

Wideband LSF Quantization by Generalized Voronoi Codes

Stéphane Ragot, Hassan Lahdili, and Roch Lefebvre

Department of Electrical and Computer Engineering, University of Sherbrooke
Sherbrooke, QC, J1K 2R1, Canada

{ragot, lahdili, lefebvre}@gel.usherb.ca

Abstract

Presented a method for quantizing the wideband line spectrum frequencies (LSF) with a specific class of near-ellipsoidal lattice codes referred to as “generalized Voronoi codes”. Optimization procedures are described with respect to a weighted mean-square error (WMSE). The lattices D_{16} , RE_{16} or RA_{16} are applied to quantize the LSF with no frequency splitting. Results indicate that near-ellipsoidal lattice quantization allows to develop efficient one-stage algebraic wideband LSF quantization at competitive bit rates.

1. Introduction

We address the problem of designing an efficient memoryless LSF quantization scheme for wideband linear predictive coding. In practice this problem is typically “solved” by employing constrained stochastic vector quantization (VQ) – e.g. split/multistage VQ. In this paper we propose an alternative technique using near-ellipsoidal lattice VQ. The proposed coding system generalizes some unpublished results on narrowband LSF quantization [1] and it is similar in some respect to a narrowband technique described in [2].

The application of lattice VQ to wideband LSF quantization is mainly motivated by:

- the need in spectrum coding to optimize performance while minimizing both computational complexity and storage requirements [3, 4],
- the intrinsic scalability of lattice VQ which allows to trade bit rate and spectral quantization quality if variable bit rate allocation is allowed [4] – this is even more important if a linear predictive coder is to be applied not only to speech but also to music signals,
- the possibility to share lattice quantization routines and tables with algebraic transform coding.

A two-stage VQ-near-spherical lattice VQ configuration was proposed in [3, 4] with these motivations in mind. Although this technique is competitive in terms of performance/complexity trade-off, it resulted in quite inefficient codebooks because the LSF source is non-uniform and has a specific geometry due to LSF ordering. To understand its inherent structural constraints the two-stage coding scheme of [3] is sketched in Figure 1 (a) as fixed ball packing (i.e. duplicating the fixed near-spherical lattice stage to represent the source distribution). In particular the structural limitations are clear in the neighborhood of the LSF stability region and in sparse regions which are not adequately covered and where there may be “holes” in between the balls. These limitations could be mitigated by employing a

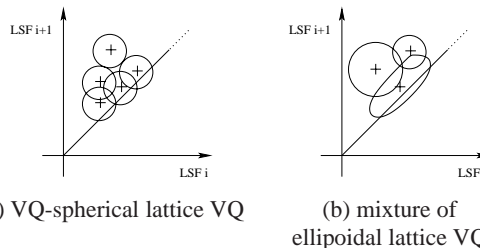


Figure 1: Ball packing as in [3, 4] vs packing of ellipsoids.

pre-processing of the LSF source such as a temporal prediction prior to quantization, or by adding some flexibility to the near-spherical lattice VQ stage – e.g. a scale factor dependent on the codevectors selected in the VQ stage. In this paper we investigate another alternative as illustrated in Figure 1 (b) : using near-ellipsoidal lattice VQ instead of fixed near-spherical lattice VQ.

The paper is organized as follows. First we define clearly the proposed coding principle in Section 2. This principle is actually very similar to that of [2], except that we apply a different system design and we use generalized Voronoi coding. The idea behind the algebraic structures used herein was originally developed in [1] based on the lattices A_2 , D_n and $\frac{1}{2}RE_8$ to quantize the narrowband LSF with one near-ellipsoidal lattice code ; for the sake of completeness we describe the definition of generalized Voronoi VQ as well as the required indexing algorithms in Section 3, and present a fast (suboptimal) nearest-neighbor search algorithm taken from [5]. Source optimization for the proposed quantization scheme is considered in Section 4. This problem is only studied in the framework of source coding based on parametric PDF estimation (as in [2, 6]). Results are presented in Section 5, before concluding in Section 6.

