THE PROBLEMS OF ROBUST LPC PARAMETRIZATION FOR SPEECH CODING

Petr Pollák & Pavel Sovka
Czech Technical University of Prague
CVUT FEL K331, 166 27 Praha 6, Czech Republic
E-mail: pollak@noel.feld.cvut.cz

Abstract

The problems and possible solutions of robust LPC parametrization in the case of noise presence are discused. The experience with preprocessing using standard speech enhancement techniques is mentioned. These techniques are based on spectral subraction and its modifications. The possibility of correlation subtraction is discussed in the second part. This problem is solved for the case of white noise becasue white noise variance seems to be efficiently estimated from correlation matrix of noisy speech.

1 INTRODUCTION

Linear Predictive Coding (LPC) [4], [3] analysis is frequently used tool for parametrization of speech signal. Nevertheless, the standard technique is not too robust to noise, the solution is biased so futher processing with biased parameters is degraded or it completely fails.

The analyzed signal $s[n]$ is assumed to be predicted by the linear predictor of order p as

$$\widehat{s}[n] = -\sum_{k=1}^{p} a[k]s[n-k]. \tag{1}$$

The solution of predictor coefficients $a[k]$ is based on minimalization of the mean square prediction error and it satisfies normal equations in the general matrix form

$$\mathbf{R}_{ss} \cdot \mathbf{a} = -\mathbf{r}_s. \tag{2}$$

We assume autocorrelation method of solution, the normal equations are called Yule-Walker equations, the matrix $\mathbf{R}_{ss}$ is the symetic equidiagonal (Toeplitz) *correlation matrix* of correlations $R[0]$, ..., $R[p-1]$, and $\mathbf{r}_s$ is the vector of correlations $R[1]$, ..., $R[p]$.

When we assume speech signal corrupted by uncorrelated additive noise $x[n] = s[n] + n[n]$ we obtaine correlation matrix of noisy speech $\mathbf{R}_{xx}$ and Yule-Walker equations yield biased solution of predictor coefficients $\mathbf{a}'$, i.e.

$$\mathbf{R}_{xx} \cdot \mathbf{a}' = -\mathbf{r}_x \tag{3}$$

We have studied two different noise removal technique which are described in following two sections.

2 PREPROCESSING WITH SPECTRAL SUBTRACTION

The preprocessing with standard speech enhancement was the first possibility of robust parametrization technique which was tested. We have studied many modifications with more detailed analysis as speech enhancement algorithms for transmission with respect to speech distortion and level of noise suppression [5], [6].

We realized the experiments with isolated word recognizer realized by HTK tools under noise condittion. We used different modifications of spectral subtraction as first step before parametrization and regognition. The resultes were published in [7] and they can be summarized:

- Generally, these speech enhancement algorithms improve input SNR of processed speech and consequently enhance the initial condition of noisy speech for further parametrization. It does not produce exactly noise level independent robust parametrization.
- The speech distortion finally decreases the rate of speech recognition. The level of correct speech recognition rate for clean speech cannot be achieved.
- The negative influence of the distortion produced by spectral subtraction can be decreased by training on speech preprocessed by same algorithm.

- The enhanced signal needs not be generated. The autocorrelation coefficients can be evaluated from power spectral density of enhanced speech and then used in the solution of Yule-Walker equations.
- These techniques are very efficient from the point of view of computational cost.

3 NOISE CORRELATION COMPENSATION TECHNIQUE

This technique of the robust pametrization is derived from the algorithm in [2] by Gesbert et al. and it is based on noise correlation compensation (correlation subtraction). The basic problem is the estimation of noise correlation again.

3.1 Theoretical solution for white noise case

The theoretical solution in the presence of white additive noise seems to be very simple. Since the noise is assume to be white, its correlation matrix is theoretically diagonal with all values equaled to $R_b[0] = \sigma^2$ and the autocorrelation coefficients satisfy $R_x[0] = R_s[0] + R_b[0]$ and $R_x[k] = R_s[k]$ for $k > 0$, i.e. the Yule-Walker equations for biased solution take form

$$(\mathbf{R}_{ss} + \sigma^2\mathbf{I}) \cdot \mathbf{a}' = -\mathbf{r}_x \,. \tag{4}$$

When we know the noise variance σ^2 we can look for the unbiased solution from the Yule-Walker equations modified as

$$(\mathbf{R}_{xx} - \sigma^2\mathbf{I}) \cdot \mathbf{a}'' = -\mathbf{r}_x \,. \tag{5}$$

In this case, the problem is simplified to estimate noise variance σ^2.

