MODIFIED SPHERE DECODING ALGORITHMS AND THEIR APPLICATIONS TO SOME SPARSE APPROXIMATION PROBLEMS

Przemysław Dymarski and Rafał Romaniuk

Institute of Telecommunications, Warsaw University of Technology
ul. Nowowiejska 15/19, 00-665, Warsaw, Poland
email: dymarski@tele.pw.edu.pl, r.romaniuk@tele.pw.edu.pl

ABSTRACT

This work presents modified sphere decoding (MSD) algorithms for optimal solution of some sparse signal modeling problems. These problems include multi-pulse excitation signal calculations for multi-pulse excitation (MPE), algebraic code excited linear predictive (ACELP) and -pulse maximum likelihood quantization (MP-MLQ) speech coders. With the proposed MSD algorithms, the optimal solution of these problems can be obtained at substantially lower computational cost compared with full search algorithm. The MSD algorithms are compared with a series of suboptimal approaches in sparse approximation of correlated Gaussian signals and low delay speech coding tasks.

Index Terms—sphere decoder, lattice, sparse approximation, speech coding, CELP, MP-MLQ, ACELP.

1. INTRODUCTION

Sparse approximation techniques are widely used in speech and audio coding [1-4], MIMO communications [5, 6], array signal processing [7], radar, etc. In this paper the following sparse approximation problem is considered:

$$\min_{\bar{\mathbf{x}}} ||\bar{\mathbf{y}} - \mathbf{F}\bar{\mathbf{x}}||^2 \tag{1}$$

where $\mathbf{F}_{N\times L}$ - codebook (dictionary, lattice-generating matrix), $\bar{\mathbf{y}} \in \Re^N$ - target vector, $\bar{\mathbf{x}} \in D_M^L \subset \Re^L$ belongs to a finite L-dimensional lattice with M elements per dimension; i.e. each component of $\bar{\mathbf{x}}$ may take one of the M values (the alphabet). An example of such an alphabet is the set of data transmission symbols in MIMO communication or the quantized gain values in a multistage shape-gain vector quantizer. If the alphabet contains a zero value, then a sparse solution may be obtained. In some applications (MIMO communication) both vectors and matrix are complex, but in this work real numbers are used.

Three cases are considered in this paper.

Case 1. Closest point in the lattice. Vector $\bar{\mathbf{x}}$ may be dense or K-sparse. If $N \geq L$, the optimal solution of (1) may be found by using a sphere decoding (SD) algorithm [5, 8]. For $N < L$ generalized SD algorithms may be used [9], at the cost of increasing the computational complexity. This model is mainly used in simulation of MIMO systems [5, 6].

Case 2. Uniformly scaled lattice (isotropic scaling). Problem (1) is generalized: now vector $\bar{\mathbf{x}}$ is multiplied by a scalar gain g. Thus $\bar{\mathbf{x}}$ and g are searched, yielding minimum $||\bar{\mathbf{y}} - g\mathbf{F}\bar{\mathbf{x}}||^2$. The most important applications are in ACELP (algebraic CELP) and MP-MLQ (multi-pulse maximum likelihood quantization) speech coders [1]. In these coders vector $\bar{\mathbf{x}}$ is K-sparse and its components are $x_i \in \{-1,0,1\}$. The sparse approximation of the target vector is the sum of the signed columns of **F** multiplied by g:

$$g\mathbf{F}\bar{\mathbf{x}} = g\sum_{i=1}^{K} x_{j(i)}\bar{\mathbf{f}}^{j(i)} \tag{2}$$

where $x_{j(i)} = \pm 1$, and $\bar{\mathbf{f}}^{j(i)}$ is a column of a square matrix **F**. Column $\bar{\mathbf{f}}^{j}$ contains the response of the predictive synthesis filter to a single pulse of a unit amplitude positioned at j. In order to find indices $j(i)$, signs $x_{j(i)}$ and gain g, suboptimal codebook search algorithms are used [1, 2, 4].

Case 3. Non-uniformly scaled lattice (anisotropic scaling). Vector $\bar{\mathbf{x}}$ is now replaced with $\mathbf{G}\bar{\mathbf{x}}$, where $\mathbf{G} = diag\{g_1, g_2, \ldots, g_L\}$ contains real gains. This is a typical speech model used in MPE (multi-pulse excitation) speech coders [10]. In these coders vector $\bar{\mathbf{x}}$ is K-sparse and its components are $x_i \in \{0,1\}$. The sparse approximation of the target vector is a linear combination:

$$\mathbf{F}\mathbf{G}\bar{\mathbf{x}} = \sum_{i=1}^{K} g_{j(i)}\bar{\mathbf{f}}^{j(i)} \tag{3}$$

In MPE coders matrix **F**, containing filtered unit pulses, is square (which is assumed in this paper) but in CELP coders L may be greater than N. Indices $j(i)$ and gains $g_{j(i)}$ are searched using suboptimal algorithms [3, 10-12]. A signal model (3) is also used to describe radar signals, sparse communication channels [13], etc.

