Acta Cybernetica 14 (1999) 315-330.

Optimal parameters of a sinusoidal representation of signals

A. Kocsor* L. Tóth* I. Bálint[†‡§]

Abstract

In the spectral analysis of digital signals, one of the most useful parametric models is the representation 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 shape of the representation functions. If using all of the above parameters, a quite difficult optimization problem arises. The applied methods are generally based on eigenvalue decomposition [3]. However this procedure is computationally expensive and works only if the sinusoids and the residual signal are statistically uncorrelated. To speed up the representation process also rather ad hoc methods occur [4]. The presented algorithm applies the newly established Homogeneous Sinus Representation Function (HSRF) to find the best representing subspace of fixed dimension N by a BFGS optimization. The optimum parameters $\{A, \omega, \varphi\}$ ensure the mean square error of approximation to be below a preset threshold.

1 Introduction

Since the invention of the telephone, speech or generally sound processing and representation have 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 speech comes from the oscillation of the vocal chord, which is modellable by an oscillating string. The voice consists of a fundamental and its harmonics, therefore it is well representable

*MTA-JATE Research Group on Artificial Intelligence, H-6720 Szeged, Aradi Vértanuk Tere 1, Hungary

†Department of Theoretical Physics, József Attila University, H-6720 Szeged, Tisza L. krt. 80-82., Hungary

‡Department of Pharmaceutical Analysis, Szent-Györgyi Albert Medical University, H-6720 Szeged, Somogyi Béla u. 4., Hungary

§Department of Natural Scienses, Polytechnic of the Miskolc University, H-2400 Dunaújváros, Táncsics M. u. 1., Hungary

in the form

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

The error of this approximation gives the 'unvoiced', noise-like part, which can be decoupled from the signal. The model fits well also musical signals, since the sound of most musical instruments (stringed-, wind instruments, etc.) consist of harmonic sinusoids. The residual signal again contains the noise-like part of the sound (e.g. drum hits), which should be modelled separately.

The above form of the model yields a complicated optimization problem enforcing some simplifications. In case of DFT (Discrete Fourier Transform), the number of sinusoids and their frequencies are fixed providing a rapid way for the computation of amplitudes and phases. However, in general, the individual sinusoidal components of this representation may significantly differ in their parameters from the real sound components. The method of McAulay-Quatieri [4] tends to deduce the real frequencies of components by looking for peaks in the DFT spectrum. A basically different approach is based on eigenvalue decomposition [3]. Here, only the dimension of approximation space is fixed, but the statistical independence of the representation functions (sinusoids) and of the residual signal is required.

The presented procedure is free from requiring any statistical condition, only the dimension of approximation space is fixed. The established optimization problem is based on a recently introduced functional [16] and it is solved very effectively by the BFGS [10,12,13] method. The efficiency of the method is illustrated by representations of artificial and natural voice patterns.

As to the structure of the report, the second section provides the usual, 'conservative' formulation of the problem, the third section deals with the introduced Homogeneous Sinus Representation Function (HSRF), the fourth section investigates the properties of HSRF, the fifth section discusses the workhorse optimization scheme BFGS, finally the sixth section delivers the numerical illustrations and conclusions.

1.1 Notational conventions

The Euclidean norm is denoted by $\| \|$, the gradient of a function $f(x) : \mathbb{R}^n \to \mathbb{R}$ by

$$\nabla f(x) = [\frac{\partial f(x)}{\partial x_1}, \cdots, \frac{\partial f(x)}{\partial x_n}]^\top,$$

and the Hessian will be denoted, as

$$\nabla \times \nabla f(x) = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_n} \end{pmatrix}.$$

Definition 1.1 *The continuous function* $f(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}$ *is homogeneous of* l*th degree, if*

$$f(k\mathbf{x}) = k^l f(\mathbf{x}), \qquad k \in \mathbb{R}.$$

2 The representation problem

A signal is sampled at points $\tau_0, \tau_1, \cdots, \tau_{K-1}$ of a closed time-interval $[0, \tau]$ and the obtained values are represented by the real sequence,

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

A function of the form

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

is sought, which approximates the measured sample with a preset error $\epsilon > 0$, by fixing the dimension of approximation space to N,

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

3 Optimization of the Homogeneous Sinusoidal Representation Function

Let be introduced the following notation,

$$w_k(\mathbf{A}, \omega, \varphi) := \sum_{n=0}^{N-1} A_n \sin(\omega_n \tau_k + \varphi_n), \quad k = 0, \cdots, K-1 \tag{2}$$

where

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

and let be applied the trigonometric identity,

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

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

where

$$a_n = A_n \cos(\varphi_n), \quad b_n = A_n \sin(\varphi_n). \tag{3}$$

By introducing new variables,

$$\frac{c_n}{\sqrt{c_n^2 + d_n^2}} := \sin(\omega_n), \quad \frac{d_n}{\sqrt{c_n^2 + d_n^2}} := \cos(\omega_n), \qquad c_n^2 + d_n^2 \neq 0, \tag{4}$$

we obtain

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

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

where

$$\mathbf{a} := [a_0, \cdots, a_{N-1}]^\top, \ \mathbf{b} := [b_0, \cdots, b_{N-1}]^\top,$$
$$\mathbf{c} := [c_0, \cdots, c_{N-1}]^\top, \ \mathbf{d} := [d_0, \cdots, d_{N-1}]^\top.$$

