]. The variance for the prior distribution follows g-prior approach, see [29] for motivation.

We assume $\mathbf{z}_k$ to be uniformly and independently distributed within the complex unit disk so $|\mathbf{z}_k| < 1$ and $p(\mathbf{z}_k) \propto 1/\pi^k$.

For the scale hyper-parameter δ^2 we adopt an inverse gamma distribution $\delta^2 \sim \mathrm{IG}(\alpha_\delta, \beta_\delta)$ with $\alpha_\delta = 2$ and $\beta_\delta > 0$. Finally, we hold Λ fixed in this paper.

This choice of prior distributions gives us an advantage of conjugacy being fairly non-informative at the same time.

The parameter space of the model can be expressed as a union of subspaces $\boldsymbol{\Theta} \triangleq \cup_{k=0}^{k_{\max}} \{k\} \times \boldsymbol{\Theta}_k$ where $\boldsymbol{\Theta}_0 \triangleq \mathbb{R}^+$ in the case there are no components in the model and the signal consists just of noise, and $\boldsymbol{\Theta}_k \triangleq \mathbb{C}^k \times \mathbb{D}^k \times \mathbb{R}^+$ for $k = 1, \ldots, k_{\max}$ where $\mathbb{D}$ is the complex unit disk. We also define $\boldsymbol{\Omega} \triangleq \cup_{k=0}^{k_{\max}} \{k\} \times \mathbb{D}^k$.

2.3. Posterior distribution

The inference of k and $\boldsymbol{\theta}_k$ is based on the joint posterior distribution $p(k, \boldsymbol{\theta}_k|\mathbf{y})$. According to Bayes' rule:

$$\begin{aligned} p(k, \boldsymbol{\theta}_k \mid \mathbf{y}) \quad &\propto \tfrac{1}{(2\pi\sigma_k^2)^N} \mathrm{e}^{\left(-(\mathbf{y}-\mathbf{D}_k\mathbf{a}_k)^H(\mathbf{y}-\mathbf{D}_k\mathbf{a}_k)/2\sigma_k^2\right)} \\ &\times \tfrac{1}{(2\pi)^k|\sigma_k^2\Sigma_k|} \mathrm{e}^{\left(-(\mathbf{a}_k^H\mathbf{a}_k)/2\sigma_k^2\Sigma_k\right)} \tfrac{\Lambda^k}{k!}\mathrm{e}^{-\Lambda} \\ &\times \sigma_k^{-2(\upsilon_0/2+1)} \mathrm{e}^{-\gamma_0/2\sigma_k^2} \delta^{-2(\alpha_\delta+1)} \mathrm{e}^{-\beta_\delta/\delta^2} \tfrac{\mathbb{I}_{\boldsymbol{\Omega}}(k,\mathbf{z}_k)}{\pi^k} \end{aligned} \tag{4}$$

By completing the square we integrate out $\mathbf{a}_k$:

$$\mathbf{a}_k \sim N(\mathbf{m}_k, \sigma_k^2\mathbf{M}_k) \tag{5}$$

where $\mathbf{M}_k = 1/(\mathbf{D}_k^H\mathbf{D}_k + \boldsymbol{\Sigma}_k^{-1})$, $\quad \mathbf{m}_k = \mathbf{D}_k^H\mathbf{y}/(\mathbf{D}_k^H\mathbf{D}_k + \boldsymbol{\Sigma}_k^{-1})$, and then σ_k^2:

$$\sigma_k^2 \sim \mathrm{IG}\left(\upsilon_0/2 + N, \left[\gamma_0 + \mathbf{y}^H(\mathbf{I}_N - \mathbf{D}_k\mathbf{M}_k\mathbf{D}_k^H)\mathbf{y}\right]/2\right) \tag{6}$$

and obtain an expression for the posterior distribution up to a normalizing constant:

$$\begin{aligned} p(\mathbf{z}_k, k \mid \mathbf{y}) &\propto \left[\gamma_0 + \mathbf{y}^H(\mathbf{I}_N - \mathbf{D}_k\mathbf{M}_k\mathbf{D}_k^H)\mathbf{y}\right]^{-(N+\upsilon_0/2)} \left|\boldsymbol{\Sigma}_k^{-1}\right| |\mathbf{M}_k| \frac{\Lambda^k}{\pi^k k!} \mathbb{I}_{\boldsymbol{\Omega}}(k, \mathbf{z}_k) \\ &\propto \left[\gamma_0 + \mathbf{y}^H(\mathbf{I}_N - \mathbf{D}_k\mathbf{M}_k\mathbf{D}_k^H)\mathbf{y}\right]^{-(N+\upsilon_0/2)} \frac{1}{(\delta^2+1)^k} \frac{\Lambda^k}{\pi^k k!} \mathrm{I}_{\boldsymbol{\Omega}}(k, \mathbf{z}_k) \end{aligned} \tag{7}$$

Clearly the posterior distribution is non-linear in the parameters $(k, \mathbf{z}_k)$ and does not admit a closed-form solution. Therefore, in the next section we employ the MCMC method to estimate the posterior $p(k, \mathbf{z}_k|\mathbf{y})$, Eq. (7), and draw $\mathbf{a}_k$ and σ_k^2 from respective full conditional distributions, Eqs. (5) and (6).

3. Bayesian computation using Reversible Jump MCMC

3.1. General formulation of Reversible Jump MCMC

The particular difficulty of our task is that the number of components in the signal is unknown so "the number of unknowns is unknown itself". Recently so-called trans-dimensional MCMC techniques have been proposed [26,30,31] which allow for joint estimation of parameters across parameter subspaces of different dimensions. In this paper, we employ the Reversible Jump (RJ) MCMC technique [26]. RJ MCMC is a random-sweep Metropolis-Hastings method adapted for sampling from a joint state space $\boldsymbol{\Theta} \triangleq \cup_{k=0}^{k_{\max}} \{k\} \times \boldsymbol{\Theta}_k$.

Let $\mathbf{x}, \mathbf{x} = (\boldsymbol{\theta}_k, k)$ denote the current state for an MCMC sampler and $\pi(\mathbf{x})$ the target probability distribution. In order to perform the next move of the Markov chain we propose a move type m and sample a candidate value $\mathbf{x}'$ from a proposal distribution $q_m(\mathbf{x}'|\mathbf{x})$. The move is accepted with probability

$$\alpha_m(\mathbf{x}', \mathbf{x}) = \min\left\{1, \frac{\pi(\mathbf{x}')q_m(\mathbf{x}' \mid \mathbf{x})}{\pi(\mathbf{x})q_m(\mathbf{x} \mid \mathbf{x}')}\right\} \tag{8}$$

Otherwise the move is rejected and the state of the chain remains the same. If we consider only move types which do not change the dimension of the parameter vector the condition, Eq. (8), reduces to a usual Metropolis-Hastings acceptance probability [26,32] and the algorithm becomes Metropolis-Hastings MCMC.

If the move changes the dimensionality of the model we need to define a deterministic, differentiable, invertible dimension matching function [26,33]. Let us, for example, consider a move from k to a higher-dimensional space $k+1$. In this case the acceptance probability has the form:

$$\begin{aligned} \alpha_{k\to k+1}(\mathbf{x}_{k+1}, \mathbf{x}_k) = \min\Bigg\{1, &\frac{p(\mathbf{x}_{k+1} \mid \mathbf{y})}{p(\mathbf{x}_k \mid \mathbf{y})} \times \frac{p(k \mid k+1)}{p(k+1 \mid k)} \\ &\times \frac{q_{k+1\to k}(\mathbf{x}_k \mid f_{k\to k+1}(\mathbf{x}_k, \mathbf{u}))}{q_{k\to k+1}(\mathbf{u} \mid \mathbf{x}_k)} \times \left|\frac{\partial f_{k\to k+1}}{\partial(\mathbf{x}_k, \mathbf{u})}\right| \Bigg\} \end{aligned} \tag{9}$$

where $p(k|k+1)$ and $p(k+1|k)$ are the probability of choosing a dimension-reducing move and vice versa, a vector $\mathbf{u}$ is independent of $\mathbf{x}$ and $q_{k\to k+1}(\cdot|\mathbf{x}_k)$ and

$q_{k+1\to k}(\cdot|\mathbf{x}_{k+1})$ are the proposal distributions for dimension-increasing and dimension-decreasing moves, respectively. The last term is the Jacobian of the transformation $f_{k\to k+1}$ as it changes variable from $(\mathbf{x}_k, \mathbf{u})$ to $\mathbf{x}_{k+1}$.

