An Iterative Linearised Solution to the Sinusoidal Parameter Estimation Problem

Jean-Marc Valin [a,d,*], Daniel V. Smith [b],
Christopher Montgomery [c,d], Timothy B. Terriberry [d]

[a] *CSIRO ICT Centre, Australia*

[b] *CSIRO Tasmanian ICT Centre, Australia*

[c] *RedHat Inc., USA*

[d] *Xiph.Org Foundation*

Abstract

Signal processing applications use sinusoidal modelling for speech synthesis, speech coding, and audio coding. Estimation of the model parameters involves non-linear optimisation methods, which can be very costly for real-time applications. We propose a low-complexity iterative method that starts from initial frequency estimates and converges rapidly. We show that for N sinusoids in a frame of length L, the proposed method has a complexity of $O(LN)$, which is significantly less than the matching pursuits method. Furthermore, the proposed method is shown to be more accurate than the matching pursuits and time-frequency reassignment methods in our experiments.

Key words: Sinusoidal modeling, iterative least-squares solution

1 Introduction

Signal processing applications such as speech synthesis [1], speech coding [2], and audio coding [3] increasingly use sinusoidal models. Estimating the model

* Corresponding author. Address: CSIRO ICT Centre, Cnr. Vimiera & Pembroke Roads, Marsfield NSW 2122, Australia. Tel: +61 (0)2 9372 4284.

Email addresses: jean-marc.valin@csiro.au (Jean-Marc Valin), daniel.v.smith@csiro.au (Daniel V. Smith), xiphmont@xiph.org (Christopher Montgomery), tterribe@xiph.org (Timothy B. Terriberry).

parameters often represents a significant fraction of their overall computational complexity. Real-time applications require a very low-complexity estimation algorithm.

This paper proposes a new parameter estimation procedure based on the linearisation of the model around an initial frequency estimate and iterative optimisation with fast convergence. For typical configurations, it is over 20 times less complex than matching pursuits [4].

We start by introducing sinusoidal modelling and prior art in Section 2. Section 3 discusses frequency estimation and our proposed linearisation. In Section 4, we present a low-complexity iterative solver for estimating sinusoidal parameters. Results are discussed in Section 5, and Section 6 concludes this paper. Unless otherwise noted, a bold uppercase symbol ($\mathbf{A}$) denotes a matrix, a bold lower case symbol ($\mathbf{a}_i$) denotes a column of the matrix, and an italic symbol ($a_{i,j}$) denotes an element of the matrix.

2 Sinusoidal Parameter Estimation

A general sinusoidal model that considers both amplitude and frequency modulation can be used to approximate a signal $\tilde{x}(t)$ as:

$$\tilde{x}(t) = \sum_{k=1}^{N} A_k(t) \cos\left(\int_0^t \omega_k(u)\, du + \phi_k\right) \ , \tag{1}$$

where $A_k(t)$ is the time-varying amplitude, $\omega_k(t)$ is the time-varying frequency and ϕ_k is the initial phase. The model in (1) has limited practical use because there are an arbitrary number of ways to approximate $A_k(t)$ and $\omega_k(t)$. Using discrete time n and normalised frequencies θ_k over a finite window $h(n)$ yields a simpler model:

$$\tilde{x}(n) = h(n) \sum_{k=1}^{N} \left(A_k + \dot{A}_k n\right) \cos\left(\theta_k n + \phi_k\right) \ , \tag{2}$$

where $\dot{A}$ is the first time derivative of the amplitude, or even

$$x(n) = h(n) \sum_{k=1}^{N} A_k \cos\left(\theta_k n + \phi_k\right) \ , \tag{3}$$

if we do not want to model amplitude variation within a frame. Although simpler, the models in (2) and (3) are still difficult to estimate because they involve a non-linear optimisation problem.

There are several methods for estimating these sinusoidal model parameters. The simplest is a standard discrete Fourier transform (DFT) over a rectangular window. This is limited by frequency leakage caused by sidelobes from the rectangular window and by its poor frequency resolution [1], which is $2\pi/L\ rad/s$ for a frame of length L.

By defining an over-complete dictionary of sinusoidal bases, matching pursuits methods [4] make it possible to increase the frequency resolution arbitrarily. Their basis functions also allow a non-rectangular window to reduce sidelobes. However, as a greedy algorithm, matching pursuits behaves sub-optimally when the basis functions are not orthogonal [5], which is usually the case for sinusoids of arbitrary frequency over a finite window length. The orthogonality problem of matching pursuits can mainly be overcome by further non-linear optimisation as in [5]. However, this increases complexity significantly, to as high as $O\left(N^4\right)$.

The time-frequency reassignment (TFR) method is another approach that improves the frequency estimate resolution. When using a spectrogram representation, phase information from the short-time Fourier transform (STFT) is exploited to reassign energy from the centre of a spectral bin (t, w) to its centre of gravity, (t^*, w^*) [6,7]. The drawback is that this approach is not well suited to noisy signal conditions, as energy becomes reassigned to noise dominated regions [7].

Other work, such as [1,8], focuses on the estimation of sinusoidal partials in harmonic signals. While these are generally low complexity methods, they are not applicable to non-harmonic signals.

3 Linearised Model

We propose another way to obtain accurate frequency estimates, by rewriting the sinusoidal model in (2) as

$$\tilde{x}\left(n\right) = h\left(n\right) \sum_{k=1}^{N} \left(A_k + n\dot{A}_k\right) \cdot \cos\left(\left(\theta_k + \Delta\theta_k\right) n + \phi_k\right) \ , \tag{4}$$

where θ_k is an initial estimate of the frequencies and $\Delta\theta_k$ is an unknown correction to the initial estimate. When both the amplitude modulation parameter $\dot{A}_k$ and the frequency correction $\Delta\theta_k$ are small, we show in Appendix

[1] Throughout this paper, "resolution" means the smallest frequency difference that can be measured for a sinusoid, not the capability to distinguish between two close sinusoids.

A that (4) can be linearised as the sum of four basis functions:

$$\tilde{x}\left(n\right) \approx h\left(n\right) \sum_{k=1}^{N} c_k \cos\theta_k n + s_k \sin\theta_k n + d_k n \cos\theta_k n + t_k n \sin\theta_k n \ , \tag{5}$$

with

$$c_k = A_k \cos\phi_k \ , \tag{6}$$
$$s_k = -A_k \sin\phi_k \ , \tag{7}$$
$$d_k = \dot{A}_k \cos\phi_k - A_k \Delta\theta_k \sin\phi_k \ , \tag{8}$$
$$t_k = -\dot{A}_k \sin\phi_k - A_k \Delta\theta_k \cos\phi_k \ . \tag{9}$$

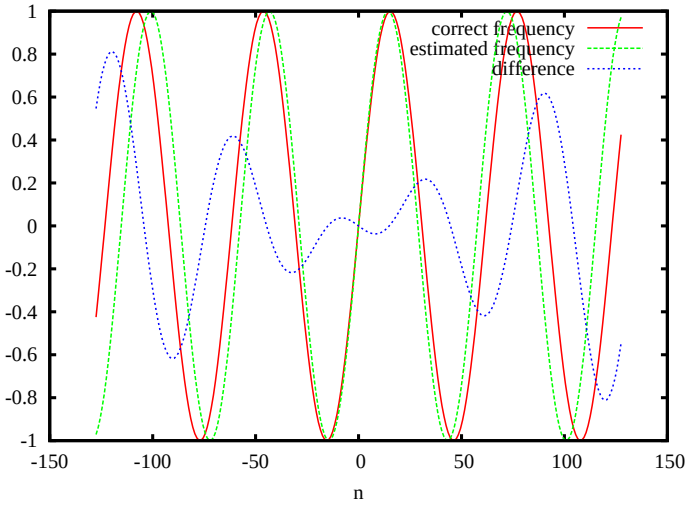

Figure 1. Difference between two sinusoids of nearly identical frequencies, resulting in an amplitude-modulated sinusoid.