3.2 Estimation of white noise variance

For strongly correlated signals the eigenvalues of correlation matrix fall rapidly to zero. We assume standard ortonormal matrix decomposition where $\mathbf{D}$ is the diagonal matrix of eigenvalues fulfilled the condition

$$\mathbf{R}_{ss} = \mathbf{U} \cdot \mathbf{D} \cdot \mathbf{U}^T \tag{6}$$

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r = \ldots \lambda_p = 0 \,. \tag{7}$$

The idea of the described technique is in the estimation of noise variance from the eigenvalues of noisy correlation matrix. Since the correlation matrix of white noise is diagonal, it must be fulfilled

$$\mathbf{R}_{xx} = \mathbf{U} \cdot (\mathbf{D} + \sigma^2\mathbf{I}) \cdot \mathbf{U}^T \tag{8}$$

so the noise variance can be estimated from minimal eigenvalue of noise correlation matrix as

$$\sigma^2 = \lambda_{\min} \tag{9}$$

Problems:

- The last eigenvalue of speech correlation matrix is not exactly equal to zero so the last eigenvalue of noisy correlation matrix should be close to σ^2 with some error.

$$\lambda_{x1} \geq \lambda_{x2} \geq \ldots \geq \lambda_{xp} \approx \sigma^2 \,. \tag{10}$$

- Let we assume that estimation of noise variance is possible to obtain from the minimal eigenvalue of noisy speech correlation matrix, eq. (9). The ratio flow-graph of minimal eigenvalue normalized to the sum of all ones is shown on fig. 1. This minimal eigenvalue seems to be very small especially for voiced parts of speech so the minimal eigenvalue can be mentioned very close to zero. It is shown on fig. 2 that the minimal eigenvalue is higher in noisy case. The dash-line shows how much it is close to noise variance (the real σ^2 is used instead of λ_{min}).

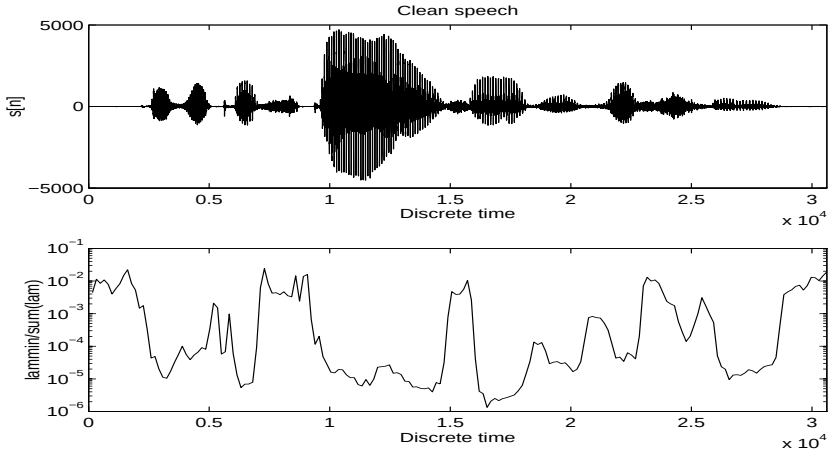

Figure 1: Order of magnitude fluctuation for clean speech.

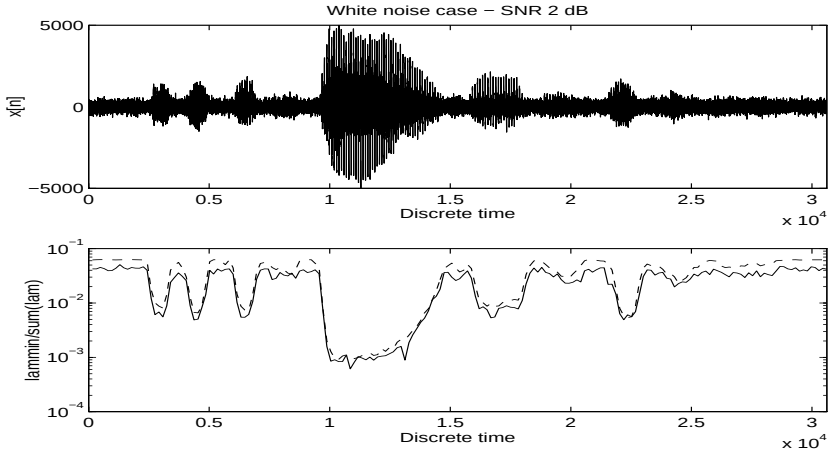

Figure 2: Order of magnitude fluctuation for noisy speech.

