Research on Weighted Least Square Optimization Algorithm of Nonlinear System Analysis

Shan Wei
Electric engineering college
Yanshan University
Qinhuangdao, Hebei China
sw@ysu.edu.cn

Zhao Wenshuang
Electric engineering college
Yanshan University
Qinhuangdao, Hebei China
Zhaowenshuang_2009@126.com

Li Lei*
Electric engineering college
Yanshan University
Qinhuangdao, Hebei China
llxiaozha@163.com
*Corresponding author

He Qun
Electric engineering college
Yanshan University
Qinhuangdao, Hebei China
hq@ysu.edu.cn

***Abstract*— In the actual control system, the controlled object is usually nonlinear. Therefore, nonlinear time series analysis has been always received considered attention from many scholars. The ability to accurately model is critical for nonlinear time series analysis. However, the challenge in system modeling is how to estimate the model parameter. Unlike the traditional parameter estimation method with the defects of low precision, poor convergence and long time optimization, this paper proposes a new optimization algorithm of nonlinear time series which combine the advantages of fast convergence near the minimum value and being able to converg for any initial value. During the calculation, the first-order derivative should be solved while the inverse matrix is not necessary. The simulation results show that the proposed method not only ensures the convergence of iteration planning but also improves the convergent speed. It can be applied the nonlinear time series analysis and provides the powerful guarantee for accurate trends prediction.**

Keywords- Nonlinear Time Series;Parameter Estimation; Optimization;Weighted Least Square;System model

I. INTRODUCTION

In the control system, such as rotating machinery system, wind turbine system, the controlled object is usually nonlinear. Through the analysis of the observing time series of object, one can achieve the goal of the nonlinear system analysis. Therefore, nonlinear time series analysis has been always received considered attention from many researchers.

With the rapid development of computer technology and signal processing technology, the theory and method of time series analysis, such as Auto Regressive (AR), Auto Regression Moving Average (ARMA), and Autoregressive Integrated Moving Average (ARIMA), have been more prefect. Wang et al. [1] suggests a new type of interval time series: interval autoregressive (IAR) model, and discuss parameter estimation. Singh et al. [2] applied a new learning method, extreme learning machine (ELM), to analyze time series.

However, the challenge of time series analysis is how to estimate model parameter. Traditional parameter estimation methods, such as maximum likelihood estimation [3], moment estimate [4] and least square estimate [5], have the defects of low precision, poor convergence and parameter estimation white noises coupling is still used for model parameter estimation. Taking this as basis the data forecasting and anomaly detection are conducted, which is hard to ensure the system's stability. Therefore, many researchers started to study new parameter estimation methods. Lemke et al. [6] study parameter estimation for time series with asymmetric α-stable innovations. Statistical analysis [7], particle swarm optimization [8], hybrid differential evolution algorithm [9], and Expectation Maximization (EM) approach [10] are also applied to estimate model parameter. This paper proposed a new optimization algorithm of nonlinear time series based on weighted least square.

This algorithm has avoided the traditional algorithm's defects of pointwise optimization, long time optimization, and linear model of relying on simplification reducing the accuracy. The optimization algorithm combines the advantages of fast convergence near the minimum value and being able to converge for any initial value, which not only ensure the convergence of iteration planning but improve the convergent speed. In the process of calculation we only need to get the first derivative without getting the inverse matrix, and it's more oblivious than the least square estimate value with better fitting model. It has many advantages of fast convergence speed, not requiring getting the inverse matrix, fewer iterations and storage unit, and superlinear convergence. It breakthroughs the linear planning method of traditional solving nonlinear unconstrained problem and provides the new thinking of solving unconstrained time series problems. It can be

applied the nonlinear time series analysis and provides the powerful guarantee for accurate trends prediction.

II. ESTABLISHMENT OF ARIMA MODEL

The useful subclass of ARIMA Model is composed of stable model. Stability reflects a class of time-invariant nature of time series, which is the necessary condition for carrying on statistical inference in certain time. However, the realistic time series data often shows the time trend (such as slowly increase) and cycling characteristics. They have no longer met the requirement of stationary ARIMA Model. The usual way of eliminating these non-stationary elements is to pre-process the data. It's the continent and effective method for eliminating trendy and the seasonal characteristic to take the difference (if necessary, can make multiple difference). After removing the time trend, we can use ARIMA Model to modeling the new reservation series. Because the original series is a combination of differential sequence, we call it Autoregressive integrated moving average ARIMA Model.

