제15회 음성통신 및 신호처리 워크샵(KSCSP '98 15권1호)

망각소자를 갖는 t-분포 강인연속추정을 이용한 음성신호추정에 관한 연구

이주헌*, 최영하**, 이기용***

*동아방송대학 정보통신과 **동아방송대학 음향제작과 ***숭실대학교 정보통신전자공학부

Robust Sequential Estimation based on t-distribution with forgetting factor for time-varying speech

Joohun Lee*, Yeongha Choi** and Ki Yong Lee***

* Dept. of Information and Telecommunication, Dong-Ah Broadcasting College, Ansung, Korea

**Dept. of Audio Production, Dong-Ah Broadcasting College, Ansung, Korea

*** School of Electronic Engineering, Soongsill University, Seoul, Korea

Abstract

In this paper, to estimate the time-varying parameters of speech signal, we use the robust sequential estimator(RSE) based on t-distribution and, for time-varying signal, introduce the forgetting factor. By using the RSE based on t-distribution with small degree of freedom, we can alleviate efficiently the effects of outliers to obtain the better performance of parameter estimation. Moreover, by the forgetting factor, the proposed algorithm can estimate the accurate parameters under the rapid variation of speech signal.

I. Introduction

The estimation and tracking of speech parameters have long been recognized as important adjuncts to speech signal processing and the several methods based on linear predictive coding(LPC) have been developed as a useful method. However, those frame-based analysis methods are known to have problems for certain type of speech signals, including source-track interaction when periodic pulse trains are the excitation, as in voiced sounds and the fast transition between vowels and consonants. To overcome the drawbacks of those methods, the Kalman filter was proposed. However, in the presence of outliers, the Kalman filter is known to show very poor performance since it is optimal only for Gaussian noise. Also, when the speech signal varies rapidly, the parameter-tracking performance of the Kalman filter is diminished by the weight which the filter gives to the history of the signal.

In this paper, to estimate the time-varying parameters of speech signal, we use the robust sequential estimator(RSE)[2] based on t-distribution and, for time-varying signal, introduce the forgetting factor. We use a loss function which assigns large weighting factor for small amplitude residuals and small weighting factor for large amplitude residuals which is for instance caused by the pitch excitations. The loss function is based on the assumption that the residual signal has an independent and identical t-distribution with α degrees of freedom. When α goes to infinite, we get the conventional LP method.

Since the t-distribution with small α has more probability on its tail than that with large α, we assume that α=3. By using small α, a better separation between the source excitation and the vocal tract system can be achieved. Therefore, by using the RSE based on t-distribution with small degree of freedom, we can alleviate efficiently the effects of outliers to obtain the better performance of parameter estimation. Moreover, in order to cope with the rapid variation of speech signal, we introduce the forgetting factor to this RSE. By the forgetting factor, the proposed algorithm can estimate the accurate parameters under the rapid variation of speech signal to base on estimation on only the most recent portion of the data.

Some experimental results performed on real speech signals, Korean sentences lasting about one second, show that the proposed algorithm achieves more accurate estimation and provides improved tracking performance with smaller variance and bias, compared to the robust Kalman filter[1] based on Huber's M-estimate for both Gaussian and heavy-tailed processes.

II. Problem Formulation

The residual signal ε_k can be expressed as a function of the linear prediction(LP) vector as

$$\varepsilon_i(\boldsymbol{a}) = s_i + \sum_{j=1}^{p} a_j \cdot s_{i-j} \qquad (2\text{-}1)$$

where $a = [a_1 a_2 \ldots a_p]^T$ and a_j are LP coefficients. In Eq. (2-1), s_i is a time-varying autoregressive model of order p, AR(p). The excitation source ε_i is considered a non-Gaussian process which is a combination of two Gaussian processes with different variances: one Gaussian process has a small variance which accounts for the modeling error caused by fitting the vocal tract structure with improper model parameters, and the other with a relatively much larger variance represents the error due to the spiky excitations. Hence, we can represents the error due to the excitation source by the δ-contaminated normal mixture model as

$$P_\delta = \{F \mid F = (1-\delta)N(0,\sigma_1^2) + \delta N(0,\sigma_2^2),\ 0 \le \delta \le 1\}, \quad (2\text{-}4)$$

where, $N(\cdot \mid \mu, \sigma^2)$ is a normal density with mean μ and variance $\sigma^2 (\sigma_2^2 >> \sigma_1^2)$, and the quantity δ is the probability of occurrence of outliers in the underlying Gaussian distribution $N(\cdot \mid \mu, \sigma_2^2)$.