If we introduce two further vectors,

$$\mathbf{x} := [x[\tau_0], \cdots, x[\tau_{K-1}]]^\top, \ \mathbf{w} := [w_0(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}), \cdots, w_{K-1}(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d})]^\top$$

the Homogeneous Sinusoidal Representation Function (HSRF) to be optimized will be

$$L_{\mathbf{xw}}(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}) := \mathbf{x}^\top \mathbf{x} \mathbf{w}^\top \mathbf{w} - (\mathbf{x}^\top \mathbf{w})^2. \tag{5}$$

4 Some properties of HSRF

Notation 4.1 *Let the parameters* $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ *of HSRF be concatenated into a single vector, as follows*

$$\mathbf{z} = [\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}]^\top = [a_0, \cdots, a_{N-1}, b_0, \cdots, b_{N-1}, c_0, \cdots, c_{N-1}, d_0, \cdots, d_{N-1}]^\top.$$

Vector $\mathbf{z}$ is of $4N$-dimension and the concatenated components occupy the following fields:

$$z_i = \begin{cases} a_i & ,if \quad 0 \leq i \leq N-1 \\ b_{i-N} & ,if \quad N \leq i \leq 2N-1 \\ c_{i-2N} & ,if \quad 2N \leq i \leq 3N-1 \\ d_{i-3N} & ,if \quad 3N \leq i \leq 4N-1 \end{cases}$$

Lemma 4.2 *HSRF exhibits the properties:*

1. $L_{\mathbf{xw}}(\mathbf{z})$ *is a homogeneous function of 2nd degree.*

2. $L_{\mathbf{xw}}$ *is a 0th degree homogeneous function of its variables* $\mathbf{c}, \mathbf{d}$*:*

$$L_{\mathbf{xw}}(\mathbf{a}, \mathbf{b}, \lambda\mathbf{c}, \lambda\mathbf{d}) = L_{\mathbf{xw}}(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}), \quad 0 \neq \lambda \in \mathbb{R}$$

3. $L_{\mathbf{xw}}$ *is a 2nd degree homogeneous function of its variables* $\mathbf{a}, \mathbf{b}$*:*

$$L_{\mathbf{xw}}(\lambda\mathbf{a}, \lambda\mathbf{b}, \mathbf{c}, \mathbf{d}) = \lambda^2 L_{\mathbf{xw}}(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}), \quad \lambda \in \mathbb{R}$$

4. $L_{\mathbf{xw}}(\mathbf{z}) = L_{\mathbf{wx}}(\mathbf{z})$

Proof. Point 4. satisfies trivially, points 1., 2., and 3. follow from the continuity of $L_{\mathbf{xw}}(\mathbf{z})$, as well as from the enumerated properties obeyed by $w_k(\mathbf{z}) \equiv w_k(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d})$:

1. $w_k(\lambda\mathbf{z}) = \lambda w_k(\mathbf{z}), \qquad 0 \neq \lambda \in \mathbb{R}.$
2. $w_k(\lambda\mathbf{a}, \lambda\mathbf{b}, \mathbf{c}, \mathbf{d}) = \lambda w_k(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}), \qquad \lambda \in \mathbb{R}.$
3. $w_k(\mathbf{a}, \mathbf{b}, \lambda\mathbf{c}, \lambda\mathbf{d}) = w_k(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}), \qquad 0 \neq \lambda \in \mathbb{R}.$ □

Theorem 4.3 *HSRF exhibits the enumerated properties:*

1. $L_{\mathbf{xw}}(\mathbf{z}) \geq 0$ *and* $L_{\mathbf{xw}}(\mathbf{z}) = 0$ *if and only if the* $\mathbf{x}$ *and* $\mathbf{w}$ *vectors are linearly dependent.*
2. *If* $\mathbf{z}$ *is an optimumpoint of* $L_{\mathbf{xw}}(\mathbf{z})$*, then* $L_{\mathbf{xw}}(\mathbf{z}) = 0$.
3. *If* $L_{\mathbf{xw}}(\mathbf{z}) = 0$*, then* $\nabla L_{\mathbf{xw}}(\mathbf{z}) = \mathbf{0}$.

Proof.

