DECOMPOSING SIGNALS INTO A SUM OF AMPLITUDE AND FREQUENCY MODULATED SINUSOIDS USING PROBABILISTIC INFERENCE

Richard E. Turner[*]

University of Cambridge,
Department of Engineering,
Cambridge, CB2 1PZ, UK.

Maneesh Sahani[†]

University College London
Gatsby Computational Neuroscience Unit,
London, WC1N 3AR, UK.

ABSTRACT

There are many methods for decomposing signals into a sum of amplitude and frequency modulated sinusoids. In this paper we take a new estimation based approach. Identifying the problem as ill-posed, we show how to regularize the solution by imposing soft constraints on the amplitude and phase variables of the sinusoids. Estimation proceeds using a version of Kalman smoothing. We evaluate the method on synthetic and natural, clean and noisy signals, showing that it outperforms previous decompositions, but at a higher computational cost.

Index Terms— frequency modulation, amplitude modulation, frequency estimation, amplitude estimation, machine learning

1. INTRODUCTION

Many engineering and scientific problems require a signal to be decomposed into a sum of amplitude modulated and frequency modulated sinusoids (AM-FM sinusoids) [1, 2]. Perhaps the most popular decomposition of this type first filters the signal and then demodulates the subbands, for example using the Hilbert Transform. Despite its popularity, subband demodulation often returns amplitudes and instantaneous frequencies (IFs) which are poorly behaved [1]. The methods encounter additional problems when each component in the signal cannot be isolated into a single subband. Sinusoidal analysis addresses some of these deficiencies by tracking the sinusoidal components present in a signal using a set of heuristics [3]. However, this analysis would benefit from an explicit signal model that simplifies the heuristics and automates how they should trade-off with one another.

We address these problems through an estimation based approach. Recently we have shown how to fit a single AM-FM sinusoid to a signal using probabilistic amplitude and frequency demodulation (PAFD, [2]). Here we generalize PAFD to the multi-sinusoid case. We begin by reinterpreting two ex-

isting algorithms as probabilistic versions of subband demodulation which serves to motivate the new approach. We end by evaluating the methods on synthetic and natural data, in clean and noisy settings. The new approach is the most accurate and robust, but also the most computationally demanding.

2. SIMPLE ALGORITHMS FOR PAFD

In this section we reinterpret two existing algorithms as methods for fitting a signal y_t with a sum of D AM-FM sinusoids,

$$y_t = \sum_{d=1}^{D} \Re\left(a_{d,t} \exp\left(i\phi_{d,t}\right)\right) + \epsilon_t. \tag{1}$$

The noise term in this expression, a zero mean Gaussian with variance σ_y^2, $p(\epsilon_t) = \text{Norm}(\epsilon_t; 0, \sigma_y^2)$, enables the methods to operate in both noisy and incomplete data settings.

A key assumption of this work is that the instantaneous frequencies of the sinusoids ($\dot{\phi}_{d,t} = \frac{1}{2\pi}(\phi_{d,t} - \phi_{d,t-1})$) undergo slow perturbations around a mean value ($\bar{\omega}_d$), so their phase can be expressed, $\phi_{d,t} = \bar{\omega}_d t + \theta_{d,t}$. The positive amplitude variables are also assumed to be slow. The ill-posed estimation problem involves inferring the amplitude and phase-perturbation variables. Later we will regularize the problem by placing soft constraints on these variables directly, but first, we use a simple alternative based on a reparameterization. Treating each phase-perturbation as a phasor, with components $\mathbf{x}_{d,t}^\top = [x_{d,t}^{(1)}, x_{d,t}^{(2)}] = a_{d,t}[\cos(\theta_{d,t}), \sin(\theta_{d,t})]$, simplifies the likelihood function,

$$y_t = \sum_{d=1}^{D} a_{d,t}\left(\cos(\bar{\omega}_d t)\cos(\theta_{d,t}) - \sin(\bar{\omega}_d t)\sin(\theta_{d,t})\right) + \epsilon_t,$$

$$= \sum_{d=1}^{D}\left(\cos(\bar{\omega}_d t)x_{d,t}^{(1)} - \sin(\bar{\omega}_d t)x_{d,t}^{(2)}\right) + \epsilon_t = \mathbf{w}_t^\top \mathbf{x}_t + \epsilon_t.$$

