JOINT SPARSITY AND FREQUENCY ESTIMATION FOR SPECTRAL COMPRESSIVE SENSING

Jesper Kjær Nielsen[†]*, Mads Græsbøll Christensen*[‡]*, Søren Holdt Jensen*[†]

[†]Aalborg University
Signal and Inf. Process., Dept. of Electronic Systems
{jkn,shj}@es.aau.dk

[‡]Aalborg University
Audio Analysis Lab, Dept. of Arch., Design & Media Tech.
mgc@create.aau.dk

ABSTRACT

Parameter estimation from compressively sensed signals has recently received some attention. We here also consider this problem in the context of frequency sparse signals which are encountered in many application. Existing methods perform the estimation using finite dictionaries or incorporate various interpolation techniques to estimate the continuous frequency parameters. In this paper, we show that solving the problem in a probabilistic framework instead produces an asymptotically efficient estimator which outperforms existing methods in terms of estimation accuracy while still having a low computational complexity. Moreover, the proposed algorithm is also able to make inference about the sparsity level of the measured signal. The simulation code is available online.

Index Terms— Compressive sensing, sinusoidal models, model order comparison, spectral estimation.

1. INTRODUCTION

In recent years, compressive sensing (CS) [1–3] has spurred a renewed interest in sparse decompositions of signals. A sparse decomposition of a noisy signal $\boldsymbol{x} \in \mathbb{C}^N$ can be written as

$$\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{s} + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{\Psi} \in \mathbb{C}^{N\times D}$ is referred to as the basis or dictionary of the signal, $\boldsymbol{s}$ is a D-dimensional l-sparse vector containing only l non-zero coefficients, and $\boldsymbol{n} \in \mathbb{C}^N$ is an error vector modelling noise and model inaccuracies. Many methods have been proposed for estimating the l non-zero coefficients in $\boldsymbol{s}$ from $\boldsymbol{x}$. However, as the sparsity level l is typically much smaller than N, these methods may suffer from a large computational overhead. In CS, this overhead is decreased considerably by utilising the sparsity during the data acquisition. That is, instead of acquiring $\boldsymbol{x}$ by sampling at at least the Nyquist rate, only an amount of data close to the sparsity level l is acquired. These data $\boldsymbol{y} \in \mathbb{C}^M$ are often referred to as the *measurements* and related to $\boldsymbol{x}$ through $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}$ where $\boldsymbol{\Phi} \in \mathbb{C}^{M\times N}$ is the so-called sensing or measurement matrix. Thus, CS may give faster algorithms, data acquisition at a lower sampling rate, and less demanding storage requirements. These properties are important in most signal processing algorithms, and CS has, therefore, become very popular. As many physical signals such as speech and musical signals [4–7] can accurately be modelled as a weighted sum of a small number of sinusoidal basis functions, an incoherent or a redundant discrete Fourier transform (DFT) basis has often been used as the dictionary $\boldsymbol{\Psi}$ for such frequency sparse signals (see [8,9] and the references therein). However, since the frequency parameter is continuous in nature, a sparse decomposition with a finite dictionary is in direct contradiction with the physics behind the sinusoidal signal model which is given by

$$x(n) = \sum_{i=1}^{l} \alpha_i \exp(j\omega_i n) + w(n)\,, n = 0, 1, \ldots, N-1 \tag{2}$$

where $\alpha_i \in \mathbb{C}$, $\omega_i \in \Omega_i \subseteq [0, 2\pi)$, and $w(n) \in \mathbb{C}$ are the i'the complex amplitude, the i'th frequency parameter in radians per sample, and a noise sample, respectively. As demonstrated and discussed in [8], a sinusoidal signal consisting of only a few sinusoidal basis function is not sparse in the DFT domain unless all the l frequencies lie on the DFT grid $\{2\pi(d-1)/D\}_{d=0}^{D-1}$. To tackle this, the redundancy in the DFT can be increased by increasing D, but this leads to that the number of measurements M must also be increased [10], thus defeating the whole purpose of using the CS framework. Therefore, various interpolation methods have recently been proposed and compared in [8, 9].