3.3 Basic technique

This method is based exactly according to the theoretical solution:

$$(\mathbf{R}_{xx} - \lambda_{\min}\mathbf{I}) \cdot \mathbf{a}'' = -\mathbf{r}_x \,. \tag{11}$$

The matrix $\mathbf{R}_{xx} - \lambda_{\min}\mathbf{I}$ is singular so it is not possible to find a solution by using simple inverse of modified matrix or using Levinson-Durbin algorithm respectively. It must be used *pseudoinverse* of the modified matrix, i.e. *inverse in Moore-Penrose sense*. More precisely, for non-singular matrices which satisfy $\mathbf{R}_{mm} = \mathbf{U} \cdot \mathbf{D} \cdot \mathbf{U}^T$ the inverse is

$$\mathbf{R}_{mm}^{-1} = \mathbf{U} \cdot \mathbf{D}^{-1} \cdot \mathbf{U}^T \,, \tag{12}$$

$$\mathbf{D}^{-1} = diag(1/\lambda_1,\, 1/\lambda_2,\, ...,\, 1/\lambda_p) \,. \tag{13}$$

For singular case a number of last eigenvalues are equal to zero, $\mathbf{D} = diag(\lambda_1,\, ...,\, \lambda_r,\, 0, ..., 0)$, and the inverse must be found as

$$\mathbf{R}_{mm}^{-1} = \mathbf{U} \cdot \mathbf{D}^{-1} \cdot \mathbf{U}^T \,, \tag{14}$$

$$\mathbf{D}^{-1} = diag(1/\lambda_1,\, ...,\, 1/\lambda_r,\, 0, ..., 0) \,. \tag{15}$$

Problems:

- The results are better for the voiced speech sequences. Absolute cepstral distance from the original clean speech is less but the distance of successive frames increases. It confirms typical behaviour demonstrated on fig. 3.

- The solution is not always stable (the reflection coefficients have modul greater than 1, see fig. 6. The polynomial stabilization with respect of unit circle must be used.
- The most important problem (the unstability of solution) appears already in the case of clean speech. But spectral characteristic of unstable model are close to the unbiased solution for clean speech. When the polynomial stabilization is used the cepstral coefficients can be computed which seem to converge to the unbiased solution.
- The exact noise variance estimation was also used instead of $\lambda_{\min}$ so the error of its estimation was eliminated but the solution remains practicaly same. It shows that the correlation matrix non-zero non-diagonal elements of noise correlation matrix have not negligible influence.

3.4 Scaled factor modification

This approach tests the behaviour when not all λ_{min} is subtracted. The scaling constant q of λ_{min} is used. Generally, it could be in the range $0 < q < 1$ but the typical value is $0.8 \div 0.9$.

$$(\mathbf{R}_{xx} - q \cdot \lambda_{\min} \cdot \mathbf{I}) \cdot \mathbf{a}'' = -\mathbf{r}_x \,. \tag{16}$$

The problem of modified correlation matrix singularity is overcome in this case. The standard technique (Levinson-Durbin) can be used.

Problems:

- For the values of q close to 1, the results start to be close to above decribe basic technique with mentioned problems.
- Some level of noise rests in the analyzed signal. This depends on parameter q and σ^2.

3.5 Experiments with modeling in the correlation domain

We assume the white noise uncorrelated with the speech signal. Since we had some problems to fulfill this condition in the environment of MATLAB we tried to model directly the correlation matrices. The correlation matrix of white noise is modeled as diagonal matrix of noise variance $\sigma^2 \cdot \mathbf{I}$ and the Toeplitz matrix of random values $\mathbf{E}$[1].

$$\mathbf{R}_{xx} = \mathbf{R}_{ss} + \sigma^2 \cdot \mathbf{I} + \mathbf{E} \,. \tag{17}$$

For the experiment which results are shown in tab. 5 the random values were scaled by $\sigma^2 \cdot 10^{-3}$.

Comments:

- The algorithm starts giving better results in the case of diagonal noise correlation matrix modeling when non-zero non-diagonal and cross-correlation elements were ommitted.
- Increasing of window length should decrease influence of undesirable cross-correlation and non-diagonal elements. But it is not possible to increase the window length without the respect of stationary parts of speech.