In the conventional LP(CLP) speech analysis, the predictor coefficients a_j, $1 \le j \le p$, are determined to minimize the sum of the squares of the prediction residuals. Therefore the result is least square fit. The same weighting function is assumed for all signal amplitude, so that the obtained estimate is very much affected by the strong signal parts and results in difficulties for the LP analysis of high-pitched voices. Also, In the CLP method, the structure of the source excitation is not taken into account. As mentioned above, when the excitation source is modeled by δ-contaminated normal mixture model, the least square method is biased and inefficient.

To cope with the drawbacks which comes from the frame-base analysis, Kalman filter can be considered to obtain the accurate a_j parameters by sequential estimation. However, the performance of conventional Kalman filter can be considerably deteriorated when the input signal is a non-Gaussian process. Also, when the speech signal varies rapidly, the parameter-tracking performance of the Kalman filter is diminished by the weight which the filter gives to the history of the signal. To overcome above problems came from the CLP method and conventional Kalman filter, the robust sequential method with forgetting factor has to be developed.

III. The Proposed RSE with Forgetting Factor

Many robust procedures can be viewed as a modified least squares(LS). Robust estimators are more efficient (lower variance) than LS when the errors are not normally distributed., and slightly less so when they are. Although there are many robust estimators in the literature, we will concentrate only on the maximum likelihood estimator(M-estimator) for a t-distribution error model.

The key step in M-estimators is to replace the square by another symmetric cost function of the residuals or to model the noise by a nonnormal, heavy-tailed distribution to account for outliers. We can then use maximum likelihood analysis to obtain robust estimates of the parameter vector. Unfortunately, the direct evaluation of maximum likelihood estimates from nonnormal distributions becomes quite complicated. But an effective means of obtaining maximum likelihood estimates for a wide class of nonnormal distributions is weighted least squares.

Let $f(\varepsilon_i)$ be any differentiable error density function which can be written in the form

$$f(\varepsilon_i) \propto \sigma^{-1} g\{(\frac{\varepsilon_i}{\sigma})^2\} \quad (3\text{-}1)$$

where σ is a scale parameter, $g\{\cdot\}$ denotes a functional form, and $\varepsilon_i = s_i - \sum_{j=1}^{p} a_j s_{i-j}$ is the i-th actual error. Given a sample s of k observations, the likelihood function for a_i and σ^2 is given by

$$L(a,\sigma|\varepsilon) = \prod_{i=1}^{k} \sigma^{-1} g\{(\frac{\varepsilon_i}{\sigma})^2\}, \quad (3\text{-}2)$$

and its logarithmic likelihood function is given as

$$l(a,\sigma|\varepsilon) = K - \sum_{i=1}^{k} [\log\sigma^{-1} + \log g\{(\frac{\varepsilon_i}{\sigma})^2\}], \quad (3\text{-}3)$$

where K is some constant.

From this, we can define the error criterion function as

$$J_k(a) = \sum_{i=1}^{k} [\log\sigma^{-1} + h\{(\frac{\varepsilon_i}{\sigma})^2\}], \quad (3\text{-}4)$$

where $h(\cdot) = \log g(\cdot)$. For heavy-tailed Gaussian process, $h(\cdot)$ is replaced by the Huber's score function $\rho_H(\cdot)$ defined as

$$\rho_H(x) = \begin{cases} x^2/2 & |x| \le c \\ c|x| - c^2/2 & |x| \le c \end{cases} \quad (3\text{-}5)$$

In this paper, heavy tailed error distribution is reasonably represented by a t-distribution defined by

$$f_\alpha(x) = \frac{1}{\sqrt{\alpha x}} \frac{\Gamma(\frac{\alpha+1}{2})}{\Gamma(\frac{\alpha}{2})} \frac{1}{(1+\frac{x^2}{\alpha})^{(\alpha+1)/2}} \quad (3\text{-}6)$$

having small degree of freedom α and scaled by a parameter σ. Therefore, $g(\cdot)$ is given by

$$g\{(\frac{\varepsilon_i}{\sigma})^2\} = \{1 + \frac{\varepsilon_i^2}{(\alpha\sigma^2)}\}^{(-1/2)(1+\alpha)}. \quad (3\text{-}7)$$

In Eq. (3-6), $t(1)$ is the Cauchy distribution and $t(\infty)$ is the Gaussian distribution with zero mean and standard deviation(SD) equal to one. For the estimation purpose, $f(x)$ has to have a finite second moment. Since $f_\alpha(x)$ for $\alpha < 3$ has an infinite second moment, here we use $\alpha \ge 3$. Research work showed that choice of small degrees of freedom $\alpha = 3$ induces to the most accurate and the efficient estimation.