Spectral interpolation is an old idea in spectral estimation [11], and this concept has also been adopted by the CS community [8, 9]. In our opinion, the various interpolation methods are rather heuristic ways of fixing a problem arising from using the signal model in (1), which is not the best choice for estimating continuous non-linear parameters such as frequencies. As we demonstrate in this paper, formulating the problem of estimating the frequency parameters from the measurements $\boldsymbol{y}$ in a consistent probabilistic framework has numerous advantages compared to the proposed interpolation methods. These include a low computational complexity, superior estimation performance, uncertainty measures for the parameter estimates, and the ability to handle an unknown noise variance and sparsity level l. Moreover, we also demonstrate that our proposed frequency estimator has optimal estimation performance for even moderate SNRs as its mean squared error attains the Cramér-Rao lower bound (CRLB) [12]. In [13], we recently analysed the related problem of jointly estimating the direction of arrivals (DOA) and the number of sources from compressively sensed array signals. The proposed algorithm was based on the MUSIC algorithm and the notion of angles between subspaces [14]. However, for compressively sensed time-series, the covariance matrix model [7] does not hold, and the subspace-based methods such as the MUSIC algorithm can therefore not be used to solve the problem considered in this paper. On the other hand, the algorithm presented here can straightforwardly be modified to solve the DOA problem as well.

2. SINUSOIDAL MODEL COMPARISON AND PARAMETER ESTIMATION

In this section, we first include compressed sensing and prewhitening into the physical signal model in (2). Based on the recently proposed framework in [15], we then formulate the probabilistic framework for jointly estimating the frequency parameters

and detecting the sparsity level l. The latter is usually referred to as model selection and comparison. In the context of CS, an unknown level of sparsity is problematic since the number of measurements M should ideally be close to this level, which therefore often is assumed known. In practice, however, the level of sparsity is often unknown and might not even be possible to upper bound.

2.1. Bayesian Model

2.1.1. Observation Model

Including the sensing matrix and writing the signal model in (2) in vector form yields

$$\boldsymbol{y} \triangleq \boldsymbol{\Phi x} = \boldsymbol{\Phi Z}_l(\boldsymbol{\omega}_l)\boldsymbol{\alpha}_l + \boldsymbol{\Phi w} \tag{3}$$

where the $(n+1,i)$'th element of $\boldsymbol{Z}_l(\boldsymbol{\omega}_l)$ and the i'th element of $\boldsymbol{\alpha}_l$ are given by $\exp(j\omega_i n)$ and α_i, respectively. Due to no prior information on the noise, entropy maximisation, and mathematical tractability [16–19], we model $w(n)$ as white Gaussian noise (WGN) with variance σ^2. The noise vector $\boldsymbol{v} = \boldsymbol{\Phi w}$ is therefore coloured Gaussian noise with the covariance matrix $\boldsymbol{\Phi\Phi}^H$. To simplify the derivation below, we first pre-whiten the signal by pre-multiplying the measurements $\boldsymbol{y}$ in (3) with the Cholesky factor $\boldsymbol{C}^H$ satisfying $(\boldsymbol{\Phi\Phi}^H)^{-1} = \boldsymbol{CC}^H$. Thus, the signal model is

$$\boldsymbol{y}_{\mathrm{f}} \triangleq \boldsymbol{C}^H\boldsymbol{y} = \boldsymbol{W Z}_l(\boldsymbol{\omega}_l)\boldsymbol{\alpha}_l + \boldsymbol{e} \tag{4}$$

where we have defined the weighting matrix as $\boldsymbol{W} \triangleq \boldsymbol{C}^H\boldsymbol{\Phi}$ and the WGN noise vector as $\boldsymbol{e} \triangleq \boldsymbol{C}^H\boldsymbol{\Phi w}$. The observation model is therefore a complex-valued normal distribution with probability density function (pdf)

$$p(\boldsymbol{y}_{\mathrm{f}}|\boldsymbol{\omega}_l, \boldsymbol{\alpha}_l, \sigma^2, l) = \mathcal{CN}(\boldsymbol{y}_{\mathrm{f}}; \boldsymbol{W Z}_l(\boldsymbol{\omega}_l)\boldsymbol{\alpha}_l, \sigma^2\boldsymbol{I}_M) \tag{5}$$

where $\boldsymbol{I}_M$ is the $M \times M$ identity matrix. We note in passing that adding the noise after or prior to taking the measurements results in essentially the same signal model if the measurements and the sensing matrix are pre-filtered in the latter case. As noted in [20], however, the signal-to-noise ratio (SNR) is not the same.

2.1.2. The Prior

Model selection in nested models such as (5) cannot solely be based on comparing candidate models in terms of their likelihood since more complex models can always be made to fit the data better than simpler models. Traditionally, this has been resolved by penalising model complexity in various ways via information criteria such as the AIC [21], the MDL [22], or the asymptotic MAP criteria [23]. Here, the model comparison is performed in the Bayesian framework recently proposed in [15] since it automatically penalises model complexity through the prior distributions and has been demonstrated to outperform the traditional information criteria.