3.2. Application of RJ MCMC to the problem of Bayesian modeling of a complex-valued damped sinusoidal signal

The algorithm is constructed using the following move types proposed by Green [26] for the sampler:

(a) Update the vector of parameters.
(b) Birth of a new component.
(c) Death of an existing component.

The moves (b) and (c) change the dimension of the model by increasing or decreasing by 1, respectively. We use a random scan form of the algorithm that means that a candidate move is chosen randomly from the three at each iteration. The transition kernel of the Markov chain is therefore a mixture of the transition kernels associated with the moves (a)–(c). Let u_k, b_k, d_k be probabilities to choose update, birth or death moves, respectively, such that

$$u_k + b_k + d_k = 1, d_0 = 0, b_{k_{\max}} = 0$$

We set

$$b_k = c\min\left\{1, \frac{p(k+1)}{p(k)}\right\}, \tag{10}$$

$$d_{k+1} = c\min\left\{1, \frac{p(k)}{p(k+1)}\right\} \tag{11}$$

where $p(k)$ is the prior probability for the model order k and parameter c adds extra flexibility to the algorithm by regulating the proportion of the birth/death moves to update moves. We empirically chose $c = 0.2$ for all k.

Let us assume that initial vector of parameters is $(\boldsymbol{\theta}_{k_0}, k_0)$. First, we consider an update move (a) which does not change dimensionality of the model. It can be summarized as a sequence of the following steps:

1. Consider an lth component z_l of the model, $l = 1, \ldots, k_0$.
2. Sample z'_l from the proposal distribution $q(z_l|\mathbf{z}_{-l})$.
3. Calculate acceptance probability α.
4. If $\alpha > \lambda$ then $z_l = z'_l$ where $\lambda \sim U_{[0,1]}$ else z_l remains the same.
5. Draw nuisance parameters $\mathbf{a}_k$ and σ_k^2 from appropriate full conditional distributions, Eqs. (5) and (6) (optional).

The choice of the proposal distribution can be critical to the success of the algorithm. We have chosen a popular random walk proposal so

$$q(z'_l \mid \mathbf{z}) \propto N(z'_l, \sigma_{zp}^2) \times \mathbb{I}_{(0,1]}(|z'_l|) \tag{12}$$

The proposal gives a random Gaussian perturbation of the current signal pole z_l given that its absolute value does not exceed 1 to satisfy non-negativity constraint on the damping factor.

The proposal has proven to be effective in simulations. However, it is quite sensitive to the choice of σ_{zp}^2. Some authors suggested the acceptance ratio of 25–75% of attempted updates [32,34]. We found that the value for σ_{zp}^2 of around $1/(3N)$ provides an acceptance ratio of about 40% in our simulations but this choice is not instructive.

The acceptance probability then is given by

$$\alpha(z'_l, z_l) = \min\left\{1, \left(\frac{\gamma_0 + \mathbf{y}^H(\mathbf{I}_N - \mathbf{D}_k\mathbf{M}_k\mathbf{D}_k^H)\mathbf{y}}{\gamma_0 + \mathbf{y}^H(\mathbf{I}_N - \mathbf{D}'_k\mathbf{M}'_k\mathbf{D}'^H_k)\mathbf{y}}\right)^{(N+\nu_0/2)} \frac{q(z_l \mid \mathbf{z}')}{q(z'_l \mid \mathbf{z})}\mathbb{I}_{\boldsymbol{\Omega}}(k, z'_l)\right\} \tag{13}$$

The nuisance parameters are drawn from their respective full conditional distributions, Eqs. (5) and (6), using Gibbs sampling steps.

As discussed by Andrieu and Doucet [28] the hyperparameters δ^2 and Λ can be roughly estimated from the data independently or sampled according to respective conditional distributions. In the latter case δ^2 has the full conditional probability distribution in the form of an inverse gamma distribution:

$$\delta^2 \sim \mathrm{IG}\left(\alpha_\delta + k, \beta_\delta + \mathbf{a}_k^H\mathbf{D}_k^H\mathbf{D}_k\mathbf{a}_k/2\sigma_k^2\right) \tag{14}$$

The hyper-hyperparameters are set as $\alpha_\delta = 2$, $\beta_\delta = 10$. (The model is quite robust to the choice of particular values of β_δ, see [28].)

We assume that value for Λ is estimated from the data independently.

The birth and death moves form a reversible pair so these were designed together. The key is to ensure that the proposal distributions conform to the requirement of dimension matching. The necessary detailed balance condition is determined by the acceptance probability, Eq. (9).

When the birth move is chosen we propose a new pole z_{k+1} by drawing from the uniform distribution $U_{\mathbb{D}}$ where $\mathbb{D}$ is the complex unit disk.

The matching death move consists of choosing an existing pole by drawing an index $ind \in (1, \ldots, k)$ at random and deleting indth component from the model.

Using Eq. (9), the Jacobian term equals 1 and after substituting expressions for the probabilities and proposal distributions and simplifications we obtain

$$\alpha_{\text{birth}} = \min\left\{1, \left(\frac{\gamma_0 + \mathbf{y}^H(\mathbf{I}_N - \mathbf{D}_k\mathbf{M}_k\mathbf{D}_k^H)\mathbf{y}}{\gamma_0 + \mathbf{y}^H(\mathbf{I}_N - \mathbf{D}_{k+1}\mathbf{M}_{k+1}\mathbf{D}_k^H)\mathbf{y}}\right)^{(N+\nu_0/2)} \times \frac{|\boldsymbol{\Sigma}_{k+1}^{-1}||\mathbf{M}_{k+1}|}{|\boldsymbol{\Sigma}_k^{-1}||\mathbf{M}_k|}\frac{1}{k+1}\right\} \tag{15}$$

and

$$\alpha_{\text{death}} = \min\left\{1, \left(\frac{\gamma_0 + \mathbf{y}^H(\mathbf{I}_N - \mathbf{D}_k\mathbf{M}_k\mathbf{D}_k^H)\mathbf{y}}{\gamma_0 + \mathbf{y}^H(\mathbf{I}_N - \mathbf{D}_{k+1}\mathbf{M}_{k+1}\mathbf{D}_k^H)\mathbf{y}}\right)^{-(N+\nu_0/2)} \times \frac{|\boldsymbol{\Sigma}_k^{-1}||\mathbf{M}_k|}{|\boldsymbol{\Sigma}_{k+1}^{-1}||\mathbf{M}_{k+1}|}(k+1)\right\} \tag{16}$$

Andrieu and Doucet [28] demonstrated that the Markov chain described above is ergodic (it is aperiodic and irreducible) and the algorithm converges to the posterior distribution uniformly geometrically.

4. Experimental results

4.1. Simulated data

The performance of the algorithm in terms of detection and estimation was first tested using synthetic data. The methodology of testing was based on a Monte Carlo study since it is not possible to obtain theoretical estimation of the algorithm's performance in a closed form. The results in detection were compared with those found using information theoretical criteria MDL and AIC. The estimation results are compared with the results given by matrix pencil algorithm [11].

