A one-dimensional search method with stable 1-norm solution for linear prediction

M. K. Jayesh, and C. S. Ramalingam

Citation: The Journal of the Acoustical Society of America **142**, EL170 (2017); doi: 10.1121/1.4996455
View online: https://doi.org/10.1121/1.4996455
View Table of Contents: https://asa.scitation.org/toc/jas/142/2
Published by the Acoustical Society of America

M. K. Jayesh and C. S. Ramalingam: JASA Express Letters [http://dx.doi.org/10.1121/1.4996455] Published Online 2 August 2017

A one-dimensional search method with stable 1-norm solution for linear prediction

M. K. Jayesh[a] and C. S. Ramalingam
Department of Electrical Engineering, IIT Madras, Chennai 600036, India
jayeshmkoroth@gmail.com, csr@ee.iitm.ac.in

Abstract: In this paper a simple iterative algorithm that is guaranteed to produce a *stable* all-pole filter when minimizing the 1-norm of the linear prediction error signal is proposed. The approach works for both the autocorrelation and covariance frameworks, involves only a one-dimensional search at each step, and obviates the need for linear programming based methods. Based on simulation studies, it was observed that the performance of the algorithm is nearly optimal, i.e., very close to the estimates obtained using interior point methods. Moreover, this method also has the ability to constrain the bandwidth of any peak. The proposed method has been applied for vocal tract estimation and, using spectral distortion as the metric, results are presented using synthetic as well as natural speech.

[DDO]
Date Received: April 5, 2017 **Date Accepted:** July 17, 2017

1. Introduction

In speech processing, the all-pole filter is the most widely used model for the vocal tract. The filter coefficients and gain are computed using the linear prediction (LP) framework.[1,2] The problem being linear in the filter coefficients and the effectiveness of this model have made this framework the workhorse for vocal tract modeling.

Let $x[n]$ be the given segment of speech that we wish to model. Let $\mathbf{x} = (x[n_1], x[n_1+1], \ldots, x[n_2])^T$, $\mathbf{a} = (a_1, a_2, \ldots, a_p)^T$, and $\mathbf{X}$ be a Toeplitz matrix with $(x[n_1-1], x[n_1-2], \ldots, x[n_1-p])$ and $(x[n_1-1], x[n_1], \ldots, x[n_2-1])^T$ as the first row and first column. Minimizing the prediction error $\|\mathbf{e}\|_2^2 = \|\mathbf{Xa} - \mathbf{x}\|_2^2$ leads to $\mathbf{a} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{x}$. Variants such as "autocorrelation method" and "covariance method" arise depending on the assumptions about the data outside the observation window.[2]

Alternatively, one can minimize the ℓ_1-norm of $\mathbf{e}$. This, however, does not admit a closed-form solution. Fortunately, the function is convex, can be cast in a linear programming framework, and solved using the interior point methods.[3,4] But the drawback is that the solution is not guaranteed to produce an all-pole filter that is stable.[5] Nevertheless, finding the ℓ_1-norm solution is important because the all-pole filter so obtained has some desirable properties, e.g., (i) the ℓ_1-norm solution is virtually pitch independent[5] (unlike the pitch locking tendency of the ℓ_2-norm solution[6]), and (ii) speech synthesized using filters based on this norm have a better Mean Opinion Score.[5] Iterative approaches have been proposed to calculate the "true envelope"[7–10] but have not been adopted in practice. In general, the problem of estimating the formants and their bandwidths is a challenging one.[11]

Recently, two algorithms guaranteeing stability for this ℓ_1-norm problem were proposed:[5] the first one works only for the autocorrelation method, whereas, as will be pointed out later in Sec. 3, only the covariance method can give the exact filter if the voiced speech data are truly from an all-pole model. The second algorithm constrains $\|\mathbf{a}\|_1 < 1$, which is sufficient but not a necessary condition for ensuring stability.[5] In Sec. 3 we point out, using a stable synthetic filter with $\|\mathbf{a}\|_1 > 1$, that the filter estimate obtained by constraining $\|\mathbf{a}\|_1 < 1$ results in unacceptable values of spectral distortion (SD) and $\|\mathrm{SD}\|_\infty$.