1. Function $L_{\mathbf{xw}}(\mathbf{z})$ stems from the Cauchy-Scwartz-Bunyakovszkij inequality applied on the vectors $\mathbf{x} \in \mathbb{R}^K$ and $\mathbf{w}(\mathbf{z}) \in \mathbb{R}^K$:

$$\mathbf{x}^\top \mathbf{x}\mathbf{w}(\mathbf{z})^\top \mathbf{w}(\mathbf{z}) \geq \left(\mathbf{x}^\top \mathbf{w}(\mathbf{z})\right)^2.$$

   The equality satisfies, if the vectors are linearly dependent,

$$\mathbf{x}^\top \mathbf{x}\mathbf{w}(\mathbf{z})^\top \mathbf{w}(\mathbf{z}) - \left(\mathbf{x}^\top \mathbf{w}(\mathbf{z})\right)^2 = 0.$$

2. If $\mathbf{z} \in \mathbb{R}^{4N}$ is an optimumpoint of $L_{\mathbf{xw}}(\mathbf{z})$, then necessarily $\nabla L_{\mathbf{xw}}(\mathbf{z}) = \mathbf{0}$. Euler's theorem ensures that the 2nd degree, homogeneous function $L_{\mathbf{xw}}(\mathbf{z})$ obeyes the equality:

$$\mathbf{z}^\top \nabla L_{\mathbf{xw}}(\mathbf{z}) = 2L_{\mathbf{xw}}(\mathbf{z}).$$

   Therefore a zerovector gradient implies a zero function value, $\nabla L_{\mathbf{xw}}(\mathbf{z}) = \mathbf{0} \Longrightarrow L_{\mathbf{xw}}(\mathbf{z}) = 0$.

3. If $L_{\mathbf{xw}}(\mathbf{z}) = 0$, the vectors $\mathbf{x}$ and $\mathbf{w}$ are linearly dependent, i.e. $\mathbf{x} = \lambda \mathbf{w}(\mathbf{z})$ without restricting generality. The following sequence of equalities proves the statement:

$$\nabla L_{\mathbf{xw}}(\mathbf{z}) =$$
$$= \mathbf{x}^\top \mathbf{x} \nabla \left(\mathbf{w}(\mathbf{z})^\top \mathbf{w}(\mathbf{z})\right) + \nabla \left(\mathbf{x}^\top \mathbf{x}\right) \mathbf{w}(\mathbf{z})^\top \mathbf{w}(\mathbf{z}) - 2 \left(\mathbf{x}^\top \mathbf{w}(\mathbf{z})\right) \nabla \left(\mathbf{x}^\top \mathbf{w}(\mathbf{z})\right) =$$
$$= \lambda^2 \left(\mathbf{w}(\mathbf{z})^\top \mathbf{w}(\mathbf{z})\right) \mathbf{w}(\mathbf{z})^\top \nabla \left(\mathbf{w}(\mathbf{z})\right) + 0 - 2\lambda^2 \left(\mathbf{w}(\mathbf{z})^\top \mathbf{w}(\mathbf{z})\right) \mathbf{w}(\mathbf{z})^\top \nabla \left(\mathbf{w}(\mathbf{z})\right) = 0$$

□

The properties of $L_{\mathbf{xw}}(\mathbf{z})$ discussed above ensure good optimization properties. The optimumpoints of this non-negative homogeneous function are global and a gradient-based optimization scheme may efficiently localize them.

Lemma 4.4 *The gradient* $\nabla L_{\mathbf{xw}}(\mathbf{z})$; $\mathbf{z} = [\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}]^\top$ *is of the form:*

$$\frac{\partial}{\partial z_i} L_{\mathbf{xw}}(\mathbf{z}) = \left(\sum_{k=0}^{K-1} x[\tau_k]^2\right) \left(2 \sum_{k=0}^{K-1} w_k(\mathbf{z}) \frac{\partial}{\partial z_i} w_k(\mathbf{z})\right) - 2 \left(\sum_{k=0}^{K-1} x[\tau_k] w_k(\mathbf{z})\right) \left(\sum_{k=0}^{K-1} x[\tau_k] \frac{\partial}{\partial z_i} w_k(\mathbf{z})\right).$$

The partial derivatives by the various sets of variables are as follow:

$$\frac{\partial}{\partial a_i} w_k(\mathbf{z}) = \sin\left(\tau_k \arcsin\left(\frac{c_i}{\sqrt{c_i^2 + d_i^2}}\right)\right)$$

$$\frac{\partial}{\partial b_i} w_k(\mathbf{z}) = \cos\left(\tau_k \arcsin\left(\frac{c_i}{\sqrt{c_i^2 + d_i^2}}\right)\right)$$

$$\frac{\partial}{\partial c_i} w_k(\mathbf{z}) = \frac{\tau_k d_i \left(a_i \cos\left(\tau_k \arcsin\left(\frac{c_i}{\sqrt{c_i^2+d_i^2}}\right)\right) - b_i \sin\left(\tau_k \arcsin\left(\frac{c_i}{\sqrt{c_i^2+d_i^2}}\right)\right)\right)}{(c_i^2 + d_i^2)}$$

$$\frac{\partial}{\partial d_i} w_k(\mathbf{z}) = -\frac{\tau_k c_i \left(a_i \cos\left(\tau_k \arcsin\left(\frac{c_i}{\sqrt{c_i^2+d_i^2}}\right)\right) - b_i \sin\left(\tau_k \arcsin\left(\frac{c_i}{\sqrt{c_i^2+d_i^2}}\right)\right)\right)}{(c_i^2 + d_i^2)}$$