2. Coding Principle

The proposed coding system is depicted in Figure 2. An arbitrary LSF vector ω is quantized by selecting the best reconstruction among KL candidates $\{\hat{\omega}_{kl}\}_{1 \leq k \leq K, 1 \leq l \leq L}$ with respect to WMSE. In the general case, the candidates are found by a sequence of operations sketched in Figure 2 (a). This can be summarized as follows:

$$\hat{\omega}_{kl} = T_k^{-1} \underbrace{g_{kl} Q_k [1/g_{kl} T_k (\omega - \mu_k)]}_{\hat{\omega}_{kl}} + \mu_k \quad (1)$$

The μ_k are appropriate points in the LSF space. The transforms T_k are unitary rotation matrices. The Q_k implement near-ellipsoidal lattice quantization. The scalar gains $g_{kl} > 0$ are

This work was supported by the NSERC and VoiceAge Corp.



lattice codebook scale factors. We can assume without loss of generality that $g_{k1} < \dots < g_{kL}$. There are K different lattice codebooks, each of which is scaled with L different values. Hence the overall quantization codebook consists of a collection of K near-ellipsoidal codes scaled by L possible values g_{kl} , centered around μ_k and rotated by T_k^{-1} with $1 \leq k \leq K$.

To reduce complexity, we can use a pre-selection $\tilde{\Omega}_k$ among the set $\{\tilde{\Omega}_{kl}\}_{1 \leq l \leq L}$ prior to reconstruction, as described in Figure 2 (c) – otherwise for optimal performance we use the full-search procedure described in Figure 2 (b).

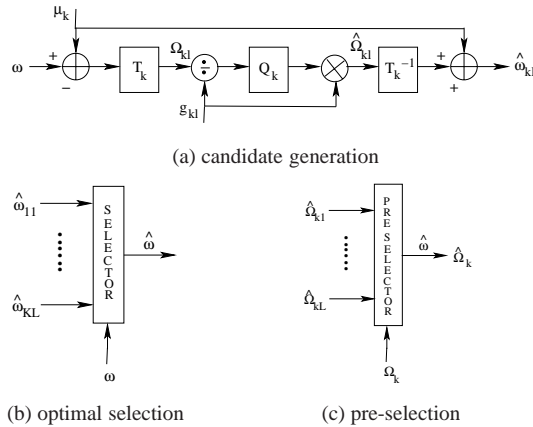


Figure 2: Coding principle.

For reference and to extend results presented in [1] we also consider in this work the special case where only one near-ellipsoidal code is applied (i.e. $K = 1$). Note that the coding principle described herein could also be applied to another source than LSF, and in such variants of VQ as split/multistage VQ, or intra-/inter-frame predictive VQ. However, we consider in this paper a direct memoryless LSF quantization with no frequency splitting.

3. Generalized Voronoi Quantization

To describe precisely generalized Voronoi quantization we need the following notations. The vector operators mod , mult and div denote element-by-element modulo, multiplication and division, respectively. The generator matrix G_Λ of a lattice Λ in n dimensions has basis vectors as its rows. Therefore $\mathbf{y} = \mathbf{k}G_\Lambda$ generates a lattice point, while $\mathbf{j} = \mathbf{y}G_\Lambda^{-1}$ retrieves the related basis expansion. The region V_Λ corresponds to the Voronoi region of Λ relative to the origin $\mathbf{0}$. This region is approximately spherical – especially as $n \rightarrow \infty$. If V_Λ is scaled according to a factor $\mathbf{m} = (m_1, \dots, m_n)$ in \mathbb{N}^{*n} , we obtain a near-ellipsoidal region $V_\Lambda(\mathbf{m}) = \text{mult}(V_\Lambda, \mathbf{m})$.

Since the linear prediction order for wideband signals (sampled at 16 kHz) is usually $n = 16$, we use the integer lattices D_{16} , E_{16} and RA_{16} defined below :

$$D_{16} = \left\{ (x_1, \dots, x_{16}) \in \mathbb{Z}^{16} : \sum_{i=1}^{16} x_i \text{ is even} \right\} \quad (2)$$

$$RE_{16} = 2D_{16} \cup \{2D_{16} + [1 \dots 1]\} \quad (3)$$

$$RA_{16} = \bigcup_{c \in \mathcal{R}(1,4)} 2D_{16} + c \quad (4)$$

In the definition of RA_{16} , $\mathcal{R}(1, 4)$ is the binary $[16, 5, 8]$ Reed-Muller code of order 1 with weight distribution $0^1 8^{30} 16^1$.