In this paper we present a simple iterative algorithm for finding the ℓ_1-norm solution that, by its very formulation, (i) guarantees stability and (ii) works for both autocorrelation and covariance methods. Our algorithm is based on Line Spectral Pairs or LSPs (also called as Line Spectral Frequencies or LSFs)[12] and takes advantage

[a]Author to whom correspondence should be addressed.

of their well-known properties.[13,14] LSPs are widely used in speech coding.[15,16] In the context of vocal tract modeling, their localization property has found use in formant modification.[17] Bäckström and Alku[18] used LSPs for all-pole modeling and proposed a method that enhances the level difference between a formant peak and the following spectral valley. This improved the quality of the decoded speech, although it results in an increase in the average residual energy when compared with conventional LPC analysis.

Our comparison metrics are the log SD and its peak. SD (in dB) between two sampled power spectra $S_1[k]$ and $S_2[k]$ is defined as[19]

$$\left(\frac{1}{K}\sum_{k=0}^{K-1}\left[10\log_{10}(S_1[k]) - 10\log_{10}(S_2[k])\right]^2\right)^{1/2}.$$

$\|\mathrm{SD}\|_\infty$ is the peak absolute difference between the dB log magnitude power spectra.

In our experiments we have used voiced speech frames taken from 1 h of the TIMIT database,[20] after downsampling it to 8 kHz. Based on SD and $\|\mathrm{SD}\|_\infty$ we report in Sec. 3 the results of our method and those obtained by using the interior point method of the Convex Optimization Toolbox[21] (denoted by CVX estimate henceforth). Our approach also has the ability to constrain the peakiness of a formant in those cases where the physics of the problem prohibits too sharp a peak (despite its optimality).

2. The proposed algorithm

A given segment of length N is to be modeled as the output of a pth order all-pole filter. The inverse filter's residual signal $e[n]$ is of length $N+p$. We seek that $A(z) = 1 + a_1 z^{-1} + \cdots + a_p z^{-p}$ that minimizes $\|\mathbf{e}\|_1 = \sum_n |e[n]|$. Including or excluding the transients at the two ends leads to the ℓ_1-norm counterparts of the autocorrelation and covariance methods. Since no closed-form solution exists for this problem, we propose an iterative algorithm that involves only a one-dimensional (1D) search at each step.

We always start off with a *stable initial guess* for $1/A(z)$. From this $A(z)$ we form the corresponding LSF polynomials $P(z)$ and $Q(z)$.[13,14] We then take advantage of their well-known properties, viz., (a) the roots of $P(z)$ and $Q(z)$ lie on the unit circle and are interlaced (forming a pair), (b) close LSF pairs correspond to narrow peaks, (c) adjusting an LSF pair alters the filter's frequency response only locally, i.e., only the corresponding peak's center frequency and bandwidth are changed, leaving others far less affected,[22] and (d) even after adjustment, as long as (a) is satisfied, the corresponding $1/A(z)$ continues to be stable. One can also adjust the roots of $P(z)$ and $Q(z)$ individually.

Our proposed algorithm to find the ℓ_1-norm optimal filter consists of *adjusting the LSF roots*. A root is adjusted only if moving it results in lowering the residual's ℓ_1-norm. Any adjustment in the LSF domain is done by *always maintaining the interleaving property of the roots of* $P(z)$ *and* $Q(z)$. This will ensure that the final all-pole filter will always be stable. Moreover, the roots at $z = \pm 1$ are left untouched because they are the constrained ones, also widely known as trivial roots.[13] By its very formulation, the search is 1D.

Adjusting individual LSF roots: A particular LSF root is moved by δ either clockwise or anticlockwise. This is repeated until any further movement causes the norm to only increase. We do the same for the next LSF root, and so on. The entire process is repeated afresh until there is no more reduction in the norm.
Adjusting LSF pairs: A particular LSF root is moved by δ either clockwise or anticlockwise. This is repeated until any further movement causes the norm to only increase. We do the same for the next LSF root, and so on. The entire process is repeated afresh until there is no more reduction in the norm.

The initial filter has LSF roots that are equi-spaced around the unit circle, which by its very formulation ensures stability. Our simulation results show that this is slightly superior to initializing using either the standard LP autocorrelation or covariance methods (an unstable covariance method solution is stabilized by reflecting the outside-unit-circle roots). In this paper, LP1 and LP2 refer to LP based on the ℓ_1 and ℓ_2 norms, respectively.

