Stable 1-norm error minimization based linear predictors for speech modeling

Daniele Giacobello, *Member, IEEE,* Mads Græsbøll Christensen, *Senior Member, IEEE,*
Manohar N. Murthi, *Member, IEEE,* Søren Holdt Jensen, *Senior Member, IEEE,* and Marc Moonen, *Fellow, IEEE*

***Abstract*—In linear prediction of speech, the 1-norm error minimization criterion has been shown to provide a valid alternative to the 2-norm minimization criterion. However, unlike 2-norm minimization, 1-norm minimization does not guarantee the stability of the corresponding all-pole filter and can generate saturations when this is used to synthesize speech. In this paper, we introduce two new methods to obtain intrinsically stable predictors with the 1-norm minimization. The first method is based on constraining the roots of the predictor to lie within the unit circle by reducing the numerical range of the shift operator associated with the particular prediction problem considered. The second method uses the alternative Cauchy bound to impose a constraint on the predictor in the 1-norm error minimization. These methods are compared with two existing methods: the Burg method, based on the 1-norm minimization of the forward and backward prediction error, and the iteratively reweighted 2-norm minimization known to converge to the 1-norm minimization with the appropriate selection of weights. The evaluation gives proof of the effectiveness of the new methods, performing as well as unconstrained 1-norm based linear prediction for modeling and coding speech.**

I. Introduction

Linear Prediction (LP) is widely used in a diverse range of speech modeling based algorithms, for instance in coding and recognition [1]. The traditional approach is to find the prediction coefficients by minimizing the 2-norm of the prediction error, i.e., the difference between the predicted and observed signal. This works well when the excitation signal is i.i.d. Gaussian [2], however, when this assumption is not satisfied, problems arise. This is the case for voiced speech where the pitch excitation is sparse and pulse-like. In this case, an alternative approach based on the 1-norm minimization of the prediction error has shown to offer a better modeling thanks to its ability to decouple the pitch excitation from the vocal tract transfer function [3].

Daniele Giacobello is with the Office of the CTO, Broadcom Corporation, Irvine, CA 92617, USA (e-mail: giacobello@broadcom.com).

Mads Græsbøll Christensen is with the Department of Architecture, Design, and Media Technology, Aalborg Universitet, 9220 Aalborg, Denmark (email: mgc@imi.aau.dk).

Manohar N. Murthi is with the Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL 33146, USA (e-mail: mmurthi@miami.edu).

Søren Holdt Jensen ia with the Department of Electronic Systems, Aalborg Universitet, 9220 Aalborg, Denmark (e-mail: shj@es.aau.dk).

Marc Moonen is with the Department of Electrical Engineering, KU Leuven, 3001 Leuven, Belgium (e-mail: marc.moonen@esat.kuleuven.be).

The work of Daniele Giacobello was supported by the Marie Curie EST-SIGNAL Fellowship (http://est-signal.i3s.unice.fr), contract no. MEST-CT-2005-021175 and was carried out at the Department of Electronic Systems, Aalborg Universitet, 9220 Aalborg, Denmark.

The improved modeling of 1-norm minimization also shows to be beneficial in speech coding. In particular, seeing the 1-norm as a convex relaxation of the 0-norm, the minimization process offers a residual that is sparser, providing tighter coupling between the multiple stages of time-domain speech coders, and thereby enabling a more efficient coding [4], [5]. Nevertheless, unlike those obtained through 2-norm minimization, the predictors obtained through 1-norm minimization are not intrinsically stable [6], [7] and, in coding applications for instance, having unstable filters may generate saturations in the synthesized speech.

The predictor stability problem in 1-norm LP has been tackled already in [7] by introducing the Burg method for all-pole parameters estimation based on 1-norm minimization of the forward and backward error. However, in this approach the sparsity is not preserved [3]. In this paper, we will introduce two novel methods to obtain intrinsically stable predictors with the 1-norm minimization. The first method is based on modifying the shift operator that generates the observation matrix from the analyzed speech segment [8], reducing the numerical range of this matrix. This one allows to restrict the zeros of the predictor polynomial to lie within the unitary circle. A similar approach has been used in weighted LP [9], [10] to modify the weighting function to guarantee stable solutions. The second method uses the alternative Cauchy bound [11], [12] to impose a constraint on the predictor in the 1-norm error minimization.