The simulated signal consisted of eight sinusoidal components. The parameters of the components are given in Table 1. The first two components overlap heavily in frequency domain and the amplitude of one is more than 60 times the amplitude of another so that the performance in simultaneous quantification of overlapping peaks in wide dynamic range can be tested. The next two components mimic the situation when there is a need to quantify a low-intensity peak obscured by baseline. The three resonances around 2 rad pose a challenging problem because they are separated just by a Nyquist step of $2\pi/N$ in frequency domain so the resonance in between two more intensive peaks is very hard to detect by visual inspection of the Fourier spectrum (Fig. 1). Therefore their separation requires improved resolution of the analysis. Finally, the last resonance has a low amplitude and relatively high damping factor that make it especially challenging to detect in a noisy spectrum, addressing the sensitivity limit of detection.

The results of three experiments with three increasingly low levels of SNR were investigated, where SNR is defined for a kth component as $10\log_{10}(A_k^2/2\sigma^2)$ where σ^2 is noise variance. For each experiment fifty different realizations of noise sequence were added to the signal. The number of iterations was 40,000 after 10,000 burn-in iterations. Each run started with a single component placed randomly. No prior information on the number of peaks or parameter values was used. The parameter σ_{zp}^2 for the random walk step was chosen equal to $1/(3N)$. This corresponded to an acceptance rate of 30–50% which is suggested as a good indicator for a random walk [35]. The parameter Λ was set to a value of 20. Preliminary studies with Λ set to 50 and 100 did not show significant difference in performance. One can assign a prior distribution to this hyper-parameter and sample from the posterior with Metropolis-Hastings steps as pointed by Andrieu and Doucet [28], but rough fixed values can give reasonable results as well.

The same data were analyzed with the combined detection–estimation algorithm presented by Lin and co-workers [11]. Two information theoretical criteria—MDL and AIC—were used to estimate the number of components. Parameter estimation was obtained by matrix pencil method which is reported to have superior performance over the popular LPSVD algorithm [11].

Considering the maximum of the posterior probability $p(k|\mathbf{y})$ as an estimation of the number of components in the signal, the algorithm confidently detects the number of the component for experiment 1 (Table 2). As SNR decreases, the data provide less support for unambiguous estimation. In the second experiment less definitive results were calculated as the detection of the components with lower SNR become increasingly difficult. Finally, the results of the third experiment show that components 1, 3, and 6 are sometimes left undetected. A confidence measure can be attached to the estimation to reflect the reliability of the result. Examples of summarized $p(k|\mathbf{y})$ values for the three experiments are presented in Fig. 2.

The AIC approach missed the sixth resonance in the first experiment. As the noise level increased all the low-amplitude components were left undetected. Performance of MDL was somewhat inferior as just four peaks were reliably detected in the first experiment and the large-amplitude fourth component was missing in the second and the third experiments.

The proposed algorithm showed good performance in detection of the number of components given quite tough experimental settings. It could be argued that it would be difficult to achieve better results by visual inspection of the magnitude spectrum, especially in the case of closely spaced frequencies or low SNR peaks.

Table 1
Experimental settings for a simulated signal

Component	Signal parameters				SNR		
	$2\pi\omega$	α	A	φ	1 ($\sigma^2=4$)	2 ($\sigma^2=16$)	3 ($\sigma^2=36$)
1	−0.1	0.006	4	$\pi/4$	3	−3	−6.5
2	−0.05	0.05	250	0	38.9	32.9	29
3	1	0.008	4	$\pi/4$	3	−3	−6.5
4	1.18	0.2	300	0	40.5	34.5	30.9
5	2	0.003	30	$\pi/3$	20.5	14.5	10.9
6	$2+2\pi/N$	0.003	6	$\pi/2$	6.5	0.5	−3
7	$2+4\pi/N$	0.0035	30	$2\pi/3$	20.5	14.5	10.9
8	3	0.016	6	3π /4	6.5	0.5	−3

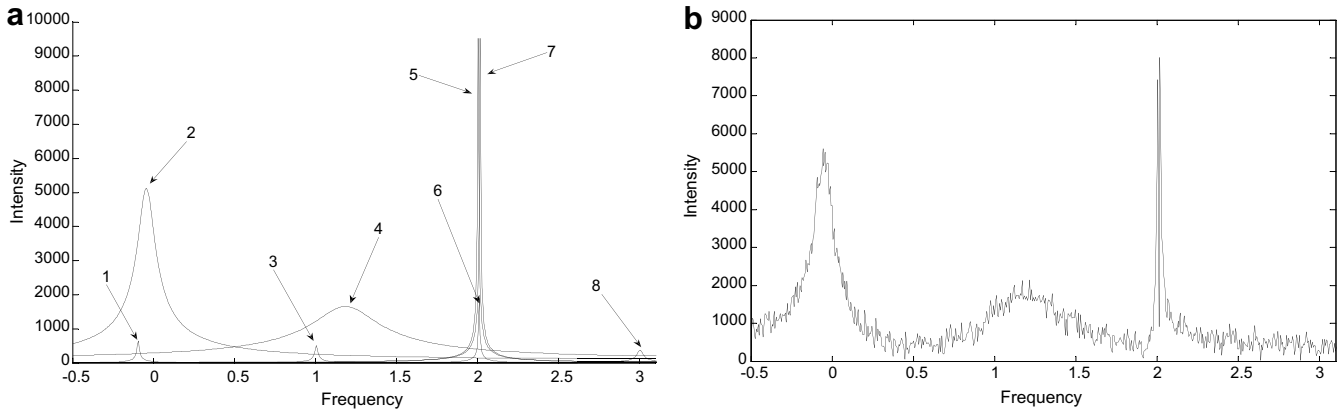

Fig. 1. Synthetic signal used in Monte Carlo study. (a) Magnitude spectra of the Lorentzian components of the signal, (b) an example spectrum of a signal realization with an added noise sequence with a noise variance equal to 36.

Table 2
Comparative performance in detection

Algorithm:	Bayesian SNR			AIC SNR			MDL SNR		
Comp.	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
1	1	**0.98**	**0.6**	1	0.06	0	0	0	0
2	1	1	0.98	1	1	**1**	1	1	1
3	1	**0.96**	**0.73**	1	0.04	0	0	0	0
4	1	1	1	1	0.98	0.94	1	0	0
5	1	1	1	1	1	1	1	1	1
6	**0.98**	**0.5**	**0.2**	0	0	0	0	0	0
7	1	1	1	1	1	1	1	1	1
8	**1**	**1**	**0.92**	0.91	0	0	0	0	0
Average total	**8**	**8**	**7**	7	4	4	4	3	3

Success rate of the techniques in the detection of the different peaks depending on SNR is given. Success rate was calculated as proportion of runs when a peak was detected to overall number of runs within a particular experiment. SNR numbers in columns refer to the three SNR levels from Table 1. Average total gives an average number of resonances detected within an experiment. Entries in bold font highlight superior results in comparison with other techniques.

An important characteristic of the MCMC sampler is its ability to visit different states and not being stuck to a fixed k value. The algorithm has shown good mixing properties in our numerical study. The proportion of accepted birth and death moves averages around 10% which is satisfactory for dimension-changing moves [27]. An example of changes in the value of k against the number of iterations is given in Fig. 3.

The problem we consider next is of parameter estimation. Obtaining parameter estimates involves summarizing the posterior distribution on the basis of the MCMC output. This problem is not trivial due to so-called 'label-switching' [27] as well as genuine multimodality of the posterior densities. A 'label-switching' problem occurs as the posterior distribution of signal poles for a fixed model order k is a mixture of $k!$ distributions up to a label permutation. One way around is to post-process output pole samples, sorting them according to their parameters such as angular frequencies [28] or amplitudes; see [27,36] for discussion. However in the case of a variable number of components the problems is more complicated as the components may appear and disappear from the chain output as well as move to another posterior mode. In this paper, we use a straightforward clustering algorithm based on pairwise distance measure to form post-processed components. The clustering algorithm gives satisfactory results in relabeling samples so components parameters' marginal posterior densities are close to unimodal.