Model with the following structure is called ARIMA Model, ARIMA(p,d,q) for short.

$$\begin{cases} \Phi(b)\nabla^d y(t) = \Theta(b)\varepsilon(t) \qquad \varepsilon(t) = \varepsilon_t \\ E(\varepsilon_t) = 0, \quad Var(\varepsilon_t) = \sigma_\varepsilon^2, \quad E(\varepsilon_t\ \varepsilon_s) = 0, \quad s \neq t \\ E[y(s)\varepsilon(t)] = 0, \quad \forall s < t \end{cases} \tag{1}$$

In Equation (1):

$$\nabla^d y(t) = \sum_{i=1}^{d} (-1)^d c_d^i y(t-1), c_d^i = \frac{d!}{i!(d-i)!}$$

$\Phi(b) = 1 - \varphi_1 b - \varphi_2 b^2 - \cdots\cdots - \varphi_p b^p$, is the autoregressive coefficient polynomial of the stationary and reversible ARMA(p,q) model.

$\Theta(b) = 1 - \theta_1 b - \theta_2 b^2 - \cdots\cdots \theta_q b^q$ is the move-smoothing coefficient polynomial of stationary and reversible ARMA(p,q) model.

Formula (1) can be simplified as:

$$\nabla^d y(t) = \frac{\Theta(b)}{\Phi(b)}\varepsilon_t \tag{2}$$

In this formula, $\{\varepsilon_t\}$ is zero mean value flat noise series.

One stable, normal, zero mean value time series $\{X(t)\}$ can fit three models of different forms, ARMA(p,q), AR(∞) and MA(∞). Among them MA(q) and AR(p) are naturally the special sample of ARMA(p,q) model. For system's description, under the pricinple of equivalent output, model of describing system characteristics is not necessarily the only. Therefore one system can be designed, taking ARMA or AR or MA model to describe respectively. ARMA model has fewer parameter, conforming to parameter saving principle. However, fitting AR model is rapid and easy. In practical application, the suitable can be chosen according to specific situation.

III. CLASSICAL ESTIMATION ALGORITHM

The most basic parameter estimation method in coefficient identification is the least square method. Because of its characteristics of requiring less priori statistical knowledge of system, easy algorithm and small calculating amount, it has been widely applied in system identification, adaptive filter, self-tuning control and other fields.

Consider AR (p) model.

$$x(t) = \varphi_1 x(t-1) + \ldots + \varphi_p x(t-p) + \varepsilon(t) \tag{3}$$

in formula $\varepsilon(t)$ is zero mean value, variance is the white noise of $\sigma_\varepsilon^2 > 0$, $x(t)$ is observation, known order p, but parameter φ_i and σ_ε^2 are unknown. The problem is the known observation data $(x(t), x(t-1), \ldots x(0), \ldots x(1-p))$ based on the time t, evaluate φ_i and σ_ε^2's estimation value $\hat{\varphi}_i(t)$ and $\hat{\sigma}_\varepsilon^2(t)$.Equation (3) can be written as vector form:

$$x(t) = s^T(t)\theta + \varepsilon(t) \tag{4}$$

Among it T is transpose No., and definite vector.

$$\theta = \left[\varphi_1, \varphi_2, \ldots, \varphi_p\right]^T \quad s^T(t) = \left[x(t-1), \ldots, x(t-p)\right] \tag{5}$$

Model residual is

$$\varepsilon(t) = x(t) - s^T(t)\theta \tag{6}$$

Least square principle is to seek the estimation value $\hat{\theta}(t)$ of unknown parameter vector θ, making its minimizing model residual sum of squares

$$I = \sum_{i=1}^{t} \varepsilon^2(i) = \sum_{i=1}^{t} \left[(x(i) - s^T(i)\theta)^2\right] \tag{7}$$

