Least 1-Norm Pole-Zero Modeling with Sparse Deconvolution for Speech Analysis

Shi, Liming; Jensen, Jesper Rindom; Christensen, Mads Græsbøll

Published in:
IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017

DOI (link to publication from Publisher):
10.1109/ICASSP.2017.7952252

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

LEAST 1-NORM POLE-ZERO MODELING WITH SPARSE DECONVOLUTION FOR SPEECH ANALYSIS

Liming Shi, Jesper Rindom Jensen and Mads Græsbøll Christensen

Audio Analysis Lab, AD:MT, Aalborg University,
{ls, jrj, mgc}@create.aau.dk

ABSTRACT

In this paper, we present a speech analysis method based on sparse pole-zero modeling of speech. Instead of using the all-pole model to approximate the speech production filter, a pole-zero model is used for the combined effect of the vocal tract; radiation at the lips and the glottal pulse shape. Moreover, to consider the spiky excitation form of the pulse train during voiced speech, the modeling parameters and sparse residuals are estimated in an iterative fashion using a least 1-norm pole-zero with sparse deconvolution algorithm. Compared with the conventional two-stage least squares pole-zero, linear prediction and sparse linear prediction methods, experimental results show that the proposed speech analysis method has lower spectral distortion, higher reconstruction SNR and sparser residuals.

Index Terms— Pole-zero model, least 1-norm cost function, sparse deconvolution, speech analysis.

1. INTRODUCTION

Speech modeling, as a fundamental speech analysis problem, has diverse applications in speech synthesis [1], speaker identification, speech recognition, etc. Based on the source-filter model of the speech production system, the speech production filter (SPF) is assumed to be time-invariant during a short-time period (frame) of approximately 20-40 ms, and excited by a pulse train or white noise for voiced or unvoiced speech.

Linear prediction (LP) with least squared error minimization criterion, based on an all-pole model, is commonly used for speech analysis [2]. The method performs well for white noise and small valued pitch harmonic excitations (aka residuals). However, for a large valued pitch, it tends to null out the input voiced speech harmonics and leads to an all-pole filter with poles close to the unit circle, and the estimated spectral envelope has a sharper contour than desired [3, 4]. Various improved schemes based on LP have been proposed, such as LP with the Itakura-Saito error criterion [5], all-pole modeling with a distortionless response at frequencies of harmonics [3] and the regularized LP [6]. More recently, motivated by the compressive sensing framework, sparse linear prediction based on the 1-norm criterion has been proposed for voiced speech analysis [7]. Unlike the conventional 2-norm method, sparse priors on the excitation signals and prediction coefficients are both utilized to offer an effective decoupling of the SPF and underlying sparse residuals. Moreover, the 1-norm method was shown to be robust against impulsive interference in all-zero plant identification [8, 9]. Fast methods and the stability of the 1-norm cost function for spectral envelope estimation are further investigated in [10, 11]. Another problem is that some sounds containing spectral zeros with voiced excitation, such as nasals, fricatives, or laterals, are poorly estimated by an all-pole model but trivial with a pole-zero model [12–14]. The estimation of the coefficients of the pole-zero model can be obtained separately [15], jointly [16] or iteratively [12]. A model identification method is proposed for time-varying stochastic pole-zero model estimation [17]. A 2-norm minimization criterion with Gaussian residual assumption is usually used to obtain the parameter estimates in these methods. Motivated by the logarithmic scale perception of the human auditory system, the logarithmic magnitude function minimization criterion has also been proposed [14, 18]. Additionally, the performance of the all-pole method deteriorates severely in noisy conditions. Various noise robust approaches based on all-pole model have been proposed [19, 20].

In this paper, a speech analysis method based on sparse pole-zero modeling is presented. Using a pole-zero model for fitting the spectral envelope compared with the all-pole model, a better approximation can be obtained. The modeling coefficients and residuals are obtained in a iterative fashion. To consider the sparse priors of residuals, instead of conventional 2-norm minimization criterion, a least 1-norm criterion is used for the coefficient estimation. Moreover, sparse deconvolution is applied for deriving sparse residuals and denoising. The effectiveness of the proposed method for the spectral envelope estimation and signal reconstruction is verified using both synthetic signals and natural speech.