This gives us an opportunity to summarize the posterior with the point estimators. A popular point estimator is a posterior mean. Its disadvantage is that it can be misleading in situations when the posterior distribution is multimodal. In this case, the mean can be located between the modes and possibly in a region of low probability. Therefore we used maximum a posteriori (MAP) estimators of the parameters obtained in a straightforward manner. Particularly the parameter values that deliver the maximum to the joint posterior distribution were found:

$$\hat{\boldsymbol{\theta}}_{\mathrm{MAP}} = \underset{\boldsymbol{\theta}^{i}_{\hat{k}}; i=1,\ldots,I}{\operatorname{argmax}}\, p(\boldsymbol{\theta}_{\hat{k}}, \hat{k} \mid \mathbf{y}) \tag{17}$$

where I is the number of draws from the posterior and $\hat{k}$ is a point estimate for k.

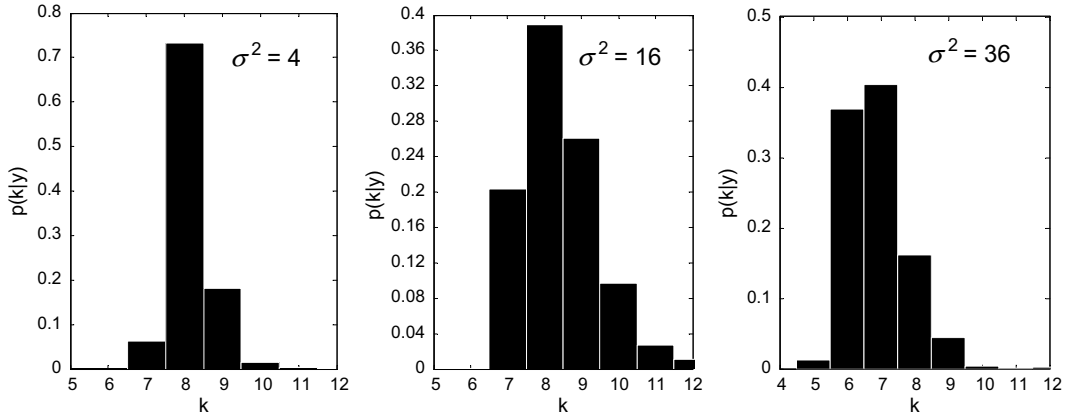

Fig. 2. Example of $p(k|y)$ estimation for three levels of noise. Naturally the lower a noise level the higher the confidence that can be associated with the point estimate of k.

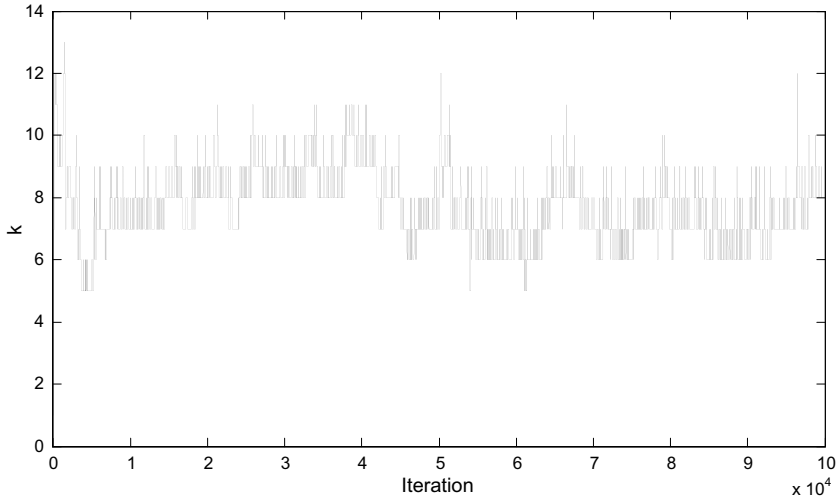

Fig. 3. Example of a chain output in respect to k over 100,000 iterations.

We note that usually it is possible to achieve much better estimation through post-processing methods discussed in Section 5. However, keeping this in mind we tried to assess 'raw' performance of the algorithm before any post-processing was applied.

The estimated values for $(\boldsymbol{\omega}_k, \boldsymbol{\alpha}_k)$ were calculated on a pooled set of MAP estimations derived from individual runs. The complex amplitudes $\mathbf{a}_k$ were simulated from the joint posterior distribution $p(\mathbf{a}_k, \sigma_k^2 \mid \mathbf{z}_k, k, \mathbf{y})$ and absolute magnitudes and phase shift were calculated according to the model, Eq. (1).

For the sake of comparison the parameters of the model were estimated with the matrix pencil given the number of components detected by AIC. The maximum a posteriori estimations and standard deviations of the model parameters obtained by the proposed algorithm along with root mean squared errors of the Bayesian estimators and the matrix pencil estimators are presented in the Tables 3–6.

As expected the estimation is more accurate in higher SNR settings. The performance of the Bayesian approach is comparable with the matrix pencil method on the high-SNR peaks that are detected by both algorithms. As the peak SNR decreases the Bayesian estimation seems to perform slightly better than the matrix pencil method. An example of several draws from posterior distribution of the eight components is given in Fig. 4. The Bayesian analysis was fully automated and no external input in the form of prior knowledge of the number of components or their parameters were made. However, in real life problems (e.g. where the number of detectable components in a biological matrix had already been calculated) such prior information would significantly decrease the time needed for convergence of the chain and therefore improve the effectiveness of the algorithm.

4.2. Real quantitative data

In order to check the quantitative properties of the proposed algorithm on an experimental NMR dataset a series of ^{1}H NMR FIDs of a dilution series of a creatine solution were used. The series consisted of 0.1, 1, 2, 5, and 10 mM creatine solution standards in D_2O solution with

0.17 mM TSP as a chemical shift and concentration reference. FIDs were collected using a pulse and collect sequence on a 500 MHz Bruker NMR spectrometer and a 5 mm TXI probe. The solvent suppression pulse sequence was based on a one-dimensional NOESY pulse sequence that saturates the residual ^{1}H water proton signal during the relaxation delay (2 s) and a mixing time (150 ms). Spectra were collected with 64 transients into 32 K data points over a spectral width of 20 ppm at 37 °C.

No pre-processing methods such as phase correction were applied. The first 80 points were removed so the signal at the first time point would have the maximal magnitude. The next 4096 complex data points were used as a data input for the algorithm. The algorithm was run for 50,000 iterations after 10,000 burn-in iterations.

Since in some cases the lineshapes of creatine singlets deviate from ideal Lorentzian form due to shimming imperfections, the algorithm used several Lorentzian components to approximate the non-Lorentzian lineshapes. In this case co-resonant components corresponding to a single peak were summed up before quantitative analysis. We measured the magnitude A of the summed peaks for the 3.02 ppm resonance as a measure of the relative creatine concentration in the time-domain, with this being proportional to the area-under-the peak in the frequency domain spectrum.

The MAP estimations for A and its standard deviations were obtained from the sampler output. The numeric results are summarized in Table 7 and illustrated in Fig. 5.

The results reveal almost perfect correlation between concentration and estimated amplitude values for both creatine peaks ($r^2 > 0.999$ with p-value $< 10^{-6}$ for 3.02 ppm peak and $r^2 > 0.999$ with p-value $< 10^{-5}$ for 3.92 ppm peak). The ratios between estimated amplitudes of the 3.02 ppm peak of consecutive experiments with increasing concentrations also give a good estimation of the ratios between real concentrations of creatine. The proposed algorithm demonstrated consistency in quantification of a series of diluted samples and therefore can be used for quantitative estimation of the concentration of underlying chemical compounds provided that the relevant resonances are identified.

Next we checked the ability of the algorithm to deal with more complex experimental data. We first analyzed the ^{1}H NMR FID data of a 10 mM L-glutamic acid standard in D_2O solution with 1 mM TSP as a chemical shift reference. The other experimental settings were the same as above.