For regression models, the Zellner's g-prior given by [24]

$$p(\boldsymbol{\alpha}_l|\boldsymbol{\omega}_l, \sigma^2, g, l) = \mathcal{CN}(\boldsymbol{\alpha}_l; \boldsymbol{0}, g\sigma^2[\boldsymbol{Z}_l^H(\boldsymbol{\omega}_l)\boldsymbol{W}^H\boldsymbol{W Z}_l(\boldsymbol{\omega}_l)]^{-1}) \tag{6}$$

has been widely adopted since it is analytically tractable and easy to understand and interpret [25, 26]. The hyperparameter g, which can be interpreted as a scaled SNR [26], is assigned the hyper-g prior

$$p(g) = (\delta - 1)(1+g)^{-\delta}, \quad \delta > 1 \tag{7}$$

where δ should be selected in the interval $1 < \delta \leq 2$ [25], and we usually set it to 3/2. Besides having some desirable analytical properties, $p(g)$ reduces to the Jeffreys' prior and the reference prior when $\delta = 1$ [27]. However, since this prior is improper, it can only be used when the prior probability of the all-noise model ($l = 0$) is zero. When this is the case, the proposed model comparison method has no user-defined parameters. Since the noise variance σ^2 has the same meaning in all models, it can be given an improper prior [15, 28]. We, therefore, use the popular Jeffreys' prior $p(\sigma^2) = (\sigma^2)^{-1}$ which is scale invariant. We assume the uniform prior

$$p(\boldsymbol{\omega}_l|l) = \frac{1}{W_l}\prod_{i=1}^{l}\mathbb{I}_{\Omega_i}(\omega_i) \tag{8}$$

for the frequencies where $\mathbb{I}_{\Omega_i}(\omega_i)$ is the indicator function on the interval $\Omega_i \subseteq [0, 2\pi)$, and W_l is the normalisation constant. The overall frequency parameter space is therefore $\boldsymbol{\Omega}_l = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_l$. Finally, the prior on the model order is also a uniform prior of the form $p(l) = (L+1)^{-1}\mathbb{I}_{\mathcal{L}}(l)$ where $\mathcal{L} = \{0, 1, \ldots, L\}$.

2.2. Bayesian Inference

As inference about the frequency parameters turns out to be made as a bi-product of comparing the various candidate model orders, the derivation below focuses on the model comparison problem. From Bayes' theorem, the posterior distribution on the model order has the probability mass function (pmf)

$$p(l|\boldsymbol{y}_{\mathrm{f}}) = \frac{\mathrm{BF}[l;0]p(l)}{\sum_{i=0}^{L}\mathrm{BF}[i;0]p(i)} \tag{9}$$

where the all noise model ($l = 0$) is the base model, all other models are compared against, and the Bayes' factor is given by

$$\mathrm{BF}[i;k] = \frac{p(\boldsymbol{y}_{\mathrm{f}}|l=i)}{p(\boldsymbol{y}_{\mathrm{f}}|l=k)} \triangleq \frac{m_i(\boldsymbol{y}_{\mathrm{f}})}{m_k(\boldsymbol{y}_{\mathrm{f}})}\,. \tag{10}$$

The function $m_l(\boldsymbol{y}_{\mathrm{f}})$ is an unnormalised marginal likelihood whose normalisation constant must be the same for all models. Working with $m_l(\boldsymbol{y}_{\mathrm{f}})$ rather than the normalised marginal likelihood $p(\boldsymbol{y}_{\mathrm{f}}|l)$ is usually simpler. Moreover, $p(\boldsymbol{y}_{\mathrm{f}}|l)$ does not even exist if an improper prior such as the Jeffreys' prior on the noise variance is used. Given g and $\boldsymbol{\omega}_l$, the marginal likelihood is given by

$$p(\boldsymbol{y}_{\mathrm{f}}|\boldsymbol{\omega}_l, g, l) = \int_0^\infty\int_{\mathbb{C}^l} p(\boldsymbol{y}_{\mathrm{f}}|\boldsymbol{\alpha}_l, \boldsymbol{\omega}_l, \sigma^2, l) \times p(\boldsymbol{\alpha}_l|\boldsymbol{\omega}_l, \sigma^2, g, l)p(\sigma^2)d\boldsymbol{\alpha}_l d\sigma^2\,. \tag{11}$$