Except $x(0), \ldots, x(1-p)$, from observation $x(1), \ldots, x(t)$ there are t equations

$$x(i) = s^T(i)\theta + \varepsilon(i) \quad i = 1, \cdots t \tag{8}$$

They can be synthetically written as matrix vector form

$$X(t) = S(t)\theta + \Sigma(t) \tag{9}$$

And definition

$$S(t) = \begin{bmatrix} s^T(1) \\ \vdots \\ s^T(t) \end{bmatrix}, X(t) = \begin{bmatrix} x(1) \\ \vdots \\ x(t) \end{bmatrix}, \Sigma(t) = \begin{bmatrix} \varepsilon(1) \\ \vdots \\ \varepsilon(t) \end{bmatrix} \tag{10}$$

Then performance index I can be written as

$$I = \left[X(t) - S(t)\theta\right]^T \left[X(t) - S(t)\theta\right] \tag{11}$$

Making $\frac{\partial I}{\partial \theta} = 0$, apply matrix differential equation, there

$$-2S^T(t)\left[X(t) - S(t)\theta\right] = 0 \tag{12}$$

Namely normal equation met by θ

$$S^T(t)S(t)\theta = S^T(t)X(t) \tag{13}$$

Then observation θ's non recursive least square estimate value based on the time t is

$$\hat{\theta}(t) = \left[S^T(t)S(t)\right]^{-1} S^T(t)X(t) \tag{14}$$

AR(p) model equation (3)'s non recursive least square estimate value is $\hat{\theta}(t) = \left[S^T(t)S(t)\right]^{-1} S^T(t)X(t)$ And suppose matrix $S^T(t)S(t)$ is non-specific.

Taking use of equation (4) to conduct parameter estimation, it needs to get the inverse matrix.

IV. WEIGHTED LEAST SQUARE OPTIMIZATION ALGORITHM

A. Constuction of the objective function

For observation time series $\{x_t\}(t=1,2,\cdots,n)$, it's necessary to fit mathematical model

$$x_t = f(X_t,\beta)+\varepsilon_t \tag{15}$$

In formula $X_t=[x_{t-1}\ x_{t-2}\cdots x_{t-k}]^T$, its k 's dimensional vector composed of observation value in different time, $\beta=[\beta_1,\beta_2,\cdots,\beta_m]^T$ is M 's dimensional vector composed of model parameter $\beta_i, i=1,\cdots,m$ to be estimated, in general $k,m<n$; ε_t is model's residual; f shows the function relationship between X_t and β .

For ARMA(p,q) model

$$x_t=\varphi_1 x_{t-1}+\varphi_2 x_{t-2}+\cdots+\varphi_p x_{t-p} +\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}-\cdots-\theta_q\varepsilon_{t-q}$$

According to formula (15) it can be written as $x_t=X_t^T\beta+\varepsilon_t$,

In formula $X_t=[x_{t-1}\ x_{t-2}\cdots x_{t-p}\ \varepsilon_{t-1}\ \varepsilon_{t-2}\cdots\varepsilon_{t-q}]^T$,

$$k=p+q;\beta=[\varphi_1\ \varphi_2\cdots\varphi_p\ \theta_1\ \theta_2\cdots\theta_q]^T$$

$$m=p+q; f(X_t,\beta)=X_t^T\beta \tag{16}$$

Superficially, X_t and β is linear relationship, however, $\varepsilon_{t-1},\varepsilon_{t-2},\cdots,\varepsilon_{t-q}$ in X_t is unknown, it's available to use ARMA(p,q) model's recursive form. Calculate from $x_{t-1},\cdots,x_{t-p}$'s recursive. Namely $\varepsilon_{t-1},\varepsilon_{t-2},\cdots,\varepsilon_{t-q}$ is also $x_{t-1},\cdots,x_{t-p}$'s function , thus leading to the appearance of β 's non-linear item. Therefore, for ARMA model X_t and β has the nonlinear relationship. Thus the target function $s(\beta)$ of ARMA model is defined as model's sum of squared residuals.

$$s(\beta)=\sum_{t=p+1}^{n}\varepsilon_t^2=\sum_{t=p+1}^{n}\left[x_t-f(X_t,\beta)\right]^2 \tag{17}$$

From the perspective of optimization theory, parameter β 's estimation value problem is the optimization (seek the minimum value) problem for $s(\beta)$.