The sparse approximation problems (1 to 3) are NP-hard. A full search algorithm, testing all possible solutions, is generally not feasible. However, the optimal tree searching algorithms may reduce the computational load significantly. In Case 1 several variants of the sphere decoding (SD) algorithm are used [5]. In Cases 2 and 3 SD and related algorithms are used for calculation of the integer vector $\bar{\mathbf{x}}$ if gains (g or **G**) are known; thus, the gains calculation is decoupled from searching the closest point in the lattice, yielding suboptimal algorithms [6, 7, 13, 14].

In this work the **optimal** tree search, based on a modified SD approach, is proposed for joint calculation of the lattice point and gain(s). The proposed algorithms are compared with suboptimal ones for two selected problems: sparse approximation of correlated Gaussian signals and low delay speech coding.

2. MODIFIED AND SPARSE SD ALGORITHMS

2.1. The standard sphere decoding algorithm

Generally, Sphere Decoding (SD) algorithm is used for solving (1) and the alphabet D_M^L is usually a set of integer values. The matrix $\mathbf{F}_{N\times L}$ is assumed to be square, but extension to $N \geq L$ is straightforward. Using QR-decomposition of the matrix **F**, (1) may be transformed to the following problem:

$$\min_{\bar{\mathbf{x}}} ||\bar{\mathbf{z}} - \mathbf{R}\,\bar{\mathbf{x}}\,||^2 \tag{4}$$

where $\bar{\mathbf{z}} = \mathbf{Q}^{-1}\bar{\mathbf{y}}$, and **R** is an upper triangular matrix.

The SD algorithm consists of N levels. In each level a new column of **R** is appended (from the right to the left side) and lattice points, generated by a corresponding component of $\bar{\mathbf{x}}$, are tested. At the k^{th} level it is the column $\bar{\mathbf{r}}^{j(k)}$ and the component $x_{j(k)}$, where $j(k) = N - k + 1$. The partial solution of problem (4), reduced to k-1 dimensions, is not changed at the k^{th} level, and thus the partial difference vector

$$\bar{\boldsymbol{\varepsilon}}^{(k-1)} = \bar{\mathbf{z}}^{(k-1)} - \mathbf{R}^{(k-1)}\,\bar{\mathbf{x}}^{(k-1)} \tag{5}$$

remains constant (see Fig.1).

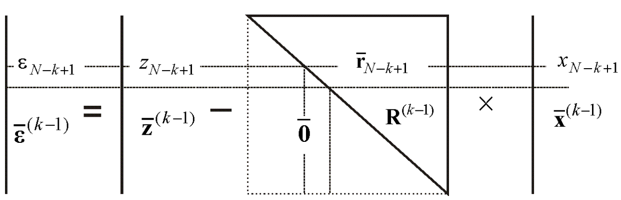

Fig.1. Evolution of the accumulated distance in SD algorithm

At the k^{th} level the partial difference equals

$$\begin{aligned}\bar{\boldsymbol{\varepsilon}}^{(k)} &= \left|\begin{matrix}\varepsilon_{N-k+1}\\ \bar{\boldsymbol{\varepsilon}}^{(k-1)}\end{matrix}\right| = \bar{\mathbf{z}}^{(k)} - \mathbf{R}^{(k)}\bar{\mathbf{x}}^{(k)} = \\ &= \left|\begin{matrix}z_{N-k+1}\\ \bar{\mathbf{z}}^{(k-1)}\end{matrix}\right| - \left|\begin{matrix}\bar{\mathbf{r}}_{N-k+1}\\ \bar{\mathbf{0}}\,\mathbf{R}^{(k-1)}\end{matrix}\right| \left|\begin{matrix}x_{N-k+1}\\ \bar{\mathbf{x}}^{(k-1)}\end{matrix}\right|\end{aligned} \tag{6}$$

where $\bar{\mathbf{r}}_{N+1-k}$ - a row of matrix **R** and $\bar{\mathbf{0}}$ - a column of zeroes. The squared norm of the partial difference vector (i.e. the squared partial distance) increases:

$$||\bar{\boldsymbol{\varepsilon}}^{(k)}||^2 = \left(z_{N-k+1} - \bar{\mathbf{r}}_{N-k+1}\bar{\mathbf{x}}^{(k)}\right)^2 + ||\bar{\boldsymbol{\varepsilon}}^{(k-1)}||^2 \tag{7}$$

Thus, when looking for a lattice point within a sphere of a radius R centered at $\bar{\mathbf{z}}$, it makes no sense to pass to the k^{th} level if $||\bar{\boldsymbol{\varepsilon}}^{(k-1)}||^2 > R^2$. The search must be continued at lower levels. That is why the SD algorithm yields far fewer arithmetical operations than the full search algorithm.