Proof. The proof is trivial by the differentiation rules. □

The next theorem provides the bridge between the function value of HSRF and the representation problem (1).

Theorem 4.5 *If for any positive number* δ *and for real vectors* $\mathbf{z} \in \mathbb{R}^{4N}$ *and* $\mathbf{x} \in \mathbb{R}^K$, $L_{\mathbf{xw}}(\mathbf{z}) < \delta$ *is satisfied, then*

$$\min_i \left\| \mathbf{x} - \frac{\mathbf{w}(\mathbf{z})}{\lambda_i} \right\|^2 < \frac{2\delta}{\|\mathbf{w}(\mathbf{z})\|^2}, \qquad \lambda_i = \frac{w_i(\mathbf{z})}{x[\tau_i]}, \; i \in \{0, \cdots K-1\}.$$

Proof. For the sake of simplicity, the argument $\mathbf{z}$ of $\mathbf{w}(\mathbf{z})$ and of $w_i(\mathbf{z})$ will be omitted and $x[\tau_i]$ will be denoted simply as x_i.

$$L_{\mathbf{xw}}(\mathbf{z}) = \mathbf{x}^\top \mathbf{x}\mathbf{w}^\top \mathbf{w} - (\mathbf{x}^\top \mathbf{w})^2 = \left(\sum_{i=0}^{K-1} x_i^2\right)\left(\sum_{i=0}^{K-1} w_i^2\right) - \left(\sum_{i=0}^{K-1} x_i w_i\right)^2 =$$

$$= \frac{1}{2}\sum_{i=0}^{K-1}\sum_{j=0}^{K-1}(w_i x_j - w_j x_i)^2 = \frac{1}{2}\sum_{i=0}^{K-1}\sum_{j=0}^{K-1}\left(w_i x_j - w_j \frac{w_i}{\lambda_i}\right)^2 =$$

$$= \frac{1}{2}\sum_{i=0}^{K-1} w_i^2 \sum_{j=0}^{K-1}\left(x_j - \frac{w_j}{\lambda_i}\right)^2 = \frac{1}{2}\sum_{i=0}^{K-1} w_i^2 \left\|\mathbf{x} - \frac{\mathbf{w}}{\lambda_i}\right\|^2 < \delta$$

If choosing the 'best' of λ_is, the inequality

$$\frac{1}{2}\|\mathbf{w}\|^2 \min_i \left\|\mathbf{x} - \frac{\mathbf{w}}{\lambda_i}\right\|^2 < \frac{1}{2}\sum_{i=0}^{K-1} w_i^2 \left\|\mathbf{x} - \frac{\mathbf{w}}{\lambda_i}\right\|^2 < \delta,$$

proves the statement. □

If the previous optimization yields a $\mathbf{z_0}$ satisfying

$$\left\|\mathbf{x} - \mathbf{w}(\mathbf{z_0})\frac{x[\tau_s]}{w_s(\mathbf{z})}\right\|^2 = \min_i \left\|\mathbf{x} - \mathbf{w}(\mathbf{z_0})\frac{x[\tau_i]}{w_i(\mathbf{z})}\right\|^2 < \frac{2L_{\mathbf{xw}}(\mathbf{z_0})}{\|\mathbf{w}(\mathbf{z_0})\|^2},$$

the difference of the Euclidean norm of the signal vector $\mathbf{x}$ and the representation vector

$$\mathbf{w}(\mathbf{z_0})\frac{x[\tau_s]}{w_s(\mathbf{z})}$$

is given by the above expression.