Fig. 1 is a visual demonstration of the linearisation for a small frequency correction. It shows that if the frequency estimate is very close to the actual frequency of the sinusoid, the error between the estimated sinusoid and the actual sinusoid can be approximated as an amplitude modulated sinusoid. Hence, that error can be modelled using the two basis functions (8) and (9).

We can express (5) in matrix form as

$$\tilde{\mathbf{x}} \approx \mathbf{A}\mathbf{w} \ , \tag{10}$$
$$\mathbf{A} = \left[\mathbf{A}^c, \mathbf{A}^s, \mathbf{A}^d, \mathbf{A}^t\right] \ , \tag{11}$$
$$\mathbf{w} = \left[\mathbf{c}, \mathbf{s}, \mathbf{d}, \mathbf{t}\right]^T \ , \tag{12}$$

where the basis components $\mathbf{A}^c$, $\mathbf{A}^s$, $\mathbf{A}^d$, and $\mathbf{A}^t$ are defined as

$$a^c_{n,k} = h(n)\cos\theta_k n \ , \tag{13}$$
$$a^s_{n,k} = h(n)\sin\theta_k n \ , \tag{14}$$
$$a^d_{n,k} = h(n)\, n\cos\theta_k n \ , \tag{15}$$
$$a^t_{n,k} = h(n)\, n\sin\theta_k n \ . \tag{16}$$

The best fit is obtained through the least-squares optimisation

$$\min_{\mathbf{w}} \left\|\mathbf{A}\mathbf{w} - \mathbf{x}_h\right\|^2 \ , \tag{17}$$

where $\mathbf{x}_h$ is the windowed input signal. This leads to the well known solution

$$\mathbf{w} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{x}_h \ . \tag{18}$$

Once the linear parameters in (5) are found, the original sinusoidal parameters can be retrieved by solving the system (6)-(9):

$$A_k = \sqrt{c_k^2 + s_k^2} \ , \tag{19}$$
$$\phi_k = \arg\left(c_k - \jmath s_k\right) \ , \tag{20}$$
$$\dot{A}_k = \frac{d_k c_k + s_k t_k}{A_k} \ , \tag{21}$$
$$\Delta\theta_k = \frac{d_k s_k - t_k c_k}{A_k^2} \ . \tag{22}$$

3.1 Frequency Modulation and Higher Order Terms

Generalising the approach to include second order basis functions yields

$$a^f_{n,k} = h(n)\, n^2\cos\theta_k n \ , \tag{23}$$
$$a^u_{n,k} = h(n)\, n^2\sin\theta_k n \ . \tag{24}$$

This allows the estimation of both the second derivative of the amplitude, $\ddot{A}$, and the derivative of the frequency, $\dot{\theta}$, resulting in the following model:

$$\tilde{x}(n) = h(n)\sum_{k=1}^{N}\left(A_k + n\dot{A}_k + n^2\ddot{A}_k\right)\cdot\cos\left(\left(\theta_k + \Delta\theta_k + \dot{\theta}_k n\right)n + \phi_k\right) \ , \tag{25}$$

Appendix B derives the second order linearised model:

$$\begin{aligned}\tilde{x}\left(n\right) \approx h\left(n\right) \sum_{k=1}^{N} & c_k \cos\theta_k n + s_k \sin\theta_k n \\ & + d_k n \cos\theta_k n + t_k n \sin\theta_k n \\ & + f_k n^2 \cos\theta_k n + u_k n^2 \sin\theta_k n \ , \end{aligned} \tag{26}$$

with

$$c_k = A_k \cos\phi_k \ , \tag{27}$$
$$s_k = -A_k \sin\phi_k \ , \tag{28}$$
$$d_k = \dot{A}_k \cos\phi_k - A_k \Delta\theta_k \sin\phi_k \ , \tag{29}$$
$$t_k = -\dot{A}_k \sin\phi_k - A_k \Delta\theta_k \cos\phi_k \ , \tag{30}$$
$$f_k = \ddot{A}_k \cos\phi_k - A_k \dot{\theta}_k \sin\phi_k \ , \tag{31}$$
$$u_k = -\ddot{A}_k \sin\phi_k - A_k \dot{\theta}_k \cos\phi_k \ . \tag{32}$$

The second order model (26) can be formulated in matrix form:

$$\tilde{\mathbf{x}} \approx \mathbf{A}\mathbf{w} \ , \tag{33}$$
$$\mathbf{A} = \left[\mathbf{A}^c, \mathbf{A}^s, \mathbf{A}^d, \mathbf{A}^t, \mathbf{A}^f, \mathbf{A}^u\right] \ , \tag{34}$$
$$\mathbf{w} = \left[\mathbf{c}, \mathbf{s}, \mathbf{d}, \mathbf{t}, \mathbf{f}, \mathbf{u}\right]^T \ , \tag{35}$$

where the basis components $\mathbf{A}^c$, $\mathbf{A}^s$, $\mathbf{A}^d$, $\mathbf{A}^t$, $\mathbf{A}^f$, and $\mathbf{A}^u$ are defined as

$$a^c_{n,k} = h\left(n\right) \cos\theta_k n \ , \tag{36}$$
$$a^s_{n,k} = h\left(n\right) \sin\theta_k n \ , \tag{37}$$
$$a^d_{n,k} = h\left(n\right) n \cos\theta_k n \ , \tag{38}$$
$$a^t_{n,k} = h\left(n\right) n \sin\theta_k n \ , \tag{39}$$
$$a^f_{n,k} = h\left(n\right) n^2 \cos\theta_k n \ , \tag{40}$$
$$a^u_{n,k} = h\left(n\right) n^2 \sin\theta_k n \ . \tag{41}$$

As with the first order model, a least-squares optimisation can be used to obtain the linear terms (27)-(32). The explicit sinusoidal parameters can then be computed with

$$A_k = \sqrt{c_k^2 + s_k^2}\ , \tag{42}$$

$$\phi_k = \arg\left(c_k - \jmath s_k\right)\ , \tag{43}$$

$$\dot{A}_k = \frac{d_k c_k + s_k t_k}{A_k}\ , \tag{44}$$

$$\Delta\theta_k = \frac{d_k s_k - t_k c_k}{A_k^2}\ , \tag{45}$$

$$\ddot{A}_k = \frac{f_k c_k + s_k u_k}{A_k}, \tag{46}$$

$$\dot{\theta}_k = \frac{f_k s_k - u_k c_k}{A_k^2}, \tag{47}$$

The first and second order models are identical, apart from the addition of the $\ddot{A}$ and $\dot{\theta}$ terms, which model quadratic amplitude modulation and linear frequency modulation, respectively. The analysis in Appendix B makes clear what the third order model and above would look like. However, the accuracy of each additional set of terms decreases with the order, limiting the usefulness of higher order models.

