TWO-STAGE SIGNAL RECONSTRUCTION UNDER UNKNOWN PARAMETERS AND MISSING DATA

Jui-Chung Hung[1], *Bor-Sen Chen*[2]

[1] Ling-Tung College
[2]Department of Electrical Engineering
National Tsing-Hua University
Hsin-Chu 300, Taiwan.
Email: hong@moti.ee.nthu.edu.tw

ABSTRACT

This paper considers the signal reconstruction problem under unknown parameters and nature missing data. The solution is divided into two stages. At first stage, the parameter estimation of autoregressive moving average(ARMA) model with nature missing data is studied. In the second stage, a robust Kalman filter to reconstruct the input signal is developed. The missing data model is based on a probabilistic structure with unknown. In this situation, the estimation becomes a highly nonlinear optimization problem with many local minima. In this paper, we combines the global search method of genetic algorithm and simulate annealing(GA/SA) to achieve a global optimal solution with fast convergent rate. After the system parameters are exactly estimated in the first stage, the problem of reconstructing the missing signal can be handled elegantly using the proposed robust Kalman filter in the second stage.

1. INTRODUCTION

Many different parameter estimation and reconstruction methods have been proposed in recent years. In most cases, however, it is assumed that the measurements always contain the signal. In fact, in practical situations, there may be a non-zero probability that any observation consists of noise alone, i.e., the measurements are not consecutive but contain missing observations. The missing observations are caused by a variety of reasons, e.g. a certain failure in the measurement, intermittent sensor, failures, accidental loss of some collected data, or some of the data may be jammed or coming from a high noise environment, etc [1]. In practice cases, it may happen that the data samples are missed in the measured signals. Fundamentally, the standard definition of the covariance in statistical analysis of data does not directly apply if some of the measurements are unavailable. If not properly taken into account, the missing measurements can seriously deteriorate the quality of the estimates. This paper considers the problem that the missing observation is random with unknown missing points and corrupted noise. In this situation, the signal reconstruction problem is very complicated, and difficult to solve by the conventional methods. At the first step, parameter estimation algorithm based on genetic algorithm/simulate annealing (GA/SA) is proposed to resolve the parameter estimation problem under missing data and then a robust Kalman filter is proposed to reconstruct the input signal.

Comparing GA [2] and SA [3], we can find that GA exhibits fast initial convergence, but its performance deteriorates as it approaches the desired global extreme. Interestingly, SA shows a complementary convergence pattern, in addition to high accuracy. We combine the selected features from GA and SA to achieve weak dependence on initial parameters, parallel search strategy, fast convergence and high accuracy. The GA/SA starts the search procedure as a pure-GA and ends as a pure-SA. The transition from GA to SA occurs when the fittest individual remains the same for Lg generations. Hence, GA/SA is very suitable to treat the global optimization problem of the nonlinear parameter estimation under the corrupted noises and missing.

After the parameter estimation is solved under missing

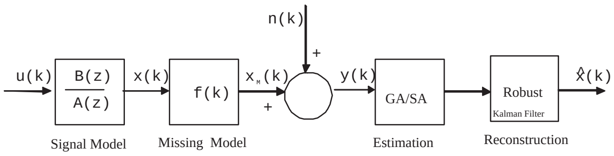

Figure 1: Signal model.

data a robust Kalman filter is proposed to treat this signal reconstruction problem. When the signal is missing, the Kalman filter's innovation signal is distorted. A nonlinear robust function is embedded in Kalman filter to reconstruct the innovation signal. From the simulation results, we have found that the reconstruction performance is improved significantly if the probability of data missing is considered in the proposed parameter method and the signal reconstruction method.