In addition to that, the forgetting factor λ, $0 < \lambda \le 1$, is employed to weight the most recent data more heavily to allow for tracking of varying parameters. Progressively smaller λ result in parameter being computed with effectively smaller windows of data that are beneficial in nonstationary situations. Then, the Eq. (3-4) is rewritten as

$$J_k(a) = \sum_{i=1}^{k} \lambda^{k-i} \cdot [\log\sigma^{-1} + \log g\{(\frac{\varepsilon_i}{\sigma})^2\}] \quad (3\text{-}8)$$

To obtain the M-estimator, by differentiating the Eq. (3-8) with respect to a and σ^2, we obtain

$$\frac{\partial J_k}{\partial a_h} = \sigma^{-2} \sum_{i=1}^{k} \lambda^{k-i} w_i (s_i - \sum_{j=1}^{k} a_i s_{i-j}) s_{i-h},$$

$$h = 1,2,\ldots,p \quad (3\text{-}9)$$

$$\frac{\partial J_k}{\partial \sigma^2} = -\frac{1}{2} k\sigma^{-2} + \frac{1}{2}\sigma^{-4} \sum_{i=1}^{k} \lambda^{k-i} w_i (s_i - \sum_{j=1}^{k} a_i s_{i-j})^2$$

where

$$w_i = w_i(a, \sigma^2) = -2\left[\frac{\partial \log g(\xi)}{\xi}\right]_{\xi = (\varepsilon_i/\sigma)^2} \quad (3\text{-}10)$$

Equating these two Eq.'s in (3-9) to zero, we get the maximum likelihood estimates $\bar{a}$ and $\bar{\sigma}^2$ as the solution of the nonlinear equations

$$\sum_i \lambda^{k-i} w_i (s_i - \sum_j \bar{a}_j s_{i-j}) s_{i-h} = 0,$$

$$h = 1,2,\ldots,p \quad (3\text{-}11)$$

$$\mu^2 = \bar{\sigma}^2 = \sum_i \lambda^{k-i} w_i (s_i - \sum_j \bar{a}_j s_{i-j})^2 \Big/ n \, .$$

We can solve these nonlinear equations using iteratively reweighted least squares(IRLS). Rewriting (3-11) in matrix form yields

$$\boldsymbol{H}^T \boldsymbol{W}(\boldsymbol{s} - \boldsymbol{H}\bar{\boldsymbol{A}}) = \boldsymbol{0},$$

where $\boldsymbol{s}$ is an n-vector of speech signals on the dependent variable, $\boldsymbol{H}$ is an $n \times p$ matrix of observed speech signals having rank p, $\boldsymbol{A}$ is a p-vector of parameters to be estimated, and $\boldsymbol{W}$ is a diagonal matrix defined such $\lambda^{n-i} w_i$ as

$$\boldsymbol{W} = \begin{bmatrix} \lambda^{n-1} w_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda^0 w_n \end{bmatrix}.$$

Substituting Eq. (3-10) for g in (3-7) gives the individual weights:

$$w_i = 1 + \alpha \Big/ \alpha + (r_i/\mu)^2 \, , \quad (3\text{-}12)$$

where the residual $r_i = s_i - \sum_j a_j s_{i-j}$. We use μ in this expression to distinguish the (unknown) true value of the scale parameter, σ, from an estimated value, μ, used in computing the weights.

The resulting iterative scheme, after simplification, is given by

$$\bar{\boldsymbol{A}}_k = \bar{\boldsymbol{A}}_{k-1} + (\boldsymbol{H}^T \boldsymbol{W}_{k-1} \boldsymbol{H})^{-1} \boldsymbol{H}^T \boldsymbol{W}_{k-1} (\boldsymbol{s} - \boldsymbol{H}\bar{\boldsymbol{A}}_{k-1}) \quad (3\text{-}13)$$

The RSE starts from an initial robust estimate of the speech parameters and these parameters are computed by IRLS with the errors assumed to be t-distributed. It then adds the remaining data sequentially, assigning weights to each new observation based on the previous estimates.

Suppose that the robust parameter estimate for the first m observations is

$$\bar{\boldsymbol{A}}_m = (\boldsymbol{H}_m^T \boldsymbol{W}_m \boldsymbol{H}_m)^{-1} \boldsymbol{H}_m^T \boldsymbol{W}_m \boldsymbol{s}_m,$$

where $\boldsymbol{W}$ is the appropriate diagonal weighting matrix computed by the maximum likelihood analysis. Now expand $\bar{\boldsymbol{A}}_{m+1}$ as

$$\boldsymbol{A}_{m+1} = [\boldsymbol{H}_m^T \boldsymbol{W}_m \boldsymbol{H}_m + w_{m+1} \boldsymbol{h}_{m+1} \boldsymbol{h}_{m+1}^T]^{-1} \times [\boldsymbol{H}_m^T \boldsymbol{W}_m \boldsymbol{s}_m + w_{m+1} s_{m+1} \boldsymbol{h}_{m+1}]$$

and define $\boldsymbol{P}_m = (\boldsymbol{H}_m^T \boldsymbol{W}_m \boldsymbol{H}_m)^{-1}$.