No pre-processing methods such as phase correction were applied. The FID data were centered and normalized to unit standard deviation prior to analysis. The algorithm was run for 50,000 iterations after 10,000 burn-in iterations. Fig. 6

Table 3
Performance in frequency estimation

Comp.	MAP point estimation (standard deviation) SNR			RMSE Bayesian SNR			RMSE matrix pencil SNR		
	1	2	3	1	2	3	1	2	3
1	−0.099 (0.0007)	−0.1008 (0.0021)	−0.09982 (0.0034)	**0.0008**	0.0022	**0.0034**	0.001	**0.0014**	N/D
2	−0.0499 (0.0006)	−0.04996 (0.00091)	−0.0509 (0.0018)	0.0006	**0.0009**	**0.0021**	**0.0004**	0.0021	0.0023
3	0.999 (0.001)	0.9997 (0.0035)	0.9985 (0.0067)	**0.0013**	0.0035	**0.0068**	0.0017	**0.0016**	N/D
4	1.1803 (0.0025)	1.179 (0.0079)	1.177 (0.011)	0.0025	**0.008**	0.0114	**0.0017**	0.0112	**0.0112**
5	1.9999 (0.0003)	2 (0.00024)	2 (0.00018)	0.0004	0.0004	0.0005	**0.0003**	0.0004	**0.0004**
6	2.0058 (0.0038)	2.005 (0.0068)	2.001 (0.01)	**0.0039**	**0.0069**	**0.0115**	N/D	N/D	N/D
7	2.0124 (0.0002)	2.013 (0.00023)	2.013 (0.00022)	0.0003	0.0005	0.0005	0.0003	**0.0004**	**0.0004**
8	2.9994 (0.0022)	3.002 (0.0059)	3.003 (0.0085)	0.0023	**0.0061**	**0.009**	**0.0018**	N/D	N/D

Point estimations for component's frequency values obtained by the MAP approach with respective standard deviation are given. Components are listed in rows. Root mean-squared errors for frequency estimations for the Bayesian algorithm and the matrix pencil technique show close performance for detected resonances. N/D, not detected. Entries in bold font highlight superior results in comparison with other techniques.

Table 4
Performance in damping factor estimation

Comp.	MAP point estimation (standard deviation) SNR			RMSE Bayesian SNR			RMSE matrix pencil SNR		
	1	2	3	1	2	3	1	2	3
1	0.00573 (0.0011)	0.0072 (0.0026)	0.00595 (0.0035)	0.0011	0.0029	**0.0124**	**0.0009**	**0.0006**	N/D
2	0.0499 (0.00039)	0.0499 (0.0012)	0.0495 (0.0013)	0.0004	0.0012	**0.0995**	0.0004	**0.0006**	0.0998
3	0.00804 (0.0015)	0.00847 (0.0024)	0.00793 (0.0043)	**0.0015**	**0.0024**	**0.0165**	0.0019	0.0028	N/D
4	0.2 (0.0027)	0.199 (0.0054)	0.196 (0.0098)	0.0027	**0.0055**	0.3965	**0.002**	0.0074	**0.3918**
5	0.003 (0.00027)	0.00311 (0.00026)	0.00322 (0.0002)	0.0003	0.0003	**0.0062**	0.0003	**0.0002**	0.0063
6	0.00303 (0.0021)	0.00249 (0.0024)	0.00203 (0.0024)	**0.0021**	**0.0025**	**0.0056**	N/D	N/D	N/D
7	0.00298 (0.00017)	0.00366 (0.00037)	0.00378 (0.00033)	**0.0002**	0.0008	0.0068	0.0003	0.0008	0.0068
8	0.0161 (0.0021)	0.0151 (0.0041)	0.0203 (0.0088)	0.0021	**0.0042**	**0.0374**	**0.002**	N/D	N/D

Point estimations for the component's damping factor values obtained by the MAP approach with respective standard deviation and root mean-squared errors for the Bayesian algorithm and the matrix pencil technique are given. Components are listed in rows. Entries in bold font highlight superior results in comparison with other techniques.

Table 5
Performance in amplitude estimation

Comp.	MAP point estimation (standard deviation) SNR			RMSE Bayesian SNR			RMSE matrix pencil SNR		
	1	2	3	1	2	3	1	2	3
1	4.39 (0.74)	4.68 (1.5)	4.21 (2.4)	0.8369	1.6659	**2.4183**	**0.4813**	**0.8261**	N/D
2	245 (1.4)	249 (3.1)	252 (5)	5.41	**5.2068**	**5.2501**	**1.1587**	5.3789	5.5456
3	3.91 (0.59)	4.36 (1.2)	4.75 (2)	0.5972	**1.2465**	**2.1597**	**0.5809**	1.3539	N/D
4	302 (2.4)	298 (4.5)	300 (8.4)	2.9266	4.7968	8.3618	**2.4358**	**3.3962**	**6.1237**
5	29.1 (1.7)	29 (1.8)	29.7 (2.1)	1.8823	2.0668	2.0978	**0.6798**	**0.8879**	**1.3257**
6	11 (7.3)	11.7 (9.3)	13 (6.2)	**8.8108**	**8.9013**	**9.3423**	N/D	N/D	N/D
7	28.7 (2.1)	29.7 (2.1)	30 (2.1)	2.4312	2.1038	2.0705	**0.5506**	**0.8783**	**1.424**
8	6.01 (0.53)	6.04 (1)	7.16 (2.2)	**0.5263**	**1.03**	**2.4601**	0.5425	N/D	N/D

Point estimations for the component's amplitude values obtained by the MAP approach with respective standard deviation and root mean-squared errors for the Bayesian algorithm and the matrix pencil technique are given. Components are listed in rows. Entries in bold font highlight superior results in comparison with other techniques.

Table 6
Performance in phase shift estimation

Comp.	MAP point estimation (standard deviation) SNR			RMSE Bayesian SNR			RMSE matrix pencil SNR		
	1	2	3	1	2	3	1	2	3
1	0.727 (0.13)	0.869 (0.32)	0.648 (0.54)	**0.1469**	0.3341	**0.5546**	0.1515	**0.2591**	N/D
2	−0.001 (0.0074)	−0.0001 (0.012)	0.0149 (0.027)	0.0077	**0.0115**	0.0306	**0.0058**	0.0221	**0.0251**
3	0.818 (0.13)	0.775 (0.28)	0.947 (0.63)	0.132	**0.28**	**0.646**	**0.1167**	0.2869	N/D
4	0.0004 (0.0069)	0.00336 (0.02)	0.0089 (0.032)	0.0069	**0.0207**	0.0332	**0.0067**	0.0284	**0.0301**
5	1.12 (0.16)	1.21 (0.085)	1.23 (0.054)	0.1769	0.18	0.1869	**0.1633**	**0.1746**	**0.1776**
6	1.43 (0.49)	1.3 (0.88)	1.4 (1.1)	**0.5105**	**0.9246**	**1.0748**	N/D	N/D	N/D
7	2.03 (0.11)	1.91 (0.073)	1.9 (0.07)	**0.132**	0.1988	0.205	0.1612	**0.1816**	**0.1789**
8	2.36 (0.11)	2.35 (0.21)	2.25 (0.21)	0.1091	**0.2148**	**0.2342**	**0.0995**	N/D	N/D

Point estimations for the component's phase shift values obtained by the MAP approach with respective standard deviation and root mean-squared errors for the Bayesian algorithm and the matrix pencil technique are given. Components are listed in rows. Entries in bold font highlight superior results in comparison with other techniques.