2. PROBLEM DESCRIPTION

Let $\{y(k)\}$ denote the observed output signal (see the Fig.1). The measurement equation is given by

$$\begin{aligned} y(k) &= x_M(k) + n(k) = f(k)x(k) + n(k) \\ &= f(k)\left[-\sum_{i=1}^{n} a_i x(k-i) + \sum_{i=0}^{q} b_i u(k-i)\right] \\ &\quad + n(k) \end{aligned} \tag{1}$$

where $\{u(k)\}$ is a zero mean unit-variance white Gaussian noise, the noise $n(k)$ is zero mean Gaussian processes with the variance σ_n^2, and $f(k)$ is binary random variable such that

$$f(k) = \begin{cases} 1 & \text{if } x(k) \text{ is measured} \\ 0 & \text{if } x(k) \text{ is missing} \end{cases} \tag{2}$$

The following assumptions are made:

(A1) The $u(k)$, $n(k)$, are mutually independent.

(A2) The sequence $f(k)$ is assumed to be asymptotically stationary and independent of $x(k)$. Furthermore, they are mutually independent. The probability p for the measurement $x_M(k)$ to be measured is assumed to be unknown and given by

$$E[f(k)] = \Pr[f(k) = 1] = p, \quad 0 < p \leq 1 \tag{3}$$

where E denotes the expectation operator, and p is a fixed probability, independent of time. Thus the probability of missing measurement is $(1-p)$.

(A3) Each measurement has the fixed probability of being missed, and for different instants, the occurrences of missing data are mutually independent. Thus, by the assumptions, the power spectral density (PSD) of the received data is

$$\begin{aligned} S_y(f) &= \sum_{l=-\infty}^{\infty} r_y(l) z^{-l} \,|_{z=e^{j2\pi f}} \\ &= p^2 \sum_{l=-\infty}^{\infty} r_x(|l|) e^{-j2\pi f l} + K \end{aligned} \tag{4}$$

where $r_x(l) = E[x(k)x(k-l)]$, $K = \sigma_n^2 + p.(r_x(0) - p.r_x(0))$. In this estimation problem, not only the system parameters $a_1, \cdots, a_n$, $b_0, \cdots, b_q$ are unknown but also the probability p and the noise's variance σ_n^2 are also needed to be estimated.

Inspection the equation(4), the series as $\sum_{l=-\infty}^{\infty} e^{-j2\pi f l}$ as a power series has the property of completeness, we can obtain

$$r_y(l) = \begin{cases} p^2\, r_x(0) + K & \text{for } l = 0 \\ p^2\, r_x(l) & \text{for } l \neq 0 \end{cases} \tag{5}$$

From the received data sequence $\{y(k)\}$, let us define the sample covariance by

$$\hat{r}_y(l) = \frac{1}{N-l} \sum_{k=l+1}^{N} y(k) y(k-l) \tag{6}$$

where N is number of received data length. Note that the sample covariances $\{\hat{r}_y(l)\}$ are unbiased estimates of the true covariances $\{r_y(l)\}$. The main idea is to search for the polynomials $\sum_{l=-\infty}^{\infty} r_y(l) z^{-l}$ such that the corresponding sequence $\{\hat{r}_y(0), \cdots, \hat{r}_y(M)\}$, from (4) and (6), a reasonable criterion is to minimize the mean square error as

$$\begin{aligned} &\min J(a_1, \cdots, a_n, b_0, \cdots, b_q, p, \sigma_n^2) \\ =\ &\min \left\{ \sum_{l=0}^{M} (r_y(l) - \hat{r}_y(l))^2 \right\} \\ =\ &\min \left\{ (r_y(0) - \hat{r}_y(0))^2 + \sum_{l=1}^{M} (r_y(l) - \hat{r}_y(l))^2 \right\} \end{aligned} \tag{7}$$

3. FIRST STAGE: GA/SA-BASED PARAMETER ESTIMATION

The combined GA/SA parameter estimation algorithm always stars the search procedure as a pure-GA and ends as a pure-SA. The transition from GA to SA occurs when the following conditions is satisfied. The fittest individual remains the same for L_g generations. The condition is satisfied whenever the algorithm converges to an intermediate solution. The solution so far constitutes a good initial guess to SA. The SA's initial and final temperature, as well the step length can adjust as a function of the fittest individual's energy.