4 EXPERIMENTS

4.1 Classification criteria.

Since the cepstral coefficients computed by the recursion from the predictors coefficients are used usually as parameters for speech recognizers the *Euclidan cepstral distance* of two cepstral vectors were used as classification criterion. Let us assume $c[k]$ - *unbiased parameters*, i.e. the vectors of parameters of standard technique for signal without any noise, $c'[k]$ - *biased parameters* for speech with noise, and $c''[k]$ - *robust parameters* of noise removal technique. Then we can compute *cepstral distance from unbiased solution*,

$$cd'_i = \sum_{k=1}^{p} (c_i[k] - c'_i[k])^2 \quad ; \qquad cd''_i = \sum_{k=1}^{p} (c_i[k] - c''_i[k])^2 \tag{18}$$

[1] Without these random values the solution of the problem is trivial, i.e. only adding and subtraction of diagonal matrix

and *cepstral distance of successive frames*

$$cds'_i = \sum_{k=1}^{p} (c'_i[k] - c'_{i+1}[k])^2 \quad ; \qquad cds''_i = \sum_{k=1}^{p} (c''_i[k] - c''_{i+1}[k])^2 \,. \tag{19}$$

The cds'' must be compared always to cds' of clean speech. Distance of successive frames for noisy speech is very low because of noise masking.

4.2 Results of experiments

All discussed techniques were tested on small database of typical speech segments. The database was created from the TIMIT database, i.e. american English database. Experiments with this datbase was realized in MATLAB. Further, the speech was mixed with white noise generated by MATLAB under different levels. Four constant noise levels were used approximatelly $14dB$, $8dB$, $2dB$, and $-6dB$ computed as long-time global SNR respectively.

The results of two basic versions of speech enhancement preprocessors are presented in tab. 1 - algorithm with PSD subtraction (Wiener filter) - and in tab. 2 - algorithm with spectral magnitude subtraction. It is shown that the distance from unbiased solution is less but still relatively high. On the other hand the cepstral distance of successive frames does not increase so much.

	Averaged value					Standard deviation				
SNR [dB]	∞	14	8	2	-6	∞	14	8	2	-6
cd'	0.0000	1.3358	2.1448	3.0706	4.2791	0.0000	0.8942	1.1351	1.3456	1.5427
cd''	0.0101	0.8100	1.4569	2.2991	3.5598	0.0490	0.6807	0.9413	1.1827	1.4612
cds'	0.1148	0.0646	0.0637	0.0633	0.0639	0.1217	0.0323	0.0257	0.0221	0.0215
cds''	0.1061	0.1676	0.1972	0.2274	0.2668	0.1077	0.0866	0.0938	0.0956	0.1026

Table 1: Global results for preprocessing by Wiener filtering

	Averaged value					Standard deviation				
SNR [dB]	∞	14	8	2	-6	∞	14	8	2	-6
cd'	0.0000	1.3358	2.1448	3.0706	4.2791	0.0000	0.8942	1.1351	1.3456	1.5427
cd''	0.0101	0.5255	1.0374	1.7921	3.0686	0.0490	0.5126	0.7780	1.0532	1.4000
cds'	0.1148	0.0646	0.0637	0.0633	0.0639	0.1217	0.0323	0.0257	0.0221	0.0215
cds''	0.1061	0.2234	0.2659	0.3151	0.3769	0.1077	0.1189	0.1298	0.1361	0.1503

Table 2: Global results for preprocessing by magnitude spectral subtraction

The results of experiments with correlation compensation techniques are in tabs. 3 - 5. It is shown that especially the distance od successive frames increased as consequence of mentioned problems. The results of experiments with modeling in the correlation domain are much more better. Especially when we see at fig. 5 - 10 the robust solution seems to converge to the unbiased one.

	Averaged value					Standard deviation				
SNR [dB]	∞	14	8	2	-6	∞	14	8	2	-6
cd'	0.0000	1.3358	2.1448	3.0706	4.2791	0.0000	0.8942	1.1351	1.3456	1.5427
cd''	0.7645	1.1752	1.6589	2.3273	3.4314	1.1315	0.9162	1.1255	1.4084	1.5528
cds'	0.1148	0.0646	0.0637	0.0633	0.0639	0.1217	0.0323	0.0257	0.0221	0.0215
cds''	0.4492	1.4438	1.7949	2.1985	2.3821	0.7462	1.0135	1.1144	1.2450	0.9185

Table 3: Global results for basic technique

	Averaged value					Standard deviation				
SNR [dB]	∞	14	8	2	-6	∞	14	8	2	-6
cd'	0.0000	1.3358	2.1448	3.0706	4.2791	0.0000	0.8942	1.1351	1.3456	1.5427
cd''	0.9181	0.7591	1.3171	2.1070	3.3374	0.5171	0.5937	0.8829	1.1577	1.4497
cds'	0.1148	0.0646	0.0637	0.0633	0.0639	0.1217	0.0323	0.0257	0.0221	0.0215
cds''	0.4043	0.5813	0.6737	0.7587	0.8274	0.4093	0.3160	0.3015	0.3148	0.3029