By performing the integration in (11), it can be shown that

$$p(\boldsymbol{y}_{\mathrm{f}}|\boldsymbol{\omega}_l, g, l) \propto m_l(\boldsymbol{y}_{\mathrm{f}}|\boldsymbol{\omega}_l, g) = \frac{m_0(\boldsymbol{y}_{\mathrm{f}})}{(1+g)^l}\left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}_l^2(\boldsymbol{\omega}_l, g)}\right)^M \tag{12}$$

where we have defined

$$\hat{\sigma}_l^2(\boldsymbol{\omega}_l, g) \triangleq M^{-1}\boldsymbol{y}_{\mathrm{f}}^H\left(\boldsymbol{I}_M - \frac{g}{1+g}\boldsymbol{P}_l(\boldsymbol{\omega}_l)\right)\boldsymbol{y}_{\mathrm{f}} = \hat{\sigma}_0^2\left(1 - \frac{g}{1+g}R_l^2(\boldsymbol{\omega}_l)\right) \tag{13}$$

$$R_l^2(\boldsymbol{\omega}_l) \triangleq \boldsymbol{y}_{\mathrm{f}}^H\boldsymbol{P}_l(\boldsymbol{\omega}_l)\boldsymbol{y}_{\mathrm{f}}(\boldsymbol{y}_{\mathrm{f}}^H\boldsymbol{y}_{\mathrm{f}})^{-1} \tag{14}$$

$$m_0(\boldsymbol{y}_{\mathrm{f}}) \triangleq \Gamma(M)(M\pi\hat{\sigma}_0^2)^{-M}\,. \tag{15}$$

The matrix $\boldsymbol{P}_l(\boldsymbol{\omega}_l)$ is the orthogonal projection matrix for the column space of $\boldsymbol{W Z}_l(\boldsymbol{\omega}_l)$, and $\hat{\sigma}_l^2(\boldsymbol{\omega}_l, g)$ is asymptotically equal to the maximum likelihood (ML) estimate of the noise variance in the limit $\hat{\sigma}_{\mathrm{ML}}^2(\hat{\boldsymbol{\omega}}_l) = \lim_{g\to\infty}\hat{\sigma}_l^2(\hat{\boldsymbol{\omega}}_l, g)$ with $\hat{\boldsymbol{\omega}}_l$ being the ML estimate of the frequency parameters. The estimate $\hat{\sigma}_0^2$ is the estimated noise variance for the all-noise model ($l = 0$) which has the unnormalised

marginal likelihood $m_0(\boldsymbol{y}_{\rm f})$. Given $\boldsymbol{\omega}_l$ and g, the Bayes' factor is therefore given by

$$\text{BF}[l;0|\boldsymbol{\omega}_l,g] = \frac{\left[\hat{\sigma}_0^2/\hat{\sigma}_l^2(\boldsymbol{\omega}_l,g)\right]^M}{(1+g)^l} = \frac{(1+g)^{M-l}}{(1+g[1-R_l^2(\boldsymbol{\omega}_l)])^M} \,. \tag{16}$$

To find the Bayes' factor in (10), we have to multiply (16) with the priors on $\boldsymbol{\omega}_l$ and g and integrate this product over these parameters. Unfortunately, this cannot be done analytically, and the integral is, therefore, evaluated using the Laplace approximation. In making this Laplace approximation, we have to approximate the posterior distribution over the frequency parameter with one or more normal distributions. To do this, we have to find the MAP estimate of the frequencies and their associated uncertainty.

2.2.1. Inference for the Frequency Parameters

Since the MAP estimate of the frequency parameters does not depend on the value of g, it is given by

$$\begin{aligned}\hat{\boldsymbol{\omega}}_l^{\text{MAP}} &= \underset{\boldsymbol{\omega}_l\in\boldsymbol{\Omega}_l}{\arg\max}\, p(\boldsymbol{\omega}_l|\boldsymbol{y}_{\rm f},l) = \underset{\boldsymbol{\omega}_l\in\boldsymbol{\Omega}_l}{\arg\max}\, p(\boldsymbol{y}_{\rm f}|\boldsymbol{\omega}_l,g,l)p(\boldsymbol{\omega}_l|l) \\ &= \underset{\boldsymbol{\omega}_l\in\boldsymbol{\Omega}_l}{\arg\max}\, R_l^2(\boldsymbol{\omega}_l) = \underset{\boldsymbol{\omega}_l\in\boldsymbol{\Omega}_l}{\arg\max}\, \boldsymbol{y}_{\rm f}^H\boldsymbol{P}_l(\boldsymbol{\omega}_l)\boldsymbol{y}_{\rm f}\end{aligned} \tag{17}$$

where the third equality follows from (8) and (12). The MAP estimate of $\boldsymbol{\omega}_l$ is therefore the same as the ML estimate which we here compute using a slightly modified version of the weighted RELAX (WRELAX) algorithm [29] originally proposed for time delay estimation. The WRELAX algorithm iteratively estimates the frequencies in L steps. In the l'th step, the signal is assumed to consist of l frequency components whose parameters are re-estimated in a circular fashion until practical convergence. Thus, the intractable and multi-dimensional optimisation problem in (17) is replaced by a series of simpler one-dimensional optimisation problems of the form