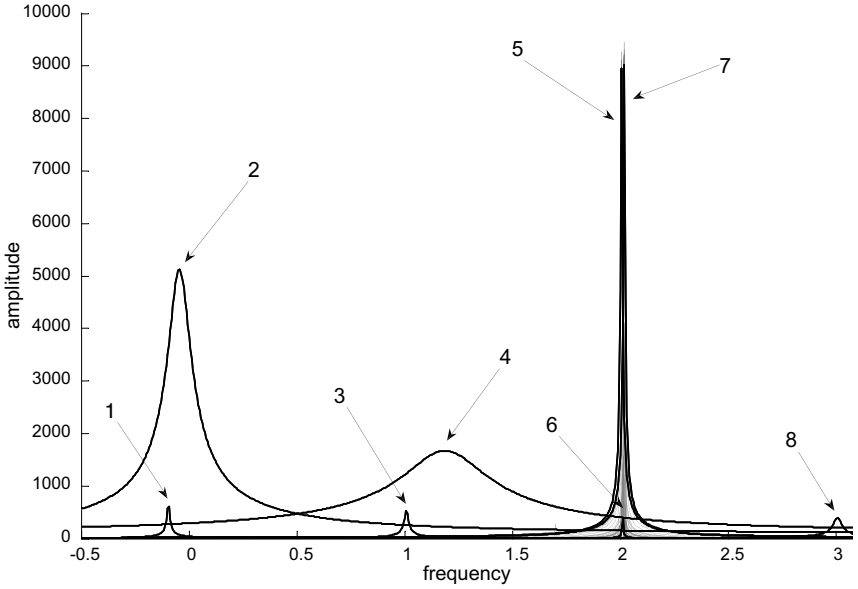

Fig. 4. Example of samples drawn from the posterior distribution. Grey-magnitude spectra of 100 samples drawn from $p(\mathbf{z}, \mathbf{a}|k, \mathbf{y})$. Black—MAP point estimates for components. Higher variation of the peak shapes around 2 rad suggests higher uncertainty in the model parameters' values.

shows the Fourier magnitude spectrum of the signal as well as a spectrum of modeled signal and residuals. Overall, about 60 components were detected. However some of them were 'used' to approximate the residual water resonance and TSP, used as an internal reference. The spectrum of glutamic acid was estimated as a set of approximately 40 peaks.

Table 7
MAP estimations and standard deviations of the amplitude estimates of creatine resonances

Creatine concentration, mM	Estimated amplitude of 3.02 ppm resonance (standard deviation)	Estimated ratios of amplitudes of consequent concentration samples (standard deviation)
0.1	5111 (243)	N/A
1	42252 (256)	8.27 (0.47)
2	81081 (267)	1.92 (0.02)
5	208306 (521)	2.57 (0.02)
10	417436 (495)	2.00 (0.01)

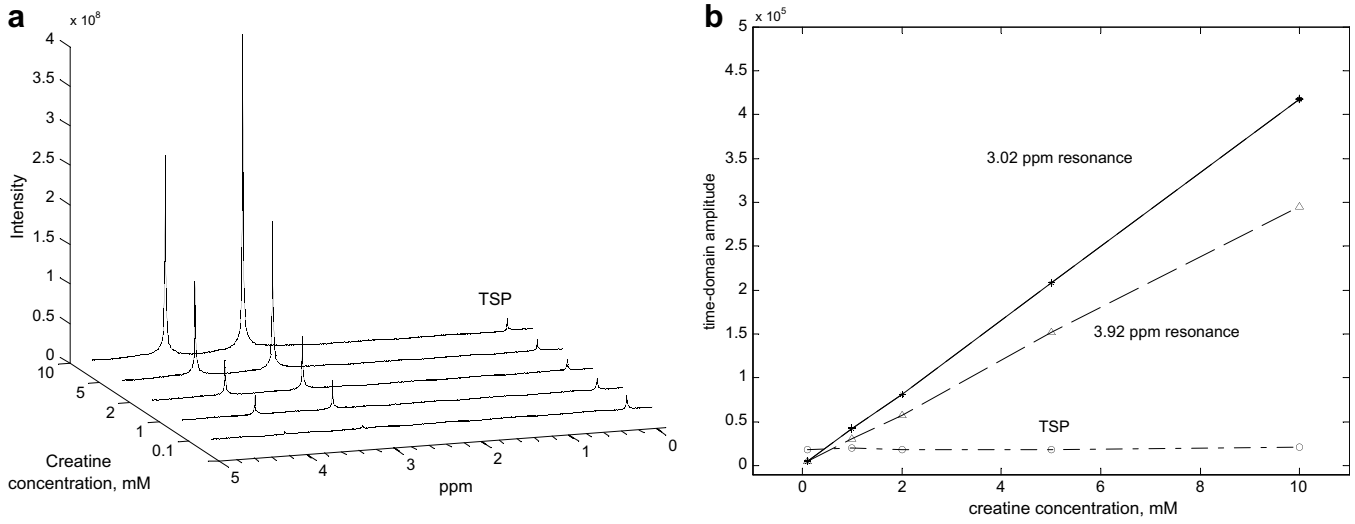

Fig. 5. A serial dilution experiment. (a) A series of magnitude spectra of creatine standard FID's with concentrations 0.1, 1, 2, 5, and 10 mM and added 0.16 mM TSP. (b) A plot of estimated amplitudes of the two creatine resonances and a TSP resonance against creatine concentration. Good linear fit suggests high correlation ($r^2 > 0.99$) between concentration and estimated amplitudes obtained with the Bayesian algorithm.

Finally, we applied the proposed method to quantification of a ^{1}H FID obtained from a human blood plasma sample. Our aim was primarily to test the ability of the algorithm to process real medical and biological samples as used in metabolomics. The sample was prepared by diluting 150 μl of blood plasma in 450 μl of 0.9% NaCl in D_2O. Other acquisition parameters were as described above.

Fig. 6 shows a general fit of the model chosen by MAP criterion after 50,000 iterations following 10,000 burn-in. Quantification was satisfactory with the residual magnitude close to the noise level. The Markov chain stabilized with a k value at ~187. Some of the resonances in the model serve for approximating non-Lorentzian and irregular lineshapes such as slow-varying background due to presence of large molecules.

The ability of the algorithm to effectively disentangle overlapping peaks invites several possible applications. One example appropriate to metabolomics is the problem of co-resonance. For example, the methyl doublet of lactate may play a significant role in medical applications of ^{1}H NMR spectroscopy as a marker for particular disease processes and therapeutic responses. However, it is co-resonant with a set of large lipid resonances. The quantitative model obtained by Bayesian modeling allows us to separate resonances according to their chemical shift, spin–spin relaxation (t_2) time and magnitude that gives clear opportunity to disentangle the lactate methyl doublet from the broad and more intensive lipid signals (Fig. 7).

In conclusion, the performance of the algorithm on both simulated and real data was highly robust. Known resonances of the standard spectrum were fitted without any interference from an experimenter. In addition to peak detection and quantification, in the case of the blood plasma sample, a 'nuisance' lipid resonances were quantified and removed without affecting components of interest.

5. Discussion

In this paper, we have presented an algorithm for the joint estimation of the number of components and their parameters within a Bayesian framework for a NMR spectrum. The algorithm is based on the method described by Andrieu and Doucet [28] and generalizes it to the case of a complex-valued signal with the presence of Lorentzian decay. This algorithm has been successfully applied to the processing of both simulated and real data. In addition, by analyzing the FID in the time domain, the approach

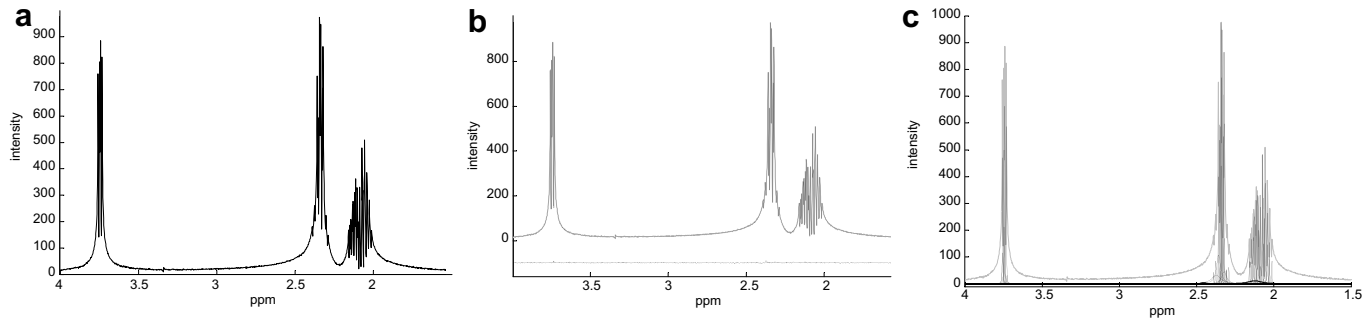

Fig. 6. Quantification of ^{1}H FID of a L-glutamic acid standard (a) Fourier magnitude spectrum of the L-glutamic acid sample. Magnitude spectrum is used to improve presentation and readability of the plot as the original spectrum is not phase-corrected. (b) Estimated magnitude spectrum obtained by MAP criterion. Grey line—original spectrum, black line—model spectrum, black line at the bottom—residuals. (c) Deconvoluted Lorentzian resonances. Grey line—original magnitude spectrum. Black lines—a real part of the Fourier transform of individual Lorentzian phase-corrected components.