Here the phasors have been collected into a vector, $\mathbf{x}_t^\top = [x_{1,t}^{(1)}, x_{1,t}^{(2)}, \ldots, x_{D,t}^{(1)}, x_{D,t}^{(2)}]$, that multiplies time-varying weights, $\mathbf{w}_t^\top = [\cos(\bar{\omega}_1 t), -\sin(\bar{\omega}_1 t), \ldots, \cos(\bar{\omega}_D t), -\sin(\bar{\omega}_D t)]$. Now the estimation problem can be regularized by placing prior distributions over the components of the phasors.

A simple choice uses independent one-step Gaussian autoregressive AR(1) priors,

$$\mathrm{x}_{d,t}^{(k)} = \lambda_d \mathrm{x}_{d,t-1}^{(k)} + \sigma_{\mathrm{x},d}\,\eta_{d,t}^{(k)}, \quad p(\eta_{d,t}^{(k)}) = \mathrm{Norm}(\eta_{d,t}^{(k)}; 0, 1).$$

This model is called Bayesian Spectrum Estimation (BSE) [4] (when $\lambda_d = 1$). As the model is a linear Gaussian state space model, estimation is a least-squares problem which can be solved using the Kalman smoothing algorithm.

An alternative version of the same model can be established using the complex representation of the sinusoids treating the imaginary part as missing data,

$$y_t = \sum_{d=1}^{D} \Re\left(z_{d,t}^{(1)} + i z_{d,t}^{(2)}\right) + \epsilon_t = [1, 0, \ldots, 1, 0]\mathbf{z}_t + \epsilon_t.$$

The new variables $\mathbf{z}_t^{\mathsf{T}} = [z_{1,t}^{(1)}, z_{1,t}^{(2)}, \ldots, z_{D,t}^{(1)}, z_{D,t}^{(2)}]$ are related to the old ones by frequency shifts. The frequency shift can be expressed as a rotation from the old phasors to the new ones, $\mathbf{z}_{d,t} = \mathrm{R}(\bar{\omega}_d t)\mathbf{x}_{d,t}$, where the rotation matrix is

$$\mathrm{R}(\theta) = \left[\begin{array}{cc} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{array}\right].$$

The relationship allows the dynamics of the new variables to be derived from the dynamics of the old variables,

$$\begin{aligned}\mathbf{z}_{d,t} &= \lambda_d \mathrm{R}(\bar{\omega}_d)\mathrm{R}(\bar{\omega}_d(t-1))\mathbf{x}_{d,t-1} + \mathrm{R}(\bar{\omega}_d t)\boldsymbol{\eta}_{d,t}, \\ &= \lambda_d \mathrm{R}(\bar{\omega}_d)\mathbf{z}_{d,t-1} + \boldsymbol{\eta}'_{d,t}.\end{aligned}$$

The new dynamics are also AR(1), but now variables at time t are generated by first rotating the variables at time $t-1$, before shrinking by λ_d and adding zero mean Gaussian noise with covariance $\langle \boldsymbol{\eta}'_{d,t}\boldsymbol{\eta}'^{\mathsf{T}}_{d,t}\rangle = \mathrm{R}(\bar{\omega}_d t)\langle\boldsymbol{\eta}_{d,t}\boldsymbol{\eta}^{\mathsf{T}}_{d,t}\rangle\mathrm{R}^{\mathsf{T}}(\bar{\omega}_d t) = \sigma_{\mathrm{x},d}^2\mathrm{I}$.