5 Solving the representation problem by NHSRF

5.1 Application of the BFGS optimization scheme

For minimizing the representation functional the most suitable procedure proved to be the gradient based Broyden-Fletcher-Goldfarb-Shanno (BFGS) scheme [10,12,13]. Since also the zero-vector is an optimumpoint of $L_{\mathbf{xw}}(\mathbf{z})$ to avoid convergence to the zero-vector, the HSRF is normalized to be a 0th degree homogeneous function of the form,

$$\frac{L_{\mathbf{xw}}(\mathbf{z})}{\|\mathbf{a}\|\,\|\mathbf{b}\|}. \tag{6}$$

This will be called Normalized HSRF, NHSRF in short. Every former obtained result are inherited by NHSRF, however the new partial derivative components of the gradient are given below,

$$\frac{\partial}{\partial a_i}\frac{L_{\mathbf{xw}}(\mathbf{z})}{\|\mathbf{a}\|\,\|\mathbf{b}\|} = \frac{1}{\|\mathbf{b}\|}\frac{\|\mathbf{a}\|\left(\frac{\partial}{\partial a_i}L_{\mathbf{xw}}(\mathbf{z})\right) - L_{\mathbf{xw}}(\mathbf{z})\frac{a_i}{\|\mathbf{a}\|}}{\|\mathbf{a}\|^2},$$

$$\frac{\partial}{\partial b_i}\frac{L_{\mathbf{xw}}(\mathbf{z})}{\|\mathbf{a}\|\,\|\mathbf{b}\|} = \frac{1}{\|\mathbf{a}\|}\frac{\|\mathbf{b}\|\left(\frac{\partial}{\partial b_i}L_{\mathbf{xw}}(\mathbf{z})\right) - L_{\mathbf{xw}}(\mathbf{z})\frac{b_i}{\|\mathbf{b}\|}}{\|\mathbf{b}\|^2},$$

$$\frac{\partial}{\partial c_i}\frac{L_{\mathbf{xw}}(\mathbf{z})}{\|\mathbf{a}\|\,\|\mathbf{b}\|} = \frac{1}{\|\mathbf{a}\|\,\|\mathbf{b}\|}\frac{\partial}{\partial c_i}L_{\mathbf{xw}}(\mathbf{z}),$$

$$\frac{\partial}{\partial d_i}\frac{L_{\mathbf{xw}}(\mathbf{z})}{\|\mathbf{a}\|\,\|\mathbf{b}\|} = \frac{1}{\|\mathbf{a}\|\,\|\mathbf{b}\|}\frac{\partial}{\partial d_i}L_{\mathbf{xw}}(\mathbf{z}).$$

Every test result displayed for illustration was obtained by the NHSRF. For terminating the line-search

$$\min_{\kappa\in\Re}\frac{L_{\mathbf{xw}}(\mathbf{z}+\kappa\mathbf{d})}{\|\mathbf{a}\|\,\|\mathbf{b}\|},$$

the Wolf-condition, for initializing the H-matrix, the unit matrix was used. For terminating the whole BFGS optimization generally the acceptable low norm of the error-vector, as well as that of the gradient was used. We also have stopped the iteration, if the condition

$$\frac{2L_{\mathbf{xw}}(\mathbf{z})}{\|\mathbf{w}(\mathbf{z})\|^2} < \delta$$

as referred in Theorem 4.5 was satisfied.

5.2 Estimation of the number of necessary operations

For one iteration step of the BFGS scheme generally the function value and gradient should be computed at several points in the line-search process. This requires to evaluate scalar products, which can be obtained with $o(NK + K^2)$ operations, because for any k,

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

can be obtained with $o(N)$ operations, and the K-dimensional vector $\mathbf{w}$ is obtainable with $o(NK)$ operations. To evaluate the function value, as well as the gradient, scalar products of K-dimensional vectors have to be computed. The update of the $4N\times 4N$-size H-matrix requires $o(N^2)$ operations in every step counting altogether $o(N^2+NK)$ operations per step. Since generally the number N of sinusoidal components is much less then the number of the components in the signal vector, the number of really required operations is of $o(NK)$.

6 Numerical illustration of the algorithm

The aim is to construct an acceptably accurate sinusoidal representation of an arbitrary (sound) signal. The procedure is based on the optimization of NHSRF starting from an approximate sinusoidal decomposition of the signal. The procedure

aims to reduce the number of sinusoidal components by retaining the accuracy of the representation. The numerical illustrations are mainly artificial examples to obtain a well-defined measure for the accuracy of approximation, however the representation of a natural sound sample is also included.

In all cases of artificial and natural sound patterns an approximate sinusoidal represenation served as initial parameters of the NHRSF optimization. The $x[\tau_0], \cdots, x[\tau_{K-1}]$ signal was decomposed by the following iterative algorithm:

- In every iteration, first the DFT of the signal was computed on a zero padded 2048 base point data set using a Hamming window. The obtained spectrum is the convolution of the transformed signal and the transformed window.
- The maximum amplitude component was selected. To remove the unpleasant effect of windowing, the Fourier-transform of the window function was subtracted from the spectrum after a suitable shifting and scaling.
- The iteration was continued until the largest amplitude was smaller than a preset positive number.

The iterations steps necessary for the NHSRF optimization algorithm to reach the required accuracy was empirically tested. Both in the case of the natural and artificial tests the number of function and gradient evaluations necessary for one iteration step was generally one or two.