4 Iterative Solver

Though solving the linear system (18) demands far less computation than a classic non-linear solver, it still requires a great amount. D'haes proposed a method that reduces that complexity from $O\left(LN^2\right)$ to $O\left(N \log N\right)$, but only for harmonic signals [8]. In this paper, we propose an $O\left(LN\right)$ solution without the restriction to harmonic signals.

Our method uses an iterative solution based on the assumption that matrix A is close to orthogonal, so that

$$\left(\mathbf{A}^T\mathbf{A}\right)^{-1} \approx \operatorname{diag}\left\{\frac{1}{\mathbf{a}_1^T\mathbf{a}_1}, \ldots, \frac{1}{\mathbf{a}_N^T\mathbf{a}_N}\right\} = \mathbf{\Phi}\ . \tag{48}$$

This way, an initial estimate can be computed as

$$\mathbf{w}^{(0)} = \mathbf{\Phi}^{-1}\mathbf{A}^T\mathbf{x}_h \tag{49}$$

and then refined as

$$\begin{aligned}\mathbf{w}^{(i+1)} &= \mathbf{w}^{(i)} + \mathbf{\Phi}^{-1}\mathbf{A}^T\left(\mathbf{x}_h - \tilde{\mathbf{x}}^{(i)}\right)\\ &= \mathbf{w}^{(i)} + \mathbf{\Phi}^{-1}\mathbf{A}^T\left(\mathbf{x}_h - \mathbf{A}\mathbf{w}^{(i)}\right)\ .\end{aligned} \tag{50}$$

The iterative method described in (49)-(50) is strictly equivalent to the Jacobi iterative method. The complexity of the algorithm is reduced to $O(LMN)$, where M is the number of iterations required for acceptable convergence. Unfortunately, while in practise the Jacobi method is stable for most matrices $\mathbf{A}$, convergence is not guaranteed and depends on the actual frequencies θ_k.

4.1 Gauss-Seidel Method

An alternative to the Jacobi method is the Gauss-Seidel method. Its main advantage is that convergence is guaranteed, since the matrix $\mathbf{A}^T\mathbf{A}$ is symmetric and positive definite [9]. Since the columns of $\mathbf{A}$ are usually nearly orthogonal, $\mathbf{A}^T\mathbf{A}$ is strongly diagonally dominant, and the Gauss-Seidel method converges quickly. The linear system can be expressed as

$$\mathbf{R}\mathbf{w} = \mathbf{b} \ , \tag{51}$$

where

$$\mathbf{R} = \mathbf{A}^T\mathbf{A} \ , \tag{52}$$

$$\mathbf{b} = \mathbf{A}^T\mathbf{x}_h \ . \tag{53}$$

Assuming $\mathbf{A}$ has been pre-normalised ($\mathbf{a}_k^T\mathbf{a}_k = 1, \forall k$), the Gauss-Seidel algorithm becomes

$$\begin{aligned} w_k^{(i+1)} =& b_k - \sum_{j<k} r_{k,j} w_j^{(i+1)} - \sum_{j>k} r_{k,j} w_j^{(i)} \\ =& \mathbf{a}_k^T\mathbf{x}_h - \sum_{j<k} \mathbf{a}_k^T\mathbf{a}_j w_j^{(i+1)} - \sum_{j>k} \mathbf{a}_k^T\mathbf{a}_j w_j^{(i)} \\ =& w_k^{(i)} + \mathbf{a}_k^T\mathbf{x}_h - \sum_{j<k} \mathbf{a}_k^T\mathbf{a}_j w_j^{(i+1)} - \sum_{j\geq k} \mathbf{a}_k^T\mathbf{a}_j w_j^{(i)} \\ =& w_k^{(i)} + \mathbf{a}_k^T\mathbf{x}_h - \mathbf{a}_k^T\left(\mathbf{A}\tilde{\mathbf{w}}_k{}^{(i+1)}\right) \\ =& w_k^{(i)} + \mathbf{a}_k^T\left(\mathbf{x}_h - \mathbf{A}\tilde{\mathbf{w}}_k{}^{(i+1)}\right) \ , \end{aligned} \tag{54}$$

where

$$\tilde{\mathbf{w}}_k{}^{(i+1)} = \Big[w_0^{(i+1)}, \ldots, w_{k-1}^{(i+1)}, w_k^{(i)}, \ldots, w_{N-1}^{(i)}\Big]^T \ . \tag{55}$$

We can further simplify the computation of (54) by noting that only one element of $\tilde{\mathbf{w}}_k{}^{(i+1)}$ changes for each step. Thus we have

$$w_k^{(i+1)} = w_k^{(i)} + \mathbf{a}_k^T\mathbf{e}_k^{(i+1)} \ , \tag{56}$$

Algorithm 1 Iterative linear optimisation

Compute basis functions (13)-(16).
$\mathbf{w}^{(0)} \leftarrow \mathbf{0}$
$\mathbf{e} \leftarrow \mathbf{x}_h$
for all iteration i=1...M **do**
  for all sinusoid component $k = 1 \dots 4N$ **do**
    $\Delta w_k^{(i)} \leftarrow \mathbf{a}_k^T \mathbf{e}$
    $\mathbf{e} \leftarrow \mathbf{e} - \mathbf{a}_k \Delta w_k^{(i)}$
    $w_k^{(i)} \leftarrow w_k^{(i-1)} + \Delta w_k^{(i)}$
  end for
end for
for all sinusoid $k = 1 \dots N$ **do**
  $A_k \leftarrow \sqrt{c_k^2 + s_k^2}$
  $\phi_k \leftarrow \arg(c_k - \jmath s_k)$
  $\dot{A}_k \leftarrow \frac{d_k c_k + s_k t_k}{A_k}$
  $\Delta\theta_k \leftarrow \frac{d_k s_k - t_k c_k}{A_k^2}$
end for

where $\mathbf{e}_k^{(i+1)}$ is the current error in the approximation, computed recursively as

$$\mathbf{e}_k^{(i+1)} = \begin{cases} \mathbf{e}_{k-1}^{(i+1)} - \\ \quad \left(w_{k-1}^{(i+1)} - w_{k-1}^{(i)}\right) \mathbf{a}_{k-1} & , \ k \neq 0 \\ \mathbf{e}_N^{(i)} & , \ k = 0 \end{cases} . \tag{57}$$

The resulting computation is summarised in Algorithm 1. If there is only one iteration, then algorithm 1 is equivalent to a simplified version of the matching pursuits algorithm, where the atoms (frequency of the sinusoids) have been pre-selected before the search. From this point of view, the proposed method relaxes the orthogonality assumption made by the matching pursuits method.