The paper is organized as follows. In Section II, we provide a brief review of LP based on the p-norm. In Section III, the core of the paper, we introduce our two new methods to obtain intrinsically stable predictors with the 1-norm minimization and also review the existing ones. In Section IV, we compare the spectral modeling and coding performances of the predictors. Finally, Section V concludes the paper.

II. Fundamentals of Linear Prediction

The problem considered in this paper is based on the following speech production model, where a sample of speech $x(n)$ at time n is written as a linear combination of K past samples:

$$x(n) = \sum_{k=1}^{K} a_k x(n-k) + e(n), \tag{1}$$

where $\{a_k\}$ are the coefficients of the predictor

$$A(z) = 1 + \sum_{k=1}^{K} a_k z^{-k}, \tag{2}$$

and $e(n)$ is the driving noise process (also referred to as prediction residual or excitation). The speech production model (1) in matrix form becomes:

$$\mathbf{x} = \mathbf{X}\mathbf{a} + \mathbf{e}. \tag{3}$$

The problem considered in this paper is associated with finding the prediction coefficient vector $\mathbf{a} \in \mathbb{R}^K$ from a set of observed real samples $x(n)$ for $n = 1, \ldots, N$ so that the prediction error is minimized [13]:

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \|\mathbf{x} - \mathbf{X}\mathbf{a}\|_p^p, \tag{4}$$

where

$$\mathbf{x} = \begin{bmatrix} x(N_1) \\ \vdots \\ x(N_2) \end{bmatrix}, \mathbf{X} = \begin{bmatrix} x(N_1 - 1) & \cdots & x(N_1 - K) \\ \vdots & & \vdots \\ x(N_2 - 1) & \cdots & x(N_2 - K) \end{bmatrix}, \tag{5}$$

and $\|\cdot\|_p$ is the p-norm defined as $\|\mathbf{x}\|_p = (\sum_{n=1}^{N} |x(n)|^p)^{\frac{1}{p}}$ for $p \geq 1$. The starting and ending points N_1 and N_2 can be chosen in various ways assuming that $x(n) = 0$ for $n < 1$ and $n > N$ [14]. We will consider the case $N_1 = 1$ and $N_2 = N + K$, which for $p = 2$ is equivalent to the autocorrelation method:

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \|\mathbf{x} - \mathbf{X}\mathbf{a}\|_2^2 = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{x}, \tag{6}$$

where $\mathbf{R} = \mathbf{X}^T\mathbf{X}$ is the autocorrelation matrix.

The case we consider here is when $p = 1$, which corresponds to minimizing the sum of absolute values:

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \|\mathbf{x} - \mathbf{X}\mathbf{a}\|_1. \tag{7}$$

This formulation is relevant particularly in LP of voiced speech signals where the prediction residual is usually modeled by an impulse train. The 1-norm, intended as a convex relaxation of the 0-norm, will offer an approximate solution to the minimization of the cardinality, i.e., the sparsest prediction prediction residual. This translates into the ability of the predictor to preserve the structure of the underlying sparse pulse-like excitation. The spectral envelope will benefit from this by avoiding the over-emphasis on peaks generated in the effort to cancel the voiced speech harmonics [3], [7].

The 1-norm minimization criterion, is also equivalent to the ML estimator when the prediction error is assumed to be i.i.d. Laplacian:

$$\hat{\mathbf{a}}_{\mathbf{ML}} = \arg\max_{\mathbf{a}} f(\mathbf{x}|\mathbf{a}) = \arg\max_{\mathbf{a}} \{\exp(-\|\mathbf{x} - \mathbf{X}\mathbf{a}\|_1)\}. \tag{8}$$

A multivariate Laplacian distribution offers in general a better model for a speech signal segment [15]. In turn, the multivariate distribution is generated by a i.i.d. univariate Laplacian vector [16], [17], thus justifing statistically the use of the 1-norm in modeling the excitation.

The minimization problem in (7) does not allow for a closed form solution and so a linear programming formulation is required [13]. In particular, interior point methods [18] have been proved to solve the minimization problem efficiently.