6.1 Representation of artificial signals

The sample to be represented was a sum of N sinusoidal components sampled at K points, where the parameters $A_0, \cdots, A_{N-1}$; $\omega_0, \cdots, \omega_{N-1}$; and $\varphi_0, \cdots, \varphi_{N-1}$ were specified:

$$x[k] = \sum_{n=0}^{N-1} A_n \sin(\omega_n k + \varphi_n), \quad k \in \{1, \cdots, K\}.$$

Using the above mentioned DFT-based decomposition of the signal,

$$x[1], \cdots, x[K],$$

the following estimate was obtained,

$$x[k] \approx \sum_{n=0}^{P-1} A'_n \sin(\omega'_n k + \varphi'_n), \quad k \in \{1, \cdots, K\}, \quad , P > N,$$

which proved to be generally unacceptably inaccurate. Without restricting generality, we can assume the following ordering of the components $A'_i \geq A'_j \Leftrightarrow i \geq j$; which selects the N components,

$$A'_0, \cdots, A'_{N-1}, \omega'_0, \cdots, \omega'_{N-1}, \varphi'_0, \cdots, \varphi'_{N-1}$$

ω	φ	A	K	iter	δ_1	δ_2
0.2	0.1	2	100	8	0.0006	10^{-17}
0.2	0.1	2	200	9	0.0031	10^{-19}
0.2	0.1	2	400	7	0.0131	10^{-15}

Table 1: Approximation of one sinusoidal component.

dominating by amplitude. The optimization process started in every case from these dominant components by constructing the initial parameter vector

$$\mathbf{z} = [a'_0, \cdots, a'_{N-1}, b'_0, \cdots, b'_{N-1}, c'_0, \cdots, c'_{N-1}, d'_0, \cdots, d'_{N-1}]^\top$$

using (3) and (4). Let be assumed, that the optimization of (6) resulted in the optimum vector

$$\mathbf{z_0} = [a''_0, \cdots, a''_{N-1}, b''_0, \cdots, b''_{N-1}, c''_0, \cdots, c''_{N-1}, d''_0, \cdots, d''_{N-1}]^\top$$

and the inverses of transformations (3) and (4) yielded the parameters

$$A''_0, \cdots, A''_{N-1}, \omega''_0, \cdots, \omega''_{N-1}, \varphi''_0, \cdots, \varphi''_{N-1}.$$

The error of the DFT-based signal representation is

$$\delta_1 = \sum_{n=0}^{N-1} (A_n - A'_n)^2 + (\omega_n - \omega'_n)^2 + (\varphi_n - \varphi'_n)^2, \tag{7}$$

while that of the NHSRF-based signal representation is

$$\delta_2 = \sum_{n=0}^{N-1} (A_n - A''_n)^2 + (\omega_n - \omega''_n)^2 + (\varphi_n - \varphi''_n)^2. \tag{8}$$

For $N = 1, 2, 3$, three examples were investigated in each case and the results are displayed in Tables 1.-9.. The rows of the tables display the parameters of the sinusoidal basis functions ω, φ, A, the number of sample points K, the number of iterations *iter* and the accuracies of DFT-based and NHRSF-based representations δ_1, δ_2. The discussion of the results will be given together with the discussion of the natural test results.

6.2 Representation of natural sound signals

To check the accuracy and efficiency of the proposed algorithm on natural sound patterns, the phone /a/ was represented and synthesised by the DFT decomposition, as well as by the NHSRF optimization based scheme. The sample was consisting in 861 points from the middle of a, the DFT algorithm described above

ω	φ	A	K	iter	δ_1	δ_2
0.2	0.1	200	100	7	0.2456	10^{-15}
0.2	0.1	200	200	10	0.1858	10^{-16}
0.2	0.1	200	400	11	0.0456	10^{-15}

Table 2: Approximation of one sinusoidal component.

ω	φ	A	K	iter	δ_1	δ_2
0.2	0.1	0.02	100	10	0.0006	10^{-28}
0.2	0.1	0.02	200	8	0.0031	10^{-20}
0.2	0.1	0.02	400	9	0.0131	10^{-18}

Table 3: Approximation of one sinusoidal component.

ω	φ	A	K	iter	δ_1	δ_2
0.1 0.2	−0.1 0.1	3 2	100	18	0.5452	10^{-17}
0.1 0.2	−0.1 0.1	3 2	200	18	0.0194	10^{-19}
0.1 0.2	−0.1 0.1	3 2	400	20	0.0727	10^{-16}

Table 4: Approximation of two sinusoidal components.

ω	φ	A	K	iter	δ_1	δ_2
0.85 0.98	−0.1 0.1	1201 1200	100	19	7.6271	10^{-10}
0.85 0.98	−0.1 0.1	1201 1200	200	17	19.470	10^{-13}
0.85 0.98	−0.1 0.1	1201 1200	400	20	77.481	10^{-11}

Table 5: Approximation of two sinusoidal components.