2. FUNDAMENTALS OF THE POLE-ZERO ESTIMATION

The pole-zero speech production filter model is considered in this paper. A sample of speech is written in the following form:

$$\begin{aligned} s\left(n\right) &= -\sum_{k=1}^{K} a_k s\left(n-k\right) + \sum_{l=0}^{L} b_l e\left(n-l\right) \\ x\left(n\right) &= s\left(n\right) + m(n) \end{aligned} \tag{1}$$

where a_k and b_l are coefficients of the pole-zero model, $b_0 = 1$, $m(n)$ is Gaussian noise, and $e\left(n\right)$ is the residual.

When $L = 0$, (1) reduces to the all-pole model and the parameter estimation can be formulated as

$$\min_{\mathbf{a}} \left\|\mathbf{x} + \mathbf{X}\mathbf{a}\right\|_p^p + \gamma \left\|\mathbf{a}\right\|_q^q \tag{2}$$

where $\mathbf{x} = \left[x\left(N_1\right), x\left(N_1+1\right)\cdots x\left(N_2\right)\right]^T$, $\mathbf{a} = \left[a_1, a_2 \cdots a_K\right]^T$, $\left[\cdot\right]^T$ denotes matrix transpose, $\left\|\cdot\right\|_p$ is the p-norm, γ is the regularization parameter and

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

This work was funded by the Danish Council for Independent Research, grant ID: DFF 4184-00056

N_1 and N_2 can be chosen in various ways. One way is setting $N_1 = 1$ and $N_2 = N + K$, which is the autocorrelation method. The covariance method is obtained by setting $N_1 = K + 1$ and $N_2 = N$ [21]. When the residual signal $e(n)$ is a Gaussian random variable, the standard 2-norm solution (i.e., $p = 2$ and $\gamma = 0$) of (1) is statistically equivalent to the maximum likelihood solution. However, the pulse train excitation for voiced speech can be better fitted as a super-Gaussian variable, known for its long-tail distribution. A sparse solution for the residuals can be obtained in principle by setting $0 \leq p \leq 1$. Moreover, prior knowledge of coefficients can be incorporated as a regularization term to improve the parameter estimation. For example, the generalized Gaussian or Laplacian distribution can be imposed to $\mathbf{a}$ by choosing $q = 2$ or $q = 1$, respectively. In [22], $p = 1$ and $\gamma = 0$ are used to obtain sparse residuals. [7] use $p = 1$ and $q = 1$ to encourage both sparse residuals and coefficients.

When $L > 0$, both poles and zeros present. The pole-zero model is known to be more effective than all-pole model, especially for fitting nasal sound [12–14, 18]. However, it inherently involves solving a nonlinear equation. A two-stage pole-zero method was proposed [15, 21]. In the first stage, a coarse estimate of $\widehat{\mathbf{e}} = \mathbf{x} + \mathbf{X}\widehat{\mathbf{a}}$ can be obtained by using (2) with a sufficiently high-order linear prediction K'. Then, replace $e(n)$ ($K^{'} + 1 \leq n \leq N$) in (1) by $\hat{e}(n)$ determined in the first stage, and solve the following minimization problem:

$$\min_{\mathbf{z}} \left\| \mathbf{x}' + \mathbf{X}'\mathbf{z} \right\|_p^p + \gamma \left\| \mathbf{z} \right\|_q^q \tag{3}$$

where $\mathbf{x}' = [x(N_1'), x(N_1'+1) \cdots x(N_2')]^T$, $\mathbf{X}' = [\bar{\mathbf{X}}, -\widehat{\mathbf{E}}]$ and

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

$$\widehat{\mathbf{E}} = \begin{bmatrix} \hat{e}(N_1'-1) & \cdots & \hat{e}(N_1'-L) \\ \vdots & & \vdots \\ \hat{e}(N_2'-1) & \cdots & \hat{e}(N_2'-L) \end{bmatrix}$$

and $\mathbf{z} = [a_1 \cdots a_K, b_1 \cdots b_L]^T$. N_1' and N_2' are usually set to $K^{'} + L + 1$ and N, respectively. When we set p to 2 and γ to 0 in (2) and (3), the above formulation is the standard two-stage least square pole-zero (TS-LS-PZ) method [21].