III. Methods for obtaining stable predictors

A. Reducing the numerical range of the shift operator

First of all, we consider a more general prediction framework where the columns of the matrix obtained concatenating $\mathbf{x}$ and $\mathbf{X}$, as defined in (5):

$$[\mathbf{x}|\mathbf{X}] = [\mathbf{x}_0\ \mathbf{x}_1\ \ldots\ \mathbf{x}_K] \in \mathbb{R}^{(N+K)\times(K+1)}, \tag{9}$$

can be generated via the formula:

$$\mathbf{x}_{k+1} = \mathbf{B}\mathbf{x}_k, \tag{10}$$

where:

$$\mathbf{x}_0 = [x_1\ x_2\ \ldots\ x_N\ 0\ \ldots\ 0]^T \in \mathbb{R}^{N+K}, \tag{11}$$

and $\mathbf{B}$ is a shift matrix of size $(N + K) \times (N + K)$:

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & \cdots & \omega \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \tag{12}$$

In our case ω plays no useful role and thus we will set $\omega = 0$, thus $\mathbf{B}$ becomes a noncirculant shift matrix. Applied to $\mathbf{x}$, $\mathbf{B}$ shifts the elements down by one position and eliminates the last element. In other words $\mathbf{B}$ is a nilpotent operator of power $n = N + K$, i.e., $\mathbf{B}^{N+K} = 0$.

Let us now consider the p-norm LP problem (4), where the column $[\mathbf{x}|\mathbf{X}]$ are constructed using the formula in (10) where $\mathbf{B}$ is generalized to any matrix in $\mathbb{R}^{(N+K)\times(N+K)}$. It has been shown that, in this case, the roots $\{z_i\}$ of the monic polynomial solution to the p-norm LP problem (4) belong to the numerical range $\eta_p(\mathbf{B})$ of the matrix $\mathbf{B}$, which, in turn, belongs to an open circular disk $\rho(\mathbf{B})$ of radius $2\|\mathbf{B}\|_2$ and center in the origin [8]. It is then clear that the roots of the predictor, obtained by solving (7), with $\mathbf{B}$ as defined in (12), will be contained in a closed circle of radius $2\|\mathbf{B}\|_2 = 2$. This result can be generalized for any shift matrix $\mathbf{B}$ with nonzero entries different from the unity:

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ B_{2,1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & 0 & B_{N+K,N+K-1} & 0 \end{bmatrix}. \tag{13}$$

In this case, the radius of the circle $\rho(\mathbf{B})$ that contain the numerical range $\eta_1(\mathbf{B})$ is defined as:

$$2\|\mathbf{B}\|_2 = 2\max|B_{i+1,i}|. \tag{14}$$

We will then change the nonzero values of $\mathbf{B}$ (and subsequently the construction of $[\mathbf{x}|\mathbf{X}]$), in order to reduce the radius of the circle containing $\eta_1(\mathbf{B})$ to be equal or less than one, therefore guaranteeing the stability of the linear predictor. In particular, having $\max|B_{i,j}| \leq 1/2$ will be sufficient for stability. We can also consider a more general formulation of the LP scheme, where we apply a weighting vector $\mathbf{w} \in \mathbb{R}_+^{N+K}$ on the analyzed speech signal segment.

The effect of the weighting can be moved into the shift matrix and the analyzed speech segment by defining:

$$\tilde{\mathbf{B}} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ w_2/w_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & 0 & w_{N+K}/w_{N+K-1} & 0 \end{bmatrix}, \quad (15)$$

and

$$\tilde{\mathbf{x}}_0 = [w_1 x_1 \; w_2 x_2 \; w_N x_N \; 0 \; \ldots \; 0]^T. \quad (16)$$

Constructing all the other columns of the new matrix $[\tilde{\mathbf{x}}|\tilde{\mathbf{X}}]$ using the relation in (10), the minimization problem (7) then becomes:

$$\min_{\mathbf{a}} \|\tilde{\mathbf{x}} - \tilde{\mathbf{X}}\mathbf{a}\|_1. \quad (17)$$

According to (18), the circle containing the numerical range of $\tilde{B}$ and, in turn, the roots of the predictor will have radius:

$$\rho(\mathbf{B}) = 2 \max \frac{w_{n+1}}{w_n}. \quad (18)$$

We can then construct a weighting vector that stabilizes the predictor. In [9] and [10], the weighting vector is chosen based on the short-time energy (STE):

$$w_n = \sqrt{\sum_{i=0}^{M-1} x_{n-i-1}^2} \quad (19)$$

where M is the length of the STE window. The STE window tends to weight more those parts of the speech signal that consist of samples of large magnitude, providing a robust signal selection especially for the analysis of voiced speech. In order to achieve intrinsically stable solutions, we can then just define the entries of the matrix $\tilde{\mathbf{B}}$ in (15) as:

$$\tilde{B}_{i+1,i} = \begin{cases} (w_{i+1}/w_i) & \text{if} \quad (w_{i+1}/w_i) \leq 1/2, \\ 1/2 & \text{if} \quad (w_{i+1}/w_i) > 1/2. \end{cases} \quad (20)$$