2.2. The modified sphere decoding algorithm

Now we introduce the modified SD (MSD) algorithm for searching for the **optimal** point in a uniformly scaled lattice (Case 2). At the level k-1, the squared norm of the partial difference vector may be expressed as follows:

$$||\bar{\boldsymbol{\varepsilon}}^{(k-1)}||^2 = ||\bar{\mathbf{z}}^{(k-1)} - g\,\mathbf{R}^{(k-1)}\,\bar{\mathbf{x}}^{(k-1)}||^2 = f^{(k-1)}(g) \tag{8}$$

In the MSD algorithm the optimal gain is calculated:

$$g_{opt}^{(k-1)} = \frac{\bar{\mathbf{z}}^{(k-1)t}\,\mathbf{R}^{(k-1)}\,\bar{\mathbf{x}}^{(k-1)}}{\bar{\mathbf{x}}^{(k-1)t}\,\mathbf{R}^{(k-1)t}\,\mathbf{R}^{(k-1)}\,\bar{\mathbf{x}}^{(k-1)}} \tag{9}$$

At the kth level, the partial difference equals:

$$\begin{aligned}\bar{\boldsymbol{\varepsilon}}^{(k)} &= \left|\begin{matrix}\varepsilon_{N-k+1}\\ \bar{\boldsymbol{\varepsilon}}^{(k-1)}\end{matrix}\right| = \bar{\mathbf{z}}^{(k)} - g\,\mathbf{R}^{(k)}\bar{\mathbf{x}}^{(k)} = \\ &= \left|\begin{matrix}z_{N-k+1}\\ \bar{\mathbf{z}}^{(k-1)}\end{matrix}\right| - g\left|\begin{matrix}\bar{\mathbf{r}}_{N-k+1}\\ \bar{\mathbf{0}}\,\mathbf{R}^{(k-1)}\end{matrix}\right| \left|\begin{matrix}x_{N-k+1}\\ \bar{\mathbf{x}}^{(k-1)}\end{matrix}\right|\end{aligned} \tag{10}$$

yielding

$$||\bar{\boldsymbol{\varepsilon}}^{(k)}||^2 = f^{(k)}(g) = \left(z_{N-k+1} - g\bar{\mathbf{r}}_{N-k+1}\bar{\mathbf{x}}^{(k)}\right)^2 + f^{(k-1)}(g) \tag{11}$$

For $g = g_{opt}^{(k-1)}$ the second term attains its minimum, but generally, gain used at k^{th} level, i.e. $g = g_{opt}^{(k)}$ differs from $g_{opt}^{(k-1)}$. Because (11) holds for any gain, the following inequality is true: $f^{(k)}(g_{opt}^{k}) > f^{(k-1)}(g_{opt}^{k}) > f^{(k-1)}(g_{opt}^{k-1})$. Thus, the squared norm of the partial distance increases and the MSD algorithm is applicable for searching for the optimal scaling factor g and the optimal lattice point $\bar{\mathbf{x}}$ in the uniformly scaled lattice.

2.3. The sparse sphere decoding algorithm

In the above described algorithm, sparsity of $\bar{\mathbf{x}}$ is not demanded, but it may be attained if the alphabet contains a symbol equal to zero. Thus the optimal number of nonzero elements in $\bar{\mathbf{x}}$ is obtained by using SD and MSD algorithms. In order to obtain a K-sparse solution some constraints are added to the search procedure. The solution is updated and the radius R is reduced at the last (N^{th}) level, if $||\bar{\boldsymbol{\varepsilon}}^{(N)}||^2 < R^2$ and if sparsity (number of nonzero components of $\bar{\mathbf{x}}$) equals K. Moreover, the k^{th} level is entered, if $||\bar{\boldsymbol{\varepsilon}}^{(k-1)}||^2 \leq R^2$ and if the sparsity of $\bar{\mathbf{x}}^{(k-1)}$ does not exceed K. The second condition yields additional reduction of computational complexity, which has been noted in [13].

Further reduction may be obtained by removing the null symbol from the alphabet and testing combinations of columns of **R** in such a way that the best K-combination is not omitted. Complexity reduction of the proposed sparse SD algorithm is owed to reduction of the alphabet.

At the k^{th} level ($k=1,\ldots,K$) of the sparse SD algorithm the columns $\bar{\mathbf{r}}^{j(k-1)},\bar{\mathbf{r}}^{j(k-2)},\ldots,\bar{\mathbf{r}}^{j(1)}$ are fixed and the column $\bar{\mathbf{r}}^{j(k)}$ is searched together with the corresponding symbol $x_{j(k)}$ and (in Case 2) the common gain g. If the partial distance $\|\bar{\boldsymbol{\varepsilon}}^{k}\| \le R$, then the algorithm passes immediately to the next level (*depth first* approach). If $\|\bar{\boldsymbol{\varepsilon}}^{k}\| > R$ for any possible position $j(k)$, symbol $x_{j(k)}$ and gain g, then the algorithm continues searching at the previous level.