This equivalent formulation of the BSE model is called the Probabilistic Phase Vocoder (PPV) [5]. The relationship between BSE and PPV is analogous to the relationship between the Short-Time Fourier Transform (STFT) and the analytic signal of a filter bank (ASFB) [6]. This suggests that BSE should be identified as a probabilistic version of the STFT, whilst PPV should be identified as a probabilistic version of an analytic filter bank. The envelopes extracted from either representation can be viewed as a probabilistic version of the spectrogram or equivalently a probabilistic version of subband demodulation (see fig. 1).
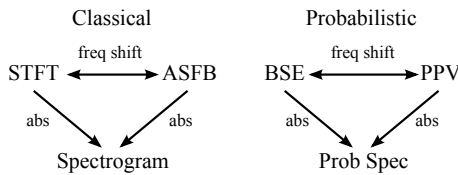
Fig. 1. Relating classical and probabilistic time-frequency representations.

2.1. Problems with the simple approach

The connection between classical subband demodulation and probabilistic methods is encouraging, but the probabilistic methods share some of the deficiencies of their classical counterparts. For example the IF estimates returned by BSE-PPV are often ill-behaved, varying erratically and taking unphysical negative values (see fig. 2B, bottom panel in blue). The source of the problem is the way we regularized the solution. First, because higher order AR processes are needed to encourage derivatives to be slow. Second, and more seriously, because placing priors on the components of the phasors favours solutions in which IF varies slowly when the amplitude is large, but extremely quickly when the amplitude is low. A second problem arises when a signal contains components which cannot be isolated in a single subband (see fig. 2B, bottom panel in blue). In such situations it would be desirable to have a method which identified a single set of amplitude and IF variables with each component in the signal. However, in the methods described above the information is scrambled across the amplitude and IF variables in multiple subbands. In the remainder of the paper, we address these problems with BSE-PPV.

3. SOPHISTICATED ALGORITHMS FOR PAFD

In this section we describe a method for fitting a sum of AM and FM sinusoids to a signal that regularizes the amplitude and phase variables separately. Two important design choices lead to a tractable optimisation problem. First, we work with the phase variables directly because they are less strongly coupled than IF variables. This stems from the fact that a change in an IF variable affects not only the current signal sample, but all subsequent samples too. Second, we regularize the amplitude and phase variables using modified AR priors because this leads to efficient estimation schemes based on Kalman smoothing. With these general principles in mind, the next section will introduce the model in detail.

3.1. Forward model

The new version of the model uses the same likelihood function as before (eq. 1). A modified AR prior is used to regularize the amplitude variables,

$$\begin{aligned}p(\mathrm{a}_{d,1:T}|\lambda_{d,1:\tau}, \sigma_d^2) &\propto \prod_{t=1}^{T} \mathbf{1}(\mathrm{a}_{d,t} \geq 0) \\ &\times \mathrm{Norm}(\mathrm{a}_{d,t}; \sum_{t'=1}^{\tau} \lambda_{d,t'}\mathrm{a}_{d,t-t'}, \sigma_d^2).\end{aligned}$$

Here, the constraint that the amplitudes are positive variables is enforced by an indicator function, $\mathbf{1}(\mathrm{a}_{d,t} \geq 0)$.

A similar regularizer can be used for phase perturbation variables, by employing the phasor representation

$c_{d,t}^{\mathsf{T}} = [c_{d,t}^{(1)}, c_{d,t}^{(2)}] = [\cos(\theta_{d,t}), \sin(\theta_{d,t})]$. We specify second order AR dynamics constrained to the unit circle,

$$p(c_{d,1:T}^{(1:2)}) \propto \prod_{t=1}^{T} \mathbf{1}((c_{d,t}^{(1)})^2 + (c_{d,t}^{(2)})^2 = 1)$$
$$\times \prod_{k=1}^{2} \mathrm{Norm}(c_{d,t}^{(k)}; \lambda_{d,1} c_{d,t-1}^{(k)} + \lambda_{d,2} c_{d,t-2}^{(k)}, \sigma_{\mathrm{x},d}^2).$$

The model parameters pertaining to a single component (the mean frequencies of the sinusoids, and the dynamical parameters of the amplitude and phase perturbation variables) will be denoted, $\alpha_d = \{\omega_d, \lambda_{d,1:2}, \sigma_{\mathrm{x},d}^2, \lambda_{d,1:\tau}, \sigma_d^2\}$.