The main difference from the Jacobi method is that the Gauss-Seidel method includes partial updates of the error term after each extracted sinusoid. Convergence follows intuitively from the fact that each individual step is an exact projection that is guaranteed to decrease the current error $\mathbf{e}$ — or at worst leave it constant if the solution is optimal. Since the error term is updated after each component k, placing the highest-energy terms first speeds up the optimisation. For this reason, we first update the $\cos\theta_k n$ and the $\sin\theta_k n$ terms, followed by the $n\cos\theta_k n$ and the $n\sin\theta_k n$ terms. This usually reduces the number of iterations required, converging in half as many iterations as sparse conjugate gradient techniques, such as LSQR [10], which cannot take advantage of the diagonal dominance of the system.

We choose $n = 0$ to lie in the centre of the frame in (13)-(16), between sample $L/2$ and sample $L/2 + 1$ if L is even, giving all the $\mathbf{a}_k^c$ and $\mathbf{a}_k^t$ vectors even symmetry and all the $\mathbf{a}_k^s$ and $\mathbf{a}_k^d$ vectors odd symmetry. This leads to the

following orthogonality properties:

$$\langle \mathbf{a}_k^c, \mathbf{a}_k^s \rangle = 0 \ , \tag{58}$$

$$\left\langle \mathbf{a}_k^c, \mathbf{a}_k^d \right\rangle = 0 \ , \tag{59}$$

$$\left\langle \mathbf{a}_k^t, \mathbf{a}_k^s \right\rangle = 0 \ , \tag{60}$$

$$\left\langle \mathbf{a}_k^t, \mathbf{a}_k^d \right\rangle = 0 \ . \tag{61}$$

Similar properties hold for the second order basis vectors. Because the even and odd bases are orthogonal to each other, we optimise them separately as

$$[\mathbf{c}, \mathbf{t}, \mathbf{f}]^T = \left(\mathbf{A}^{evenT}\mathbf{A}^{even}\right)^{-1}\mathbf{A}^{evenT}\mathbf{x} \ , \tag{62}$$

$$[\mathbf{s}, \mathbf{d}, \mathbf{u}]^T = \left(\mathbf{A}^{oddT}\mathbf{A}^{odd}\right)^{-1}\mathbf{A}^{oddT}\mathbf{x} \ , \tag{63}$$

$$\mathbf{A}^{even} = \left[\mathbf{A}^c, \mathbf{A}^t, \mathbf{A}^f\right] \ , \tag{64}$$

$$\mathbf{A}^{odd} = \left[\mathbf{A}^s, \mathbf{A}^d, \mathbf{A}^u\right] \ . \tag{65}$$

Not only does the orthogonality accelerate convergence, but it allows us to split the error $\mathbf{e}$ into half-length even and odd components, reducing the complexity of each iteration by half.

4.2 Non-Linear Optimisation

If the initial frequency estimates θ_k^0 are close to the real frequencies of the sinusoids θ_k, then the error caused by the linearisation (5) is very small. In this case, Algorithm 1 should result in values of $\theta_k^0 + \Delta\theta_k$ that are very close to the real frequencies. However, if the initial estimates deviate significantly from the real values, then it may be useful to restart the optimisation with

$$\theta_k \leftarrow \theta_k + \alpha\Delta\theta_k \ .$$

where α is the update rate. Typically $\alpha = 1$. Repeating the operation several times, we obtain a non-linear iterative solver for A_k, θ_k, $\dot{A}_k$, and ϕ_k, and optionally for $\dot{\theta}_k$ and $\ddot{A}_k$.

It is not necessary to wait for Algorithm 1 to converge before updating the frequencies θ_k. We can let both the linear part and the non-linear part of the solution run simultaneously. To do that, we must first subtract the solution of the previous iteration before restarting the linear optimisation.

The non-linear method we propose is detailed in Algorithm 2 and shares some similarities with the Gauss-Newton method [11]. However, the reparametrisation in (6)-(9) allows updates to A_k, $\dot{A}_k$, and ϕ_k to be incorporated into the linear model immediately when solving the normal equations. This greatly

Algorithm 2 Non-linear iterative optimisation, including the second order terms. Steps marked with † are only applied for the second order model.

$\forall k,\ \theta_k = \theta_k^0$
$\forall k,\ [A_k, \phi_k, \dot{A}_k, \ddot{A}_k, \dot{\theta}_k] \leftarrow 0$
$\mathbf{w}^{(0)} \leftarrow \mathbf{0}$
$\mathbf{e} \leftarrow \mathbf{x}_h$
for all non-linear iteration i=1...M **do**
  for all sinusoid k **do**
    $c_k \leftarrow A_k \cos\phi_k$
    $s_k \leftarrow -A_k \sin\phi_k$
    $d_k \leftarrow \dot{A}_k \cos\phi_k$
    $t_k \leftarrow -\dot{A}_k \sin\phi_k$
    $^\dagger f_k \leftarrow \ddot{A}_k \cos\phi_k - A_k\dot{\theta}_k \sin\phi_k$
    $^\dagger u_k \leftarrow -\ddot{A}_k \sin\phi_k + A_k\dot{\theta}_k \cos\phi_k$
  end for
  $\mathbf{e} \leftarrow \mathbf{x} - \mathbf{A}\mathbf{w}^{(i-1)}$ (result of the last iteration with updated frequency)
  for all sinusoid component $k = 1 \ldots 4N$ **do**
    $\Delta w_k^{(i)} \leftarrow \mathbf{a}_k^T \mathbf{e}$
    $\mathbf{e} \leftarrow \mathbf{e} - \mathbf{a}_k \Delta w_k^{(i)}$
    $w_k^{(i)} \leftarrow w_k^{(i-1)} + \Delta w_k^{(i)}$
  end for
  for all sinusoid $k = 1 \ldots N$ **do**
    $A_k \leftarrow \sqrt{c_k^2 + s_k^2}$
    $\phi_k \leftarrow \arg(c_k - \jmath s_k)$
    $\dot{A}_k \leftarrow \frac{d_k c_k + s_k t_k}{A_k}$
    $\Delta\theta_k \leftarrow \frac{d_k s_k - t_k c_k}{A_k^2}$
    $\theta_k \leftarrow \theta_k + \alpha\Delta\theta_k$
    $^\dagger \ddot{A}_k \leftarrow \frac{f k c_k + s_k u_k}{A_k}$
    $^\dagger \dot{\theta}_k \leftarrow \frac{f_k s_k - u_k c_k}{A_k^2}$
  end for
end for

improves convergence compared to a standard Gauss-Newton iteration in the original parameters. Just like Algorithm 1, it is possible to reduce the complexity of Algorithm 2 in half by taking advantage of the even-odd symmetry of the basis functions.