The step-size δ is a crucial part of the procedure. If it is too large, changing an LSF root or pair by δ will not reduce the norm; if it is too small, a needlessly large number of iterations will be needed. Moreover, for a given step size, it is easy to see that we will reach a point where any change will only cause an increase in the residual

norm. At this point, the step size is halved as many times as needed to help us jump down from the plateau. The whole procedure is repeated until the next saturation point. Such a procedure will end when the change in the error norm falls below the desired threshold.

The starting step size is randomly chosen, i.e., $\delta \in [0, \pi]$. The halving procedure will ensure that large, unhelpful step sizes will quickly give way to ones that start contributing to lowering the residual norm. In fact, more than one starting step size can be chosen: if L different starting step sizes are chosen, leading to L different estimates, the one that results in the lowest norm can be taken as the optimal filter.

In the algorithm given below, p denotes the assumed LP model order, L denotes the number of random starting step sizes, and `tol` is the threshold for the change in the residual error norm (after LSF root adjustment). The error norm is defined as $\|\mathbf{e}\|_1 = \|\mathbf{Xa} - \mathbf{x}\|_1$. The computational steps involved for a given analysis frame at every iteration are: (i) conversion of the LSF roots to $A(z)$, (ii) filtering the frame using $A(z)$, and (iii) computing $\|\mathbf{e}\|_1$ (range of summation depends on whether it is autocorrelation or covariance method).

Algorithm Stable LP1 Solution.

Initialization: p, L, `tol`, equi-spaced LSF roots $\rightsquigarrow A(z)$
for $i = 1 \rightarrow L$ **do**
  Choose $\delta \in [0, \pi]$ randomly
  while reduction in norm $>$`tol` **do**
    for each single root and root pair **do** ▷ *Roots at $z = \pm 1$ are untouched*
      Adjust single root by δ ▷ *Clockwise, anticlockwise*
      Adjust root pair's central angle by δ ▷ *Clockwise, anticlockwise*
      Adjust root pair's angular difference by δ ▷ *Clockwise, anticlockwise*
      Error norm reduced? Retain Adjustment: Discard ▷ *Maintain interleaving property*
    while in plateau region **do** ▷ *See Fig. 2(a)*
      $\delta \leftarrow \delta/2$
Choose $A(z)$ with least error norm.

There is no guarantee that the above procedure will converge to the global minimum because only a 1D search is carried out at each step. However, simulation results on natural speech segments given in Sec. 3 lead us to conjecture that the suboptimal solution involving only 1D searches is nearly optimal.

If the estimated all-pole filter has roots very close to the unit circle, it is well-known that the corresponding LSF roots lie close to each other.[14] However, since the actual vocal tract cannot have formants that are too narrow,[23] ℓ_1-norm estimates with poles very close to the unit circle are in dissonance with the physics of the problem despite being optimal in the given sense. To avoid this we merely place the constraint that the LSF roots cannot get closer to each other beyond a certain limit, thereby preventing estimates with sharp peaks. Simulation results indicate that placing such a constraint can lead to better estimates.

Using LSFs for ℓ_1-norm minimization has been considered for vocal tract estimation by Díaz[24] but in a different manner: (i) the LSFs and **a** are related by a first-order Taylor series approximation and the resulting minimization is *solved using interior point methods*, and (ii) constraints on the closeness of the LSF pairs are passed on as an input to this routine. Moreover, for a given frame, this algorithm requires multiple calls to an interior point method for achieving convergence.

In the weighted sum of LSP polynomials method[18] the authors consider a sum of the form $\lambda P(z) + (1 - \lambda) Q(z)$ with $\lambda \in (0, 1)$ and minimize over the single parameter λ. They too employ the ℓ_1-norm. However, the minimization of the absolute error is between the (normalized) autocorrelations of the given signal and that implied by the estimated filter, whereas in our case the residual error's ℓ_1-norm is minimized.