3. LEAST 1-NORM POLE-ZERO MODELING WITH SPARSE DECONVOLUTION (SD-L1-PZ)

There are problems with the above TS-LS-PZ method. First, to obtain an accurate estimate of the parameter vector $\mathbf{z}$, $K^{'}$ should be very large. As a result, a large segment of data $e(n)$ ($1 \leq n \leq K^{'}$) cannot be estimated, and the observations are not sufficiently used for estimating model parameters. Second, in presence of pulse train residuals, especially for a high valued pitch, the 2-norm based criterion is inappropriate for the estimation in both stages. Besides, this method does not consider the influence of the additive noise $m(n)$, which affects the reconstruction performance. To overcome these problems, we proceed to design a least 1-norm based pole-zero modeling with sparse deconvolution algorithm for speech analysis.

3.1. Finding a sparse residual

Instead of estimating the residuals using a sufficiently high-order all-pole model, a deconvolution method can be used [23]. First, (1) can be reformulated in a matrix form as follows

$$\mathbf{A}\mathbf{x} = \mathbf{B}\mathbf{e} + \mathbf{A}\mathbf{m} \tag{4}$$

where $\mathbf{e} = [e(1) \cdots e(N)]^T$, $\mathbf{m} = [m(1) \cdots m(N)]^T$, $\mathbf{A}$ and $\mathbf{B}$ have the form

$$\mathbf{A} = \begin{bmatrix} 1 & & & & & \\ a_1 & 1 & & & & \\ \vdots & \cdots & \ddots & & & \\ a_K & \cdots & a_1 & 1 & & \\ & \ddots & & & \ddots & \\ & & a_K & \cdots & a_1 & 1 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 1 & & & & & \\ b_1 & 1 & & & & \\ \vdots & \cdots & \ddots & & & \\ b_L & \cdots & b_1 & 1 & & \\ & \ddots & & & \ddots & \\ & & b_L & \cdots & b_1 & 1 \end{bmatrix}$$

When $L = 0$, $\mathbf{B}$ reduces to the identity matrix and (4) becomes a standard AR estimation problem. The usage of both poles and zeros is to obtain a better fitting to sounds containing spectral zeros. With a known or estimated parameter vector $\mathbf{z}$, the estimation of residuals $\mathbf{e}$ can be formulated as

$$\min_{\mathbf{e}} \left\| \mathbf{e} \right\|_{q_1}^{q_1} \text{ s.t. } \left\| \mathbf{x} - \mathbf{A}^{-1}\mathbf{B}\mathbf{e} \right\|_{p_1}^{p_1} \leq C \tag{5}$$

Assuming Gaussian noise and pulse train residuals, a reasonable choice is to set $p_1 = 2$ and $q_1 = 1$ in (5). A good approximation between observations and modeling fitting is guaranteed by the 2-norm constraint. The 1-norm cost is to deemphasize the spiky residuals associated with pulse train excitation. Note that setting C larger yields an estimate with poorer data fitting but sparser residuals, and vice versa. When the variance of the noise is known, a good choice is to set $C = N\sigma_m^2$. The reconstruction of speech signal can be obtained as

$$\widehat{\mathbf{x}} = \mathbf{A}^{-1}\mathbf{B}\widehat{\mathbf{e}} \tag{6}$$

where $\widehat{\mathbf{e}}$ denotes the residual estimate obtained by (5).

3.2. Estimation of pole-zero modeling coefficients

With known or estimated residuals, we estimate the pole-zero modeling coefficients $\mathbf{z}$ using the second stage of the TS-LS-PZ algorithm but with $p = 1$ and $N_1' = \max(K, L) + 1$, instead of the conventional $p = 2$ and $N_1' = K' + L + 1$, i.e.,

$$\min_{\mathbf{z}} \left\| \mathbf{x}' + \mathbf{X}'\mathbf{z} \right\|_1^1 + \gamma \left\| \mathbf{z} \right\|_q^q \tag{7}$$

To account for the non-Gaussian distribution characteristics of the residuals, a 1-norm minimization criterion is again used here instead of conventional 2-norm. Also, since complete estimates of residuals are available, observations can be sufficiently used with a smaller N_1'. Furthermore, as noted before, prior knowledge of coefficients can be incorporated as regularization term. Especially, when $q = 1$, and high orders of K and L are used, it will lead to sparse pole-zero coefficient estimates (see [7] for details about sparse linear prediction). The estimation of residuals and pole-zero modeling coefficients are repeated until convergence. To guarantee the causality