5 Results And Discussion

In this section, we characterise the proposed algorithm and compare it to other sinusoidal parameter estimation algorithms. We attempt to make the comparison as fair as possible despite the fact that the methods we are com-

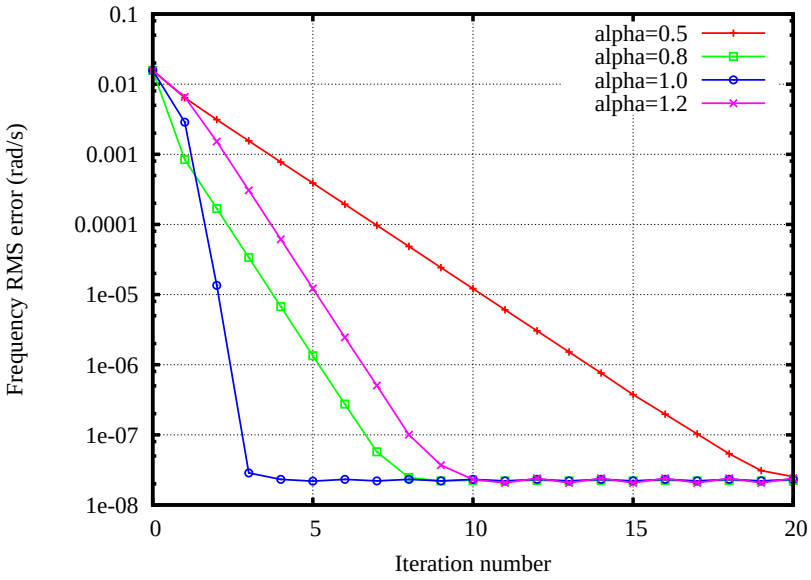

Figure 2. Convergence of the non-linear optimisation procedure for various values of α. For $\alpha = 1$, convergence is achieved in only 3 iterations. The floor at $2 \times 10^{-8} rad/s$ is due to the finite machine precision.

paring do not have exactly the same assumptions or output. Both the linear and the non-linear versions of the proposed algorithm are evaluated. For all algorithms, we use a *sine window*:

$$h(n) = \cos \pi \frac{n - (L+1)/2}{L} , \tag{66}$$

so that the result of applying the window to both the input signal $\mathbf{x}$ and the basis functions $\mathbf{a}_k$ is equivalent to a Hanning analysis window. Unless otherwise noted, we use a frame length $L = 256$.

5.1 Convergence

We first consider the case of a single amplitude-modulated sinusoid of normalised angular frequency $\theta = 0.1\pi$. We start with an initial frequency estimate of $\theta = 0.095\pi$, which corresponds to an error of slightly more than one period over the 256-sample frames we use. The non-linear optimisation Algorithm 2 is applied with different values of α, using only the first-order terms. The convergence speed in Figure 2 shows that for $\alpha = 1$, convergence becomes much faster than for other values of α, indicating that convergence is super-linear.

If we let the linear part of the algorithm converge at each iteration, the result is equivalent to the second order Newton's method, since as shown in Appendix A, the terms in our linearisation are equal to a first-order Taylor

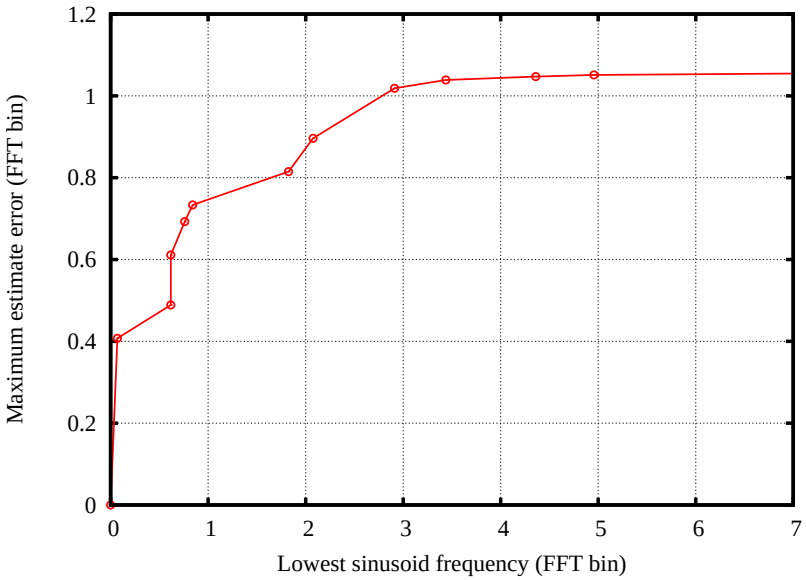

Figure 3. Region of convergence as a function of the sinusoid frequency. The algorithm never converges when the initial estimate is off by more than 1.05 DFT bins ($2\pi/L\,rad/s$).

expansion in the original variables. Using the chain rule, one can show that Algorithm 2 is only super-linear if the Gauss-Seidel iteration is super-linear. Since Gauss-Seidel is an iterative linear method, this can only happen if the basis vectors in $\mathbf{A}$ are orthogonal. In practise, so long as the separation between frequencies is larger than the sidelobe of the windowing function, these basis vectors are approximately orthogonal, although in practise they are never truly orthogonal. However, with a good choice of windowing function and well-separated frequencies, convergence is quasi-second order.

If we include second-order terms, then convergence becomes linear, since the frequency modulation term is not "recentered" like the frequency is. While such recentring is possible, it unnecessarily increases the complexity of the algorithm while making it more susceptible to numerical errors.

As stated in Section 3, the proposed algorithm depends on an initial approximation sufficiently close to the true frequency of a sinusoid. Fig. 3 shows the maximum error in the initial estimate for which the non-linear algorithm converges to the true frequency. For most frequencies, that maximum error is equivalent to 1.05 DFT bins. However, for low frequencies, the tolerance to error is reduced. This is due to the fact that $a^c_{n,k}$ becomes highly correlated with $a^t_{n,k}$ and $a^s_{n,k}$ becomes highly correlated with $a^d_{n,k}$, making it harder to estimate the frequency offsets $\Delta\theta_k$ accurately.

5.2 Chirps

Next, we measure the frequency estimation accuracy and the energy of the residual signal for known signals. We use a synthetic signal that is the sum of five chirps with white Gaussian noise. The chirps have linear frequency variations starting at 0.05, 0.1, 0.15, 0.2, and 0.25 rad/s and ending at 2.0, 2.2, 2.4, 2.6, and 2.8 rad/s, respectively. The relative amplitudes of the chirps are 0 dB, -3 dB, -6 dB, -9 dB, and -12 dB. We consider the following algorithms:

- Time-frequency reassignment (**TFR**),
- Matching pursuits (32x over-sampled dictionary) (**MP**),
- Proposed algorithm with linear optimisation (**linear**),
- Proposed algorithm with non-linear optimisation (**non-linear**), and
- Proposed algorithm with non-linear optimisation and second order model (**second order**).

The time-frequency reassignment method is implemented as in [6]. The matching pursuits algorithm uses a dictionary of non-modulated sinusoids with a resolution of $\pi/8192$. We also compare to the theoretical resolution obtained from the picking the highest peaks in the DFT. These are used as the initial seeds for our algorithm and TFR. To make sure that algorithms are compared fairly, all algorithms are constrained to frequencies within one DFT bin of the initial seed, i.e. there are no outliers. MP does not consider any dictionary elements outside this range, and any step by the optimisation algorithms is clamped to lie within it. This occurs only rarely when the SNR is low.

Fig. 4 shows the RMS energy of the residual $(\tilde{\mathbf{x}} - \mathbf{x}_h)$ as a function of the number of iterations for both the linear optimisation and the non-linear optimisation. The linear version converges after only 2 iterations, while the non-linear version requires 3 iterations. These are the iteration limits we use for the experiments that follow. In the case of the second order non-linear version, the convergence continues until limited by numerical precision, so we limit it to 5 iterations, which already significantly improves on the first order model.