Table 4: Global results for scaled subtraction method

	Averaged value					Standard deviation				
SNR [dB]	∞	14	8	2	-6	∞	14	8	2	-6
cd'	0.0000	1.3353	2.1435	3.0663	4.2692	0.0000	0.8973	1.1377	1.3405	1.5203
cd''	0.7645	0.5937	0.6366	0.8058	1.3413	1.1315	0.6517	0.6843	0.7607	1.1475
cds'	0.1148	0.0286	0.0191	0.0123	0.0067	0.1217	0.0306	0.0229	0.0173	0.0109
cds''	0.4492	0.5472	0.7203	1.0150	1.5260	0.7462	0.6201	0.7544	0.9576	1.0194

Table 5: Global results for experiments with modeled correlation matrix

5 CONCLUSIONS

The problems and comments connected to the particular parts of this study were mentioned above. They can be summarized by following points.

- However some global criteria did not give the satisfied results the described approach seems to converge to the unbiased solution.
- Great differences between signal in the generated database may deteriorate the results. Correct tests seem to be realized only by the application in the recognition system and by comparison with results for clean, noisy, and robust parametrization technique.
- The robust parameters gives the unstable model which had to be stabilized with the respect of unit circle. The more detailed analysis of robust parametrization for speech without any noise with the focuse to the unstability of the solution.

REFERENCES

[1] S. F. Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-27(2):113–120, April 1979.

[2] D. Gesbert, P. Duhamel, and M. Unser. Régression linéare en présence de bruit: Une solution adaptative non-biasée. In *GRETSI*, France, 1995.

[3] J. Makhoul. Linear prediction: A tutorial review. *Proceedings of the IEEE*, 63(4):561–580, April 1975.

[4] J. D. Markel and A. H. Gray Jr. *Linear Prediction of Speech.* Springer-Verlag, 1976.

[5] P. Pollák, P. Sovka, and J. Uhlíř. Noise suppression system for a car. In *Proceedings of the 3rd European Conference on Speech, Communication, and Technology - EUROSPEECH'93*, pages 1073–1076, Berlin, September 1993.

[6] P. Sovka, P. Pollák, and J. Kybic. Extended spectral subtraction. In *EUSIPCO'96*, Trieste, September 1996.

[7] T.Kreisinger, P. Pollák, P. Sovka, and J. Uhlíř. Study of speech recognition in noisy environment. In *The European Conference on Signal Analysis and Prediction*, Praha, June 1997.

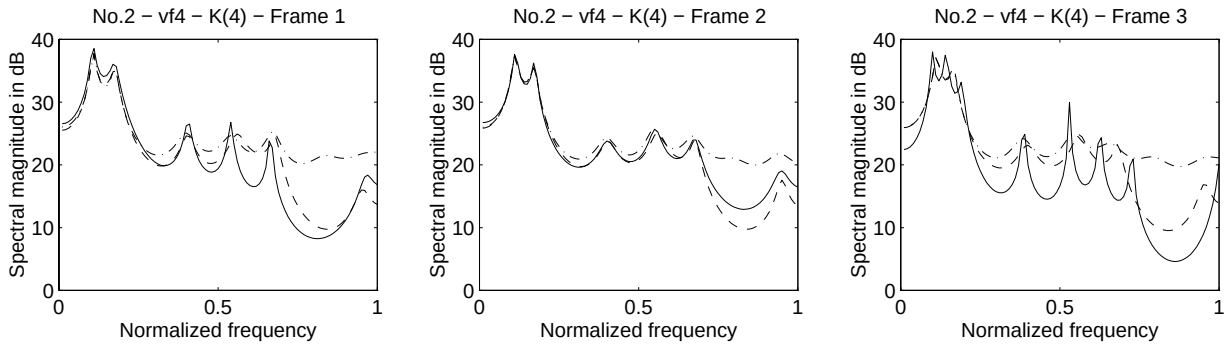

Figure 3: One signal demonstration for basic technique: dashed curve - original clean speech, dash-dotted curve - noisy speech, solid curve - robust modeled noisy speech