Finally, we can solve our modified 1-norm problem in (17) obtaining an intrinsically stable predictor. Clearly, the window, and thus the weights, can be chosen *ad libitum*; we will use the STE windowing that provides important signal selection properties to retrieve the underlying spiky structure of the speech signal, as done in [10].

B. Constrained 1-norm minimization

Let us consider the univariate polynomial $A(z)$ in (2). According to [19], the alternative Cauchy bound states that all zeros of (2) lie in the disk:

$$|z| \leq \lambda, \quad \text{where} \quad \lambda = \max \left\{ 1, \sum_{k=1}^{K} |a_k| \right\}. \quad (21)$$

This bound, a refinement of the famous Cauchy bound [11], gives precious hints on how to modify the formulation of (7) to guarantee a stable predictor. In particular, we can rewrite the problem as:

$$\min_{\mathbf{a}} \|\mathbf{x} - \mathbf{X}\mathbf{a}\|_1, \quad \text{s.t.} \quad \|\mathbf{a}\|_1 < 1, \quad (22)$$

Algorithm 1 Iteratively Reweighted 2-norm Minimization

Inputs: speech segment $\mathbf{x}$
Outputs: predictor $\hat{\mathbf{a}}^i$, residual $\hat{\mathbf{r}}^i$
$i = 0$, initial weights $\mathbf{W}^{i=0} = \mathbf{I}$
while halting criterion false **do**
1. $\hat{\mathbf{a}}^i \leftarrow \arg\min_{\mathbf{a}} \|\mathbf{W}^i(\mathbf{x} - \mathbf{X}\mathbf{a})\|_2^2$
2. $\mathbf{W}^{i+1} \leftarrow \text{diag}\left(\left|\mathbf{x} - \mathbf{X}\hat{\mathbf{a}}^i\right| + \epsilon\right)^{-1/2}$
3. $i \leftarrow i + 1$
end while

where the constraint $\|\mathbf{a}\|_1 < 1$, according to (21), provides a sufficient (not necessary) condition for the zeros of (2) to belong to the open unit disk, and can be easily incorporated in the linear program to solve (7) [13]. Now let us consider some previously proposed methods for obtaining stable solutions.

C. Iteratively Reweighted 2-norm minimization [20]

A known method to obtain a stable predictor based on 1-norm minimization is based on iteratively reweighted 2-norm minimization [20]. The algorithm is shown in Algorithm 1. It is guaranteed to output a stable predictor since the only difference to the original formulation is the projection in the weighted domain by the matrix $\mathbf{W}^i$, leaving $\mathbf{x}$ and $\mathbf{X}$ untouched, as discussed in Section III-A. In [20] a proof that $\|\hat{\mathbf{r}}^{i+1}\|_2 \leq \|\hat{\mathbf{r}}^i\|_2$ (where $\hat{\mathbf{r}}^i = \mathbf{x} - \mathbf{X}\hat{\mathbf{a}}^i$) is provided, meaning that this is a descent algorithm. In Algorithm 1, the halting criterion can be chosen as either a maximum number of iterations or as a convergence criterion. The parameter $\epsilon > 0$ is used to avoid problems when a component of $\hat{\mathbf{r}}$ goes to zero. The weighting with the square root of the inverse of the amplitude of the residual increases the influence of the small values in the residual while the influence of the large residual values decreases, which is consistent with the Laplacian probability density functions, cfr (7).

D. Burg method based on 1-norm minimization [7]

The Burg method based on 1-norm minimization was first proposed in [7]. This method stands as a generalization of the Burg method where the reflection coefficients of the lattice filter are obtained by minimizing the 2-norm of the forward and backward prediction error. In this case the 1-norm is minimized instead and the algorithm is shown in Algorithm 2. Once the K reflection coefficients are found, the prediction polynomial and the prediction error can be easily calculated. This method is also guaranteed to provide a stable predictor since all the reflection coefficients obtained have amplitude less than one. A simple proof is shown in [7]. This method is, however, suboptimal due to the decoupling of the main K-dimensional minimization problem (7) in K one-dimensional minimization sub-problems.