Fig. 5 shows the frequency RMS estimation error as a function of the SNR for each of the four algorithms. At very low SNR, all algorithms perform similarly. However, as the SNR increases above 20 dB, matching pursuits stops improving. This is likely due to the fact that the frequencies are not orthogonal, which makes its greedy approach sub-optimal. Both the proposed linear and non-linear approaches provide roughly the same accuracy up to 30 dB, after which the non-linear approach provides superior performance. For this scenario, the only limitation of the non-linear algorithm at infinite SNR is the fact that it does not account for frequency modulation within a frame.

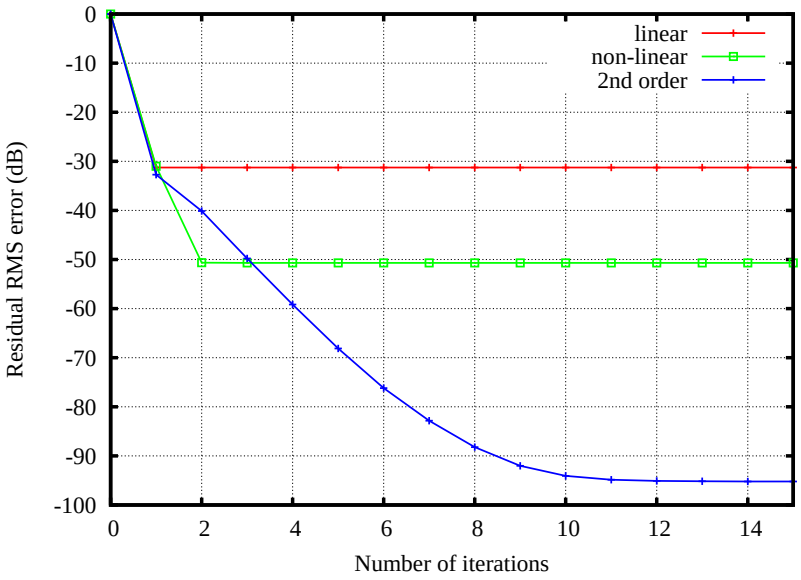

Figure 4. Reconstruction RMS error as a function of the number of iterations in clean conditions (linear vs. non-linear)

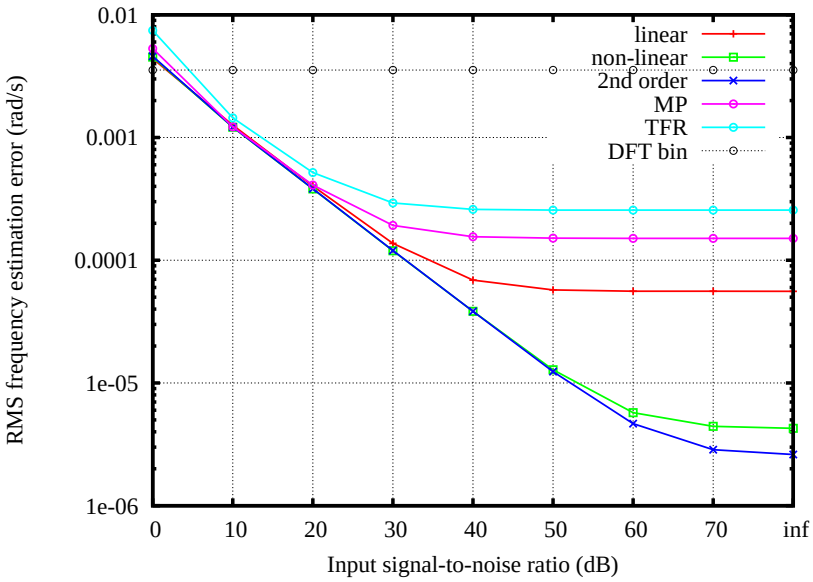

Figure 5. Frequency RMS estimation error as a function of the SNR.

Fig. 6 shows the reconstruction error for all algorithms except the time-frequency reassignment method, which cannot estimate the amplitude and thus cannot provide a reconstructed signal. The reconstruction error is measured against the noise-free version of the chirps. The performance mirrors that of Fig. 5, with the notable exception that the non-linear optimisation's reconstruction error plateaus long before the second order method, even though it is able to accurately estimate the frequency.

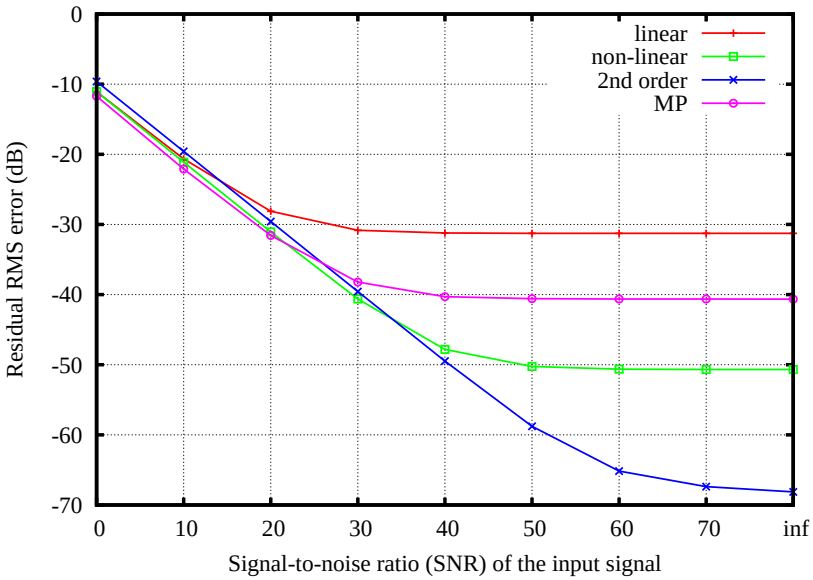

Figure 6. Reconstruction RMS error as a function of the SNR (the input noise is not considered in the error).

The performance of our algorithm is slightly worse than matching pursuits at low SNR. This is caused by some slight over-fitting due to the inclusion of an amplitude modulation term. The difference disappears if this term is forced to zero.

In the chirp experiments our proposed non-linear algorithms out-perform both matching pursuits and time-frequency reassignment overall. The linear version has performance similar to previous methods, but it does not perform as well as non-linear optimisation. In all cases (Fig. 5 to Fig. 6), all the algorithms behave similarly. Their error at low SNR is similar, and the slope of the error curve is the same. The main differentiator between algorithms is how far they improve with SNR before reaching a plateau.

5.3 Audio

We apply our proposed algorithm to a 90-second collage of diverse music clips sampled at 48 kHz, including percussive, musical, and amusical content. In this case, we cannot compare to matching pursuits because the lack of ground truth prevents us from forcing a common set of initial sinusoid frequencies. We select the initial frequency estimates required for the proposed algorithm using peaks in the standard DFT. The number of sinusoids is variable (depends on the number of peaks) and a 256-sample window is used.