3. Results and discussion

Consider an eighth order all-pole filter with poles at $0.9902\, e^{\pm j0.1\pi}$, $0.9910\, e^{\pm j0.18\pi}$, $0.9783 e^{\pm j0.34\pi}$, and $0.9782\, e^{\pm j0.5\pi}$, which is a stable filter with $\|\mathbf{a}\|_1 = 79.15$ (illustrating that $\|\mathbf{a}\|_1 < 1$ is sufficient but not necessary). To simulate the high pitch case we excited it with an impulse train having f_0 of 370.37 Hz (period = 27 samples with $F_s = 10$ kHz). The parameters were chosen to ensure that the formants are located in between the pitch harmonics. For analysis, we considered a 25 ms window. In this and all other experiments the input frame was normalized to have unit energy, i.e.,

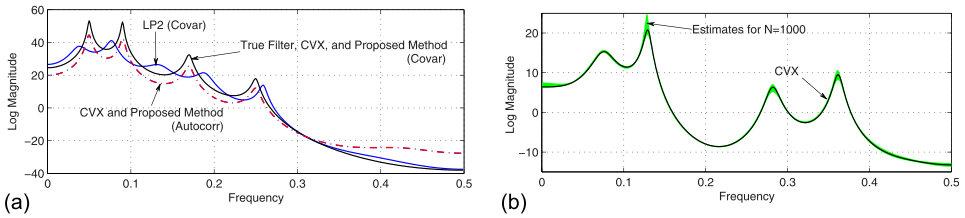

Fig. 1. (Color online) (a) Log magnitude frequency response of true filter and estimates using CVX and the proposed method (autocorrelation and covariance) for an impulse train excitation. The LP2 covariance estimate tends to lock on to the pitch harmonics. (b) For a single frame of natural voiced speech, the estimates for 1000 random starting step sizes using the proposed method lie in the shaded region. The CVX estimate's magnitude response is shown for comparison.

$\|\cdot\|_2 = 1$; the filter gain was also set to unity and this true gain used in subsequent processing. In this work we do not consider the problem of estimating the gain.

The log magnitude frequency response of the true filter, the covariance ℓ_1-norm solutions obtained using the CVX toolbox and our method are shown in Fig. 1(a) for an assumed model order of $p = 12$. Both CVX and our method give the *exact solution*. However, the CVX autocorrelation framework with Hamming window does not, resulting in $\mathrm{SD} = 0.17$ and $\|\mathrm{SD}\|_\infty = 11.29$ with respect to the true filter (rectangular windowed data results are poorer). The proposed method's autocorrelation estimate is close to its CVX counterpart, with the distortion between them being $\mathrm{SD} = 1.7 \times 10^{-4}$ and $\|\mathrm{SD}\|_\infty = 0.25$. Also shown is the estimate obtained using LP2 covariance, which is poor and along expected lines. For this example, constraining $\|\mathbf{a}\|_1 < 1$ resulted in an estimate with $\mathrm{SD} = 23.4$ and $\|\mathrm{SD}\|_\infty = 45.0$, which is clearly unacceptable.

In our simulations, for the synthetic unvoiced case, we found that covariance is better than autocorrelation, with not much difference between the ℓ_1- and ℓ_2-norm solutions.[25] Considering both voiced and unvoiced cases, the covariance ℓ_1-norm framework is the best overall, and the remaining experiments are based on it.

For an impulse train excitation, the estimate is independent of the starting step size. However, in the case of synthetic unvoiced sounds and natural speech examples, it influences the final estimate, albeit only slightly. Figure 1(b) shows the estimates obtained using CVX and the proposed method when analyzing a single frame of natural voiced speech. For our method, 1000 random initial step sizes were chosen in the range $[0, \pi]$. As can be seen from the figure, the variation in the final estimate is small. The CVX residual error was 3.5866, whereas its range in our method is 3.5866–3.5905. The average values of SD and $\|\mathrm{SD}\|_\infty$ are 0.1375 and 0.5893, respectively. This example illustrates the fact that the suboptimal solution is nearly optimal, a point that we will return to later.

