Convergence Results of the Analytic Center Estimator

Er-Wei Bai[1], Minyue Fu[2], Roberto Tempo[3] and Yinyu Ye[4]

[1]Dept. of Electrical and Computer Engineering
University of Iowa, Iowa City, Iowa 52242
erwei@icaen.uiowa.edu
[2]Dept. of Electrical and Computer Engineering,
University of Newcastle, N.S.W., 2308, Australia
eemf@hartley.newcastle.edu.au
[3]CENS-CNR
Politecnico di Torino, Torino, Italy
tempo@polito.it
[4]Dept. of Management Science
University of Iowa, Iowa City, Iowa 52242
yyye@dollar.biz.uiowa.edu

Abstract

The analytic center approach for bounded error parameter estimation was recently proposed as an alternative to the well-known Chybyshev and least squares estimates. In this paper, we show the asymptotic performance of this approach and prove that the analytic center converges to the true parameter under mild conditions.

This work was supported in part by grants of NSF(USA)[1,4], ARC(Australia)[2] and CENS-CNR(Italy)[1,3]. The first author especially thanks the support of CNR short-term mobility program 1998.

1 Introduction

Consider a single input-single output discrete-time system

$$y_i = \phi_i^T \theta + v_i, \quad i = 1, 2, \ldots, n \tag{1.1}$$

where $y_i \in \mathbf{R}$ is the system output, $\phi_i \in \mathbf{R}^m$ the measurable regressor, $\theta \in \mathbf{R}^m$ the unknown parameter vector to be identified and $v_i \in \mathbf{R}$ the noise. In the bounded error parameter estimation setting, it is assumed that the noise v_i is bounded by $\epsilon > 0$, i.e.,

$$|v_i| \leq \epsilon \tag{1.2}$$

for $i = 1, 2, ..., n$. Then, the membership set is defined by

$$\Omega^n = \bigcap_{i=1}^{n} \{\hat{\theta} \in \mathbf{R}^m \ : \ -\epsilon \leq y_i - \phi_i^T \hat{\theta} \leq \epsilon\}. \tag{1.3}$$

The membership set includes all the parameter estimates that are consistent with the equation (1.1), the input-output data and the noise bound (1.2). The goal is to compute a specific estimate in the membership-set enjoying some optimality properties. There is a large body of research (see for example [6, 7]) in this area. The most popular estimate along this direction is the Chebyshev center θ_c of the set Ω^n

$$\theta_c = \arg \min_{\hat{\theta} \in \Omega^n} \max_{\eta \in \Omega^n} \|\hat{\theta} - \eta\|$$

where $\|\cdot\|$ can be any l_p norm. This is the best worst-case estimate of the true but unknown system parameter vector in the sense that it minimizes the maximum "distance" between θ_c and the unknown parameter vector that generated the data. However, it is well-known that the Chebyshev center is very sensitive to outliers and moreover on-line sequential implementation of the Chebyshev center does not seem feasible. To overcome these difficulties, an analytic center estimate θ_a was proposed in [1, 2, 3]. It was shown in [1] that the analytic center θ_a is the one which minimizes the logarithmic average output error among all the estimate and allows an implementable sequential algorithm. The complexity of this sequential algorithm for computing a sequence of analytic centers upto observation time n is linear in terms of the maximum number of Newton iterations.

To be more specific, let us re-write the membership set

$$\Omega^n = \bigcap_{i=1}^{n} \{\hat{\theta} \in \mathbf{R}^m \ : \ -\epsilon \le y_i - \phi_i^T \hat{\theta} \le \epsilon\} = \{\hat{\theta} \in \mathbf{R}^m \ : \ (A^n)^T \hat{\theta} \le c^n\} \tag{1.4}$$

where

$$A^n = (a_1, a_2, ..., a_{2n}) \doteq (\phi_1, -\phi_1, \phi_2, -\phi_2, ..., \phi_n, -\phi_n) \in \mathbf{R}^{m \times 2n},$$

$$\text{and} \quad c^n = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_{2n-1} \\ c_{2n} \end{pmatrix} \doteq \begin{pmatrix} \epsilon + y_1 \\ \epsilon - y_1 \\ \vdots \\ \epsilon + y_n \\ \epsilon - y_n \end{pmatrix} \in \mathbf{R}^{2n}.$$

and the inequality applies element-wise. Then, the analytic center θ_a of Ω^n is an interior point of the set Ω^n maximizing the (dual) potential function

$$\theta_a = \arg \max_{\hat{\theta} \in \Omega^n} \psi(\hat{\theta}) = \arg \max_{\hat{\theta} \in \Omega^n} \frac{1}{2n} \sum_{i=1}^{2n} \ln(c_i - a_i^T \hat{\theta}). \tag{1.5}$$

The following theorem was given in [1] concerning the analytic center.