The energy of the residual is plotted as a function of the number of iterations in Fig. 7. Both algorithms converge quickly and we can see that the linear opti-

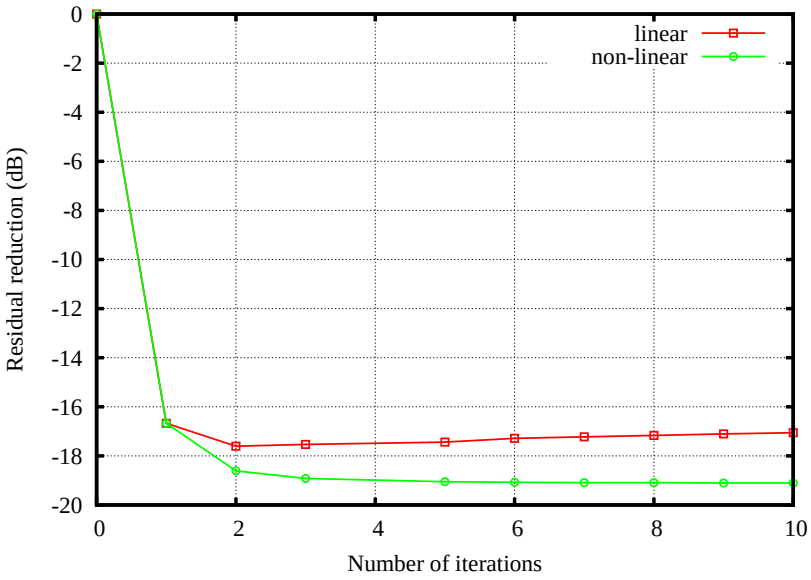

Figure 7. Reduction in residual energy as a function of the number of iterations.

misation only requires 2 iterations, while the non-linear optimisation requires 3 iterations.

5.4 Algorithm complexity

In this section, we compare the complexity of the proposed algorithms to that of other similar algorithms. For the sake of simplicity, we discard some terms that are deemed negligible, e.g., we discard $O(LN)$ terms when $O(LN^2)$ terms are present.

In Algorithm 1, we can see that each iteration requires $8LN$ multiplications and $8LN$ additions. Additionally, computation of the $4N$ basis functions $\mathbf{a}_k$ prior to the optimisation requires LN additions and $3LN$ multiplications. It is possible to further reduce the complexity of each iteration by taking advantage of the fact that all of our basis functions have either even or odd symmetry. By decomposing the residual into half-length even and odd components, only one of these components needs to be updated for a given basis function. This reduces the complexity of each iteration in Algorithm 1 by half without changing the result. The complexity of each iteration is thus $4LN$ multiplications and $4LN$ additions. For M iterations, this amounts to a total of $(8M + 5)\,LN$ operations per frame.

The complexity of the proposed non-linear optimisation algorithm (Algorithm 2) is similar to that of the linear version, with two exceptions. First, because the frequencies change every iteration, the basis functions need to be re-computed each time. Second, when starting a new iteration, the resid-

Table 1
Complexity comparison of various parameter estimation algorithms. *The typical complexity of [5] is not given, but we estimate it to be at least 500 Mflops, probably much higher.

Algorithm	Complexity	Typical (Mflops)
Matching pursuits (direct)	$2L^2NP$	14,000
Matching pursuits (FFT-based)	$\frac{5}{2}LNP\log_2 LP$	960
Direct non-linear ([5])	$O\left(N^4 + LN^2\right)$	>500*
Proposed (linear)	$(8M+5)\,LN$	9
Proposed (non-linear)	$(17M-4)\,LN$	20
Proposed (2^{nd} order)	$(24M-6)\,LN$	49

ual must be updated using the new basis functions. The total complexity is thus $(17M-4)\,LN$ operations per frame. For a single iteration, the linear and non-linear versions are strictly equivalent.

By comparison, a simple matching pursuits algorithm that does not consider modulation requires $2LN^2P$ operations per frame, where P is the oversampling factor, i.e. the increase over the standard DFT resolution. Using a fast FFT-based implementation [5] reduces the complexity to $5/2LNP\log_2 LP$.

Table 1 summarises the complexity of several algorithms. Because the algorithms have different dependencies on all the parameters, we also consider the total complexity in Mflops for real-time estimation of sinusoids in a *typical* scenario, where we have

- frame length: $L = 256$,
- number of sinusoids: $N = 20$,
- oversampling: $P = 64$ (matching pursuits only),
- number of iterations: $M = 2$ (linear), $M = 3$ (non-linear), $M = 5$ (2^{nd} order)
- sampling rate: 16 kHz,
- frame offset: 192 samples (25% overlap).

Table 1 shows that the proposed algorithms, both linear and nonlinear, reduce the complexity by more than an order of magnitude when compared to matching pursuits algorithms. However, while matching pursuits can estimate the sinusoidal parameters directly from the input signal, the proposed method requires initial frequency estimates. The cost of producing these estimates is not included in the table but is generally small (e.g., 0.4 Mflops for performing an FFT).

6 Conclusion

We have presented a method for estimating sinusoidal parameters with very low complexity. It is based on a linearisation of the sinusoidal model followed by an iterative optimisation of the parameters. The algorithm converges quickly, requiring only 2 iterations for the linear optimisation and 3 iterations for the non-linear optimisation. We showed that the frequency estimation of the non-linear version of our algorithm is more accurate than the matching pursuits and time-frequency reassignment methods. In addition, we demonstrated computational complexities considerably lower than matching pursuits. For applications that require it, we have also proposed a second order algorithm that estimates the frequency modulation within a frame. The total complexity of our approach is more than an order of magnitude less complex than other proposed methods for estimating sinusoid parameters. Consequently, our approach could offer significant benefits to areas such as audio and speech coding, which require sinusoidal modeling to be performed in real time.

Like other non-linear optimisation methods, ours requires a good initial estimate of the sinusoids' frequencies. Therefore, low-complexity sinusoid selection is an important area of future work.

A Linearisation of the Sinusoidal Model

Consider a sinusoidal model with piecewise linear amplitude modulation and a frequency offset from an initial estimate:

$$\tilde{x}\left(n\right) = \sum_{k=1}^{N} \left(A_k + n\dot{A}_k\right) \cdot \cos\left(\left(\theta_k + \Delta\theta_k\right) n + \phi_k\right) \ , \tag{A.1}$$

where θ_k is known in advance and $\Delta\theta_k$ is considered small. Using trigonometric identities, we can expand the sum in the cosine term into

$$\begin{aligned}
\tilde{x}\,(n) &= \sum_{k=1}^{N} \left(A_k + n\dot{A}_k\right) \cos\phi_k \cos\left(\theta_k + \Delta\theta_k\right) n \\
&\quad - \sum_{k=1}^{N} \left(A_k + n\dot{A}_k\right) \sin\phi_k \sin\left(\theta_k + \Delta\theta_k\right) n \qquad \text{(A.2)} \\
&= \sum_{k=1}^{N} \left(A_k + n\dot{A}_k\right) \cos\phi_k \cos\Delta\theta_k n \cos\theta_k n \\
&\quad - \sum_{k=1}^{N} \left(A_k + n\dot{A}_k\right) \cos\phi_k \sin\Delta\theta_k n \sin\theta_k n \\
&\quad - \sum_{k=1}^{N} \left(A_k + n\dot{A}_k\right) \sin\phi_k \cos\Delta\theta_k n \sin\theta_k n \\
&\quad - \sum_{k=1}^{N} \left(A_k + n\dot{A}_k\right) \sin\phi_k \sin\Delta\theta_k n \cos\theta_k n \ . \qquad \text{(A.3)}
\end{aligned}$$