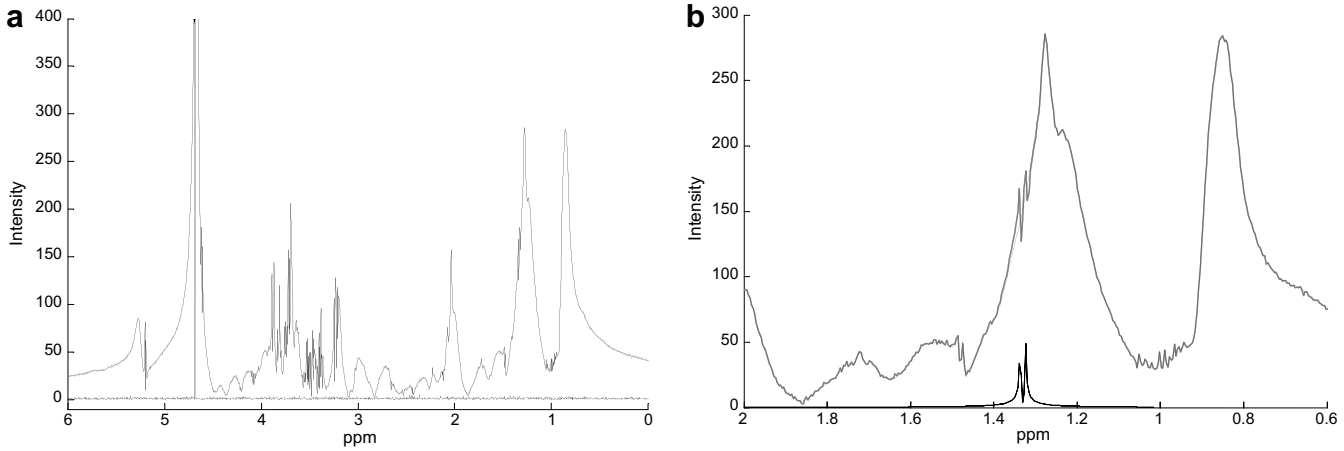

Fig. 7. Quantification of a blood plasma spectrum. (a) Magnitude Fourier spectrum of a blood plasma sample. Grey line—original spectrum, black line—maximum aposteriory estimation. Black line at the bottom shows residuals. (b) Methyl doublet of lactate isolated from broad lipid resonances. Grey line—original spectrum, black dashed line—model fit with the lactate doublet removed, black line—isolated lactate resonances.

has provided improved resolution in the time domain of complex mixtures such as blood plasma.

The Bayesian approach has been applied previously to the problem of sinusoidal signal estimation. Bretthorst presented a convincing Bayesian treatment to the problem [19–23]. However, this approach required a number of assumptions and may be complicated to implement in certain situations. In addition, Barone and Ragona [37] employed a Metropolis MCMC method for sampling from the posterior distribution of signal poles. However, this algorithm needs prior knowledge e.g. approximate values of the prior means and variances of $\mathbf{z}$ and $\mathbf{a}$ in order to reach a better estimation results. Dou and Hodgson offered an algorithm based on Gibbs sampling [24,25]. They linearize a non-linear posterior distribution around a maximum which requires optimization and may be difficult to implement in the case of low SNR.

The major difference with our approach is that for the majority of other algorithms the number of signal components has to be predefined. This is not always known in many problems associated with mixtures, especially in applications such as metabolomics. The model selection problem is difficult, and is addressed here within a consistent Bayesian framework. Additional benefit of the trans-dimension approach presented here is the better mixing of the Markov chain achieved through varying k [27].

The model we chose to use in the paper is probably the most widely used for modeling NMR spectroscopic signals but is not the only possible approach. Other approaches could include Gaussian and more generally Voigt models for signal decay, and these should not present any serious problem in being implemented within our general approach. However, the basic model presented in this paper performed satisfactorily, and could approximate non-Lorentzian components by a mixture of Lorentzian components as can be seen in the case of the water resonance.

In the current implementation the algorithm can be computationally intensive. The processing of one simulated dataset (Section 4.1) on a Pentium 4 with 1GB of RAM and 64-bit Linux operating system took typically about 10 min. The algorithm was implemented within Matlab 7.3. There are several ways to improve the processing time including algorithm modifications mentioned below as well as optimizing the code and the possible use of a more efficient programming environment.

Our choice of prior and proposal distributions and sampler moves are quite general and these choices are not necessarily optimal to the problem. The use of more specific proposal distributions for $\mathbf{z}$ can potentially improve the convergence time of the algorithm. For example, the power spectrum can be used for proposing angular frequencies [28]. The set of sampler moves clearly can be extended by split/merge moves as proposed in [27,28]. A more interesting extension though would be a generalization of dimension-switching moves to allow the introduction of not just single components but more complex entities as multiplets or even fingerprints of chemical compounds. This could lead to a probabilistic solution of the assignment problem, i.e. relating model components to chemical compounds, that is crucial for identification and quantification of NMR samples.

Furthermore, the estimator used in the experimental section could be improved. MCMC methods are not specifically aimed at deriving MAP estimators as the sampler can fail to visit the regions near posterior density modes [38]. Moreover, modes of the joint posterior density may not necessary provide good estimations for the marginal densities of the parameters of interest, especially in high-dimensional parameter space. In the cases where the closed-form solution is known for the marginal posterior densities the expectation–maximization (EM) algorithm is the standard technique for finding local posterior modes. As expectation and maximization steps cannot be performed analytically in our case, numerical non-linear optimization algorithms can be used such as the well-known Newton–Raphson algorithm [39]. Another possibility is to employ MCMC-based techniques for obtaining a marginal maximum as a posteriori estimation [38].

6. Conclusions

An algorithm for joint detection and estimation of a complex exponentially damped sinusoidal model was proposed. The algorithm uses Bayesian inference and Reversible Jump MCMC technique to obtain estimations for the model order and sinusoidal component parameters. The Monte Carlo simulations have shown that the proposed approach performs well in the difficult cases of low SNR and closely spaced frequencies. The algorithm has been tested on real ^{1}H NMR time-domain data of a L-glutamic acid standard and a blood plasma sample and has shown good performance in resonance detection and estimation.

The most significant benefit of the algorithm is that the analysis is essentially automatic and demands little operator intervention. The quantified model of the spectrum produced by this algorithm produces the parameters of interest directly for problems associated with the detection and quantification of components in mixtures. Thus, most of the data processing steps used in NMR spectroscopy of mixtures such as nuisance peak removal, baseline and phase correction are automated. This approach could greatly facilitate the quantitative use of NMR as well as comparative studies and pattern recognition based techniques such as those used in metabolomics.

Acknowledgments

This work was funded by BBSRC Grant BB/D01638X/1. Julian L. Griffin is a Royal Society University Research Fellow.

Appendix A

Notation

$\mathbb{I}_E(z)$—the indicator function of the set E, $\mathbb{I}_E(z) = 1$ if $z \in E$, 0 otherwise.
$\mathbf{I}_N$—identity matrix of dimension $N \times N$.
If $\mathbf{z} = [z_1, \ldots, z_l, \ldots, z_k]$ then $\mathbf{z}_{-l} = [z_1, \ldots, z_{l-1}, z_{l+1}, \ldots, z_k]$.