Some insight into convergence in these cases can be obtained from Fig. 2(a). Here, for a different voiced segment of natural speech, the residual signal's ℓ_1-norm has been plotted as a function of iteration number. The curves correspond to updating single roots, root pairs, and both. Also shown is the error norm for the CVX solution. The presence of local minima is illustrated with a different natural voiced speech example. For this frame, Fig. 2(b) shows the LSF roots at convergence, along with those obtained using CVX. The circled pair corresponds to the roots with the largest difference between our method and CVX. The error norm in our method was 3.9111, whereas for CVX it was 3.9108. If the roots that are farthest away from the CVX solution are made to coincide with the latter, the error norm *increases* to 3.9180, which

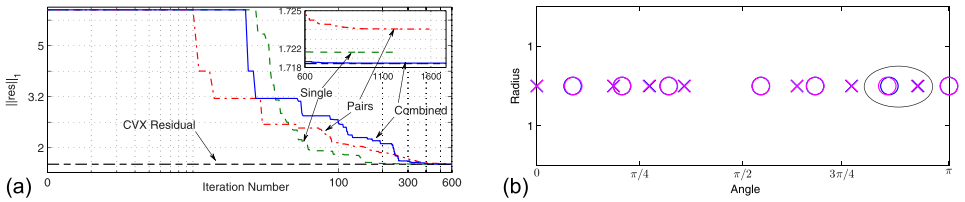

Fig. 2. (Color online) (a) Reduction in residual norm for a single frame of natural voiced speech as a function of iteration number. The step size is halved as many times as needed to help us jump down from a plateau. (b) LSF roots of the estimate from the proposed method and that from CVX. The pair with the largest difference is circled. Making them coincide *increases* the error norm, pointing to the presence of a local minimum.

Table 1. Average values of SD, $\|\mathrm{SD}\|_\infty$, number of steps ($\mathtt{tol} = 10^{-7}$), and residual error difference for 144 000 frames of natural voiced speech.

Update strategy	SD	$\|\mathrm{SD}\|_\infty$	Avg. steps	Avg. Err. diff.
Single root	0.126	0.400	7.55×10^3	7.17×10^{-4}
Root pair	0.189	0.569	7.75×10^3	1.90×10^{-3}
Combined	0.112	0.354	9.06×10^3	5.39×10^{-4}

shows that a multidimensional search is needed to reach the minimum obtained by CVX.

We now present results for 144 000 frames of natural voiced speech taken from TIMIT (8 kHz downsampled version). The analysis frame size was 25 ms and a tenth order all-pole model was fitted using both the CVX Toolbox and the proposed method. The natural speech segments were pre-emphasized by filtering them using $1 - z^{-1}$. The frames were chosen such that no unstable filter was obtained when using CVX. For our method, the convergence tolerance was $\mathtt{tol} = 10^{-7}$. For each frame, $L = 10$ random step sizes were chosen. In Table 1 we give the values of SD and $\|\mathrm{SD}\|_\infty$ between our method and CVX (averaged over 144 000 frames). Also shown are the average differences in the time-domain norm. The convergence depends on the update strategy, with the best solution obtained by combining single-root and pair-updating strategies. The corresponding histograms of SD and $\|\mathrm{SD}\|_\infty$ are given in Fig. 3 for all three update strategies. For all the natural speech frames that were considered we observed that both CVX and our method gave stable filter estimates with $\|\mathbf{a}\|_1 > 1$. Using CVX, constraining $\|\mathbf{a}\|_1 < 1$ (as proposed by Giacobello *et al.*[5]), resulted in much larger average values of SD (=4.28) and $\|\mathrm{SD}\|_\infty$ (=10.0), where the reference spectrum corresponds to the CVX filter with no constraint.

Additional results not reported here due to lack of space but detailed elsewhere[25] are: (i) initialization using estimates from the LP2 autocorrelation and covariance methods gave a slightly higher SD compared to initializing with a filter whose LSF roots distribution is equi-spaced; (ii) SD increases with model order; hence, at higher sampling frequencies, the distortion will be more because of the higher model order needed; (iii) as p was increased, the computational time also increases roughly linearly, whereas for CVX it increases much more slowly; (iv) as L was increased, both SD and $\|\mathrm{SD}\|_\infty$ decreased, as one would expect (for this database $L = 10$ was found to be a good compromise between getting closer to the CVX estimate versus computational burden).