Combinations $j(1) > j(2) > \cdots > j(K)$ are generated in K nested loops. In the outer loop $\bar{\mathbf{r}}^{j(1)}$ is placed from $j(1)=N$ to $j(1)=K$. The possible positions of $\bar{\mathbf{r}}^{j(2)}$ range from $j(2)=j(1)-1$ to $j(2)=K-1$, etc. Testing occurs from the right to the left, as in standard SD algorithms.

An SD algorithm may be used if the partial distance at level k is always greater than the partial distance at level k-1. First it will be shown for Case 1. At level k-1 the partial solution of problem (4), reduced to $j(k$-1) dimensions, yields the difference vector $\bar{\boldsymbol{\varepsilon}}^{(k-1)}$ (5). At the next level this vector equals (Fig.2):

$$\bar{\boldsymbol{\varepsilon}}^{(k)}=\begin{vmatrix}\Delta\bar{\boldsymbol{\varepsilon}}^{(k)}\\ \bar{\boldsymbol{\varepsilon}}^{(k-1)}\end{vmatrix}=\bar{\mathbf{z}}^{(k)}-\mathbf{R}^{(k)}\bar{\mathbf{x}}^{(k)}=$$
$$=\begin{vmatrix}\Delta\bar{\mathbf{z}}^{(k)}\\ \bar{\mathbf{z}}^{(k-1)}\end{vmatrix}-\begin{vmatrix}\Delta\mathbf{R}^{(k)}\\ \mathbf{0}\,\mathbf{R}^{(k-1)}\end{vmatrix}\begin{vmatrix}\Delta\bar{\mathbf{x}}^{(k)}\\ \bar{\mathbf{x}}^{(k-1)}\end{vmatrix}=\begin{vmatrix}\Delta\bar{\mathbf{z}}^{(k)}-\Delta\mathbf{R}^{(k)}\bar{\mathbf{x}}^{(k)}\\ \bar{\mathbf{z}}^{(k-1)}-\mathbf{R}^{(k-1)}\bar{\mathbf{x}}^{(k-1)}\end{vmatrix} \quad (12)$$

where $\mathbf{0}$ – null matrix.

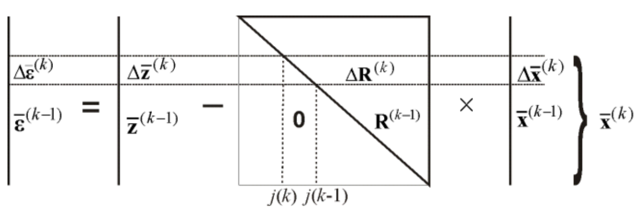

Fig.2. Evolution of the accumulated distance in sparse SD algorithm

Similarly to (7), the squared partial distance equals

$$\|\bar{\boldsymbol{\varepsilon}}^{(k)}\|^2=\left\|\Delta\bar{\mathbf{z}}^{(k)}-\Delta\mathbf{R}^{(k)}\,\bar{\mathbf{x}}^{(k)}\right\|^2+\|\bar{\boldsymbol{\varepsilon}}^{(k-1)}\|^2 \quad (13)$$

Since $\|\bar{\boldsymbol{\varepsilon}}^{(k)}\|^2 > \|\bar{\boldsymbol{\varepsilon}}^{(k-1)}\|^2$, the k^{th} level may be abandoned if $\|\bar{\boldsymbol{\varepsilon}}^{(k-1)}\|^2 > R^2$. The same holds for Case 2 (the proof is almost identical to that presented in Section 2.2) and the sparse SD algorithm yielding the optimal solution of problem (2) will be called the sparse modified sphere decoding algorithm (SMSD).

The proposed SMSD algorithm may be programmed in the form of a self-calling procedure, similarly to the standard SD algorithm [15]: SMSD(k) where k is a level. At the first call k=1 and R is an initial radius, obtained with some suboptimal search procedure.