IV. Performance Analysis

In this section, we analyze and compare the performance of the stable predictors obtained with the methods presented in the previous section with traditional 2-norm LP and 1-norm

Algorithm 2 1-norm Burg Method

Inputs: speech segment $\mathbf{x}$
Outputs: reflection coefficients $\{k_i\}$
Initialize forward $\mathbf{f}_0 = \mathbf{x}$ and backward $\mathbf{b}_0 = \mathbf{x}$ error
for $i = 1, \ldots, K$ **do**
1. $k_i \leftarrow \arg\min_{k_i} \|\mathbf{f}_{i-1} + k_i \mathbf{b}_{i-1}\|_1 + \|k_i \mathbf{f}_{i-1} + \mathbf{b}_{i-1}\|_1$
update forward error
2. $f_i(n) \leftarrow f_{i-1}(n) + k_i b_{i-1}(n-1)$
update backward error
3. $b_i(n) \leftarrow k_i f(n)_{i-1} + b_{i-1}(n-1)$
end for

TABLE I
DESCRIPTION OF THE DIFFERENT PREDICTION METHODS COMPARED IN OUR EVALUATION.

METHOD	DESCRIPTION
LP2	Traditional 2-norm minimization (6) with 10Hz bandwidth expansion ($\gamma = 0.996$) and Hamming windowing.
LP1	Unconstrained 1-norm minimization (7). Stability is imposed by pole reflection if unstable. No windowing is performed.
STW	Stable 1-norm minimization through reduction of the numerical range of the shift operator (17). The weigths in (15) and (16) are chosen from the STE (19).
CT1	Constrained 1-norm minimization as shown in (22). No windowing is performed.
BU1	Burg method based on the 1-norm minimization of forward and backward error (as shown in Algorithm 2). No windowing is performed.
RW2	Reweighted 2-norm minimization (as shown in Algorithm 1). No bandwidth expansion is performed. No windowing is performed.

LP. An overview of the methods compared is shown in Table I. In the case of 1-norm LP, a stability check takes place once the predictor is obtained, the stabilization is performed through pole reflection when the predictor is unstable. Notice that pole reflection is the only way to obtain an amplitude response for the stabilized predictor that is exactly the same as the one of the unstable predictor. In all other methods, no stability check has to be performed.

A. Modeling Performance

In this section we analyze the modeling performance of the predictors in case of voiced speech. The experimental analysis has been done on 5,000 segments of length $N = 40$ (5 ms) of clean voiced speech coming from several different speakers with different characteristics (gender, age, pitch, regional accent) taken from the TIMIT database, downsampled at 8 kHz. As a reference, we have used the envelope obtained through a cubic spline interpolation between the harmonics peaks of the logarithmic periodogram. This method was presented in [21] and provides an approximation of the vocal tract transfer function, "cleaned" from the fine structure belonging to the pitch excitation. We then calculate the log spectral distortion between our reference envelope $S_{int}(\omega)$ and the estimated model of the all-pole model corresponding to the inverse of the predictor $S(\omega, \mathbf{a})$ as:

$$\mathrm{SD}_m = \sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} \left[10 \log_{10} S_{int}(\omega) - 10 \log_{10} S(\omega, \mathbf{a})\right]^2 d\omega}. \quad (23)$$

In general, the linear predictors obtained through 1-norm minimization provide smoother all-pole models of the vocal tract, therefore more robust to quantization. We will then also compare the log spectral distortion between our reference envelope $S_{int}(\omega)$ and the quantized AR model $S(\omega, \hat{\mathbf{a}})$:

$$\mathrm{SD}_q = \sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} \left[10 \log_{10} S_{int}(\omega) - 10 \log_{10} S(\omega, \hat{\mathbf{a}})\right]^2 d\omega}. \quad (24)$$

The quantizer used is the one presented in [22], with the number of bits fixed at 20 for the different prediction orders. A critical analysis of the results in Table II shows how 1-norm based LP (**LP1**) offers substantially better modeling of the envelope than traditional 2-norm LP (**LP2**). All the other methods achieve a performance similar to **LP1**, nevertheless **STW** offers even better modeling performance, thanks also to the choice of weights. It should be noted that **CT1** increases its performances considerably from order $K = 10$ to $K = 12$. This is due to the stringent constraint on the prediction coefficients ($\|\mathbf{a}\|_1 < 1$) that necessarily needs a larger K in order to grasp the spectral information as well as the other methods.