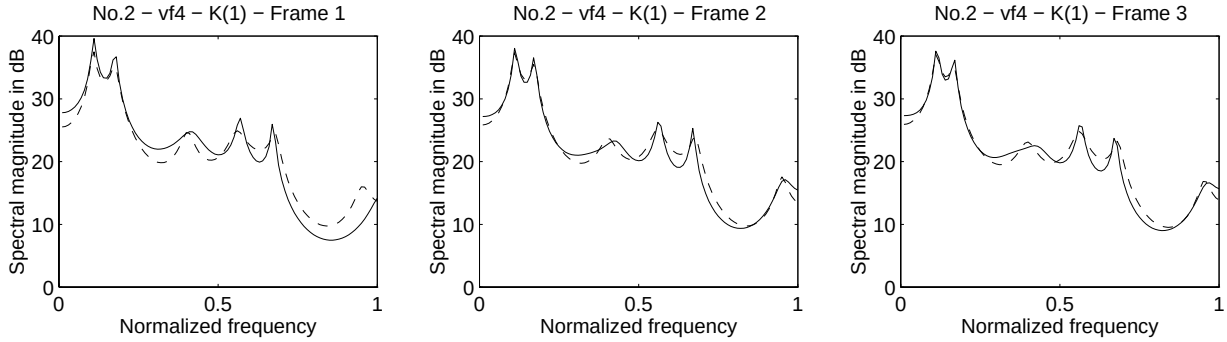

Figure 4: Demostration for clean speech and basic technique: dashed curve - original clean speech, solid curve - robust modeled noisy speech

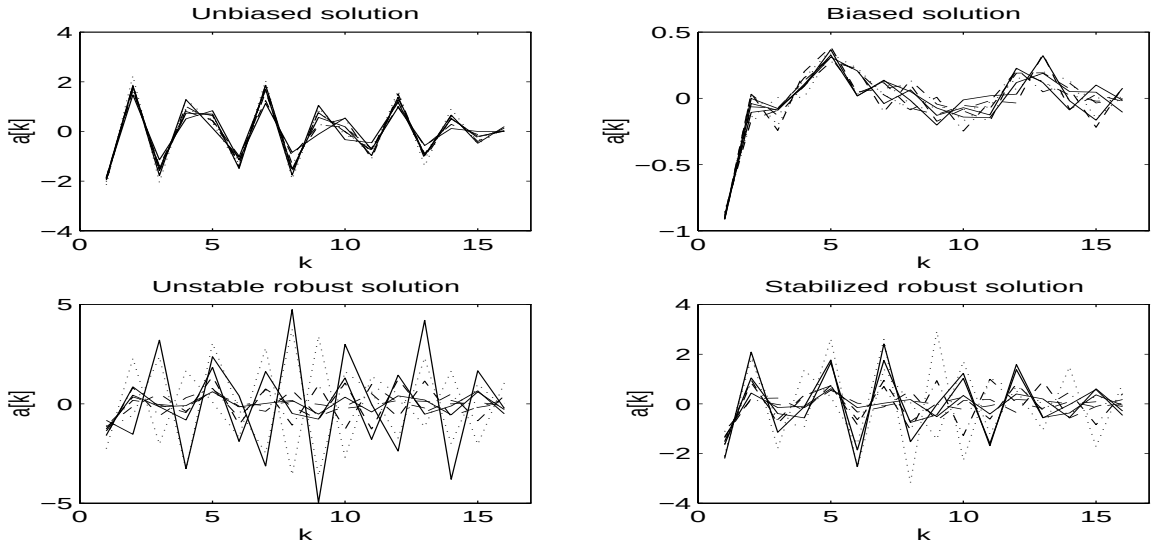

Figure 5: Basic technique - predictor coefficients over one signal.

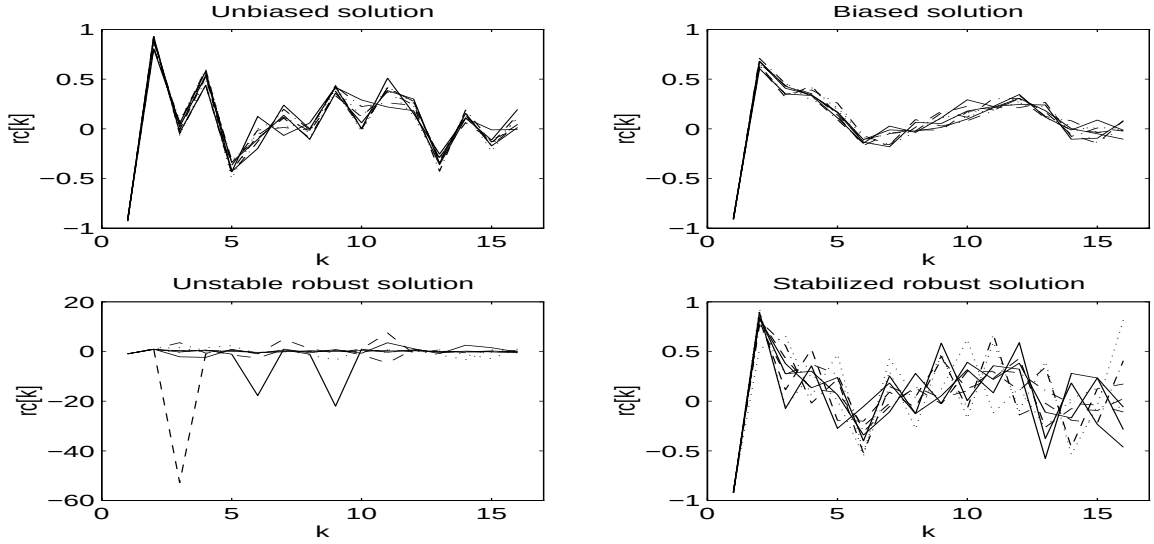

Figure 6: Basic technique - reflection coefficients over one signal.