For Case 2 and $k=1,2,\ldots,K-1$ two nested loops are needed:

Outer loop: $j(k)=j(k-1)-1,\ldots,K+1-k$ ($j(0)=N+1$)

  Inner loop: for all nonzero symbols $x_{j(k)}$ (e.g. ± 1)

    $\bar{\mathbf{v}}^{(k)}=\mathbf{R}^{(k)}\bar{\mathbf{x}}^{(k)}$, ($\bar{\mathbf{x}}^{(k)}$ is a k-sparse vector)

    $g_{opt}^{(k)}=\dfrac{\bar{\mathbf{z}}^{(k)t}\,\bar{\mathbf{v}}^{(k)}}{\bar{\mathbf{v}}^{(k)t}\,\bar{\mathbf{v}}^{(k)}}$, $\|\bar{\boldsymbol{\varepsilon}}^{(k)}\|^2=\left\|\bar{\mathbf{z}}^{(k)}-g_{opt}^{(k)}\,\bar{\mathbf{v}}^{(k)}\right\|^2$

    If $\|\bar{\boldsymbol{\varepsilon}}^{(k)}\|^2 \le R^2$ call SMSD(k+1)

  End

  If, for $x_{j(k)}=0$, $\|\bar{\boldsymbol{\varepsilon}}^{(k)}\|^2 > R^2$ break the outer loop

End

The last condition stems from the fact that the partial distance increases with increasing vector dimension.

At the last level ($k=K$) the N-dimensional problem is considered:

Outer loop: $j(K)=j(K-1)-1,\ldots,1$

  Inner loop: for all nonzero symbols $x_{j(K)}$ (e.g. ± 1)

    $\bar{\mathbf{v}}=\mathbf{R}\,\bar{\mathbf{x}}$, ($\bar{\mathbf{x}}$ is an N-dim K-sparse vector)

    $g_{opt}=\dfrac{\bar{\mathbf{z}}^{t}\,\bar{\mathbf{v}}}{\bar{\mathbf{v}}^{t}\,\bar{\mathbf{v}}}$, $\|\bar{\boldsymbol{\varepsilon}}\|^2=\left\|\bar{\mathbf{z}}-g_{opt}\,\bar{\mathbf{v}}\right\|^2$

    If $\|\bar{\boldsymbol{\varepsilon}}\|^2 \le R^2$: $\bar{\mathbf{x}}^{out}=\bar{\mathbf{x}}$, $R=\|\bar{\boldsymbol{\varepsilon}}\|$, $g^{out}=g_{opt}$

  End

  If, for $x_{j(K)}=0$, $\|\bar{\boldsymbol{\varepsilon}}^{(K)}\|^2 > R^2$ break the outer loop

End

Note that, in general, $\bar{\boldsymbol{\varepsilon}}^{(K)}$ is not an N-dimensional vector, so $\|\bar{\boldsymbol{\varepsilon}}\|^2 \ge \|\bar{\boldsymbol{\varepsilon}}^{(K)}\|^2$. The SMSD algorithm returns the optimal vector $\bar{\mathbf{x}}^{out}$ and the optimal gain g^{out} in Case 2. Extension to Case 1 is straightforward (gain g=1), and Case 3 is analyzed below.

2.4. The sparse sphere decoding algorithm and non-uniformly scaled lattice

The sparse sphere decoding algorithm may be also applied to solve the "classic" sparse approximation problems (3). After a QR decomposition of $\mathbf{F}$, problem (3) is transformed to

$$\min_{\bar{\mathbf{g}}}\|\bar{\mathbf{z}}-\mathbf{R}\,\bar{\mathbf{g}}\|^2 \quad (14)$$

where $\bar{\mathbf{g}}\in\Re^N$ is a K-sparse vector.

First it must be proved that the partial distance increases at subsequent levels. At level k-1, the problem is reduced to $N-j(k-1)+1$ dimensions (Fig.2): vector $\bar{\mathbf{z}}^{(k-1)}$ is approximated as a linear combination of k-1 columns of matrix $\mathbf{R}^{(k-1)}$: $\bar{\mathbf{z}}^{(k-1)}\approx\mathbf{R}^{(k-1)}\bar{\mathbf{g}}^{(k-1)}$, where $\bar{\mathbf{g}}^{(k-1)}$ is a k-1-sparse vector. By extraction of these columns matrix $\widetilde{\mathbf{R}}^{(k-1)}$ is obtained, yielding $\bar{\mathbf{z}}^{(k-1)}\approx\widetilde{\mathbf{R}}^{(k-1)}\widetilde{\mathbf{g}}^{(k-1)}$, where $\widetilde{\mathbf{g}}^{(k-1)}$ is k-1-

dimensional, dense vector of gains. The error of this approximation is a function of $\tilde{\mathbf{g}}^{(k-1)}$:

$$||\bar{\boldsymbol{\varepsilon}}^{(k-1)}||^2 = ||\bar{\mathbf{z}}^{(k-1)} - \tilde{\mathbf{R}}^{(k-1)}\tilde{\mathbf{g}}^{(k-1)}||^2 = f^{(k-1)}(\tilde{\mathbf{g}}^{(k-1)}) \quad (15)$$

The optimal vector of gains equals:

$$\tilde{\mathbf{g}}_{opt}^{(k-1)} = [\tilde{\mathbf{R}}^{(k-1)t}\tilde{\mathbf{R}}^{(k-1)}]^{-1}\tilde{\mathbf{R}}^{(k-1)t}\bar{\mathbf{z}}^{(k-1)} \quad (16)$$