In the linearisation process, we further assume that $\Delta\theta_k n \ll 1$ and $\dot{A}_k n \ll A_k$, so we can neglect all terms second order and above. This translates into the following approximations:

$$\sin\Delta\theta_k n \approx \Delta\theta_k n \ , \qquad \text{(A.4)}$$

$$\cos\Delta\theta_k n \approx 1 \ , \qquad \text{(A.5)}$$

$$n\dot{A}_k \sin\Delta\theta_k n \approx 0 \ . \qquad \text{(A.6)}$$

When substituting the above approximations into (A.3), we obtain

$$\begin{aligned}
\tilde{x}\,(n) &\approx \sum_{k=1}^{N} \left(A_k + n\dot{A}_k\right) \cos\phi_k \cos\theta_k n \\
&\quad - \sum_{k=1}^{N} A_k \cos\phi_k \Delta\theta_k n \sin\theta_k n \\
&\quad - \sum_{k=1}^{N} \left(A_k + n\dot{A}_k\right) \sin\phi_k \sin\theta_k n \\
&\quad - \sum_{k=1}^{N} A_k \sin\phi_k \Delta\theta_k n \cos\theta_k n \ . \qquad \text{(A.7)}
\end{aligned}$$

Reordering the terms in (A.7) leads to the following formulation:

$$\begin{aligned}\tilde{x}\left(n\right) \approx & \sum_{k=1}^{N} A_k \cos\phi_k \cos\theta_k n \\ & - \sum_{k=1}^{N} A_k \sin\phi_k \sin\theta_k n \\ & + \sum_{k=1}^{N} \left(\dot{A}_k \cos\phi_k - A_k \Delta\theta_k \sin\phi_k\right) n \cos\theta_k n \\ & - \sum_{k=1}^{N} \left(\dot{A}_k \sin\phi_k + A_k \Delta\theta_k \cos\phi_k\right) n \sin\theta_k n \ , \end{aligned} \tag{A.8}$$

which is a linear combination of four functions. The result in (A.8) is in fact equivalent to a first-order Taylor expansion.

B Derivation For the Second Order Model

Keeping second order terms allows us to model both the first derivative of the frequency and the second derivative of the amplitude with respect to time:

$$\begin{aligned}\tilde{x}\left(n\right) &= \sum_{k=1}^{N} \left(A_k + \dot{A}_k n + n^2 \ddot{A}_k\right) \cos\phi_k \cos\left(\theta_k + \Delta\theta_k + \dot{\theta}_k n\right) n \\ &- \sum_{k=1}^{N} \left(A_k + \dot{A}_k n + \ddot{A}_k n^2\right) \sin\phi_k \sin\left(\theta_k + \Delta\theta_k + \dot{\theta}_k n\right) n \qquad \text{(B.1)} \\ &= \sum_{k=1}^{N} \left(A_k + \dot{A}_k n + \ddot{A}_k n^2\right) \cos\phi_k \cos\left(\Delta\theta_k n + \dot{\theta}_k n^2\right) \cos\theta_k n \\ &- \sum_{k=1}^{N} \left(A_k + \dot{A}_k n + \ddot{A}_k n^2\right) \cos\phi_k \sin\left(\Delta\theta_k n + \dot{\theta}_k n^2\right) \sin\theta_k n \\ &- \sum_{k=1}^{N} \left(A_k + \dot{A}_k n + \ddot{A}_k n^2\right) \sin\phi_k \cos\left(\Delta\theta_k n + \dot{\theta}_k n^2\right) \sin\theta_k n \\ &- \sum_{k=1}^{N} \left(A_k + \dot{A}_k n + \ddot{A}_k n^2\right) \sin\phi_k \sin\left(\Delta\theta_k n + \dot{\theta}_k n^2\right) \cos\theta_k n \ . \qquad \text{(B.2)}\end{aligned}$$

This time, we neglect third order terms in n. Non-linear terms involving the

parameters (e.g. $\dot{A}_k \Delta\theta_k$) are discarded as well. This leads to

$$\sin\left(\Delta\theta_k n + \dot{\theta}_k n^2\right) \approx \Delta\theta_k n \ + \dot{\theta}_k n^2 \ , \tag{B.3}$$

$$\cos\left(\Delta\theta_k n + \dot{\theta}_k n^2\right) \approx 1 \ , \tag{B.4}$$

$$\left(\dot{A}_k n + \ddot{A}_k n^2\right) \sin\left(\Delta\theta_k n + \dot{\theta}_k n^2\right) \approx 0 \ , \tag{B.5}$$

Substituting into (B.2), we obtain

$$\begin{aligned}
\tilde{x}\,(n) \approx & \sum_{k=1}^{N} A_k \cos\phi_k \cos\theta_k n \\
& - \sum_{k=1}^{N} A_k \sin\phi_k \sin\theta_k n \\
& + \sum_{k=1}^{N} \left(\dot{A}_k \cos\phi_k - A_k \Delta\theta_k \sin\phi_k\right) n \cos\theta_k n \\
& - \sum_{k=1}^{N} \left(\dot{A}_k \sin\phi_k + A_k \Delta\theta_k \cos\phi_k\right) n \sin\theta_k n \\
& + \sum_{k=1}^{N} \left(\ddot{A}_k \cos\phi_k - A_k \dot{\theta}_k \sin\phi_k\right) n^2 \cos\theta_k n \\
& - \sum_{k=1}^{N} \left(\ddot{A}_k \sin\phi_k + A_k \dot{\theta}_k \cos\phi_k\right) n^2 \sin\theta_k n \ .
\end{aligned} \tag{B.6}$$

References

[1] Y. Stylianou. Applying the harmonic plus noise model in concatenative speech synthesis. *IEEE Transactions on Speech and Audio Processing*, 9(1):21–29, 2001.

[2] P. Hedelin. A tone oriented voice excited vocoder. In *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Processing*, volume 6, pages 205–208, April 1981.

[3] S. N. Levine. *Audio Representations for Data Compression and Compressed Domain Processing*. PhD thesis, Stanford University, 1998.

[4] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

[5] K. Vos, R. Vafin, R. Heusdens, and W. B. Kleijn. High-quality consistent analysis-synthesis in sinusoidal coding. In *AES 17th International conference on high quality audio coding*, 2005.

[6] F. Auger and P. Flandrin. Improving the readability of time-frequency and time-scale representations by the reassignment method. *IEEE Transactions on Signal Processing*, 43(5):1068–1089, 1995.

[7] F. Plante, G. Meyer, and W.Ainsworth. Improvement of speech spectrogram accuracy by the method of reassignment. *IEEE Trans SAP*, 6(3):282–287, 1998.

[8] W. D'haes. A highly optimized method for computing amplitudes over a windowed short time signal: from $O(K^2N)$ to $O(N\log(N))$. In *IEEE Signal Processing Symposium (SPS)*, 2004.

[9] Y. Saad. *Iterative methods for sparse linear systems*. SIAM, 2003.

[10] C. C. Paige and M. A. Saunders. LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software*, 8(1):43–71, March 1982.

[11] P. E. Frandsen, K. Jonasson, H. B. Nielsen, and O. Tingleff. *Unconstrained Optimization*. Informatics and Mathematical Modelling, Technical University of Denmark, DTU, 3rd edition, 2004.