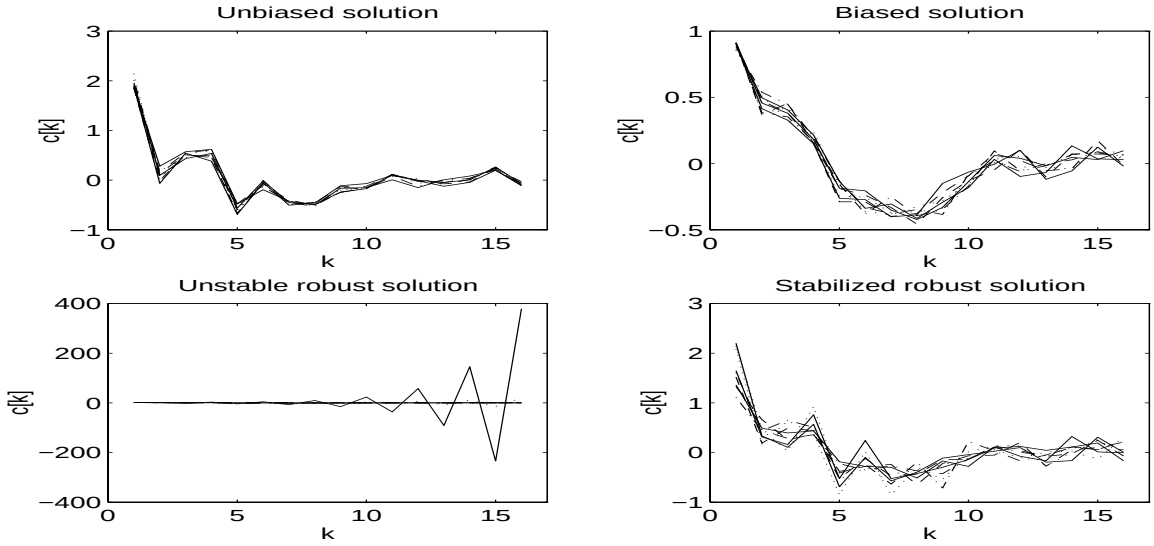

Figure 7: Basic technique - cepstral coefficients over one signal.

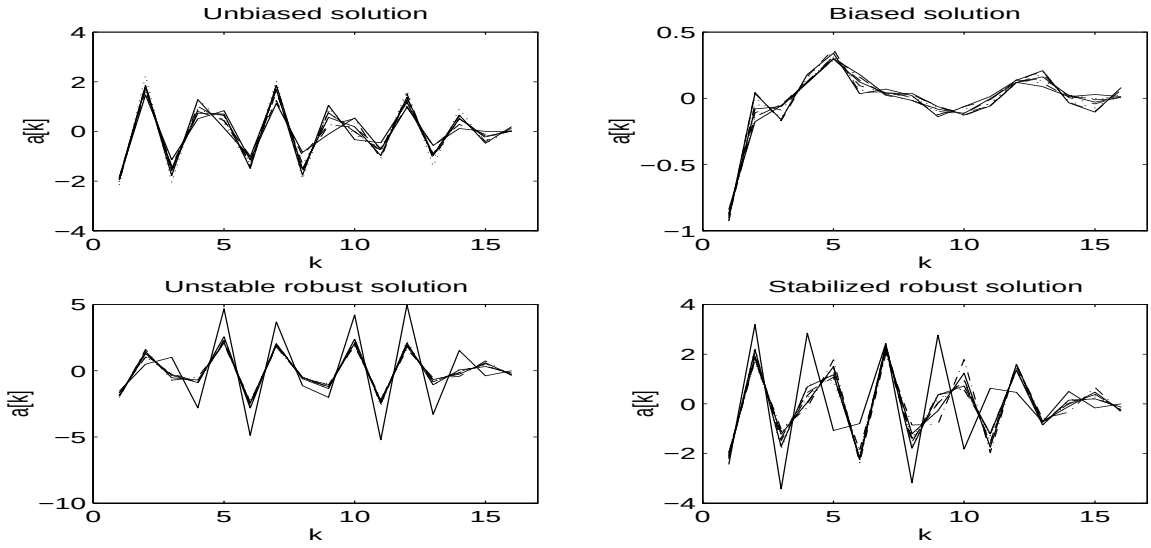

Figure 8: Correlation modeling - predictor coefficients over one signal.