$$\hat{\omega}_i = \underset{\omega\in\Omega_i}{\arg\max}\, \frac{|\boldsymbol{z}^H(\omega)\boldsymbol{W}^H\boldsymbol{r}_i|^2}{\boldsymbol{z}^H(\omega)\boldsymbol{W}^H\boldsymbol{W}\boldsymbol{z}(\omega)} \tag{18}$$

where the $(n+1)$'th element of $\boldsymbol{z}(\omega)$ is $\exp(j\omega n)$ and $\boldsymbol{r}_i = \boldsymbol{y}_{\rm f} - \boldsymbol{W}\boldsymbol{Z}_{l\backslash i}(\hat{\boldsymbol{\omega}}_l)\hat{\boldsymbol{\alpha}}_{l\backslash i}$ with $l\backslash i$ denoting without column or element i. For every $i = l, l-1,\ldots,1$, the ML estimate of the complex amplitudes are computed, and this is repeated until the relative change of $\|\boldsymbol{y}_{\rm f} - \boldsymbol{W}\boldsymbol{Z}_l(\hat{\boldsymbol{\omega}}_l)\hat{\boldsymbol{\alpha}}_l\|_2^2$ at $i=1$ is below some treshold ϵ. The optimisation problem in (18) is non-convex, but can be solved efficiently by first finding a coarse frequency estimate on a DFT grid and then by refining it using a golden section search. The coarse estimate can be computed efficiently by pre-computing the DFT of the columns of $\boldsymbol{W}^H$ using M FFTs. From these DFTs, the denominator in (18) can also be pre-computed leaving only a single matrix-vector product in the numerator to be recomputed in every iteration of the WRELAX algorithm.

Given g, the posterior uncertainty about $\boldsymbol{\omega}_l$ is represented by the Hessian matrix $\boldsymbol{H}(\boldsymbol{\omega}_l|g)$ of $\ln p(\boldsymbol{\omega}_l|\boldsymbol{y}_{\rm f},g,l)$ at the mode $\hat{\boldsymbol{\omega}}_l^{\text{MAP}}$. Finding an analytical expression for $\boldsymbol{H}(\boldsymbol{\omega}_l|g)$ is a tedious task, and we refer the interested reader to the derivation in [15]. However, when the SNR is large enough, the frequencies are well-separated (relative to N), and the projection matrix $\boldsymbol{W}^H\boldsymbol{W}$ is replaced by its expectation, a very simple approximation to $\boldsymbol{H}(\hat{\boldsymbol{\omega}}_l^{\text{MAP}}|g)$ is

$$\boldsymbol{H}(\hat{\boldsymbol{\omega}}_l^{\text{MAP}}|g) \approx -\frac{gMN^2}{6(1+g)\hat{\sigma}_l^2(\hat{\boldsymbol{\omega}}_l^{\text{MAP}},g)}\,\text{diag}(|\hat{\boldsymbol{\alpha}}_l|^2) \,. \tag{19}$$

In our experience, this approximation works very well in practice. Thus, a normal approximation around the mode $\hat{\boldsymbol{\omega}}_l^{\text{MAP}}$ is $\mathcal{N}(\boldsymbol{\omega}_l;\hat{\boldsymbol{\omega}}_l^{\text{MAP}}, -\boldsymbol{H}^{-1}(\hat{\boldsymbol{\omega}}_l^{\text{MAP}}|g))$. However, since the posterior distribution $p(\boldsymbol{\omega}_l|\boldsymbol{y}_{\rm f},g,l)$ is invariant to permutations of the frequency parameters, $p(\boldsymbol{\omega}_l|\boldsymbol{y}_{\rm f},g,l)$ consists of $l! = \Gamma(l+1)$ identical peaks which all contribute in the Laplace approximation which we consider next.

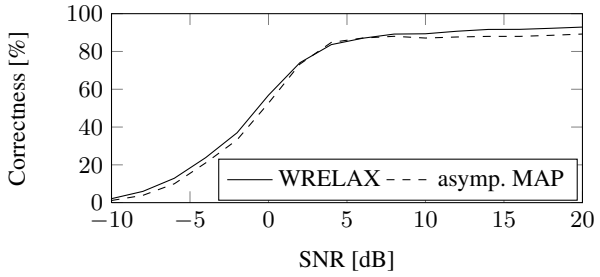

Fig. 2. The model detection correctness of the WRELAX algorithm and asymptotic MAP criterion.