At the next level the partial difference vector equals:

$$\bar{\boldsymbol{\varepsilon}}^{(k)} = \begin{vmatrix}\Delta\bar{\boldsymbol{\varepsilon}}^{(k)} \\ \bar{\boldsymbol{\varepsilon}}^{(k-1)}\end{vmatrix} = \bar{\mathbf{z}}^{(k)} - \tilde{\mathbf{R}}^{(k)}\tilde{\mathbf{g}}^{(k)} = \\ = \begin{vmatrix}\Delta\bar{\mathbf{z}}^{(k)} \\ \bar{\mathbf{z}}^{(k-1)}\end{vmatrix} - \begin{vmatrix}\Delta\tilde{\mathbf{R}}^{(k)} \\ \bar{\mathbf{0}}\,\tilde{\mathbf{R}}^{(k-1)}\end{vmatrix}\begin{vmatrix}g_{j(k)} \\ \tilde{\mathbf{g}}^{(k-1)}\end{vmatrix} = \begin{vmatrix}\Delta\bar{\mathbf{z}}^{(k)} - \Delta\tilde{\mathbf{R}}^{(k)}\tilde{\mathbf{g}}^{(k)} \\ \bar{\mathbf{z}}^{(k-1)} - \tilde{\mathbf{R}}^{(k-1)}\tilde{\mathbf{g}}^{(k-1)}\end{vmatrix} \quad (17)$$

where $\bar{\mathbf{0}}$ - a column of $N - j(k-1)+1$ zeroes, $\Delta\tilde{\mathbf{R}}^{(k)}$ - matrix of dimensions $j(k-1) - j(k)+1 \times k$. Therefore, for $\tilde{\mathbf{g}}^{(k)} = \begin{vmatrix}g_{j(k)} \\ \tilde{\mathbf{g}}^{(k-1)}\end{vmatrix}$, the norm $||\bar{\boldsymbol{\varepsilon}}^{(k)}||^2 = f^{(k)}(\tilde{\mathbf{g}}^{(k)})$ equals

$$f^{(k)}(\tilde{\mathbf{g}}^{(k)}) = \left\|\Delta\bar{\mathbf{z}}^{(k)} - \Delta\tilde{\mathbf{R}}^{(k)}\tilde{\mathbf{g}}^{(k)}\right\|^2 + f^{(k-1)}(\tilde{\mathbf{g}}^{(k-1)}) \quad (18)$$

For the optimal gains, i.e. $\tilde{\mathbf{g}}_{opt}^{(k)} = \begin{vmatrix}g_{j(k),opt} \\ \tilde{\mathbf{g}}_{opt}^{(k-1)} + \Delta\tilde{\mathbf{g}}\end{vmatrix}$, $f^{(k)}(\tilde{\mathbf{g}}_{opt}^{(k)}) \geq f^{(k-1)}(\tilde{\mathbf{g}}_{opt}^{(k-1)})$ because of the first term of (18) and because the last k-1 components of $\tilde{\mathbf{g}}_{opt}^{(k)}$ are not necessarily equal to $\tilde{\mathbf{g}}_{opt}^{(k-1)}$. This justifies use of the SD algorithm.

The corresponding self-calling procedure (SMSD(k)) does not differ much from that described in Section 2.3. The main difference is a suppression of the inner loop, testing a set of symbols $x_{j(k)}$. At each level, the optimal gain vector $\tilde{\mathbf{g}}^{(k)}$ is calculated, similarly to (16), and the squared partial distance $||\bar{\boldsymbol{\varepsilon}}^{(k)}||^2$ is evaluated. If $||\bar{\boldsymbol{\varepsilon}}^{(k)}||^2 \leq R^2$, the level is increased, if not, the outer loop is broken. This does not concern the last level, in which N-dimensional system is tested. Here, if $||\bar{\boldsymbol{\varepsilon}}||^2 \leq R^2$, the radius R is updated, and the indices $j(1), j(2), \ldots, j(K)$ and gains $\tilde{\mathbf{g}}_{opt}^{(K)}$ are stored.

3. COMPARATIVE EVALUATION

In this section, the proposed SMSD algorithms are compared with full codebook search and with some suboptimal algorithms. Two sparse signal models are used: they are described in equations (2) and (3). In accordance with the codebook **F** and the target vector $\bar{\mathbf{y}}$ two examples are analyzed.

