On the Optimal Parameters of a Sinusoidal Representation of Signals

András Kocsor, László Tóth, and Imre Bálint

One of the most useful parametric models in the spectral resolution of digital signals is the modelling by a sum of phase-shifted sinusoids in form of $\sum_{n=0}^{N-1} A_n \sin(\omega_n t + \varphi_n)$, where A_n, ω_n and φ_n are the component's amplitude, frequency and phase, respectively. This model generally fits well speech and most musical signals due to the intrinsic nature of the representation functions. However using all of the above parameters leads to a very difficult optimization problem. The solutions are generally based on eigenvalue decomposition, but this is computationally very expensive and works only if the sinusoids and the residual signal is statistically uncorrelated. To find the best approximation in a subspace of the least dimension N is a major problem. To speed up the representation process some authors use rather ad hoc methods for determining the parameters. Such is the model of McAulay and Quatieri, which looks for peaks in the FFT spectrum. In the present paper the BFGS optimization method is applied to find the best approximating subspace of minimal dimension N, which is determined by parameters $\{A, \omega, \varphi\}$ and ensures a mean square error of approximation below a preset threshold.

From the invention of the telephone, speech or more generally voice processing especially voice representation has a paramount importance in electrical engineering. In the last years the rapid development of multimedia and computer networks brought a revival of the high-effective coding and representation problem. By the classic model of speech generation, the voiced part of the speech comes from the oscillation of the vocal chord, which can be modelled by an oscillating string. The voice is consisting of a fundamental and it's harmonics, which is well representable in the form $\sum_{n=0}^{N-1} A_n \sin(\omega_n t + \varphi_n)$ used in the paper. The error of approximation gives the 'unvoiced', noise-like part, which can be decoupled from the signal. The model fits well also musical signals, since the voices of most musical instruments (stringed-, wind instruments, etc.) consist of harmonic sinusoids. The rest-signal contains again the noise-like parts of the voice (as the drum), which should be modelled separately. The voice representation is used for data compressing, as well as detecting voiced parts, pitch estimation, or modification of the time-scale of the music, etc.. The above form of the model yields a complicated optimization problem enforcing some simplifications. In case of DFT, the number of sinusoids and their frequencies are fixed providing a rapid way of computing amplitudes and phases. However the individual components (sinusoids) of the representation generally differ from the really voice forming components. The method of [McAulay-Quatieri] tends to deduce the real frequencies of components from the results of DFT. A basically different approach characterizes the methods using eigenvalue decomposition. Here only the dimension N of the approximation subspace is fixed, but the algorithms are generally very slow and can be used only if the representation functions (sinusoids) and the rest-signals are statistically independent.

A detailed definition of the problem First the signal is sampled at points of a time-interval

$$\tau_0, \tau_1, \cdots, \tau_{K-1} \in [0, \tau]$$

and the obtained values are represented by the real sequence,

$$x[\tau_0], \cdots, x[\tau_{K-1}].$$

We are looking for the smallest number N, for which the function

$$\sum_{n=0}^{N-1} A_n \sin(\omega_n t + \varphi_n)$$

approximates the measured sample with the preset error $\epsilon > 0$,

$$\min_{N} \left[\min_{\substack{A_1, \cdots, A_{N-1} \\ \omega_1, \cdots, \omega_{N-1} \\ \varphi_1, \cdots, \varphi_{N-1}}} \sum_{k=0}^{K-1} \left(\sum_{n=0}^{N-1} A_n \sin(\tau_k \omega_n + \varphi_n) - x[\tau_k] \right)^2 < \epsilon \right]$$

The optimization problem with the Homogeneous Sinusoidal Representation Function A function will be given, which significantly simplifies the optimization problem defined in (1.1.1). Let be introduced the following notations,

$$w_k(\mathbf{A}, \omega, \varphi) := \sum_{n=0}^{N-1} A_n \sin(\omega_n \tau_k + \varphi_n), \quad k = 0, \cdots, K-1,$$

where

$$\mathbf{A} := [A_0, \cdots, A_{N-1}]^\top, \quad \omega := [\omega_0, \cdots, \omega_{N-1}]^\top, \quad \varphi := [\varphi_0, \cdots, \varphi_{N-1}]^\top.$$

Applying the identities,

$$w_k(\mathbf{A}, \omega, \varphi) := \sum_{n=0}^{N-1} A_n \sin(\tau_k \omega_n + \varphi_n) =$$

$$\sum_{n=0}^{N-1} A_n (\sin(\omega_n k) \cos(\varphi_n) + \cos(\omega_n k) \sin(\varphi_n)) = \sum_{n=0}^{N-1} a_n \sin(\omega_n k) + b_n \cos(\omega_n k),$$

where

$$a_n = A_n \cos(\varphi_n), \quad b_n = A_n \sin(\varphi_n).$$

Notice that,

$$\frac{b_n}{a_n} = \tan(\varphi_n), \quad \varphi_n = \arctan\left(\frac{b_n}{a_n}\right), \quad A_n = \frac{a_n}{\cos(\arctan\left(\frac{b_n}{a_n}\right))}.$$

Let be introduced,

$$\frac{c_n}{\sqrt{c_n^2 + d_n^2}} := \sin(\omega_n), \quad \frac{d_n}{\sqrt{c_n^2 + d_n^2}} := \cos(\omega_n).$$

By using these notations,

$$\sum_{n=0}^{N-1} a_n \sin(\omega_n k) + b_n \cos(\omega_n k) =$$

$$\sum_{n=0}^{N-1} a_n \sin(\tau_k \arcsin(\frac{c_n}{\sqrt{c_n^2 + d_n^2}})) + b_n \cos(\tau_k \arcsin(\frac{c_n}{\sqrt{c_n^2 + d_n^2}})) = w_k(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}),$$

where

$$\mathbf{c} = [c_0, \cdots, c_{N-1}]^\top, \quad \mathbf{d} = [d_0, \cdots, d_{N-1}]^\top.$$

After all these steps, let be used

$$\mathbf{x} := [x[\tau_0], \cdots, x[\tau_{K-1}]]^\top, \quad \mathbf{w} = [w_0, \cdots, w_{K-1}]^\top$$

and the function to be optimized is

$$L_{\mathbf{xw}} := \mathbf{x}^\top \mathbf{x} \mathbf{w}^\top \mathbf{w} - (\mathbf{x}^\top \mathbf{w})^2.$$

This function will be called Homogeneous Sinusoidal Representation Function (HSRF). The properties of HSRF will be investigated by optimizing several artificial and natural test-functions based on the BFGS algorithm.