The model described above will be termed multi-component Probabilistic Amplitude and Frequency Demodulation (mPAFD) because it is a generalisation of PAFD [2]. PAFD is itself a generalisation of convex amplitude demodulation [7, 8]. In the next section we discuss how to estimate the variables using a recursive version of the Kalman smoother.

3.2. Estimation

PAFD contains three different non-linear contributions; the product in the likelihood, the positivity constraint on the amplitudes, and the circular constraint on the phasor variables. These non-linearities make estimation difficult, but in earlier work we described how to handle them when there is just one component $D = 1$ [2]. The main idea is to replace the non-linear likelihood function and the hard constraints with soft Gaussian approximations. The approximations are iteratively refined using a moment-matching algorithm called expectation propagation [9] which proceeds via a modified version of the Kalman Smoother because of the AR structure of the regularisation and the Gaussian form of the approximation.

We extend the old method to handle multiple sinusoids via a pursuit style algorithm. Initially the amplitude and phase perturbation variables are set to zero. To update the estimate of the dth component, we subtract from the signal the contribution from the other components, $y_{d,t}^{\mathrm{eff}} = y_t - \sum_{d' \neq d} a_{d',t}(\cos(\bar{\omega}_{d'}t)c_{d',t}^{(1)} - \sin(\bar{\omega}_{d'}t)c_{d',t}^{(2)})$, and run PAFD on this residual, $y_{d,t}^{\mathrm{eff}}$, returning $a_{d,t}$ and $c_{d,t}^{(k)}$. This process is repeated for each component multiple times until convergence. Annealing prevents local optima. The algorithm is outlined below.

> $a_{d,t} \leftarrow 0, c_{d,t}^{(k)} \leftarrow 0$ {initialize envelopes and carriers}
> **for** $k = 1 \rightarrow K$ **do**
> $\sigma_{\mathrm{y}}^2 \leftarrow \mathrm{ANNEAL}(k, \sigma_{\mathrm{y},1}^2, \sigma_{\mathrm{y},K}^2)$ {anneal σ_{y}^2}
>   **for** $d = 1 \rightarrow D$ **do**
> $y_{d,t}^{\mathrm{eff}} \leftarrow y_t - \sum_{d' \neq d} a_{d',t}(\cos(\bar{\omega}_{d'}t)c_{d',t}^{(1)} - \sin(\bar{\omega}_{d'}t)c_{d',t}^{(2)})$
> $[a_{d,t}, c_{d,t}^{(k)}] \leftarrow \mathrm{PAFD}(y_{d,t}^{\mathrm{eff}}, \alpha_d, \sigma_{\mathrm{y}}^2)$ {re-estimate}
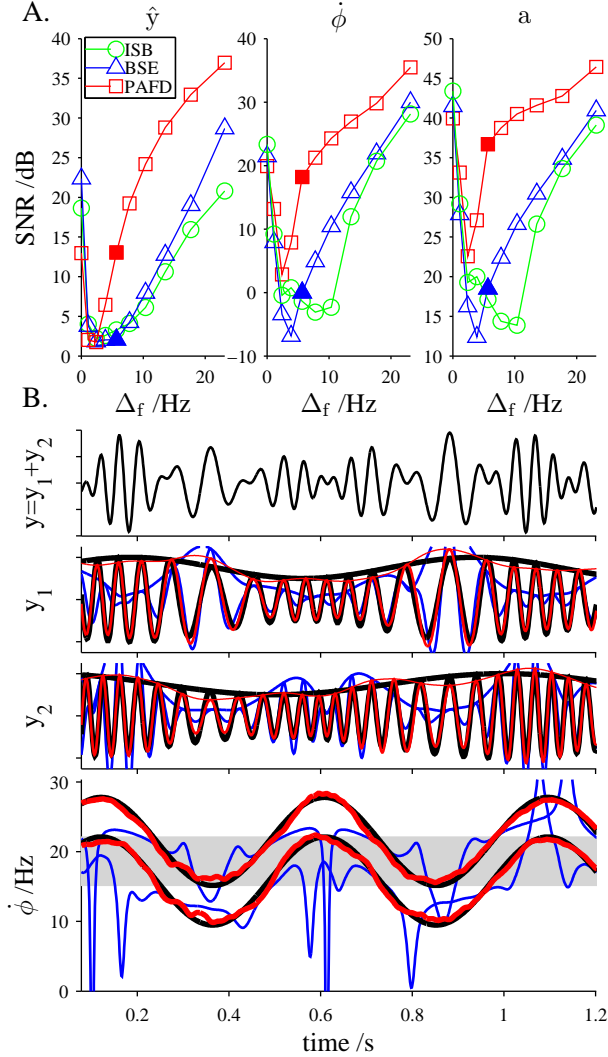>   **end for**
> **end for**