- **Synthetic signals:** $\bar{\mathbf{y}}$ is a vector of dimension N=20, obtained by passing a Gaussian noise through an AR filter $H(z) = \frac{1}{1-2\rho\cos\varphi z^{-1} + \rho^2 z^{-2}}$, where $\rho = 0.98$ and $\varphi = \pi/16$. The columns of **F** are obtained by passing single pulses through the same filter, i.e. they contain shifted impulse response of $H(z)$. This yields a coherent codebook, making search procedures difficult [16].
- **Speech signals** processed in a nonstandard low-delay CELP coder. In this case $\bar{\mathbf{y}}$ is a perceptual signal vector (filtered speech) of dimension N=16 (delay 2 ms at sampling frequency 8 kHz), and **F** is obtained as before, but $H(z)$ is an adaptive predictive synthesis filter. The gain coefficient g is not quantized, in order to compare only the sparse approximation algorithms. At the end, however, the fully quantized coder is simulated, transmitting speech at 13.5 kbit/s.

Suboptimal algorithms for solving the uniformly scaled lattice search problem (2) are compared in [4]. These algorithms may deliver a starting point (the initial radius R) for SMSD algorithms. The following algorithms are considered.

- **Sparsity-forcing**: calculation of a dense solution $\bar{\mathbf{g}} = \mathbf{F}^{-1}\bar{\mathbf{y}}$ and choice of K components of greatest absolute values.
- **Minimum angle**: a simple greedy algorithm, selecting K codebook vectors in K steps by minimizing the angle between the target vector $\bar{\mathbf{y}}$ a and its model $\mathbf{F}\bar{\mathbf{g}}$ [17].
- **Global replacement**: the initial solution is found as above, then all vectors, one by one, are replaced by others, if such exchange yields reduction of error [2].
- **M-best** implementation of minimum angle algorithm. The M-best algorithm calculates, in a parallel way, M sequences of codebook vectors (here, $M=10$). At the end the best sequence is retained [4].
- **M-best + replacement**: an M-best algorithm is executed, and then global replacement is performed.

Suboptimal algorithms for solving the non-uniformly scaled lattice search problem (3) are compared in [3, 12]. Here the following algorithms are considered.

- **Sparsity-forcing**: indices $j(1), j(2), \ldots, j(K)$ are chosen as above, and then gains are calculated as in (16).
- **OOMP**: (optimized orthogonal matching pursuit) [10-12]. Here the fast implementation of this algorithm is used, namely the RMGS (recursive modified Gram-Schmidt) [12].
- **M-best** implementation of the OOMP algorithm.

Results for problem (2) and synthetic signals show considerable reduction of computational effort (Tab.1): SMSD visits less than 0.6% of nodes tested with full search approach, yielding the same optimal result. Further reduction may be obtained if gain is forced to be positive (SMSD g>0), but in some cases the optimal solution is skipped. The other suboptimal search algorithms yield much worse results.

The SMSD algorithm is much more efficient than the full search approach in solving problem (3) (Tab.2). Similar conclusions stem from simulations of the LD-CELP coder without gain quantization (in Fig.3 and Fig.4 mean SNR values for four phrases and more than 10000 segments are given). Using signal model (2) with K=10 and 4-bit predictive gain quantizer, we obtain a bit rate of 13.5 kbit/s. Comparing mean opinion score values for 10 speech phrases, we observe a systematic advantage of the SMSD algorithm over the other ones.

Algorithm	SNR [dB]	nodes tested	% of full search
Full search	25.187	32.25 10^6	100
MSD	25.187	588 10^3	1.82
SMSD/sparsity-forcing	25.187	191 10^3	0.59
SMSD/min.angle	25.187	189 10^3	0.59
SMSD/M-best	25.187	181 10^3	0.56
SMSD g>0	25.11	51 10^3	0.16
M-best+replacement	20.31	3.3 10^3	0.01
Global replacement	19.67	810	0.002
M-best (M=10)	18.68	2.8 10^3	0.008
Minimum angle	15.42	288	0.001
Sparsity-forcing	10.01	1	0.00003

Tab.1. Comparison of optimal and suboptimal algorithms in solving (2) for K=8, N=20 (mean values for 1000 runs). SMSD/min.angle means that the initial radius is obtained using minimum angle algorithm, etc. (confidence interval for SNR values: 0.15 dB)

Algorithm	SNR [dB]	nodes tested	% of full search
Full search	32.574	125970	100
SMSD/OOMP	32.574	9523	7.6
M-best	31.85	1240	1
OOMP	29.94	124	0.1
Sparsity-forcing	27.03	1	0.0008

Tab.2. Comparison of several sparse approximation algorithms in solving (3) for K=8, N=20 (mean values for 1000 runs, confidence interval for SNR: 0.15 dB)

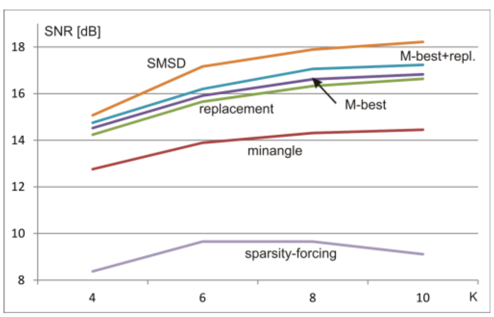

Fig.3. SNR for LD-CELP coder with signal model (2) - confidence interval for all SNR values: 0.05 dB

