# ANALYTIC CENTER APPROACH TO PARAMETER ESTIMATION: CONVERGENCE ANALYSIS

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

[1]Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242 USA
[2]Department of Electrical and Computer Engineering, University of Newcastle, N.S.W., 2308, Australia
[3]CENS-CNR, Politecnico di Torino, Torino 10129 ITALY
[4]Department of Management Science, University of Iowa, Iowa City, IA 52242 USA

## ABSTRACT

The so-called *analytic center approach* to parameter estimation has been proposed recently as an alternative to the well-known least squares approach. This new approach offers a parameter estimate that is consistent with the past data observations, has a simple geometric interpretation, and is computable sequentially. In this paper, we study the asymptotic performance of the analytic center approach and show that the resulting estimate converges to the true parameter asymptotically, provided some mild conditions are satisfied. These conditions involve some weak persistent excitation and independence between noise and regressor, similar to the least squares case. This result is used to derive a new parameter estimation approach which offers both good transient and asymptotic performances.

## 1. INTRODUCTION

It is well-known that the commonly used least squares based parameter estimation algorithms often suffer from a poor transient performance. This is caused by the lack of observation data for establishing reliable noise statistics or non-stationarity of noises and/or parameters.

An alternative approach called **analytic center approach** to parameter estimation has been proposed recently by Bai, Ye and Tempo [2]. To understand this, we consider the following system:

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

where $y_i \in \mathbf{R}$ is the output, $q_i \in \mathbf{R}^m$ is the regressor, $\theta \in \mathbf{R}^m$ is the unknown parameter vector, $v_i \in \mathbf{R}$ is the measurement noise with $|v_i| \leq \bar{v}$. We assume that the regressor is bounded, i.e.,

$$||q_i|| \leq \alpha \tag{1.2}$$

for all $i = 1, 2, \ldots, n$. This assumption is necessary for the output $y_i$ to be bounded.

Define the membership set [1]

$$\Omega_n = \bigcap_{i=1}^{n} \{\hat{\theta} : (y_i - q_i^T\hat{\theta})^2 \leq \bar{v}^2\}. \tag{1.3}$$

The analytic center of $\Omega_n$ is defined to be

$$\theta_n^a = \arg\min_{\hat{\theta}\in\Omega_n} f_n(\hat{\theta}) \tag{1.4}$$

where

$$f_n(\hat{\theta}) = -\frac{1}{n}\sum_{i=1}^{n} \ln(\bar{v}^2 - (y_i - q_i^T\hat{\theta})^2) \tag{1.5}$$

An obvious observation of the analytic center is that the estimate $\theta_n^a$ at any $n$ is consistent with the past observations, i.e., the estimated noise samples $\hat{v}_i = y_i - q_i^T\theta_n^a, i = 1, \cdots, n$ do not violate the prescribed bound $\bar{v}$. Due to its consistency with the past observations, the analytic center approach gives a better transient performance than the least squares approach in general.

A simple interpretation of the analytic center is that this is a "geometric center" because $\theta_n^a$ can be rewritten as follows.

$$\theta_n^a = \arg\max_{\hat{\theta}\in\Omega_n} \prod_{i=1}^{n} (\bar{v}^2 - (y_i - q_i^T\hat{\theta})^2) \tag{1.6}$$

In contrast, the least squares estimate computes

$$\theta_n^{LS} = \arg\min_{\hat{\theta}} \sum_{i=1}^{n} (y_i - q_i^T\hat{\theta})^2 \tag{1.7}$$

which is an "arithmetic center".

The main results of [2] can be summarized as follows: The analytic center estimate is an estimate inside the membership set which maximizes the complementary average output error (1.6). Moreover, contrary of other centers like Chebyshev, it allows for an easy-to-compute sequential algorithm. The maximum number of Newton iterations required to compute a sequence of analytic centers is linear in the number of observed data points and it is comparable to the complexity of off-line algorithms for estimating a single analytic center.

In this paper, we are interested in the asymptotic performance of the analytic center. The motivations are as follows. First, we know that the least squares estimate asymptotically converges to the true parameter under some mild conditions, i.e., persistent excitation and independence between the noise and regressor [4]. Hence, it is important to know if the analytic center possesses a similar property or not. Secondly, the least square estimate is very efficient to compute using a recursive formula while the computation of the analytic center is in general much more involved when the number of data observations becomes large. Knowing the asymptotic performance of the analytic center will help us design an efficient algorithm to reduce its computational complexity.