3.1. Definition

Generalized Voronoi codes are index-optimized near-ellipsoidal lattice codes. They were originally introduced in [1]. They consist of a subset of a lattice obtained after lattice truncation by a Voronoi region $V_\Lambda(\mathbf{m})$ with different scaling per dimension. Formally, they are defined as the codes $C_\Lambda(\mathbf{m}, \mathbf{a}) = \Lambda \cap (V_\Lambda(\mathbf{m}) + \mathbf{a})$ where \mathbf{a} is an appropriate offset. The code size is $\prod_{i=1}^n m_i$. An example is shown in Figure 3 where $\Lambda = A_2$.

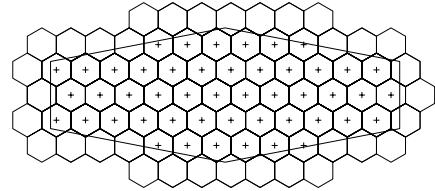


Figure 3: Example of generalized Voronoi codebook derived from the A_2 lattice.

Interestingly Voronoi codes – defined in [7] – are found as a subclass: they correspond to the case $\mathbf{m} = 2^r [1 \dots 1]$ with $r > 0$.

3.2. Fast indexing algorithms

Indexing algorithms for generalized Voronoi codes derived from the lattices A_2 , D_n and $\frac{1}{2}RE_8$ are presented in [1]. Their common framework is summarized in Figure 4. Contrary to [1], we have incorporated an offset \mathbf{a} in the description of the inverse mapping.

Forward mapping: codevector $\mathbf{y} \rightarrow$ index \mathbf{i}

1. Compute $\mathbf{j} = \mathbf{y}G^{-1}$
2. Modify conditionally \mathbf{j}
3. Compute $\mathbf{i} = \text{mod}(\mathbf{j}, \mathbf{m})$

Inverse mapping: index $\mathbf{i} \rightarrow$ codevector \mathbf{y}

1. Compute $\mathbf{u} = \mathbf{k}G$
2. Compute $\mathbf{z} = \text{div}(\mathbf{u} - \mathbf{a}, \mathbf{m})$
3. Find the nearest neighbor \mathbf{v} of \mathbf{z} in Λ
4. Compute $\mathbf{y} = \mathbf{u} - \text{mult}(\mathbf{m}, \mathbf{v})$

Figure 4: Generalized Voronoi indexing.

The index \mathbf{i} is actually not an integer but a vector of integers. More details regarding the choice of \mathbf{a} and the conditional modification of \mathbf{j} can be found in [1, 5].

3.3. Fast (suboptimal) codebook search algorithm

The nearest-neighbor search in a lattice codebook generated by a convex shaping region is typically broken into 2 steps :

1. a search in the infinite lattice Λ (also known as “lattice decoding”),
2. outlier saturation if an overload is detected.

The outlier saturation depends on the truncation used to generate the lattice code. We present in Figure 5 a fast algorithm in the case of a generalized Voronoi code $V_\Lambda(\mathbf{m}, \mathbf{a})$.

The underlying principle consists of saturating an outlier \mathbf{x} by iterative scaling until the nearest neighbor of \mathbf{x} in Λ falls in



Nearest neighbor search: point $\mathbf{x} \rightarrow$ codevector \mathbf{y}

1. Find the nearest neighbor \mathbf{z} of \mathbf{x} in Λ
2. If \mathbf{z} is a codevector, $\mathbf{y} = \mathbf{z}$ and stop.
3. Project \mathbf{x} on $\mathbf{h} \in \{V_\Lambda(\mathbf{0}, \mathbf{m}) + \mathbf{a}\}$
(alternative used herein: set directly $\mathbf{h} = \beta(\mathbf{h} - \mathbf{a}) + \mathbf{a}$)
4. Find the nearest neighbor \mathbf{z} of \mathbf{h} in Λ
5. If \mathbf{z} is a codevector, $\mathbf{y} = \mathbf{z}$ and stop.
6. Otherwise scale \mathbf{h} by a factor β : $\mathbf{h} = \beta(\mathbf{h} - \mathbf{a}) + \mathbf{a}$.
Goto 4.