Theorem 1.1 *Consider the system (1.1), the noise bound (1.2) and the membership set (1.3). Then, the analytic center is the solution to the following optimization problem*

$$\theta_a = \arg \max_{\hat{\theta} \in \Omega^n} \Pi_{i=1}^n (\epsilon^2 - (y_i - \phi_i^T \hat{\theta})^2) = \arg \max_{\hat{\theta} \in \Omega^n} \sum_{i=1}^{n} \ln(\epsilon^2 - (y_i - \phi_i^T \hat{\theta})^2). \tag{1.6}$$

The main result of the present paper is to show the convergence of the analytic center to the true parameter under various conditions. The preliminary version of this paper was previously presented in a conference paper [4].

2 Convergence analysis

To avoid some unnecessary complications, we modify the noise bound and assume that

$$|v_i| \le \epsilon - \xi < \epsilon \tag{2.1}$$

for some arbitrarily small constant $\xi > 0$. Further, we define two conditions:

Condition 1: There exist $n_0 > 0$ and $\beta_1 \geq \beta_2 > 0$ such that for all $n \geq n_0$,

$$\beta_1 I \geq \frac{1}{n} \sum_{i=1}^{n} \phi_i \phi_i^T \geq \beta_2 I. \tag{2.2}$$

In the literature this condition is referred to as the Weak Persistent Excitation condition [5].

Condition 2:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \frac{v_i}{\epsilon^2 - v_i^2} \phi_i = 0. \tag{2.3}$$

Condition 2 is satisfied if v_i and ϕ_i are independent random variables with $\mathbf{E}(\phi_i) = 0$ or v_i is symmetric; i.e, $q(v_i) = q(-v_i)$ which implies $\mathbf{E}(\frac{v_i}{\epsilon^2 - v_i^2}) = 0$. The above also holds when ϕ_i is a deterministic time function and v_i is a symmetric independent random variable or when v_i is a deterministic time function but ϕ_i is an independent random variable with $\mathbf{E}(\phi_i) = 0$.

Three results are presented below. The first one answers the question when the analytic center gives a correct estimate, i.e., $\theta_a = \theta$. The second theorem, which is the main result in this section, studies the convergence property of the analytic center. The third theorem is a consequence of the second theorem which shows the convergence of θ_a to the unknown θ with probability one if v_i of zero mean.

Lemma 2.1 *Under Condition 1, $\theta_a = \theta$ if and only if*

$$\frac{1}{n} \sum_{i=1}^{n} \frac{v_i}{\epsilon^2 - v_i^2} \phi_i = 0. \tag{2.4}$$

Proof: Condition 1 guarantees the existence and uniqueness of the analytic center θ_a for $n \geq n_0$. Further, θ_a is the solution of

$$\frac{d\psi(\hat{\theta})}{d\hat{\theta}} = 0$$

where the potential function $\psi(\hat{\theta})$ is defined in (1.5), which is equivalent to

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\phi_i (y_i - \phi_i^T \theta_a)}{\epsilon^2 - (y_i - \phi_i^T \theta_a)^2} = 0. \tag{2.5}$$

If $\theta_a = \theta$, then this equation reduces to (2.4). Conversely, if (2.4) holds, then $\theta_a = \theta$ is a solution to the above. By the uniqueness of θ_a under Condition 1, this is the only solution. Hence, $\theta_a = \theta$ if and only if (2.4) holds.

Theorem 2.1 *The parameter estimation error given by the analytic center has the following bound for $n \geq n_0$:*

$$||\theta_a - \theta|| \leq || \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\phi_i \phi_i^T}{\epsilon^2 - v_i^2} \right)^{-1} || \cdot || \frac{1}{n} \sum_{i=1}^{n} \frac{2v_i}{\epsilon^2 - v_i^2} \phi_i ||. \tag{2.6}$$

In particular, $\theta_a = \theta$ if Conditions 1 and 2 hold.