2.2.2. The Laplace Approximation

Since the marginal posterior pdf over g is not symmetric and in order to avoid edge effect near $g=0$, the re-parametrisation $\tau = \ln g$ with the Jacobian $dg/d\tau = \exp(\tau)$ is made [25]. This re-parametrisation suggests that the posterior distribution over g is approximately a log-normal distribution. Thus, if we define the integrand

$$q(\boldsymbol{\omega_l},\tau) \triangleq \text{BF}[l;0|\boldsymbol{\omega}_l,\exp(\tau)]p(\boldsymbol{\omega_l}|l)p(\tau) \,, \tag{20}$$

the Laplace approximation to the Bayes' factor in (10) is [15]

$$\text{BF}[l;0] = \int_{\boldsymbol{\Omega}_l}\int_{-\infty}^{\infty} q(\boldsymbol{\omega}_l,\tau)d\tau d\boldsymbol{\omega}_l \tag{21}$$

$$\approx \text{BF}[l;0|\hat{\boldsymbol{\omega}}_l^{\text{MAP}},\hat{g}]\frac{\hat{g}(\delta-1)(2\pi)^{(l+1)/2}}{(1+\hat{g})^{\delta}W_l} \times\sqrt{\gamma(\hat{g}|\hat{\boldsymbol{\omega}}_l^{\text{MAP}})}\left|-\boldsymbol{H}(\hat{\boldsymbol{\omega}}_l^{\text{MAP}}|\hat{g})\right|^{-1/2}\Gamma(l+1) \tag{22}$$

where we have defined

$$\hat{g} = -(\beta_\tau + \sqrt{\beta_\tau^2 - 4\alpha_\tau})/(2\alpha_\tau) \tag{23}$$

$$\alpha_\tau = (1-R_l^2(\hat{\boldsymbol{\omega}}_l^{\text{MAP}}))(1-l-\delta) \tag{24}$$

$$\beta_\tau = (M-1)R_l^2(\hat{\boldsymbol{\omega}}_l^{\text{MAP}}) + (2-l-\delta) \tag{25}$$

$$\gamma(\hat{g}|\hat{\boldsymbol{\omega}}_l^{\text{MAP}}) = \frac{1}{\hat{g}}\left[\frac{M(1-R_l^2(\hat{\boldsymbol{\omega}}_l^{\text{MAP}}))}{[1+\hat{g}(1-R_l^2(\hat{\boldsymbol{\omega}}_l^{\text{MAP}}))]^2} - \frac{M-l-\delta}{(1+\hat{g})^2}\right]^{-1} \,. \tag{26}$$

Inserting the Bayes' factor in (22) into (9) allows us to select the most likely model order and to compute the posterior probabilities of all model orders.

3. SIMULATIONS

We here compare our proposed joint frequency and model order estimator to the Band-Excluded Interpolating Subspace Pursuit (BISP) algorithm recently proposed in [9] and to the CRLB with and without CS [30]. The BISP algorithm is based on polar interpolation between the dictionary elements [31], and it was demonstrated to outperform other baseline and state-of-the-art interpolation methods in terms of estimation accuracy. Moreover, the BISP[1] algo-

[1] We thank the authors of [9] for making their implementation of the BISP algorithm publicly available.

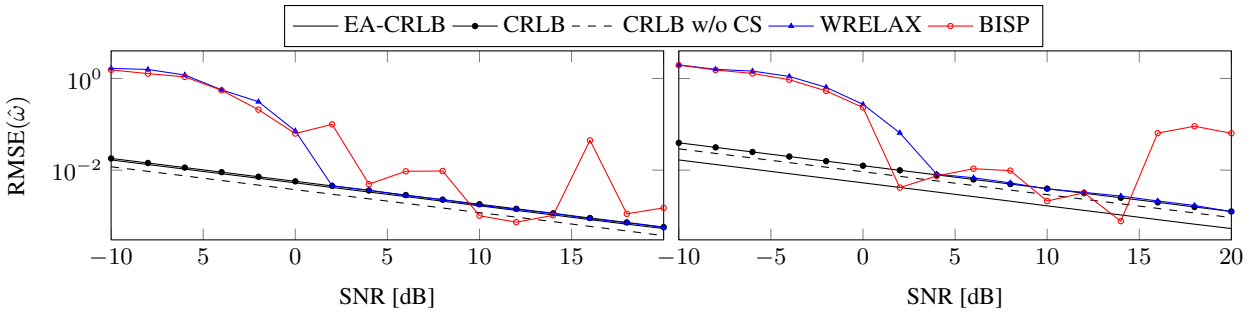

Fig. 1. The RMSE of first (left) and third (right) frequency estimates produced by the WRELAX and the BISP algorithms.

rithm has a lower computational complexity than many of the other methods. Contrary to our proposed algorithm, the BISP algorithm assumes the noise variance to be known, that $\boldsymbol{\alpha}_l$ consists of non-negative and real-valued amplitudes, and that the number of sinusoids is known. The simulation code used for generating the results presented here can be found at `http://kom.aau.dk/~jkn/publications/publications.php`.