TABLE II
AVERAGE SPECTRAL DISTORTION FOR THE CONSIDERED METHODS IN THE UNQUANTIZED CASE SD_m AND QUANTIZED CASE SD_q FOR DIFFERENT PREDICTION ORDERS K. A 95% CONFIDENCE INTERVALS IS GIVEN FOR EACH VALUE.

METHOD	K	SD_m	SD_q
LP2	10	1.97±0.03	2.95±0.09
	12	1.98±0.05	2.72±0.12
LP1	10	1.78±0.01	2.53±0.02
	12	1.61±0.01	2.31±0.04
STW	10	1.71±0.02	2.47±0.01
	12	1.52±0.01	2.19±0.09
CT1	10	1.88±0.01	2.64±0.01
	12	1.65±0.01	2.22±0.01
BU1	10	1.91±0.06	2.71±0.09
	12	1.84±0.11	2.59±0.10
RW2	10	1.83±0.01	2.51±0.02
	12	1.69±0.03	2.37±0.05

B. Coding Performance

The second objective is to adopt the presented methods in the speech coding context. The experimental analysis has been conducted on about one hour of clean speech (both voiced and unvoiced) coming from several different speakers with different characteristics (gender, age, pitch, regional accent) taken from the TIMIT database, re-sampled at 8 kHz. We propose a simple scheme to evaluate the coding performance of the proposed methods. A 10^{th} order predictive analysis is first done on a segment of speech of $N = 40$. Then a multipulse encoding procedure [23] is performed to code T pulses in the residual, with $T = 5$ and $T = 10$. Multipulse

TABLE III
COMPARISON BETWEEN THE CONSIDERED PREDICTORS ADOPTED IN MULTIPULSE ENCODING (T PULSES). A 95% CONFIDENCE INTERVAL IS GIVEN FOR EACH VALUE.

METHOD	T	$\hat{\mathbf{a}}$	SSNR
LP2	5	19	14.1±3.2
	10	19	19.1±2.9
LP1	5	18	15.3±2.1
	10	18	20.1±1.7
STW	5	17	14.9±1.6
	10	17	20.6±0.9
CT1	5	15	13.9±1.9
	10	15	19.2±1.5
BU1	5	19	14.2±0.9
	10	19	19.4±0.4
RW2	5	21	15.2±1.2
	10	21	20.9±1.7

encoding is used to obtain a sparse residual, rather than a pseudo-random one, as obtained through algebraic codes, matching the characteristics of the 1-norm minimization. In Table III, we present the results in terms of segmental SNR and number of bits necessary to encode the prediction vector $\hat{\mathbf{a}}$ within the well-known 1 dB distortion [24] using the method presented in [22]. The best coding performance is achieved by **RW2**, consistently with the "guidance" in the reweighting algorithm based on the square root of the inverse of the residual amplitude, although it requires a larger number of bits to transparently encode the predictor. As mentioned in the introduction, **BU1** does not preserve the sparsity of the residual and the coding characteristics of the 1-norm, performing very similarly to the 2-norm. The methods we have introduced seem to offer a good coding perfomance. The very smooth spectrum obtained with **CT1** allows considerably less bits than any other methods to achieve transparent coding of the prediction coefficients, achieving a performance comparable to **LP2** and **BU1**. **STW** performs just slightly worse than **RW2**, but with a significant saving in the bit budget of the predictor.

V. CONCLUSIONS

In this paper, we have presented two new methods for finding intrinsically stable predictors based on 1-norm error minimization. The methods introduced, one based on constrained 1-norm minimization and one on the reduction of the numerical range of the shift operator, have both shown to offer a valid alternative to the original 1-norm linear prediction preserving the properties of the 1-norm error minimization criterion. In particular, the experimental analysis has shown that both methods offer attractive modeling and coding performance without any significant increase in complexity. The two methods have also been shown to offer a slightly better modeling performance compared to the Burg method based on the 1-norm minimization and the 2-norm reweighted minimization method.