Finally, a set of experiments using synthetic unvoiced speech frames illustrates the benefits of constrained minimization. The synthetic all-pole filter given earlier was excited with Gaussian noise and 10 000 frames of length 25 ms were used for analysis with sampling frequency 10 kHz, $L = 10$, $p = 12$, and equi-spaced LSF roots initialization. For the true filter's LSF roots, the smallest angular separation was 0.032. We employed the constraint that no two roots can get closer than 0.02. (For natural speech, the narrowest formant bandwidth is 30 Hz,[23] which can be used for determining the corresponding constraint.)

Table 2 gives the results for SD and $\|\mathrm{SD}\|_\infty$. With the constraint in place, the proposed method performs better than CVX for both the stable and unstable cases. It was observed that CVX gave unstable estimates in 12.4% of the frames that were analyzed, whereas for the LP2 covariance method this number was 1.9%. For both stable and unstable estimates, SD is lower for the covariance estimate compared to the

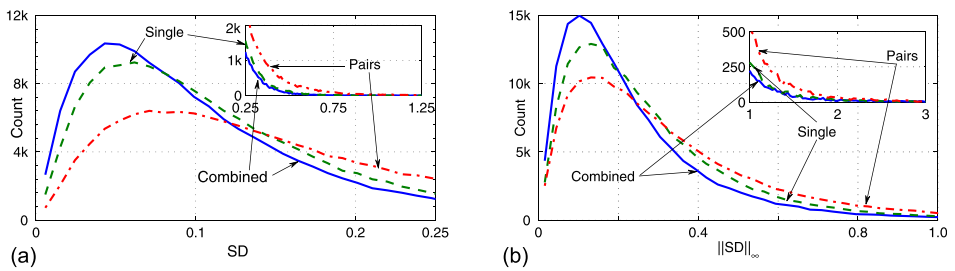

Fig. 3. (Color online) Natural speech results: (a) Histograms of SD with ten different random starting step sizes and choosing the best solution. (b) Corresponding histograms for $\|\mathrm{SD}\|_\infty$. Combined adjustment gives the best results.

Table 2. Values of average SD and $\|SD\|_\infty$ (between the estimated and true filters) for a synthetic unvoiced example using formant bandwidth constraint, averaged over 10 000 trials.

	Stable			Unstable (CVX, LP2 Cov)			
	CVX	LP2 Cov	New	CVX	New	LP2 Cov	New
SD	1.72	1.39	1.65	1.85	1.64	1.51	1.59
$\|SD\|_\infty$	11.13	8.90	8.64	17.93	10.18	19.46	10.26

proposed one, whereas based on the $\|SD\|_\infty$ measure, the proposed method is superior. These results are similar to those reported[24] using CVX, where the constraints are passed on as a condition to the optimization toolbox.

For $p > 6$ it was observed that CVX was faster than the proposed method. For example, the average computation time for a single frame of natural speech for $p = 10$ and $L = 1$ was 0.42 s compared to CVX's 0.23 s [running on Intel's Xeon CPU X3470 2.93 GHz, averaged over 1000 frames; MATLAB version 7.14.0.739 (R2012a)]. For large model orders, the proposed method is limited by the numerical problems that arise in converting LSFs to $A(z)$. Since our focus is on vocal tract analysis, the order p will be dictated by the number of formants to be modeled (typically not more than five[23]). For natural speech segments, despite the presence of local minima (and hence the inability to match the CVX solution exactly) the average SD is small, which will not give rise to any perceptual difference[26] in the reconstructed speech.

4. Conclusion

In this paper, we have presented a method that does not use linear programming to solve the ℓ_1-norm LP equations. It involves a 1D search algorithm that guarantees a stable solution, and is based on adjusting the LSF roots. The method works for both autocorrelation and covariance frameworks, with the latter producing the exact solution if the voiced speech segment were truly from an all-pole model. When applied to vocal tract estimation, simulation results using synthetic and natural speech indicate that this method produces estimates that are very close to the ones given by using interior point methods. We presented different update strategies for adjusting the LSF roots. Our method can incorporate naturally the ability to control the peakiness of any formant; when analyzing synthetic unvoiced examples we found that placing constraints on the bandwidth leads to better estimates compared with the unconstrained problem. Because the method requires only a 1D search at each step, it can easily be implemented on a DSP processor.