3.1. Estimation Accuracy

To evaluate the estimation accuracy of the proposed frequency estimator, we considered the same parameter setup as in [9]. That is, a data vector of length $N = 100$ was compressed with a WGN sensing matrix to $M = 50$ measurements, and the signal was set to consist of $l = 4$ sinusoids. The estimation performance was evaluated using a Monte Carlo simulation consisting of 500 runs at SNRs from -10 dB to 20 dB in steps of 2 dB. Since the exact CRLB depends on the sensing matrix and the model parameters, these were fixed to[2]

$$\boldsymbol{\omega}_4 = \angle\boldsymbol{\alpha}_4 = \begin{bmatrix}0.5 & 1.5 & 1.5+2\pi/N & 3\end{bmatrix}^T \tag{27}$$

$$|\boldsymbol{\alpha}_4| = \begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}^T/\sqrt{N} \tag{28}$$

in all runs. As the order of the estimated frequencies is not necessarily the same as the order of the frequencies in $\boldsymbol{\omega}_4$, the estimated frequencies were permuted using the Hungarian algorithm [32] before the estimation errors were computed.

In Fig. 1, the estimation performance is shown for the first (left plot) and the third (right plot) frequency parameters. The first frequency parameter was well-separated from the rest whereas the third parameter was only $2\pi/N$ radians/sample larger than the second. When sinusoidal components are close, they are harder to estimate and this is also reflected in the increased CRLB. The expected asymptotic CRLB (EA-CRLB) [30] is derived under the assumption that the sinusoids are well-separated (relative to N) and is therefore too optimistic for closely spaced sinusoids. In the two plots, we therefore observe that the CRLB and the EA-CRLB nearly coincided for the first component, but were distinct for the third component. We also observe that the CRLB with CS was $N/M = 2$ times the CRLB without CS which is consistent with our previous findings in [13, 30] that a penalty is paid in terms of the estimation accuracy by employing CS. However, when we are only given the measured samples in $\boldsymbol{y}$ and the SNR is large enough, we see that the proposed WRELAX estimator is an asymptotically efficient estimator since it attained the CRLB for SNRs above 2 dB. On the other hand, the estimation accuracy of the BISP algorithm was above the bound for most SNRs. Since the bound was derived under the assumption that the phases and the noise variance are unknown, an algorithm assuming the same strong prior knowledge as the BISP estimator should be able to have an RMSE below the bound. However, this is not the case and the BISP estimator is therefore suboptimal. Moreover, the computation time of our implementation of the proposed WRELAX algorithm is noticeable lower than that of the BISP algorithm.

[2] In [9], the phase is assumed to be zero for a time index running from 1 to N. To compensate for that the time index here runs from 0 to $N-1$, the phase and the frequency are therefore identical. When running the simulations, we noticed that the performance of the BISP algorithm was sensitive to how the ground truth amplitudes were selected, and we therefore used the same values for the amplitudes as in [9].

3.2. Model Detection Performance

To evaluate the model detection performance of the proposed WRELAX algorithm, we conducted another Monte Carlo simulation consisting of 5000 runs for each SNR from -10 dB to 20 dB in steps of 2 dB. As recommended in [33], the model, the model parameters, and the noise realisation were generated at random in each run. The model order was generated uniformly from the set $\{1, \ldots, 5\}$, the amplitudes were set to ones, and the phases and frequencies were generated uniformly from the interval $[0, 2\pi)$. We compared the performance of the proposed WRELAX algorithm to the asymptotic MAP criterion for frequency estimation [23] and in both of these methods, the range of candidate model orders were the set $\mathcal{L} = \{0, 1, \ldots, 8\}$. In Fig. 2, the amount of correctly detected model orders are shown. Clearly, both model selection criteria performed well with the WRELAX algorithm performing better than the asymptotic MAP criterion for most SNRs. In our experience, the performance difference between the two methods increases for shorter data length and decreases for longer data length.

4. CONCLUSION

We have here considered the problem of jointly estimating the frequency parameters and the sparsity level of a frequency sparse signal from compressively sensed measurements. In contrast to previously proposed interpolation methods, we have analysed the problem in a probabilistic framework and based the inference on the physical signal model. Not only did this enable us to easily cope with unknown nuisance parameters such as the complex amplitudes and the noise variance, but it also produced the WRELAX frequency estimator which outperformed a state-of-the-art interpolation based method in terms of both estimation accuracy and computation time. Moreover, inference was also made about the sparsity level or model order, and we demonstrated that the proposed WRELAX estimator outperformed the popular asymptotic MAP criterion.

5. REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[2] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[3] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[4] R. J. Sluijter, *The Development of Speech Coding and the First Standard Coder for Public Mobile Telephony*, Ph.D. thesis, Techniche Universiteit Eindhoven, 2005.

[5] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, Springer, 2 edition, Jun. 1998.

[6] S. M. Kay, *Modern Spectral Estimation: Theory and Application*, Prentice Hall, 1988.

[7] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*, Englewood Cliffs, NJ, USA: Prentice Hall, May 2005.

[8] M. F. Duarte and R. G Baraniuk, "Spectral compressive sensing," *Appl. Comput. Harmon. Anal.*, vol. 35, no. 1, pp. 111–129, July 2012.

[9] K. Fyhn, H. Dadkhahi, and M. F. Duarte, "Spectral compressive sensing with polar interpolation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2013, pp. 6225–6229.

[10] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Appl. Comput. Harmon. Anal.*, vol. 31, no. 1, pp. 59–73, July 2011.

[11] M. D. Macleod, "Fast nearly ML estimation of the parameters of real or complex single tones or resolved multiple tones," *IEEE Trans. Signal Process.*, vol. 46, no. 1, pp. 141–148, Jan. 1998.

[12] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, Englewood Cliffs, NJ, USA: Prentice Hall PTR, Mar. 1993.

[13] M. G. Christensen and J. K. Nielsen, "Joint direction-of-arrival and order estimation in compressed sensing using angles between subspaces," in *Proc. IEEE Workshop on Stat. Signal Process.*, June 2011, pp. 449–452.

[14] M. G. Christensen, A. Jakobsson, and S. H. Jensen, "Sinusoidal order estimation using angles between subspaces," *EURASIP J. on Advances in Signal Process.*, vol. 2009, pp. 1–11, Nov. 2009.

[15] J. K. Nielsen, M. G. Christensen, A. T. Cemgil, and S. H. Jensen, "Bayesian model comparison with the g-prior," *IEEE Trans. Signal Process.*, vol. 62, no. 1, pp. 225–238, 2014.

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

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

[18] P. Stoica and P. Babu, "The Gaussian data assumption leads to the largest Cramér-Rao bound," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 132–133, May 2011.

[19] K. Kim and G. Shevlyakov, "Why Gaussianity?," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 102–113, Mar. 2008.

[20] E. Arias-Castro and Y. C. Eldar, "Noise folding in compressed sensing," *IEEE Signal Process. Lett.*, vol. 18, no. 8, pp. 478–481, Aug. 2011.

[21] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, Dec. 1974.

[22] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, Sep. 1978.

[23] P. M. Djuric, "Asymptotic MAP criteria for model selection," *IEEE Trans. Signal Process.*, vol. 46, no. 10, pp. 2726–2735, Oct. 1998.

[24] A. Zellner, "On assessing prior distributions and Bayesian regression analysis with g-prior distributions," in *Bayesian Inference and Decision Techniques*. New York, NY, USA: Elsevier, 1986.

[25] F. Liang, R. Paulo, G. Molina, M. A. Clyde, and J. O. Berger, "Mixtures of g priors for Bayesian variable selection," *J. Amer. Statistical Assoc.*, vol. 103, pp. 410–423, Mar. 2008.

[26] J. K. Nielsen, M. G. Christensen, and S. H. Jensen, "Bayesian model comparison and the BIC for regression models," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2013, pp. 6362–6366.

[27] R. Guo and P. L. Speckman, "Bayes factor consistency in linear models," in *The 2009 International Workshop on Objective Bayes Methodology*, Jun. 2009.

[28] J. O. Berger and L. R. Pericchi, "Objective Bayesian methods for model selection: Introduction and comparison," *Institute of Mathematical Statistics Lecture Notes – Monograph Series*, vol. 38, pp. 135–207, 2001.

[29] J. Li and R. Wu, "An efficient algorithm for time delay estimation," *IEEE Trans. Signal Process.*, vol. 46, no. 8, pp. 2231–2235, Aug. 1998.

[30] J. K. Nielsen, M. G. Christensen, and S. H. Jensen, "On compressed sensing and the estimation of continuous parameters from noisy observations," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2012.

[31] C. Ekanadham, D. Tranchina, and E. P. Simoncelli, "Recovery of sparse translation-invariant signals with continuous basis pursuit," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4735–4744, 2011.

[32] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval research logistics quarterly*, vol. 2, no. 1–2, pp. 83–97, Mar. 1955.

[33] P. Stoica, Y. Selén, and J. Li, "Multi-model approach to model selection," *Digital Signal Process.*, vol. 14, no. 5, pp. 399–412, Sep. 2004.