REFERENCES

[1] J. H. L. Hansen, J. G. Proakis, and J. R. Deller, Jr., *Discrete-Time processing of speech signals*, Prentice-Hall, 1987.

[2] J. Makhoul, "Linear prediction: a tutorial review", *Proc. IEEE*, vol. 63, no. 4, pp. 561–580, 1975.

[3] D. Giacobello, M. G. Christensen, M. N. Murthi, S. H. Jensen, and M. Moonen "Sparse linear prediction and its applications to speech processing," *IEEE Trans. Speech, Audio and Language Processing*, vol. 20, no. 5, pp. 1644–1657, 2012.

[4] D. Giacobello, M. G. Christensen, M. N. Murthi, S. H. Jensen, and M. Moonen, "Speech Coding Based on Sparse Linear Prediction," *Proc. European Signal Proc. Conf.*, pp. 2524–2528, 2009.

[5] D. Giacobello, M. G. Christensen, M. N. Murthi, S. H. Jensen, and M. Moonen, "Retrieving sparse patterns using a compressed sensing framework: applications to speech coding based on sparse linear prediction," *IEEE Sig. Proc. Letters*, vol. 17, no. 1, pp. 103–106, 2010.

[6] W. F. G. Mecklenbrauker, "Remarks on the minimum phase property of optimal prediction error filters and some related questions," *IEEE Signal Processing Letters*, vol. 5, no. 4, pp. 87–88, 1998.

[7] E. Denoël and J.-P. Solvay, "Linear prediction of speech with a least absolute error criterion," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, no. 6, pp. 1397–1403, Dec. 1985.

[8] L. Knockaert, "Stability of linear predictors and numerical range of shift operators in normed spaces," *IEEE Trans. on Information Theory*, vol. 38, no. 5, pp. 1483–1486, 1992.

[9] C. Ma, Y. Kamp and L. F. Willems, "Robust signal selection for linear prediction analysis of voiced speech," *Speech Communication*, vol. 12, no. 1, pp. 69–81, 1993.

[10] C. Magi, J. Pohjalainen, T. Bäckström, and P. Alku, "Stabilized weighted linear prediction," *Speech Communication*, vol. 51, pp. 401–411, 2009.

[11] A. L. Cauchy, *Exercise de mathematique*, Oeuvres 2, vol. 19, 1829.

[12] M. Marden, *Geometry of Polynomials*, Mathematical Surveys and Monographs, American Mathematical Society, 1966.

[13] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[14] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Pearson Prentice Hall, 2005.

[15] S. Gazor and W. Zhang, "Speech probability distribution," *IEEE Sig. Proc. Letters*, vol. 10, no. 7, pp. 204–207, 2003.

[16] A. W. Marshall and I. Olkin, "Maximum likelihood characterizations of distributions," *Statistica Sinica*, vol. 3, pp. 157–171, 1993.

[17] C. Fernandez, J. Osiewalski, and M. F. J. Steel, Modeling and Inference with v-spherical distributions, *J. Amer. Statist. Assoc*, vol. 90, pp. 1331–1340, 1995.

[18] S. J. Wright, *Primal-Dual Interior-Point methods*, SIAM, 1997.

[19] H. P. Hirst and W. T. Macey, "Bounding the roots of polynomials," *The College Mathematics Journal*, vol. 18, no. 4, pp. 292–295, 1997.

[20] Y. Li, "A globally convergent method for L_p problems," *SIAM J. Optimization*, vol. 3, no. 3, pp. 609–629, 1993.

[21] L. A. Ekman, W. B. Kleijn, and M. N. Murthi, "Regularized Linear Prediction of Speech," *IEEE Trans. Audio, Speech, Language Processing*, vol. 16, no. 1, pp. 65–73, 2008.

[22] A. D. Subramaniam and B. D. Rao, "PDF optimized parametric vector quantization of speech line spectral frequencies," *IEEE Trans. Speech and Audio Processing*, vol. 11, no. 2, pp. 87–89, 2003.

[23] B. S. Atal and J. R. Remde, "A new model of LPC excitation for producing natural sounding speech at low bit rates," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 7, pp. 614–617, 1982.

[24] W. B. Kleijn and A. Ozerov, "Rate Distribution Between Model and Signal," IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pp. 243–246, 2007.