Based on the above analysis, the design procedure of GA/SA-based parameter estimation of time series with noises and missing data is divided into the following steps.

Step 0: Given the received data $y(k)$. Compute the $\left\{\hat{r}_y(0), \cdots, \hat{r}_y(M)\right\}$ from $y(k)$.

Step 1: Generate random Γ chromosomes.

Step 2: Find $r_x(1), \cdots, r_x(M)$.

Step 3: Compute the minimum mean square error $\left\{\sum_{l=0}^{M}(r_y(l)-\hat{r}_y(l))^2\right\}$ by the genetic algorithm

Step 4: Compute the corresponding fitness value

Step 5: Remain the best chromosome intact into next generation.

Step 6:Use the GA operators (reproduction, crossover, mutation) to produce chromosomes of next generation.

Step 7:Repeat the produce from step 2 to step 7 until the fittest individual remains the same for L_g generations.

Step 8: Given the best chromosome, C_d, C_i, C_t, k, T, B and δ_a^2 of SA initial values.

Step 9: Generate the new-state $(a_1, \cdots, a_n, b_0, \cdots, b_q, p, \sigma_n^2)' = (a_1, \cdots, a_n, b_0, \cdots, b_q, p, \sigma_n^2) + Norm(0, \delta_a^2)$ and Compute the new-state energy

$$J' = \left\{\sum_{l=0}^{M}(r_y(l)-\hat{r}_y(l))^2\right\}$$

Step 10: Let $p_s(J') = 1/(1+exp(-J'/BT))$; if $p_s(J') >$random [0,1) or new-state energy $(J')<$ original-tate energy (J) then the current-state = new-state.

Step 11: if the temperature T is the same in k iterations, then decrease the temperature as $T = C_t T$.

Then, repeat the procedure from step 9 to step 11 until a suitable parameter set is obtained.

4. SECOND STAGE:ROBUST KALMAN FILTER-BASED SIGNAL RECONSTRUCTION

From the state space form the equation (1) can as

$$\begin{aligned} \mathbf{w}(k+1) &= \mathbf{F}\mathbf{w}(k) + \mathbf{q}u(k+1) \\ y(k) &= \mathbf{h}\mathbf{w}(k)+n(k) \end{aligned} \quad (8)$$

Since the signal is embedded in the state vector $\mathbf{w}(k)$, there the signal reconstruction problem becomes a state estimation problem. From the state equation (8), the state $\mathbf{w}$ can be optimally estimate in the minimum mean-square-error sense by using the robust Kalman filter. Thus, the optimal state estimate $\hat{\mathbf{w}}$ based on the received measurement time series $y(k)$ can be obtained by the following robust Kalman filter

$$\begin{aligned} \hat{\mathbf{w}}(k \mid k) &= \hat{\mathbf{w}}(k \mid k-1) + \mathbf{G}(k)e_r(k) \\ e_r(k) &= \hat{y}(k-1)\Phi\left(\frac{e(k)}{\hat{y}(k-1)}\right) \\ \hat{\mathbf{w}}(k \mid k-1) &= \mathbf{F}\,\hat{\mathbf{w}}(k-1 \mid k-1) \\ \mathbf{G}(k) &= \mathbf{P}(k \mid k-1)\mathbf{h}^t\left[\sigma_{n_g}^2 + \mathbf{h}\mathbf{P}(\mathbf{k} \mid \mathbf{k}-\mathbf{1})\mathbf{h}^t\right]^{-1} \\ \mathbf{P}(k \mid k) &= (\mathbf{I} - \mathbf{G}(k)\mathbf{h})\mathbf{P}(k \mid k-1) \\ \mathbf{P}(k \mid k-1) &= \mathbf{F}\,\mathbf{P}(k-1 \mid k-1)\mathbf{F}^t + \mathbf{q}\mathbf{q}^t \end{aligned} \quad (9)$$

where$e(k) = y(k) - \widehat{y}(k) = y(k) - \mathbf{h}\,\hat{\mathbf{w}}(k \mid k-1)$, and the nonlinearity Φ is given as

$$\Phi(k) = \begin{cases} -l\ \exp(-\frac{(k+l)^2}{\sigma^2}),\ if \quad k \leq -l \\ l, \qquad\qquad\qquad\quad if \quad k < l \\ l\ \exp(-\frac{(k-l)^2}{\sigma^2}),\ if \quad k \geq l \end{cases} \qquad (10)$$

The robust Kalman filter is designed by $e(k)$ passing through the nonlinearity Φ to suppress the effects of missing data and impulse noise that might be present, and the next step is to reconstruct the signal by Kalman filter via robust residual error $e_r(k)$.