References and links

[1]B. S. Atal and S. L. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," J. Acoust. Soc. Am. **50**(2), 637–655 (1971).

[2]D. O'Shaughnessy, *Speech Communications: Human and Machine*, 2nd ed. (IEEE Press, Piscataway, NJ, 2000).

[3]S. J. Wright, *Primal-Dual Interior-Point Methods* (SIAM, Philadelphia, PA, 1997).

[4]S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, United Kingdom, 2004).

[5]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, Lang. Process. **22**(5), 912–922 (2014).

[6]J. Makhoul, "Linear prediction: A tutorial review," Proc. IEEE **63**(4), 561–580 (1975).

[7]S. Imai and Y. Abe, "Spectral envelope extraction by improved cepstral method," IEICE Trans. A **J62-A**(4), 217–223 (1979) (in Japanese).

[8]F. Villavicencio, A. Robel, and X. Rodet, "Improving LPC spectral envelope extraction of voiced speech by true-envelope estimation," in *IEEE International Conference on Acoustics Speech and Signal Processing Proceedings* (2006), Vol. 1, pp. I-869–I-872.

[9]H. Kameoka, N. Ono, and S. Sagayama, "Speech spectrum modeling for joint estimation of spectral envelope and fundamental frequency," IEEE Trans. Audio, Speech, Lang. Process. **18**(6), 1507–1516 (2010).

[10]L. Shi, J. R. Jensen, and M. G. Christensen, "Least 1-norm pole-zero modeling with sparse deconvolution for speech analysis," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, New Orleans, LA (2017).

[11]D. D. Mehta and P. J. Wolfe, "Statistical properties of linear prediction analysis underlying the challenge of formant bandwidth estimation," J. Acoust. Soc. Am. **137**(2), 944–950 (2015).

[12]F. Itakura, "Line spectrum representation of linear predictor coefficients of speech signals," J. Acoust. Soc. Am. **57**, S35 (1975).

[13]T. Bäckström and C. Magi, "Properties of line spectrum pair polynomials—A review," Signal Process. **86**(11), 3286–3298 (2006).
[14]I. V. McLoughlin, "Review: Line spectral pairs," Signal Process. **88**(3), 448–467 (2008).
[15]W. C. Chu, *Speech Coding Algorithms: Foundation and Evolution of Standardized Coders* (John Wiley & Sons, New York, 2003).
[16]A. M. Kondoz, *Digital Speech*, 2nd ed. (John Wiley & Sons, West Sussex, England, 2005).
[17]R. W. Morris and M. A. Clements, "Modification of formants in the line spectrum domain," IEEE Signal Process. Lett. **9**(1), 19–21 (2002).
[18]T. Bäckström and P. Alku, "All-pole modeling technique based on weighted sum of LSP polynomials," IEEE Signal Process. Lett. **10**(6), 180–183 (2003).
[19]L. A. Ekman, W. B. Kleijn, and M. N. Murthi, "Regularized linear prediction of speech," IEEE Trans. Audio, Speech, Lang. Process. **16**(1), 65–73 (2008).
[20]J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett, "Darpa TIMIT acoustic-phonetic continous speech corpus cd-rom. NIST speech disc 1-1.1," NASA STI/Recon Technical Report No. 93 (1993).
[21]M. Grant and S. Boyd, CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx (2014) (Last viewed July 2017).
[22]K. Paliwal, "On the use of line spectral frequency parameters for speech recognition," Digital Signal Process. **2**(2), 80–87 (1992).
[23]J. R. Deller, J. H. L. Hansen, and J. G. Proakis, *Discrete-Time Processing of Speech Signals* (Wiley-IEEE Press, New York, 2000).
[24]A. C. Díaz, "New insights on speech signal modeling in a Bayesian framework approach," Master's degree project, KTH Royal Institute of Technology, Stockholm, Sweden, 2015.
[25]M. K. Jayesh, "New methods for vocal tract analysis," Ph.D. thesis, IIT Madras, 2017.
[26]T. L. Jensen, D. Giacobello, T. van Waterschoot, and M. G. Christensen, "Fast algorithms for high-order sparse linear prediction with applications to speech processing," Speech Commun. **76**, 143–156 (2016).