Figure 5: Suboptimal search in a generalized Voronoi code.

the codebook $V_\Lambda(\mathbf{m}, \mathbf{a})$. To reduce the worst-case complexity (i.e. avoid unnecessary loops), a detected outlier \mathbf{x} is pre-processed: \mathbf{x} is projected onto the relevant Voronoi facet of $\{V_\Lambda(\mathbf{0}, \mathbf{m}) + \mathbf{a}\}$. The scale $0 < \beta < 1$ ensures to saturate any outlier in the codebook.

4. Sequential Design Procedures

The design problem for the proposed quantization system consists of optimizing jointly all parameters μ_k , T_k , Q_k and g_{kl} . For simplicity, we apply a sequential design by selecting first μ_k , T_k and Q_k in open-loop, and then optimizing the g_{kl} in closed-loop. We discuss this design problem considering the cases $K = 1$ and $K > 1$ separately.

The optimization target is the WMSE criterion estimated in practice as:

$$\bar{d}_{wmse} = \frac{1}{|\mathcal{B}|} \sum_{\omega \in \mathcal{B}} \sum_{i=1}^n s_i(\omega) (\omega_i - \hat{\omega}_i)^2 \quad (5)$$

where \mathcal{B} denotes the LSF database, $|\cdot|$ the set cardinality, and the weighting factor $s_i(\omega)$ represents the ‘‘spectral sensitivity’’ of the i th LSF component ω_i .

4.1. Quantization by $K = 1$ near-ellipsoidal code

The design problem for $K = 1$ was first considered in [1] in the case of narrowband LSF quantization – however in [1] automatic optimization of the Voronoi modulo and closed-loop optimization of the codebook scale factors were not addressed. To simplify the notation in this specific case, the subscript k will be dropped.

Under the assumption that the LSF source is a stationary and ergodic process, μ can be set as the mean of the LSF database, and T can be estimated through a singular value decomposition (SVD) of the LSF covariance matrix Γ . The SVD also gives the eigenvalues $\sigma_1^2 > \dots > \sigma_n^2$. These eigenvalues correspond to the variances in the adequate principal axes of the ellipsoid related to Γ .

Optimization of \mathbf{m} :

The optimization of Q amounts to the selection of a modulo vector $\mathbf{m} = (m_1, \dots, m_n)$. This can be done based on the value of the eigenvalues σ_i^2 . The goal is to find the optimal values m_i , subject to the constraint $\prod m_i \leq R$ where R is a specific bit budget. To come up with an automatic selection procedure, we can use an integer-constrained allocation based on the intuitive rule: the number of quantization levels of a quantizer (say, m_i) should be proportional to the standard deviation (say, σ_i). Note that the allocation depends on the lattice Λ and shall verify the constraints on \mathbf{m} . An example is given in Figure 6 for $\Lambda = D_{16}$.

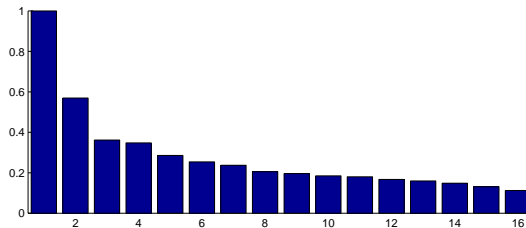


Figure 6: Distribution of normalized standard deviations (σ_i) for the wideband LSF source used herein. For a budget of $R = 47$ bits, the corresponding optimized Voronoi modulo for D_{16} is $\mathbf{m} = [34, 18, 12, 12, 10, 10, 8, 6, 6, 6, 6, 6, 6, 4, 4, 4, 4]$. The effective Voronoi code size is close to $2^{46.996}$.