Algorithm 1 SD-L1-PZ

Intiate γ, q, $p_1 = 2$ and $q_1 = 1$.
Initialization with the TS-L1-PZ:
  Solve (9) with a large K', and $\widehat{\mathbf{e}} = \mathbf{x} + \mathbf{X}\widehat{\mathbf{a}}$
  Coefficients estimation by (7) with ${N_1}' = K' + L + 1$
for $k = 1, \cdots$ **do**
  Calculate $\mathbf{A}^{-1}\mathbf{B}$
  Obtain refined sparse residual $\widehat{\mathbf{e}}_k$ using (5)
  Solve coefficients using (7) with ${N_1}' = \max(K, L) + 1$
  while poles or zeros are larger than 1 **do**
    Compute re-estimated coefficients using (8)
  end while
end for

and stability of the proposed method for both estimation and reconstruction, the parameter vector can be re-estimated using

$$\min_{\mathbf{z}} \left\|\mathbf{x}' + \mathbf{X}'\mathbf{z}\right\|_1^1 + \gamma \left\|\mathbf{z}\right\|_q^q \quad \text{s.t.} \quad \|\mathbf{a}\| \leq 1, \|\mathbf{b}\| \leq 1 \tag{8}$$

when the poles or zeros are outside of the unit circle.

Furthermore, for the initialization of this iteration procedure, we modify the first stage of the TS-LS-PZ to the 1-norm formulation [22]

$$\min_{\mathbf{a}} \|\mathbf{x} + \mathbf{X}\mathbf{a}\|_1^1 \tag{9}$$

In the second stage, (7) is used to replace the original cost function (3) with $p = 2$. Due to the 1-norm cost function, we refer this initialization approach to the two-stage least 1-norm pole-zero (TS-L1-PZ). We summarize the SD-L1-PZ in Algorithm 1.

4. RESULTS

In this section, we test the performance of the proposed TS-L1-PZ and SD-L1-PZ in both synthetic and real speech signals analysis scenarios.

4.1. Synthetic signal analysis

Synthetic speech signals are generated by convolving an input excitation with a constructed filter to estimate the performance of the proposed method. The input excitation is a pulse train with the fundamental frequency between 300-500 Hz. The filter we used here has the following characteristics

$$H(z) = \frac{\sum_{i=1}^{4}(1 - \beta_i z^{-1})}{\sum_{j=1}^{5}(1 - \alpha_j z^{-1})} \tag{10}$$

where $\beta_1 = \beta_2^* = 0.5348 + 0.5529j$, $\beta_3 = \beta_4^* = -0.0263 + 0.7688j$, $\alpha_1 = \alpha_2^* = -0.5026 + 0.5976j$, $\alpha_3 = \alpha_4^* = 0.4449 + 0.7928j$, $\alpha_5 = 0.8602$. The SNR is set to 30 dB for additive Gaussian noise. As a measure for the accuracy of the estimated spectral envelope, the spectral distortion (SD) is defined as [24]

$$\text{SD} = \frac{1}{\text{S}} \sum_{s=1}^{\text{S}} (\log \left|H(e^{j\omega_s})\right| - \log \left|\hat{H}(e^{j\omega_s})\right|)^2 \tag{11}$$

where $\hat{H}(e^{j\omega_s})$ denotes an estimate of the true envelope $H(e^{j\omega_s})$, S is the number of spectral samples. The experimental results are obtained by the ensemble averages over 2 s with 30 ms frame length. The SD curves for the SD-L1-PZ, TS-L1-PZ, TS-LS-PZ, 1-norm linear prediction (1-norm LP), 2-norm linear prediction (2-norm LP) with different excitation frequencies are shown in Fig. 1, where 5 iterations are used for the SD-L1-PZ, $\gamma = 0$, $C = 0.01\left\|\mathbf{x}\right\|_2^2$, K and L are set to 10. As can be seen, the SD of the 1-norm LP is lower than the 2-norm LP. Moreover, the plot of the TS-LS-PZ has more fluctuations than the TS-L1-PZ for different frequencies. Furthermore, by utilizing the 1-norm cost function and pole-zero modeling with sparse deconvolution together, the proposed SD-L1-PZ achieves the lowest spectral distortion compared with others.