ω	φ	A	K	iter	δ_1	δ_2
0.1 0.2	−0.1 0.1	1800 179	100	20	568.43	10^{-12}
0.1 0.2	−0.1 0.1	1800 179	200	20	14.969	10^{-10}
0.1 0.2	−0.1 0.1	1800 179	400	20	118.35	10^{-12}

Table 6: Approximation of two sinusoidal components.

ω	φ	A	K	iter	δ_1	δ_2
1 0.1 0.2	0 −0.1 0.1	30 3 2	100	27	0.7118	10^{-16}
1 0.1 0.2	0 −0.1 0.1	30 3 2	200	21	0.0723	10^{-13}
1 0.1 0.2	0 −0.1 0.1	30 3 2	400	22	0.0706	10^{-15}

Table 7: Approximation of three sinusoidal components.

ω	φ	A	K	iter	δ_1	δ_2
1 1.2 1.1	0 0.5 −0.9	1200 200 129	100	28	2178.4	10^{-12}
1 1.2 1.1	0 0.5 −0.9	1200 200 129	200	21	41.797	10^{-12}
1 1.2 1.1	0 0.5 −0.9	1200 200 129	400	25	57.899	10^{-10}

Table 8: Approximation of three sinusoidal components.

ω	φ	A	K	iter	δ_1	δ_2
1	0	130				
0.1	−0.1	129	100	29	0.3769	10^{-12}
0.2	0.1	128				
1	0	130				
0.1	−0.1	129	200	25	0.3253	10^{-13}
0.2	0.1	128				
1	0	130				
0.1	−0.1	129	400	26	0.4165	10^{-12}
0.2	0.1	128				

Table 9: Approximation of three sinusoidal components.

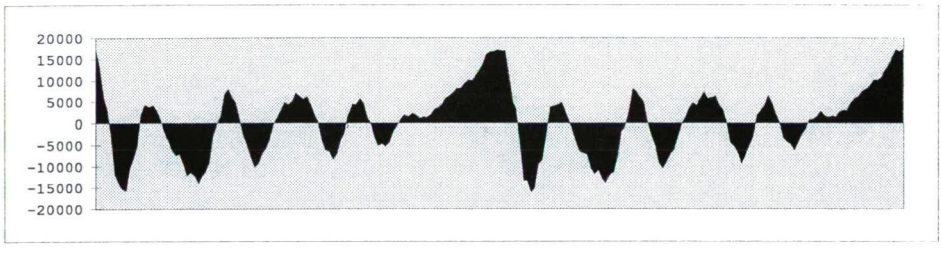

Figure 1:

provided the sinusoidal decomposition of the signal and the sound pattern was synthesised from the obtained sinusoidal components. The parameters of the first 20 dominant sinusoidal components of the DFT decomposition were used as initial parameters of the NHSRF optimization scheme and the minimization of (6) yielded the optimum decomposition of /*a*/, by the NHSRF-based procedure. The sound signal was synthesised again from the obtained components. Unfortunately the quality of sound synthesis is not easy to measure, since the metric is not Euclidean, but a 'perceptual' distance function would be necessary to measure the 'goodness' of the representation procedure. Therefore the 'comparison by listening' of the original and synthesised sounds had a decisive role in the judgement. However to give an easily noticeable impression on the accuracies of approximations of the natural sound pattern, figure 1-3 display the original signal, the signal synthesised from the 50 largest amplitude components of the DFT-based decomposition and the signal synthesised from the 20 components of the NHSRF-based decomposition.

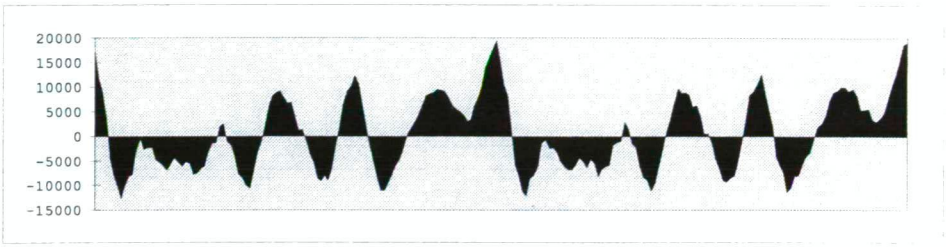

Figure 2:

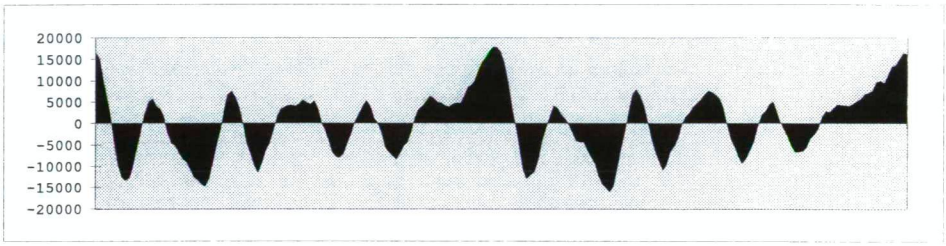

Figure 3:

7 Discussion

7.1 Artificial signals

The application of the proposed algorithm is especially important if either the components of the signal are required with high accuracies, or the usual DFT-based technics are not suitable to provide acceptably accurate results at all. This case occurs if the sample is too short, the components are too close, or the amplitudes differ too much. Our test functions were therefore of these types.