B. Optimization algorithm

For ARMA model parameter's estimation value problem , use Newton method's improved one in optimization theory —weighted least-squares, and derive iterative algorithm of parameter estimation value.

Determination of Parameter Initial Value β^0 the selection of parameter initial value is very important. It relates to iterative computation convergence speed. First estimate the parameter of AR(p_0) model and present the parameter of ARMA (p,q) model , then , there is $p_0\ge p+q$ in general, so AR(p_0) is called long auto-regressive model.

For ARMA(p,q) model equation (15), assume Initial Value is given

$$\beta^0=\left(\beta_1^0,\beta_2^0,\cdots,\beta_m^0\right)^T=\left[\varphi_1^0,\cdots,\varphi_p^0,\theta_1^0,\cdots,\theta_q^0\right]^T, m=p+q$$

according to equation (17), the gradient of model's sum of squared residuals $s(\beta)$ in the β^0 is

$$g_i^0=\left.\frac{\partial S(\beta)}{\partial\beta}\right|_{\beta=\beta_0}=-2\sum_{t=p+1}^{n}\left.\left\{\left[x_t-f(X_t-\beta)\right],\frac{\partial f(X_t,\beta)}{\partial\beta_i}\right\}\right|_{\beta=\beta_0}$$

$$i=1,2,\ldots,m \tag{18}$$

In vector form

$$g^0=\left.\frac{\partial S(\beta)}{\partial\beta}\right|_{\beta=\beta_0}=\left[\left.\frac{\partial S(\beta)}{\partial\beta_1}\right|_{\beta=\beta_0},\cdots,\left.\frac{\partial S(\beta)}{\partial\beta_m}\right|_{\beta=\beta_0}\right]^T \tag{19}$$

For the convenience of computing, x_t of the ARMA model can be Taylor expanded in the β^0

$$\begin{aligned}x_t=&f\left(X_t,\beta^0\right)+\left(\beta_1-\beta_1^0\right)\left.\frac{\partial f\left(X_t,\beta\right)}{\partial\beta_1}\right|_{\beta=\beta^0}\\&+\left(\beta_2-\beta_2^0\right)\left.\frac{\partial f\left(X_t,\beta\right)}{\partial\beta_2}\right|_{\beta=\beta^0}\\&+\cdots+\left(\beta_m-\beta_m^0\right)\left.\frac{\partial f\left(X_t,\beta\right)}{\partial\beta_m}\right|_{\beta=\beta^0}\\&+o\left(\left(\beta_1-\beta_1^0\right)^2+\left(\beta_2-\beta_2^0\right)^2+\cdots+\left(\beta_m-\beta_m^0\right)^2\right)\end{aligned} \tag{20}$$

Note $\eta_t=o\left(\left(\beta_1-\beta_1^0\right)^2+\left(\beta_2-\beta_2^0\right)^2+\cdots+\left(\beta_m-\beta_m^0\right)^2\right)$

Then move the first item of equation (20) on the right to left,

$$\begin{aligned}\varepsilon_t^0=x_t-f\left(X_t,\beta^0\right)=&\left(\beta_1-\beta_1^0\right)\left.\frac{\partial f\left(X_t,\beta\right)}{\partial\beta_1}\right|_{\beta=\beta^0}\\&+\left(\beta_2-\beta_2^0\right)\left.\frac{\partial f\left(X_t,\beta\right)}{\partial\beta_2}\right|_{\beta=\beta^0}\\&+\cdots+\left(\beta_m-\beta_m^0\right)\left.\frac{\partial f\left(X_t,\beta\right)}{\partial\beta_m}\right|_{\beta=\beta^0}+\eta_t\end{aligned} \tag{21}$$