4. EXPERIMENTS

It is difficult to evaluate AFD because ground truth estimates for signals of interest, like speech, are unknown. Evaluation must therefore take an indirect approach. Here we first use simple synthetic examples, before moving to a natural signal.

The first set of evaluations use a synthetic signal comprising two AM-FM sinusoids with additive Gaussian noise, $y_t = y_{1,t} + y_{2,t} + \epsilon_t = \sum_{d=1}^{2} a_{d,t}\cos(\phi_{d,t}) + \epsilon_t$. An example signal is shown in fig. 2B in black. The top panel shows the signal, the middle two panels the two sinusoidal components, and the bottom panel the IF of the components. Both the amplitude and IF variables undergo sinusoidal modulation. The difference between the mean IFs of the components is denoted Δ_{f}. It is simple to estimate the two sinusoidal components when they occupy different subbands (large Δ_{f}), but it is difficult when there is partial overlap (indicated by the gray area in the bottom panel). When Δ_{f} is zero the signal is equivalent to a single AM-FM sinusoid so estimation is again simple. The first experiment evaluates performance as Δ_{f} varies, comparing mPAFD (fig. 2B, red estimates), BSE-PPV (fig. 2B, blue estimates) and an ideal version of subband demodulation (ISB) which optimizes the parameters of half-cosine filters to maximize performance on a signal by signal basis. The performance of this method represents an upper bound for a blind version. As fig. 2A shows, mPAFD is able to estimate the components most reliably as indicated by the superior signal to noise ratio (SNR), (defined as $\mathrm{SNR}(x, \hat{x}) = 10\log_{10}\mathrm{var}(x)/\mathrm{var}(\hat{x} - x)$ where var is the variance). Similarly, when noise is added to the signal, mPAFD is again the most accurate method (see fig. 3, which uses large Δ_{f} because this is most favourable to the ISB method).

Finally we evaluate mPAFD on a noisy song containing glissandi (see fig. 4). As a surrogate to ground truth, IF tracks were labeled by hand using a spectrographic representation. These tracks were compared those estimated using the automated methods. We find SNRs of 20.7dB for ISB, 23.3dB for BSE-PPV and 32.3dB for mPAFD.
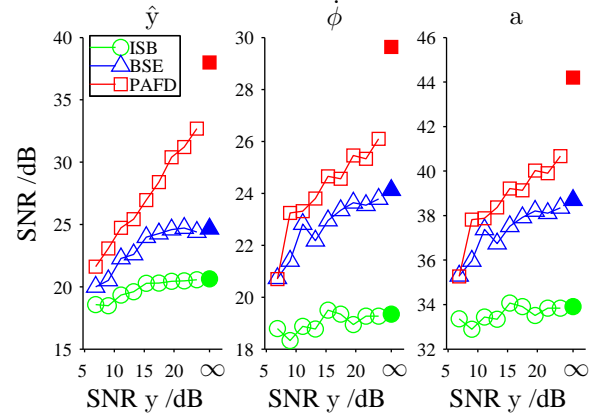
5. CONCLUSION