The proposed algorithm proved to be powerful in correcting the estimations of the DFT-based decomposition procedure even in those cases where the former algorithm provided quite acceptable results. Notice that the DFT-based scheme estimated the phases very poorly, while the NHSRF-based algorithm corrected these values.

7.2 Real sound signals

The proposed algorithm allowed to reconstruct the analysed sound signal from less sinusoidal components (20 components) at higher accuracy than the DFT-based method (50 components), as seen clearly on the figures. The DFT-based decomposition of voiced sounds is generally unable to provide the sinusoidal fundamental and overtones accurately enough and yields more overtone components than the sample really contains. However the high-fidelity modelling of voiced sounds re-

quires to synthesise the signal by the least sinusoidal components defined with accurately determined parameters. This is a necessary condition for developing and using efficient data compression technics, too.

7.3 Future Work

On the basis of the presented results, the NHSRF-optimizing algorithm proved to be robust and efficient in applications of speech- and audio-processing. The aim is to use the algorithm in sound coding and to develop a more advanced pitch estimation method than the ones used nowadays. The momentarily fixed dimension of the approximation subspace will be handled as variational parameter in the future. This feature will help to separate the sinusoidal and the noiselike components of the sound allowing to screen noise, to detect the unvoiced/voiced parts of the sound, furthermore the upgraded procedure would be a candidate for being applied in those sound coding methods, which are based on the 'sinusoidal + noise'-type decomposition of the signal (e.g. Quatieri-McAulay).

References

[1] Alan V. Oppenheim, Ronald W. Schafer: Discrete-Time Signal Processing, PRENTICE HALL

[2] L.R. Rabiner, R.W Schafer: Digital Processing of Speech Signals, PRENTICE HALL

[3] S. Lawrence MArple, Jr. : Digital Spectral Analysis with applications, PRENTICE HALL

[4] McAulay, R.J. and T.F. Quatieri. 1986. "Speech Analysis/Synthesis based on a Sinusoidal Representation." IEEE Transactions on Acoustics, Speech and Signal Processing, 34(4):744–754

[5] Allen, J.B 1977. "Short Term Spectral Analysis, Synthesis, and Modification by Discrete Fourier Transformation. " IEEE Transactions on Acoustics, Speech and Signal Processing, 25(3):235-238

[6] Hess, W. 1983. Pitch Determination of Speech Signals. New york: Springer-Verlang

[7] Harrs, F.J. 1978. " On the use of windows for harmonic analysis with the discrete Fourier transform." Proceedings IEEE, vol. 66, pp.51-83.

[8] Goodwin, M. and X.Rodet. 1994. "Efficient Fourier Synthesis of Nonstacionary Sinusoids." Proceedings of the 1994 International Computer Music Conference. San Francisco: Computer Music Association.

[9] Mather, R. C. and J. W. Beauchamp. 1994. "Fundamental Frequency Estimation of Musical Signals using a two-way Mismatch Procedure." Journal of Acoustical Society of America 95(4):2254–2263

[10] Mokhtar S. Bazaraa, Hanif D. Sherali, C. M. Shetty, *NONLINEAR PROGRAMMING theory and algorithms*, John Wiley & Sons, [1993]

[11] J. L. Nazareth, *Conjugate Gradient Methods Less Dependency on Conjugacy*, SIAM Review, 28(4), PP. 501-511, 1986.

[12] J. Nocedal, *The Performance of Several Algorithms for Large Scale Unconstrained Optimization, in Large-Scale Numerical Optimization*, T. F. Coleman and Y. Li (Eds.), SIAM, Philadelphia, pp. 138-151, 1990.

[13] Gill, P.E., W. Murray, and P.A. Pitfield, *The implementation of Two Revised Quasi-Newton Algorithms for Unconstrained Optimization*, Report NAC-11, National Physical Lab., 1972

[14] Usmani, R. A.: Applied Linear Algebra. Marcel Dekker, New York, 1987

[15] Lippmann, Stanley B.: C++ Primer. Addison Wesley, 1991

[16] A. Kocsor, J. Dombi, I. Bálint, An Optimization Algorithm for Determining Eigenpairs of Large Real Matrices, (submitted to SIAM Journal On Scientific Computing)