Note $u_i^0=\beta_i-\beta_i^0$, $w_{ti}^0=\left.\frac{\partial f\left(X_t,\beta\right)}{\partial\beta_1}\right|_{\beta=\beta^0}$, $i=1,2,\ldots,m$,

written in vector form

$$u^0=[u_1^0,u_2^0,\cdots,u_m^0]^T \tag{22}$$

$$w_t^0=[w_{1,t}^0,w_{2,t}^0,\cdots,w_{m,t}^0]^T \tag{23}$$

Then ε^0 can be expressed as follows:

$$\varepsilon^0=w^0\left[u^0\right]^T+\eta \tag{24}$$

when $t=p+1,p+2,\cdots,n$, then

$$W^0=\begin{bmatrix}w_{1,p+1}^0 & w_{2,p+1}^0 & \cdots & w_{n,p+1}^0\\ w_{1,p+2}^0 & w_{2,p+2}^0 & \cdots & w_{n,p+2}^0\\ \cdots & \cdots & \cdots & \cdots\\ w_{1,n}^0 & w_{2,n}^0 & \cdots & w_{n,n}^0\end{bmatrix} \tag{25}$$

$$\eta=\left[\eta_{p+1},\eta_{p+2},\cdots,\eta_n\right]^T \tag{26}$$

$$\varepsilon^0=\left[\varepsilon_{p+1}^0,\varepsilon_{p+2}^0,\cdots,\varepsilon_n^0\right]^T \tag{27}$$

Thus, Algorithm steps are as follows:

Step 1: choose $\mu\in(0,1)$, initial parameter α^0 , growth factor $\nu>1$ (2 or 10);

Step 2: Given initial value β^0 , make $k=0$;

Step 3: Calculate $\varepsilon^k=\varepsilon(\beta^k)$, $S^k=S(\beta^k)=(\varepsilon^k)^T\varepsilon^k$;

Step 4: Calculate $g^k = (w^k)^T \varepsilon^k$;

Step 5: Calculate $p^k = [(w^k)^T w^k - \alpha^k I]^{-1} (w^k)^T \varepsilon^k]$;

Step 6: Calculate $\beta^{k+1} = \beta^k + p^k$, $\varepsilon^{k+1} = \varepsilon(\beta^{k+1})$, $S^{k+1} = (\varepsilon^{k+1})^T \varepsilon^{k+1}$;

Step 7: Meet H-convergence criteria, calculate β^{k+1}, S^{k+1};

Step 8: See if $S\beta^{k+1} < S\beta^k + \mu(g^k)^T p^k$: if yes, let $\alpha^{k+1} = \dfrac{\alpha^k}{v}$, and turn to Step 9; if no, let $\alpha^{k+1} = v\alpha^k$, and turn to Step 5;

Step 9: make $k = k+1$, turn to Step 3.

V. USING THE TEMPLATE

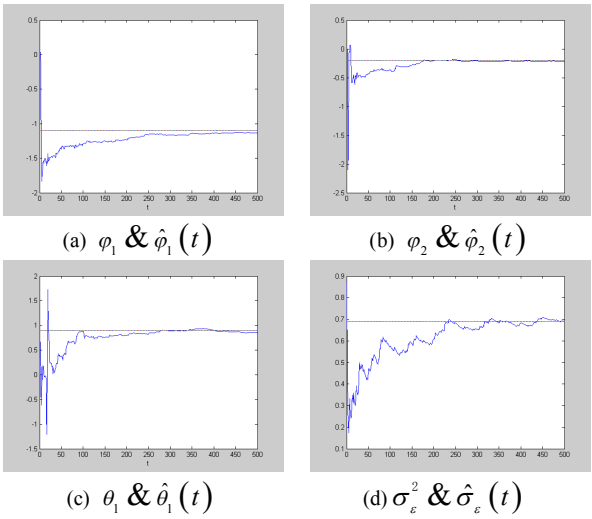
From AIC criteria and BIC criteria the order p=2, d=1, q=1 of forecast model is determined. Fit as ARIMA (2,1,1) model.

$$\left(1+1.1b+0.2b^2\right)(1-b)y(t) = (1-0.9b)\varepsilon(t)$$

Make $x(t) = (1-b)y(t)$ get ARMA(2, 1) model

$$\left(1+1.1b+0.2b^2\right)x(t) = (1-0.9b)\varepsilon(t)$$

Where $\varepsilon(t)$ is normal white noise whose mean value is zero, and variance is $\sigma_\varepsilon^2 = 0.69$, utilize output $x(t)$ to give the estimation value of model parameter and noise variance. Parameter estimation convergence based on weighted iteration stage is shown as Figure 1, where line represents the real value of parameter, curve represents parameter estimation value, and t is iteration step. When t=500, the corresponding parameter estimation value is $\hat{\varphi}_1 = -1.089593$, $\hat{\varphi}_2 = -0.200421$, $\hat{\theta}_1 = 0.902739$, $\hat{\sigma}_\varepsilon{}^2 = 0.692365$.