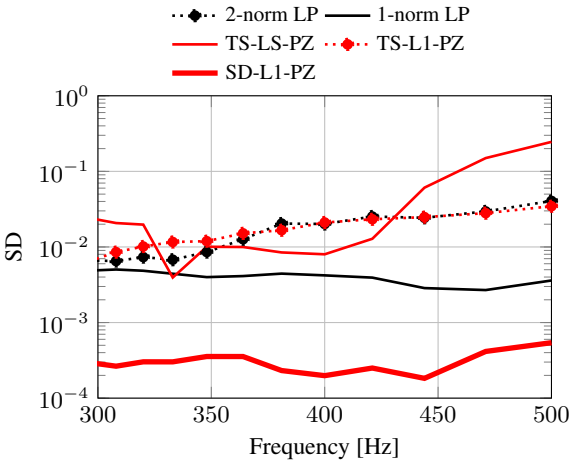

Fig. 1: Spectral distortion for different excitation frequencies

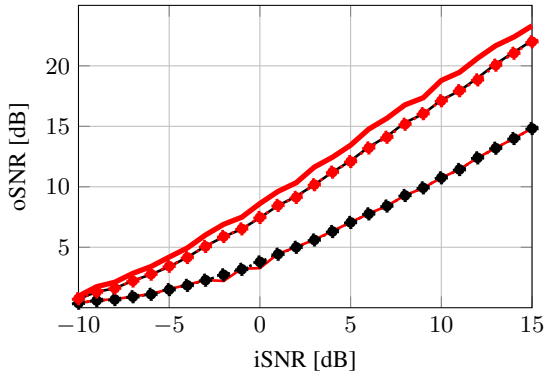

Fig. 2: SNR of reconstructed signals over different input SNR (iSNR)

Then, the reconstruction performance is tested in terms of the output SNR (oSNR). Synthetic speech signals are generated by convolving $e(n) = \delta(n - 50) + 0.5\delta(n - 80) - 0.3\delta(n - 100)$ with a filter with transfer function $H(z) = (1 + 0.8z^{-1})/(1 - 0.9z^{-1} + 0.81z^{-2})$ [23]. The oSNR is defined as

$$\text{oSNR} = E(\bar{x})^2 / E((\bar{x} - \hat{x})^2) \tag{12}$$

where $\bar{x}$ denotes the noise-free signal. The reconstructed signals $\hat{x}$ are obtained using (5) and (6) with $q_1 = 1$ for 1-norm based methods (i.e. the 1-norm LP, TS-L1-PZ and SD-L1-PZ), but with $q_1 = 2$ for the TS-LS-PZ. The experimental results are obtained by the ensemble averages over 100 Monte Carlo simulations. The oSNR curves for different algorithms are shown in Fig. 2, where $\gamma = 0$, $C = \|\mathbf{m}\|_2^2$, $N = 300$, $K = 10$ and $L = 5$. As can be