Optimization of g_l :

A simple (joint) gradient search with respect to WMSE provides the optimal values of g_1, \dots, g_L . This can be done on a closed-loop fashion based on the LSF database and with an empirical estimation of the gradient and small search steps.

4.2. Quantization by $K > 1$ near-ellipsoidal codes

The open-loop part of the design procedure relies in this case on the E-M algorithm in a similar way as described in [2] and extends the case $K = 1$. The source database is used to train a parametric PDF model (namely, a Gaussian mixture model). This procedure comes up with the weights, the means and the covariance matrices of the Gaussian components. These outputs are depicted in 2-D in Figure 7 for $K = 8$.

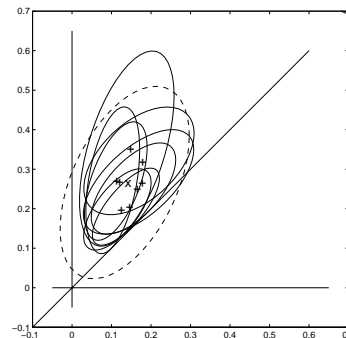


Figure 7: Projection in the (ω_1, ω_2) plane of the data models estimated by principal component analysis ($K = 1$, dashed lines and ‘ \times ’) and Gaussian mixture modelling ($K = 8$, solid lines and ‘ $+$ ’) – ω_1 and ω_2 are expressed in radians.

We set μ_k to the means derived by the E-M algorithm. After diagonalizing the covariance matrices, we obtain the rotation matrices T_k . We then could allocate bits based on the variances and the mixture weights as in [2] – to do so, we would have to neglect the overlap between components. In this paper, we assume that the weights are of comparable order of magnitude (i.e. K is small, typically $K = 8$), and we allocate the same bit budget R to each component. The modulo vectors \mathbf{m}_k specifying Q_k can be selected through the integer-constrained optimization designed in the case $K = 1$. And the gradient search developed for the case $K = 1$ can be easily extended to optimize jointly all the scale factors g_{kl} .



5. Application to LSF Quantization

5.1. Experimental setup

The setup for generating the LSF database is the same as in [8]. For this reason we omit here the very details regarding the fine tuning of the linear predictive analysis (i.e. pre-filtering, white noise correction, lag windowing). The database comprising 353,345 vectors was obtained by processing a predefined set of speech and music samples by the front end of a linear predictive wideband coder. The frame length is 20 ms with a 20 ms lookahead. A Hamming window of length 40 ms was used and located at the center of the current frame. A standard linear predictive analysis of fixed order 16 based on the autocorrelation method and Levinson Durbin recursion was applied.

The whole database was used in the design stage as well as for the performance characterization. Unstable filters were not rejected to measure the distortion statistics. Instead a LSF reordering with a 50 Hz gap was applied. The standard log-spectral distortion SD was estimated by a FFT of length 512, and considering frequencies in the range 50 Hz – 7 kHz.

5.2. Quantization results

Results for direct LSF quantization in 16 dimensions based on the lattices D_{16} , RE_{16} and $R\Lambda_{16}$ are shown in Table 1 for $K = 1$ and in Table 2 for $K = 8$. They were obtained by applying the MSE criterion only – for design and quantization. However, experiments showed that WMSE with appropriate weighting allows to reduce the amount of outliers while maintaining a good quantization quality.

For $K = 1$, we allocated $R = 47$ bits to the generalized Voronoi code. If L scale factors are used, the total bit budget is then $R + \log_2 L$. It turns out that $R\Lambda_{16}$ allows to save near 1 bit for the same quality compared to D_{16} , but RE_{16} gives the best performance/complexity trade-off.

Table 1: Log spectral distortion statistics for a single near-ellipsoidal code ($K = 1$ and $L = 1, 2, 4$).