We have developed a probabilistic estimation based approach for decomposing a signal into a sum of AM-FM sinusoids called mPAFD. Based on Kalman Smoothing, mPAFD is more computationally intensive than previous approaches. However, unlike many other approaches, it is able to track components when they cannot be isolated into a single subband and it is also robust to noise. We hope that the probabilistic formulation will lead to automatic methods for learning the parameters of the model, therefore making the representation adaptive.
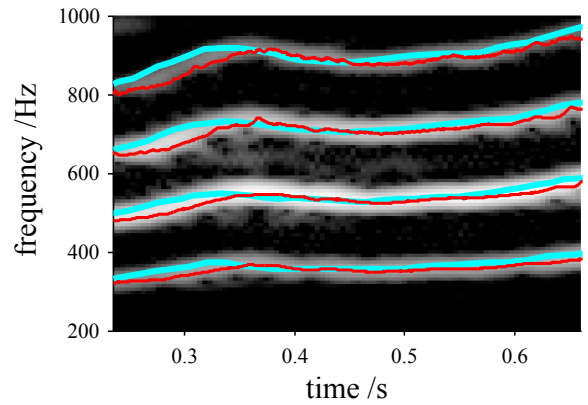
Fig. 3. Denoising a signal comprising a pair of noise corrupted sinusoids. The panels show the SNRs of the reconstructed component signals (left) their IFs (middle) and their amplitudes (right). Solid markers indicate estimates derived from a noiseless signal.

Fig. 4. Tracking IF in noisy song. A close up of a section of the spectrogram of the song. Overlaid are the IF estimates derived from mPAFD (red) and the labeled IF tracks (cyan).

Fig. 2. Estimating the amplitudes and IFs of a signal containing two AM-FM sinusoids. A) SNR of the estimated components (left) IFs (middle) and amplitudes (right) as the separation between the two components (dF) increases. The solid markers indicate the example shown below. B) An example signal (black) with mPAFD estimates (red) and BSE-PPV (blue). See text for more details.

6. REFERENCES

[1] P. Clark and L.E. Atlas, "Time-frequency coherent modulation filtering of nonstationary signals," *Trans. Sig. Proc.*, vol. 57, no. 11, pp. 4323 –4332, 2009.

[2] R. E. Turner and M. Sahani, "Probabilistic amplitude and frequency demodulation," in *NIPS*, 2012.

[3] R. McAulay and T. Quatieri, "Speech analysis/Synthesis based on a sinusoidal representation," *TASLP*, vol. 34.

[4] Y. Qi, T. Minka, and R. Picard, "Bayesian spectrum estimation of unevenly sampled nonstationary data," in *ICASSP*, 2002.

[5] A. T. Cemgil and S. J. Godsill, "Probabilistic Phase Vocoder and its application to Interpolation of Missing Values in Audio Signals," in *Euro. Sig. Proc. Conf.*, 2005.

[6] J. L. Flanagan, "Parametric coding of speech spectra," *JASA*, vol. 68, pp. 412–419, 1980.

[7] G. Sell and M. Slaney, "Solving demodulation as an optimization problem," *TASLP*, vol. 18, 2010.

[8] R. E. Turner and M. Sahani, "Demodulation as probabilistic inference," *TASLP*, 2011.

[9] T. Minka, *A family of algorithms for approximate Bayesian inference*, Ph.D. thesis, MIT Media Lab, 2001.