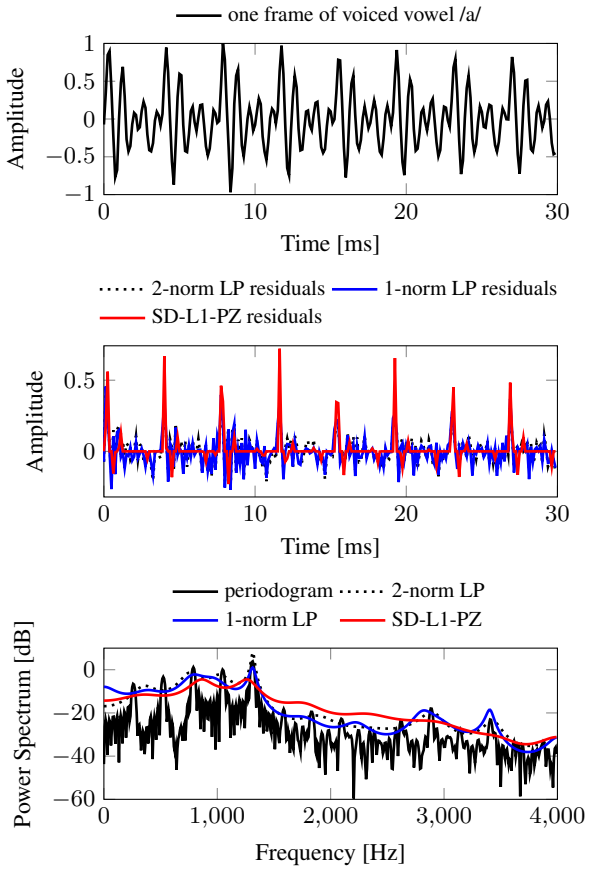

Fig. 3: Residuals and spectral envelope estimates for the voiced vowel /a/

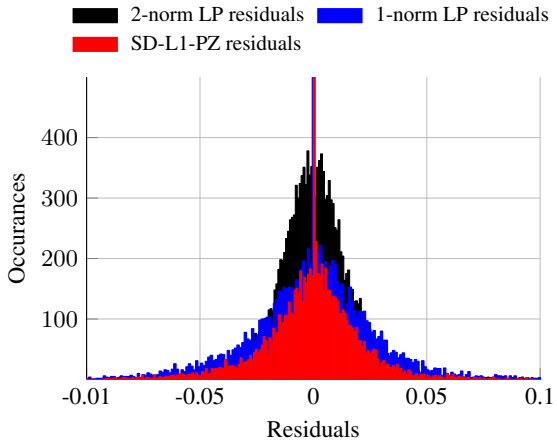

Fig. 4: The histogram of residuals of the 2-norm LP, 1-norm LP and SD-L1-PZ

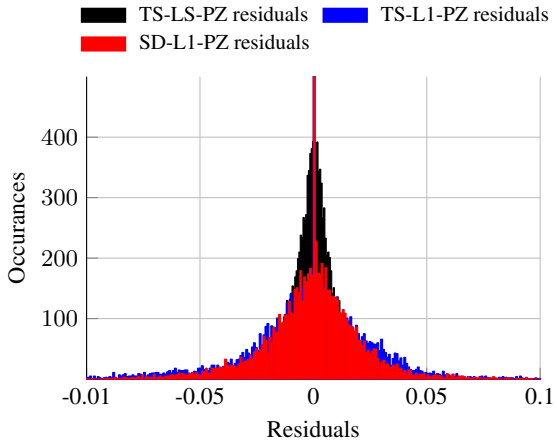

Fig. 5: The histogram of residuals of the TS-LS-PZ, TS-L1-PZ and SD-L1-PZ

seen, the performance of the 2-norm and TS-LS-PZ, 1-norm LP and TS-LS-PZ are similar. The SD-L1-PZ presents a higher oSNR.

4.2. Speech signal analysis

This work also examines the performance of the SD-L1-PZ for a real voiced vowel /a/ sampling of 8000 Hz, as shown in Fig. 3, where $\gamma = 0$, $C = 0.1\,\|\mathbf{x}\|_2^2$, $K = 20$, $L = 10$, the SNR for Gaussian white noise is set to 30 dB. As can be seen, the residuals of the SD-LS-PZ are sparser than both the 2-norm and 1-norm LP methods. Moreover, since we admit the existence of the pitch and harmonics, the spectral envelope estimate of the 1-norm LP and SD-L1-PZ is smoother than the conventional 2-norm LP, which tends to null out the harmonics [3]. In fact, due to the sparser residual estimates, the estimated spectral envelope of the SD-L1-PZ tends to be the smoothest one. Above all, due to the usage of the pole-zero model and 1-norm cost function, compared with all-pole model and 2-norm cost, the SD-L1-PZ presents sparser residuals and smoother spectral envelope estimation performance for voiced speech.

The residual estimates of the proposed approach are further tested on real speech signals "Why were you away a year, Roy?" uttered by a female speaker sampled at 8000 Hz. The histograms of the residuals for the 2-norm LP, 1-norm LP, TS-LS-PZ, TS-L1-PZ and SD-L1-PZ are shown in Fig. 4 and Fig. 5, where $\gamma = 0$, $C = \|\mathbf{m}\|_2^2$, $K = 10$, $L = 5$, and the SNR for Gaussian white noise is set to 20 dB. Analysis is performed every 30 ms without overlap. The residuals are obtained using (5) with $q_1 = 1$ for the 1-norm LP, TS-L1-PZ and SD-L1-PZ, but with $q_1 = 2$ for the TS-LS-PZ. As can be seen, the 1-norm-based approach, such as the 1-norm LP, or TS-L1-PZ and SD-L1-PZ, is thinner than the corresponding 2-norm method, which is the 2-norm LP or TS-LS-PZ, respectively. The SD-L1-PZ is the thinnest and highest among all the others, which means the residuals of the SD-L1-PZ are the sparsest.

5. CONCLUSION

A least 1-norm based pole-zero speech analysis method is proposed in this paper. By using the pole-zero model, it can fit the spectral zeros of speech signals easily than all-pole methods. Moreover, sparse residuals are encouraged by applying 1-norm criterion compared with the 2-norm methods. By iteratively updating parameters and residuals using the 1-norm cost and sparse deconvolution, robust coefficient estimates in noisy conditions can be obtained. Simulation results in both synthetic and real speech scenarios show that improved analysis performance in terms of lower spectral distortion, higher reconstruction SNR and sparser residuals can be obtained.