Probability distributions:
Inverse Gamma

$$\mathrm{IG}(z \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} z^{-\alpha-1} \mathrm{e}^{(-\beta/z)}$$

Gaussian

$$N(\mathbf{z} \mid \mathbf{m}, \Sigma) = \frac{1}{|2\pi\Sigma|^{1/2}} \mathrm{e}^{\left(-\frac{1}{2}(\mathbf{z}-\mathbf{m})^H \Sigma^{-1}(\mathbf{z}-\mathbf{m})\right)}$$

Uniform

$$U_A(\mathbf{z}) = \left[\int_A \mathrm{d}\mathbf{z}\right]^{-1}$$

References

[1] J.K. Nicholson, J. Connelly, J.C. Lindon, E. Holmes, Metabonomics: a platform for studying drug toxicity and gene function, Nature Reviews Drug Discovery 1 (2002) 153–161.
[2] J.L. Griffin, Metabolic profiles to define the genome: can we hear the phenotypes? Philosophical Transactions of the Royal Society of London Series B, Biological Sciences 359 (2004) 857–871.
[3] R.E. Hoffman, G.C. Levy, Modern methods of NMR data-processing and data evaluation, Progress in Nuclear Magnetic Resonance Spectroscopy 23 (1991) 211–258.
[4] P. Barone, L. Guidoni, R. Ragona, V. Viti, E. Furman, H. Degani, Modified Prony method to resolve and quantify in-vivo P-31 NMR-spectra of tumors, Journal of Magnetic Resonance Series B 105 (1994) 137–146.
[5] Y.B. Hua, T.K. Sarkar, On SVD for estimating generalized eigenvalues of singular matrix pencil in noise, IEEE Transactions on Signal Processing 39 (1991) 892–900.
[6] K. Konstantinides, K. Yao, Statistical-analysis of effective singular-values in matrix rank determination, IEEE Transactions on Acoustics Speech and Signal Processing 36 (1988) 757–763.
[7] J.A. Cadzow, Signal enhancement—a composite property mapping algorithm, IEEE Transactions on Acoustics Speech and Signal Processing 36 (1988) 49–62.
[8] M. Shinnar, S.M. Eleff, A generalized Pisarenko technique for analyzing the NMR FID, Journal of Magnetic Resonance 76 (1988) 200–208.
[9] H. Akaike, New look at statistical-model identification, IEEE Transactions on Automatic Control AC19 (1974) 716–723.

[10] J. Rissanen, Modeling by shortest data description, Automatica 14 (1978) 465–471.
[11] Y.Y. Lin, P. Hodgkinson, M. Ernst, A. Pines, A novel detection–estimation scheme for noisy NMR signals: applications to delayed acquisition data, Journal of Magnetic Resonance 128 (1997) 30–41.
[12] L. Vanhamme, T. Sundin, P.V. Hecke, S.V. Huffel, MR spectroscopy quantitation: a review of time-domain methods, NMR Biomedical 14 (2001) 233–246.
[13] M.R. Wall, D. Neuhauser, Extraction, through filter-diagonalization, of general quantum eigenvalues or classical normal-mode frequencies from a small number of residues or a short-time segment of a signal. 1. Theory and application to a quantum-dynamics model, Journal of Chemical Physics 102 (1995) 8011–8022.
[14] V.A. Mandelshtam, FDM: the filter diagonalization method for data processing in NMR experiments, Progress in Nuclear Magnetic Resonance Spectroscopy 38 (2001) 159–196.
[15] J.W.C. Vanderveen, R. Debeer, P.R. Luyten, D. Vanormondt, Accurate quantification of in vivo P-31 NMR signals using the variable projection method and prior knowledge, Magnetic Resonance in Medicine 6 (1988) 92–98.
[16] L. Vanhamme, A. van den Boogaart, S. Van Huffel, Improved method for accurate and efficient quantification of MRS data with use of prior knowledge, Journal of Magnetic Resonance 129 (1997) 35–43.
[17] F.S. Digennaro, D. Cowburn, Parametric-estimation of time-domain NMR signals using simulated annealing, Journal of Magnetic Resonance 96 (1992) 582–588.
[18] G.J. Metzger, M. Patel, X.P. Hu, Application of genetic algorithms to spectral quantification, Journal of Magnetic Resonance Series B 110 (1996) 316–320.
[19] G.L. Bretthorst, Bayesian-analysis. 1. Parameter-estimation using quadrature NMR models, Journal of Magnetic Resonance 88 (1990) 533–551.
[20] G.L. Bretthorst, Bayesian-analysis. 2. Signal-detection and model selection, Journal of Magnetic Resonance 88 (1990) 552–570.
[21] G.L. Bretthorst, Bayesian-analysis. 5. Amplitude estimation for multiple well-separated sinusoids, Journal of Magnetic Resonance 98 (1992) 501–523.
[22] G.L. Bretthorst, Bayesian-analysis. 3. Applications to NMR signal-detection, model selection, and parameter-estimation, Journal of Magnetic Resonance 88 (1990) 571–595.
[23] G.L. Bretthorst, Bayesian-analysis. 4. Noise and computing time considerations, Journal of Magnetic Resonance 93 (1991) 369–394.
[24] L. Dou, R.J.W. Hodgson, Bayesian-inference and Gibbs sampling in spectral-analysis and parameter-estimation. 1, Inverse Problems 11 (1995) 1069–1085.
[25] L.X. Dou, R.J.W. Hodgson, Bayesian inference and Gibbs sampling in spectral analysis and parameter estimation. 2, Inverse Problems 12 (1996) 121–137.
[26] P.J. Green, Reversible Jump Markov Chain Monte Carlo computation and Bayesian model determination, Biometrika 82 (1995) 711–732.
[27] S. Richardson, P.J. Green, On Bayesian analysis of mixtures with an unknown number of components, Journal of the Royal Statistical Society Series B-Methodological 59 (1997) 731–758.
[28] C. Andrieu, A. Doucet, Joint Bayesian detection and estimation of harmonic signals via Reversible Jump MCMC, Tech. Rep. CUED-F-TR318, University of Cambridge, Cambridge, 1998.
[29] P.K. Goel, A. Zellner, B.D. Finetti, Bayesian Inference and Decision Techniques: Essays in Honour of Bruno de Finetti, North-Holland, Amsterdam, Oxford, 1986.
[30] B.P. Carlin, S. Chib, Bayesian model choice via Markov-Chain Monte-Carlo methods, Journal of the Royal Statistical Society Series B-Methodological 57 (1995) 473–484.
[31] P.J. Green, N.L. Hjort, S. Richardson, Highly Structured Stochastic Systems, Oxford University Press, Oxford, 2003.
[32] W.R. Gilks, S. Richardson, D.J. Spiegelhalter, Markov Chain Monte Carlo in Practice, Chapman & Hall, London, 1996.
[33] C. Andrieu, N. de Freitas, A. Doucet, M.I. Jordan, An introduction to MCMC for machine learning, Machine Learning 50 (2003) 5–43.
[34] J. Besag, P. Green, D. Higdon, K. Mengersen, Bayesian computation and stochastic-systems, Statistical Science 10 (1995) 3–41.
[35] J.M.E. Bernardo, Bayesian Statistics 5, Oxford University Press, 1996.
[36] A. Jasra, C.C. Holmes, D.A. Stephens, Markov Chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling, Statistical Science 20 (2005) 50–67.
[37] P. Barone, R. Ragona, Bayesian estimation of parameters of a damped sinusoidal model by a Markov Chain Monte Carlo method, IEEE Transactions on Signal Processing 45 (1997) 1806–1814.
[38] A. Doucet, S.J. Godsill, C.P. Robert, Marginal maximum a posteriori estimation using Markov Chain Monte Carlo, Statistics and Computing 12 (2002) 77–84.
[39] A. Gelman, Bayesian Data Analysis, Chapman & Hall/CRC, Boca Raton, Fla., London, 2004.