Proof: Define the parameter estimation error $e = \theta - \theta_a$ and prediction error $\rho_i = y_i - \phi_i^T \theta_a$. Then,

$$\rho_i = v_i + \phi_i^T e.$$

Subsequently, (2.5) gives the unique solution to θ_a and is equivalent to

$$-\sum_{i=1}^{n} \frac{v_i \phi_i}{\epsilon^2 - \rho_i^2} = \sum_{i=1}^{n} \frac{\phi_i \phi_i^T}{\epsilon^2 - \rho_i^2} e. \tag{2.7}$$

Define

$$\eta_i = \frac{v_i \phi_i}{\epsilon^2 - \rho_i^2} - \frac{v_i \phi_i}{\epsilon^2 - v_i^2}.$$

Then, it is straightforward to check that

$$\eta_i = \frac{\phi_i \phi_i^T}{\epsilon^2 - \rho_i^2} \frac{v_i^2 + v_i \rho_i}{\epsilon^2 - v_i^2} e.$$

Adding $\sum_{i=1}^{n} \eta_i$ to the both sides of (2.7), we obtain

$$\begin{aligned} -\sum_{i=1}^{n} \frac{v_i \phi_i}{\epsilon^2 - v_i^2} &= \sum_{i=1}^{n} \frac{\phi_i \phi_i^T}{\epsilon^2 - \rho_i^2} \left(1 + \frac{v_i^2 + v_i \rho_i}{\epsilon^2 - v_i^2} \right) e \\ = & \left(\sum_{i=1}^{n} \frac{\epsilon^2 + v_i \rho_i}{(\epsilon^2 - v_i^2)(\epsilon^2 - \rho_i^2)} \phi_i \phi_i^T \right) e = \left(\sum_{i=1}^{n} \frac{d_i}{\epsilon^2 - v_i^2} \phi_i \phi_i^T \right) e \end{aligned}$$

where

$$d_i = \frac{\epsilon^2 + v_i \rho_i}{\epsilon^2 - \rho_i^2}.$$

Noting that $|v_i| < \epsilon$ and $|\rho_i| < \epsilon$, we get

$$d_i \geq \frac{1}{2}.$$

Thus,

$$e = -\left(\sum_{i=1}^{n} \frac{2d_i}{\epsilon^2 - v_i^2}\phi_i\phi_i^T\right)^{-1}\left(\sum_{i=1}^{n} \frac{2v_i}{\epsilon^2 - v_i^2}\phi_i\right) \tag{2.8}$$

and

$$||\theta - \theta_a|| \leq ||\left(\frac{1}{n}\sum_{i=1}^{n} \frac{\phi_i\phi_i^T}{\epsilon^2 - v_i^2}\right)^{-1}|| \cdot ||\frac{1}{n}\sum_{i=1}^{n} \frac{2v_i}{\epsilon^2 - v_i^2}\phi_i||.$$

This completes the proof.

Corollary 2.1 *Consider the system (1.1). Suppose that the noise v_i is a sequence of independently (not necessarily identically) distributed random variables of zero mean with the bound given by (2.1). Further, assume that the regressor Φ satisfies Condition 1 and is independent of v_i. Then, the analytic center θ_a converges to the true but unknown θ with probability one as $n \to \infty$.*

Proof: The hypothesis implies that the random variables $\frac{v_i}{\epsilon^2 - v_i^2}$ are independent with zero mean and finite convariance. Then, by the law of large numbers, Condition 2 is satisfied with probability one as $n \to \infty$. Therefore the conclusion follows.

3 Some remarks

It is interesting to compare the convergence conditions for the analytic center approach with the recursive least squares (RLS) given by

$$\begin{aligned} \theta_{ls} &= \left(\frac{1}{n}\sum_{i=1}^{n} \phi_i\phi_i^T\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} \phi_i y_i \\ &= \theta + \left(\frac{1}{n}\sum_{i=1}^{n} \phi_i\phi_i^T\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} \phi_i v_i. \end{aligned} \tag{3.9}$$

For the convergence of the RLS estimate, Condition 1 remains the same while Condition 2 is replaced with a slightly simpler condition

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} v_i\phi_i^T = 0. \tag{3.10}$$