6. REFERENCES

[1] D. Erro, I. Sainz, E. Navas, and I. Hernaez, "Harmonics plus noise model based vocoder for statistical parametric speech synthesis," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 2, pp. 184–194, 2014.

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

[3] M. N. Murthi and B. D. Rao, "All-pole modeling of speech based on the minimum variance distortionless response spectrum," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 3, pp. 221–239, 2000.

[4] T. Drugman and Y. Stylianou, "Fast inter-harmonic reconstruction for spectral envelope estimation in high-pitched voices," *IEEE Signal Process. Lett.*, vol. 21, no. 11, pp. 1418–1422, 2014.

[5] A. El-Jaroudi and J. Makhoul, "Discrete all-pole modeling," *IEEE Trans. Signal Process.*, vol. 39, no. 2, pp. 411–423, 1991.

[6] L. A. Ekman, W. B. Kleijn, and M. N. Murthi, "Regularized linear prediction of speech," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 16, no. 1, pp. 65–73, 2008.

[7] 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. Audio, Speech, and Language Process.*, vol. 20, no. 5, pp. 1644–1657, jul 2012.

[8] T. Shao, Y. R. Zheng, and J. Benesty, "An affine projection sign algorithm robust against impulsive interferences," *IEEE Signal Process. Lett.*, vol. 17, no. 4, pp. 327–330, apr 2010.

[9] L. Shi, Y. Lin, and X. Xie, "Combination of affine projection sign algorithms for robust adaptive filtering in non-gaussian impulsive interference," *Electronics Lett.*, vol. 50, no. 6, pp. 466–467, 2014.

[10] D. Giacobello, M. G. Christensen, T. L. Jensen, M. N. Murthi, S. H. Jensen, and M. Moonen, "Stable 1-norm error minimization based linear predictors for speech modeling," *IEEE/ACM Trans. Audio, Speech, and Language Process.*, vol. 22, no. 5, pp. 912–922, may 2014.

[11] T. L. Jensen, D. Giacobello, T. V. Waterschoot, and M. G. Christensen, "Fast algorithms for high-order sparse linear prediction with applications to speech processing," *Speech Commun.*, vol. 76, pp. 143–156, 2016.

[12] K. SteTiange, "On the simultaneous estimation of poles and zeros in speech analysis," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 25, no. 3, pp. 229–234, 1977.

[13] K. H. Song and K. U. Chong, "Pole-zero modeling of speech based on high-order pole model fitting and decomposition method," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 31, no. 6, pp. 1556–1565, 1983.

[14] D. Marelli and P. Balazs, "On pole-zero model estimation methods minimizing a logarithmic criterion for speech analysis," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 18, no. 2, pp. 237–248, 2010.

[15] J. Durbin, "The fitting of time-series models," *Rev. Int'l Statistical Inst.*, vol. 28, no. 3, pp. 233–244, 1960.

[16] E. Levy, "Complex-curve fitting," *IRE Trans. Automat. Contr.*, vol. AC-4, no. 1, pp. 37–43, 1959.

[17] Y. Miyanaga, N. Miki, and N. Nagai, "Adaptive identification of a time-varying ARMA speech model," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 3, pp. 423–433, 1986.

[18] T. Kobayashi and S. Imai, "Design of IIR digital filters with arbitrary log magnitude function by WLS techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 2, pp. 247–252, 1990.

[19] C. Magi, J. Pohjalainen, T. Backstrom, and P. Alku, "Stabilised weighted linear prediction," *Speech Commun.*, vol. 51, no. 5, pp. 401–411, 2009.

[20] J. Pohjalainen, H. Kallasjoki, K. J. Palomäki, M. Kurimo, and P. Alku, "Weighted linear prediction for speech analysis in noisy conditions," *Proc. Interspeech*, pp. 1315–1318, 2009.

[21] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Prentice Hall, Upper Saddle River, New Jersey, 2004.

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

[23] I. Selesnick, "Sparse deconvolution (an MM algorithm)," Available: http://cnx.org/content/m44991/1.4/, 2012 [Online].

[24] H. Kameoka, N. Ono, and S. Sagayama, "Speech spectrum modeling for joint estimation of spectral envelope and fundamental frequency," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 18, no. 6, pp. 1507–1516, 2010.