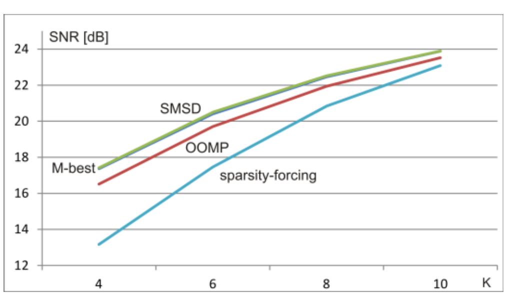

Fig.4. SNR for LD-CELP coder with signal model (3) - confidence interval for all SNR values: 0.05 dB

4. CONCLUSIONS

It is shown that the sphere decoding algorithm may be extended to uniformly or non-uniformly scaled lattices. The resulting modified SD and sparse modified SD algorithms yield the optimal solution of some sparse approximation problems at substantially reduced computational cost, compared with the full search approach. The proposed algorithms are compared with suboptimal solutions.

5. REFERENCES

[1] F. K. Chen, J. F. Yang, "Maximum take precedence ACELP – a low complexity search method", *Proc. ICASSP'01*, pp.693-696

[2] F. K. Chen, G. M. Chen, B. K. Su, Y. R. Tsai, "Unified pulse-replacement search algorithms for algebra codebooks of speech code", *IET Signal Proc.*, 2010, Vol. 4, Iss. 6, pp. 658-665

[3] P. Dymarski, R. Romaniuk, "Sparse Signal Approximation Algorithms in a CELP Coder", *Proc. EUSIPCO 2011*, pp. 201-205

[4] P.Dymarski, R.Romaniuk "Sparse signal modeling in a scalable CELP coder", *Proc. of EUSIPCO 2013*, Marrakesh, pp. 1-5

[5] A.D. Murugan, H. El Gamal, M.O. Damen, G. Caire, "A unified framework for tree search decoding: rediscovering the sequential decoder", *IEEE Trans. Information Theory*, Vol. 52, pp. 933-953, March 2006

[6] E. Viterbo, J. Boutros, "A Universal Lattice Code Decoder for Fading Channels", *IEEE Trans. on Information Theory*, Vol. 45, Iss. 5, July 1999, pp. 1639-1642

[7] T. Yardibi, J. Li, P. Stoica, L.N. Cattafesta III, "Sparse representations and sphere decoding for array signal processing", *Digital Signal Processing* 2011, doi:10.1016/j.dsp.2011.10.006

[8] E. Agrell, T. Eriksson, A. Vardy, K. Zeger, "Closest point search in lattices", *IEEE Transactions on Information Theory*, Vol. 48, Iss. 8, 2002, pp. 1620-1625

[9] P.Wang, T. Le-Ngoc, "Selection of Design Parameters for Generalized Sphere Decoding Algorithms", *Int. J. Communications, Network and System Sciences,* Vol. 3, No=2, 2010, pp. 126-132

[10] P. Kroon, E. Deprettere, "Experimental evaluation of different approaches to the multi-pulse coder", *Proc. ICASSP'84*, pp. 396-399

[11] L. Rebollo-Neira, D. Lowe, "Optimized orthogonal matching pursuit approach", *IEEE Sig. Proc. Letters*, Vol. 9/2002, pp.137-140.

[12] P. Dymarski, N. Moreau, G. Richard, "Greedy sparse decompositions: a comparative study", *Eurasip Journal on Advances in Signal Processing*, 2011:34, doi:10.1186/1687-6180-2011-34

[13] S. Barik, H. Vikalo "Sparsity-aware sphere decoding: algorithms and complexity" *IEEE Trans. on Signal Proc*. Vol. 62 , May 2014, pp.2212-2225

[14] S. Sparrer, R.F.H. Fischer "Discrete Sparse Signals: Compressed Sensing by Combining OMP and the Sphere Decoder", in Proceedings of CoRR, 2013, arXiv:1310.2456

[15] X. Guo "Sphere Decoder for MIMO systems", *Matlab Central, File Exchange*, Feb. 2009

[16] Z. Ben-Haim, Y. C. Eldar, M. Elad "Coherence-Based Performance Guarantees for Estimating a Sparse Vector Under Random Noise", *IEEE Trans. on Signal Proc.*, Vol. 58, Iss. 10, Oct. 2010, pp. 5030-5043

[17] P.Dymarski, N.Moreau, "Algorithms for the CELP coder with ternary excitation", *Proc. Eurospeech 1993*, pp.241-244.