(a) φ_1 & $\hat{\varphi}_1(t)$ (b) φ_2 & $\hat{\varphi}_2(t)$

(c) θ_1 & $\hat{\theta}_1(t)$ (d) σ_ε^2 & $\hat{\sigma}_\varepsilon(t)$

Figure 1. The convergence simulation figure of parameter estimation in the weighted iteration stage parameter estimation algorithm

Simulation results show that the algorithm has better convergence. After 200 steps of iteration, the trends of parameters estimation curve have been steady, and the errorsbetween the true value and the estimation value has been less than 3%.

VI. CONCLUSION

On the basis of analyzing the least square algorithm and nonlinear time series data, ARIMA model is established. The new type of recursive iterative parameter estimation optimization algorithm is emphasized and algorithm's convergence proof and example simulation is given. The new algorithm has avoided the coupling with ARMA model parameter white noise estimation value and improved the accuracy of parameter estimation with better convergence. Simulation results show that this algorithm has fast convergence, stability and high accuracy. Under the condition of calculating smaller steps, this model has achieved better convergence effects and accuracy, which lays a solid foundation for the forecast of nonlinear time series, anomaly detection and the application of nonlinear system load forecast with theoretical significance and practical value.

ACKNOWLEDGMENT

This work is supported by the Natural Science Foundation of Hebei Province (F2011203149) and Scientific Research Project of the Higher Education Institutions of Hebei Province (ZD20131080).

REFERENCES

[1] Wang X, Li S M. The interval autoregressive time series model, Fuzzy Systems (FUZZ), 2011 IEEE International Conference on. IEEE, 2011, pp. 2528-2533

[2] Singh R, Balasundatam S. Application of extreme learning machine method for time series analysis [J]. International Journal of Intelligent Technology, 2007, 2(4): 256-262.

[3] G. Melard, Roc. Roy, A. Saidi. Exact maximum likelihood estimation of structured or unit root multivariate time series models, Computational Statistics and Data Analysis, 2006, 50(11): 2958-2986.

[4] B. M. Keel, E. G. J.Baxa. Overdetermined approach to autocorrelation based spectral moment estimators for use in Doppler weather radar, IEEE National Radar Conference - Proceedings, 1997: 300-305.

[5] B. Ye, C. Guo, Y. Cao. Identification of fuzzy model using evolutionary programming and least squares estimate, 2004 IEEE International Conference on Fuzzy Systems - Proceedings, 2004, 2: 593-598.

[6] Lemke T, Godsill S J. Linear gaussian computations for near-exact Bayesian Monte Carlo inference in skewed alpha-stable time series models, Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on. IEEE, 2012, pp. 3737-3740.

[7] Thiraviam A, Foley T, Malone L. Development of an acceleration model for subsea pressure, 2010 Proceedings of Reliability and Maintainability Symposium (RAMS), IEEE, 2010, pp. 1-6.

[8] Gonzalez L F, Regulski P, Wall P, et al. Fixed speed wind generator model parameter estimation using improved particle swarm optimization and system frequency disturbances, Renewable Power Generation (RPG 2011), IET Conference on. IET, 2011, pp. 1-5.

[9] Zhao C, Xu Q L, Aimin A, et al. Kinetic model parameter estimation by hybrid differential evolution algorithm, Control Conference (CCC), 2012 31st Chinese. IEEE, 2012, pp. 1788-1793.

[10] Jezierska A, Chaux C, Pesquet J C, et al. An EM Approach for Time-Variant Poisson-Gaussian Model Parameter Estimation [J], Signal Processing, IEEE Transactions on, 2014, 62(1): 17-30.