After rearranging the terms, and simplifying, we obtain the following recursive equations for the robust sequential algorithm:

$$\bar{\boldsymbol{A}}_{m+1} = \bar{\boldsymbol{A}}_m + \gamma_{m+1} \boldsymbol{P}_m \boldsymbol{h}_{m+1} \left[s_{m+1} - \boldsymbol{h}_{m+1}^T \bar{\boldsymbol{A}}_m \right],$$

$$\boldsymbol{P}_{m+1} = \boldsymbol{P}_m - \gamma_{m+1} \boldsymbol{P}_m \boldsymbol{h}_{m+1} \boldsymbol{h}_{m+1}^T \boldsymbol{P}_m,$$

where

$$\gamma_{m+1} = w_{m+1} \Big/ 1 + w_{m+1} \boldsymbol{h}_{m+1}^T \boldsymbol{P}_m \boldsymbol{h}_{m+1} \, .$$

The crucial problem in robust estimators is the estimation of the scale parameter. As an efficient, robust and simple approach to simultaneous scale and parameter estimation for a wide class of nonnormal distributions is obtained using maximum likelihood analysis:

$$\mu_m^2 = \sum_i^m w_i (s_i - \sum_j \tilde{a}_j s_{i-j})^2 \Big/ m$$

IV. Simulation Results

The proposed algorithm has been tested on natural Korean speech, 'sa', and its results has been compared to those by the conventional Kalman filter and the robust Kalman filter proposed in [1]. The results are shown in Fig. 1. In Fig. 1, from the top, a_1's of the conventional Kalman, the robust Kalman and the proposed algorithm are presented, respectively. They are given a little bias to show their differences exactly. As shown in this figure, the proposed algorithm can estimate the trajectory of the parameter more accurately, while the others are much affected by the outliers, the pitch excitations. In this simulation, λ is given by 0.98 for all method and c is given by 1.5 for the robust Kalman filter.

V. Conclusions

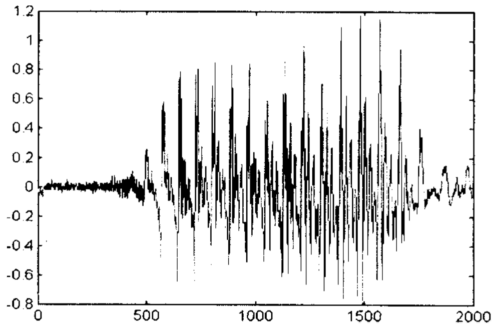
We proposed the Robust Sequential Estimator based on t-distribution with forgetting factor. The proposed algorithm can alleviate efficiently the effects of outliers to obtain the better parameter estimation by introducing t-distribution to sequential estimation. Also, by introducing the forgetting factor, it can estimate the rapid varying parameters which brings some drawbacks to the Kalman filter. The simulation results show that, by the proposed method, the better estimation performance can obtained.

References

[1] T. Yang, J. Lee, K. Lee and K. Sung, "On Robust Kalman Filtering with Forgetting Factor for Sequential Speech Analysis", Signal Processing, vol.63, no.2, pp.1151-1156, Dec. 1997

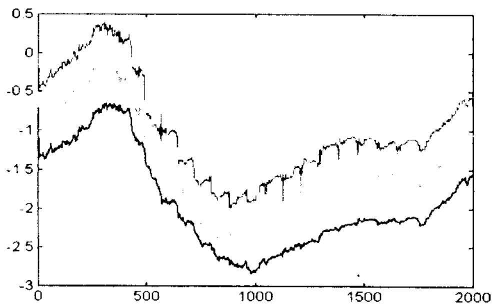
[2] K. Boyer, J. Mizra and G. Ganguly, "The Robust Sequential Estimator: A General Approach and its Application to Surface Organization in Range Data", IEEE Trans. On Pattern Analysis and Machine Intelligence, vol.16, no.10, Oct. 1994

[3] J. Sanubari, K. Tokuda and M. Onoda, "Speech Analysis Based on AR Model Driven by t-distribution Process", IEICE Trans. Fundamentals, vol.E75-A, no.9, pp.1159-1169, Sep. 1992



Time [msec]

(a)



Time [msec]

(b)

Fig. 1. Real speech and a_1 coefficient trajectories. (a) Real speech 'sa', (b) estimated trajectories of a_1 coefficients obtained by the conventional Kalman filter, the robust Kalman filter and the proposed robust sequential algorithm, respectively, from the top.