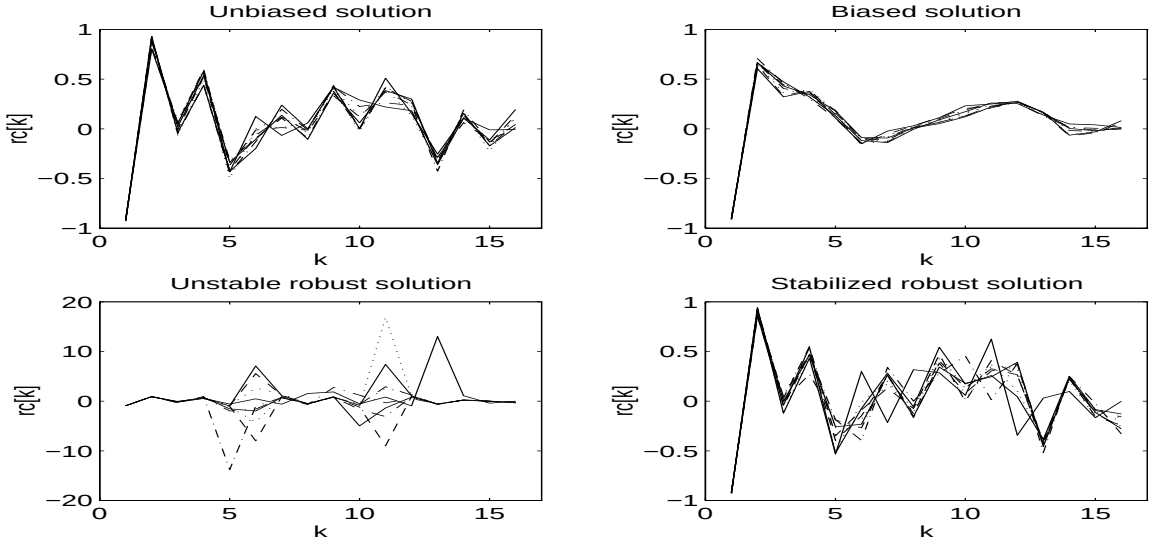

Figure 9: Correlation modeling - reflection coefficients over one signal.

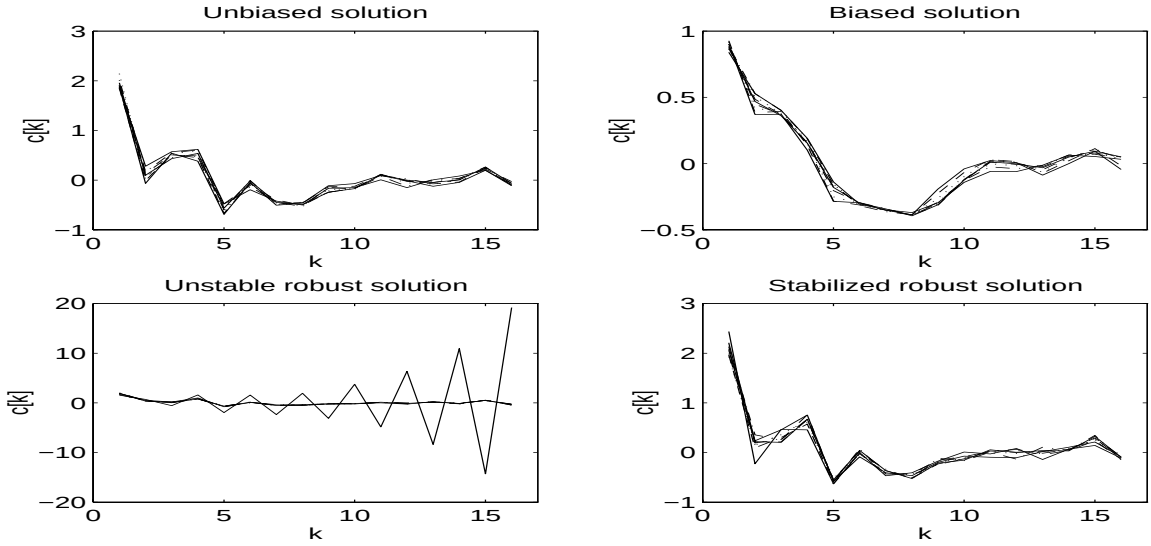

Figure 10: Correlation modeling - cepstral coefficients over one signal.