The main results of this paper are summarized as follows. In Section 2, we show that the analytic center estimate $\theta_n^a$ converges to the true parameter $\theta$ provided that the noise and the regressor satisfy a weak persistent excitation condition and an independence condition. The second condition is also shown to be necessary for asymptotic convergence. These conditions are analogous (but different) to those required for the asymptotic convergence of the least squares estimate. In Section 3, we extend the result above by proposing a mixed approach to parameter estimation. This approach combines the analytic center and the least squares, and is expected to help give both good transient and asymptotic performances. A similar asymptotic convergence property is established. A simulation example is given in Section 4 to illustrate the asymptotic behaviours of the least squares approach and the analytic center approach. This example also demonstrates the tradeoff between them.

## 2. CONVERGENCE ANALYSIS

In this section, we analyze the asymptotic performance of the analytic center, i.e., the behaviour of $\theta_n^a$ as $n \to \infty$. To this end, we assume in the rest of the paper that the noise $v_i$ is "normalized" such that its bound $\bar{v} < 1$ but close to 1. Accordingly, we modify the analytic center a bit by using a slightly larger membership set

$$\Omega_n = \bigcap_{i=1}^{n} \{\hat{\theta} : (y_i - q_i^T \hat{\theta})^2 < 1\}. \tag{2.8}$$

and a slightly different objective function

$$f_n(\hat{\theta}) = -\frac{1}{n} \sum_{i=1}^{n} \ln(1 - (y_i - q_i^T \hat{\theta})^2). \tag{2.9}$$

We define two conditions:

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

$$\frac{1}{n} \sum_{i=1}^{n} q_i q_i^T \geq \beta I. \tag{2.10}$$

In the literature the above condition is referred to as the Weak Persistent Excitation condition [3].
**Condition 2:**

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \frac{v_i}{1 - v_i^2} q_i^T = 0. \tag{2.11}$$

**Remark 2.1** *Condition 2 is satisfied if $v_i$ and $q_i$ are independent ergodic random variables with $\mathrm{E}(q_i) = 0$ or $v_i$ is symmetric; i.e, $p(x) = p(-x)$ which implies $\mathrm{E}(\frac{v_i}{1-v_i^2}) = 0$. The above also holds when $q_i$ is a deterministic time function and $v_i$ is a symmetric independent random variable or when $v_i$ is a deterministic time function but $q_i$ is an independent random variable with $\mathrm{E}(q_i) = 0$.*

Two theorems are presented below. The first one answers the question when the analytic center gives a correct estimate, i.e., $\theta_n^a = \theta$. The second theorem, which is the main result in this section, studies the convergence property of the analytic center.

**Theorem 2.1** *Suppose Condition 1 holds and $n \geq n_0$. Then, $\theta_n^a = \theta$ if and only if the following condition holds:*

$$\sum_{i=1}^{n} \frac{v_i}{1 - v_i^2} q_i^T = 0. \tag{2.12}$$

**Proof:** Condition 1 guarantees the existence and uniqueness of the analytic center $\theta_n^a$ for $n \geq n_0$. Further, $\theta_n^a$ is the solution to

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

which is equivalent to

$$-\frac{1}{n} \sum_{i=1}^{n} \frac{q_i (y_i - q_i^T \theta_n^a)}{1 - (y_i - q_i^T \theta_n^a)^2} = 0. \tag{2.13}$$

If $\theta_n^a = \theta$, then the above reduces to (2.12). Conversely, if (2.12) holds, then $\theta_n^a = \theta$ is a solution to the above. By the uniqueness of $\theta_n^a$, this is the only solution. Hence, $\theta_n^a = \theta$ if and only if (2.12) holds. ∎

**Theorem 2.2** *The parameter estimation error given by the analytic center has the following bound for $n \geq n_0$:*

$$||\theta_n^a - \theta|| \leq || \left( \frac{1}{n} \sum_{i=1}^{n} \frac{q_i q_i^T}{1 - v_i^2} \right)^{-1} || \cdot || \frac{1}{n} \sum_{i=1}^{n} \frac{2v_i}{1 - v_i^2} q_i ||. \tag{2.14}$$

*In particular, $\theta_n^a \to \theta$ as $n \to \infty$ if Conditions 1 and 2 hold. Conversely, if $\theta_n^a \to \theta$ as $n \to \infty$, then Condition 2 must hold.*

**Proof:** The proof is rather lengthy, and will be included in the full version of the paper. ■

## 3. MIXED APPROACH TO PARAMETER ESTIMATION

It is interesting to compare the convergence conditions for the analytic center approach with the recursive least squares (RLS) approach. In the LS case Condition 1 remains the same while Condition 2 is replaced with a slightly simpler condition

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} v_i q_i^T = 0. \tag{3.15}$$

The solution to the LS problem (3.16) is given by

$$\begin{aligned}\theta_n^{\text{LS}} &= \left(\frac{1}{n}\sum_{i=1}^{n} q_i q_i^T\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} q_i y_i \\ &= \theta + \left(\frac{1}{n}\sum_{i=1}^{n} q_i q_i^T\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} q_i v_i. \end{aligned}\tag{3.16}$$

The main tradeoff is that RLS is much simpler but does not guarantee that the solution lies in the membership set all the time, while the analytic center is much more difficult to compute [2]. In particular, the computing time grows as $n$ increases. Therefore, it would be nice to start with the analytic center and then switch to RLS when $n$ becomes large. To this end, we propose a mixed approach which takes the advantages of the both approaches.