L	rate (bits)	Λ	SD (dB)	SD 2-4 dB (%)	SD >4 dB (%)
1	47	D_{16}	1.14	3.51	0.48
		RE_{16}	1.04	2.75	0.36
		$R\Lambda_{16}$	1.00	2.08	0.25
2	48	D_{16}	0.94	1.90	0.16
		RE_{16}	0.86	1.28	0.13
		$R\Lambda_{16}$	0.82	0.93	0.10
4	49	D_{16}	0.83	1.15	0.08
		RE_{16}	0.77	0.98	0.02
		$R\Lambda_{16}$	0.74	0.71	0.004

For $K > 1$, storage requirements for the transformation matrices T_k typically limit the value of K . We set $K = 8$ and allocated $R = 40$ bits to each near-ellipsoidal lattice code. Hence if L scale factors are used, the total bit budget is then $R + \log_2(KL)$. Results show that we can save around 4 bits compared to $K = 1$ for the same quantization quality.

In consequence, the number of near-ellipsoidal codes and scale factors, the choice of lattice and the optimization target have a significant impact on the overall performance.

Table 2: Log spectral distortion statistics for 8 near-ellipsoidal codes ($K = 8$ and $L = 1, 2, 4$).

L	rate (bits)	Λ	SD (dB)	SD 2-4 dB (%)	SD >4 dB (%)
1	43	D_{16}	1.04	3.75	0.33
		RE_{16}	1.01	2.55	0.21
2	44	D_{16}	0.92	1.65	0.13
		RE_{16}	0.87	1.41	0.10

6. Conclusions

Results show that the wideband LSF source can be efficiently represented by generalized Voronoi codes. The use of several scale factors per code improves the performance significantly. For the experimental setup described herein, the shaping gain obtained when packing 8 ellipsoids instead of using a single ellipsoid can be estimated to 4 bits or so.

Several aspects were not completely considered in this paper. In particular, the same integer rate was allocated to each near-ellipsoidal code (contrary to [2]), and the proposed optimization target could be modified to take outliers explicitly into account. Perceptual aspects may also be studied.

Note that the proposed system can be interpreted as cell-conditioned residual VQ. In this respect, the centers μ_k can be viewed as forming a VQ codebook (the “1st stage”), while the rotated near-ellipsoidal codes correspond to a conditional residual VQ stage (the “2nd stage”). The interpretation to be valid implies some conditions on the “overlap” between the near-ellipsoidal codes, but it allows to reuse the iterative optimization procedure developed in [3, 5]. This would eliminate the mismatch between the open-loop source model and the distortion criteria.

7. Acknowledgements

The authors are grateful to Vesa T. Ruoppila for providing the wideband LSF database.

8. References

- [1] M. Xie, *Quantification vectorielle algébrique et codage de parole en bande élargie*, Ph.D. thesis, University of Sherbrooke, Quebec, Canada, Feb. 1996, (in French).
- [2] A.D. Subramanian and B.D. Rao, “PDF optimized parametric vector quantization of speech line spectral frequencies,” IEEE Workshop on Speech Coding, Delavvan, Wisconsin, U.S.A., Sept. 2000.
- [3] S. Ragot, J.-P. Adoul, R. Lefebvre, and R. Salami, “Low complexity LSF quantization for wideband speech coding,” IEEE Workshop on Speech Coding, Porvoo (Finland), June 1999.
- [4] S. Ragot, R. Lefebvre, R. Salami, and J.-P. Adoul, “Stochastic-algebraic wideband LSF quantization,” IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Istanbul, Turkey, June 2000.
- [5] S. Ragot, M. Xie, and R. Lefebvre, “Near-ellipsoidal lattice quantization by generalized Voronoi shaping,” Canadian Workshop on Information Theory, Vancouver, B.C., Canada, June 2001.
- [6] P. Hedelin and J. Skoglund, “Vector quantization based on Gaussian mixture models,” *IEEE Trans. on Speech and Audio Processing*, vol. 8, no. 4, July 2000.
- [7] J.H. Conway and N.J.A. Sloane, “A fast encoding method for lattice codes and quantizers,” *IEEE Trans. Inform. Th.*, vol. IT-29, no. 6, pp. 820–824, Nov. 1983.
- [8] V.T. Ruoppila and S. Ragot, “Index assignment for predictive wideband LSF quantization,” IEEE Workshop on Speech Coding, Delavan, Wisconsin, U.S.A., Sept. 2000.