5. DESIGN EXAMPLES

Consider a segment of speech signal ”welcome” (spoken in mandarin by a female), which is digitized at 8 kHz represented by the 8 bit u-law PCM(pulse code modulation). A segment of the representative speech signal ”welcome” contains 3800 points. This speech signal is corrupted by an additive white Gaussian noise n with $SNR = 20$ db. A 7^{th}-order ($n = 7, q = 0$) AR signal generation model is fitted to this data and the model parameters are obtained by first stage with frame-length 15 ms. The reconstruction performance is compared with method 1(first stage by conventional method [4], second stage by robust Kalman filter), and method 2 (first stage by conventional method [4], second stage by Kalman filter).The reconstruction of the 17th frame of speech signal is shown in Fig. 2. The signal to error ratio (SNR_r) with various missing probability is shown in Fig. 3. By inspection of Fig. 3, the proposed method provides a good reconstruction performance; the SNR_r gains are about 4 db over the method 1 and . The performance of method 1 is lower than the proposed method, because the parameter estimation of method 1 is not accuracte in the first stage.

6. REFERENCES

[1] M. J. R. Healy and M. Westmacott, ”*Missing values in experiments analyzed on automatic computers.*” Appl. Statist. 5, pp. 203-206, 1956.

[2] D. E. Goldberg, Genetic algorithms in search, optimization, and machine learning, reading, MA: Addition Wesley, 1989.

[3] M. P. Vecchi and S. Kirkpatrick, ”*Global Wiring by Simulated Annealing,*” IEEE, Trans. on Computer-Aided DEsign CAd-2, pp. 215-222, 1987.

[4] S. M. Kay, Modern Spectral Estimation: Theory and Application Prentice-Hall, Englewood Chiffs, NJ 1987.

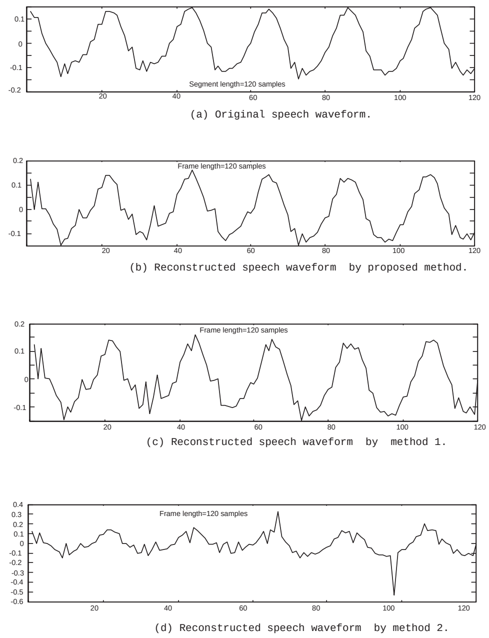

Figure 2: Reconstructions of 17th frame of speech by different methods.

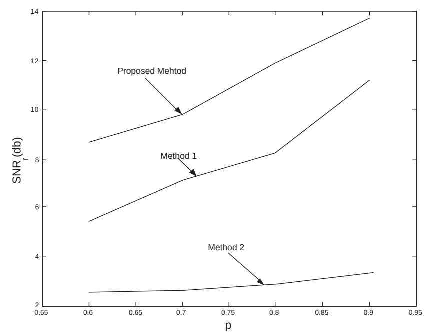

Figure 3: Signal to noise ratios versus missing probability of speech.