Our approach computes a mixed analytic-arithmetic center

$$\begin{aligned}\theta_n^m &= \arg\min_{\hat\theta}\{\sum_{i=1}^{n} w_{i,n}^a \ln(1-(y_i - q_i^T\hat\theta)^2)^{-1} \\ &\quad + \sum_{i=1}^{n} w_{i,n}^{\text{LS}}(y_i - q_i^T\hat\theta)^2\}\end{aligned}\tag{3.17}$$

where $w_{i,n}^a$ and $w_{i,n}^{\text{LS}}$ are weights satisfying the following condition:

**Condition 3**: $0 \le w_{i,n}^a, w_{i,n}^{\text{LS}} \le 1$; there exists $\gamma > 0$ such that for all $n > 0$,

$$\frac{1}{n}\sum_{i=1}^{n}(w_{i,n}^a + w_{i,n}^{\text{LS}}) \ge \gamma \tag{3.18}$$

**Remark 3.1** *The condition above allows both the analytic center approach and the LS approach as special cases. It also accommodates exponential forgetting factors, i.e., the weights decay exponentially in reverse time. The membership set for $\hat\theta$ in (3.18) is not explicitly defined. In fact, this set is simply defined by the constraints $(y_i - q_i^T\hat\theta)^2 < 1$ for those $i$ at which $w_{i,n}^a \neq 0$.*

Since the function to be minimized in (3.18) is strictly convex, the optimal solution $\theta_n^m$ is implicitly given by

$$\frac{1}{n}\sum_{i=1}^{n}\left(w_{i,n}^a \frac{q_i(y_i - q_i^T\theta_n^m)}{1-(y_i - q_i^T\theta_n^m)^2} + w_{i,n}^{\text{LS}} q_i (y_i - q_i^T\theta_n^m)\right) = 0 \tag{3.19}$$

We again obtain a good asymptotic performance.

**Theorem 3.1** *Suppose Condition 3 holds. Then the mixed center $\theta_n^m$ has the following convergence property:*

$$\begin{aligned}\|\theta_n^m - \theta\| \le &\left\|\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{w_{i,n}^a}{1-v_i^2} + w_{i,n}^{LS}\right) q_i q_i^T\right)^{-1}\right\| \\ &\cdot \left\|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{2w_{i,n}^a}{1-v_i^2} + w_{i,n}^{LS}\right) v_i q_i\right\|\end{aligned}\tag{3.20}$$

*In particular, $\theta_n^m \to \theta$ as $n \to \infty$ if Condition 1 holds and that*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{2w_{i,n}^a}{1-v_i^2} + w_{i,n}^{LS}\right) v_i q_i = 0 \tag{3.21}$$

*Conversely, if $\theta_n^m \to \theta$ as $n \to \infty$, then (3.21) must hold.*

**Proof:** The proof is rather lengthy, and will be included in the full version of the paper. ■

Because the analytic center part contributes significantly to the computation, it is desirable to keep the number of nonzero terms $w_{i,n}^a$ as little as possible. One option is to keep only a few terms corresponding to the most recent sampling points. The second option is to keep those weights which correspond to large $(y_i - q_i^T\theta_{n-1}^m)^2$. The third option, which is to be exploited below, is simply to begin with an analytic center estimator and then switch to a LS estimator when the estimate starts to converge.