The main tradeoff is that RLS is much simpler but does not guarantee that the solution lies in the membership set all the time. On the other hand, the analytic center requires more computation [1], in particular when n is large, but at the same time it is guaranteed to lie in the membership set. Therefore, it would be nice to start with the analytic center and then switch to RLS when n becomes large. This mixed approaches is simply to begin with an analytic center estimator and then switch to a RLS estimator when the estimate begins to converge. The key point, however, is to determine the switching time. To further elaborate on this, we assume zero mean and independence of the noise (not necessarily identical distributions). Then, the covariance of the RLS estimate error becomes

$$\mathrm{Cov}(\theta_{ls}-\theta) = R_n^{-1}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{k=1}^{n}\phi_i\phi_k^T\mathbf{E}(v_iv_k)R_n^{-1}$$

where

$$R_n = \frac{1}{n}\sum_{i=1}^{n}\phi_i\phi_i^T.$$

Denote

$$\sigma^2 = \max_i \mathbf{E}(v_i^2) = \max_i \int_{-\epsilon}^{\epsilon} x^2 q_i(v_i)dv_i$$

where $q_i(x)$ is the probability density function of v_i. Then,

$$\mathrm{Cov}(\theta_{ls}-\theta) \leq \frac{\sigma^2}{n}\left(\frac{1}{n}\sum_{i=1}^{n}\phi_i\phi_i^T\right)^{-1} \leq \frac{\sigma^2}{\beta_2 n}I \tag{3.11}$$

where β_2 is defined in (1.5). Note that $\sigma^2 \leq \epsilon^2$ because $|v_i| \leq \epsilon$ and

$$\int_{-\epsilon}^{\epsilon} x^2 q_i(x)dx \leq \epsilon^2\int_{-\epsilon}^{\epsilon} q_i(x)dx \leq \epsilon^2.$$

Accordingly, $\mathrm{Cov}(\theta_{ls}-\theta) \leq \frac{\epsilon^2}{\beta_1 n}I$ and

$$\mathbf{E}(\|\theta_{ls}-\theta\|^2) = \mathrm{Trace}\{\mathrm{Cov}(\theta_{ls}-\theta)\} \leq \frac{m\epsilon^2}{\beta_2 n}$$

where m is the dimension of θ. Consequently, for any given tolerances $\rho > 0$ and $\delta > 0$, we compute the switching iteration n

$$n \geq \frac{m\epsilon^2}{\beta_2\delta\rho^2}. \tag{3.12}$$

By the Chebyshev inequality, we obtain

$$\mathrm{Prob}\{\|\theta_{ls}-\theta\| \geq \rho\} \leq \frac{\mathbf{E}(\|\theta_{ls}-\theta\|^2)}{\rho^2} \leq \frac{m\epsilon^2}{\rho^2\beta_2 n} \leq \delta. \tag{3.13}$$

Clearly, there is no guarantee that if $n \geq \frac{m\epsilon^2}{\beta_2 \delta \rho^2}$, the RSL estimate lies in the membership set. However, with a high probability at least $1-\delta$, the RLS estimate θ_{ls} is ρ-close to the true θ provided that (3.12) is satisfied and thus for small enough $\rho > 0$, θ_{ls} is very likely to be in the membership set. Therefore, we can safely switch the estimate from the analytic center to least squares. The above idea is easily implementable.

References

[1] Bai, E.W., Y. Ye and R. Tempo (1997). "Bounded error parameter estimation: a sequential analytic center approach", *Proc. of 1997 CDC, San Diego*, also to appear in *IEEE Trans. on Automatic Control*, 1999.

[2] Bai, E.W., R. Tempo and Y. Ye (1998). "Recent Results on the Analytic Center Approach for Bounded Error Parameter Estimation", in Learning Control and Hybrid Systems, Y. Yamamoto and S. Hara Editors, Springer-Verlag, pp.245-254, London.

[3] Bai, E.W., R. Tempo and Y. Ye (1998). "Open Problems in Sequential Parametric Estimation", in Open Problems in Mathematical Systems Theory and Control, V. Blondel, E. Sontag, M. Vidyasagar and J. Willems Editors, Springer-Verlag, pp. 19-22, London.

[4] Bai, E.W., M. Fu, R. Tempo and Y. Ye (1998). "Analytic Center Approach to Parameter Estimation: Convergence Analysis", *Proc. of 1998 ICASSP*, Seattle.

[5] Goodwin, G. and K.S. Sin (1984). ADAPTIVE FILTERING, PREDICTION AND CONTROL, Prentice-Hall, New York.

[6] Milanese, M. and A. Vicino (1991). "Optimal estimation theory for dynamic systems with set membership uncertainty: an overview", *Automatica*, **27**, 997-1009.

[7] Piet-Lahanier, H. and E. Walter (1994). "Exact description of feasible parameter sets and minimax estimator", *Int. J. Adap. Contr. and Signal Proc.*, **8**, 5-14.