To elaborate this last option further, we assume zero mean and independence of the noise (not necessarily identical distributions). Then, the covariance of the LS estimate error

$$\text{Cov}(\theta_n^{\text{LS}} - \theta) = R_n^{-1}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{k=1}^{n} q_i q_k^T \mathbf{E}(v_i v_k) R_n^{-1}$$

where

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

Denote

$$\sigma^2 = \max_i \mathbf{E}(v_i^2)$$

$$\mathbf{E}(v_i^2) = \int_{-1}^{1} x^2 p_i(x)dx$$

and $p_i(x)$ is the probability density function for $v_i$. Then,

$$\text{Cov}(\theta_n^{LS} - \theta) = \frac{\sigma^2}{n}\left(\frac{1}{n}\sum_{i=1}^{n} q_i q_i^T\right)^{-1} \leq \frac{\sigma^2}{\beta n} I \quad (3.22)$$

Note that $\sigma^2 \leq 1$ because $|v_i| < 1$ and

$$\int_{-1}^{1} x^2 p_i(x) dx \leq \int_{-1}^{1} p_i(x) dx \leq 1.$$

Thus, $\text{Cov}(\theta_n^{LS} - \theta) \leq \frac{1}{\beta n} I$ and

$$\mathbf{E}(\|\theta_n^{LS} - \theta\|^2) = Trace\{\text{Cov}(\theta_n^{LS} - \theta)\} \leq \frac{m}{\beta n}.$$

where $m$ is the dimension of $\theta$.

Consequently, for any given $\epsilon > 0$ and $\delta > 0$, we compute the switch iteration $n$

$$n \geq \frac{m}{\beta \delta \epsilon^2} \quad (3.23)$$

so by the Chebyshev inequality, we obtain

$$\text{Prob}\{\|\theta_n^{LS} - \theta\| \geq \epsilon\} \leq \frac{\mathbf{E}(\|\theta_n^{LS} - \theta\|^2)}{\epsilon^2} \leq \frac{m}{\epsilon^2 \beta n} \leq \delta. \quad (3.24)$$

In other words, with probability at least $1 - \delta$, the LS estimate is $\epsilon$-close to the true $\theta$ provided (3.23) is satisfied. Thus, we can safely switch the estimate from analytic center to least squares.

## 4. SIMULATION EXAMPLE

The simulation example is a second order system

$$y_i = (u_i, u_{i-1}) \begin{pmatrix} a \\ b \end{pmatrix} + v_i$$

where the unknown parameter vector $\theta = \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 5 \\ 5 \end{pmatrix}$, the disturbance $v_i$'s are independent random variables uniformly in the interval $[0.95, 0.95]$ and the input sequence $\{u_i\} = \{1, 0, 1, 0, ...\}$. Subsequently, both Conditions 1 and 2 are satisfied. Figure 1 shows the membership sets, the analytic center estimates and the least squares estimates. The least squares estimates are computed using a standard recursive formula with initial estimate $\theta_0^{LS} = (4, 4)^T$ and inverse covariance matrix $P(0) = I$.

Several observations are in order. First, we note that a least squares estimate is poorer than the analytic center estimate in the sense that it may lie outside of the membership sets, e.g. when $n = 15$ and 25. In contrast, the analytic center, by definition, always lies inside of the membership set. Secondly, As $n$ increases, both estimates converge to the true parameter. Therefore, it is desirable to start with the analytic center for a good transient performance and then switch to the least squares for numerical efficiency. This can be done by adjusting the weights in the mixed approach.

## 5. REFERENCES


[1] Special Issue on Bounded Error Estimation(Part II), *Int. J. Adap. Control and Signal Proc*, **9**, 1995.

[2] Bai, E. W., Y. Ye and R. Tempo, "Bounded error parameter estimation: a sequential analytic center approach," *Proc. of IEEE Conf on Decision and Control, 1997*, to appear

[3] Goodwin, G. and K.W. Sin, *Adaptive Filtering, Prediction and Control*. Prentice-Hall, Engle-wood Cliffs, N.J., 1984.

[4] Ljung, L., *System Identification: Theory for the Users*. Prentice-Hall, Engle-wood Cliffs, N.J., 1987.


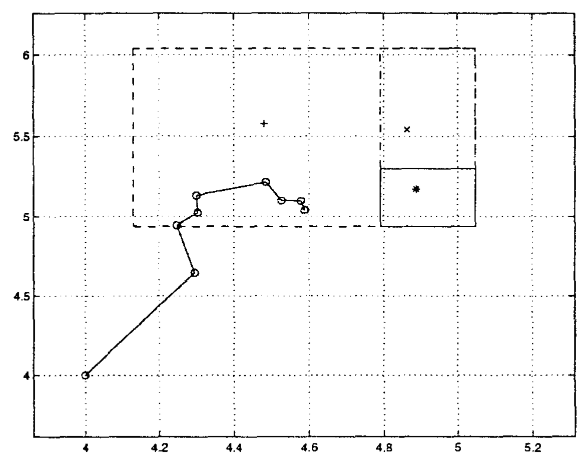


Figure 1: Membership sets(- - n=10, -.- n=15, solid n=25), Analytic Center(+ n=10, x n=15, * n=25) and RLS(o at n=1,4,7,10,13,15,18,22,25)