Bounded Error Parameter Estimation: A Sequential Analytic Center Approach

Er-Wei Bai[+], Yinyu Ye[*] and Roberto Tempo[!]

[+]Dept. of Electrical and Computer Engineering
University of Iowa, Iowa City, Iowa 52242
erwei@icaen.uiowa.edu

[*]Dept. of Management Science
University of Iowa, Iowa City, Iowa 52242
yyye@dollar.biz.uiowa.edu

[!]CENS-CNR
Politecnico di Torino, Torino, Italy
tempo@polito.it

February 1997

Abstract

In this paper, a sequential analytic-center approach for bounded error parameter estimation is proposed. The analytic center minimizes the "average" output error and allows an easy-to-compute sequential algorithm. With little computational effort two ellipsoids centered at the analytic center can be obtained as well: One inscribes and the other outscribes the so-called membership set. Finally, a sequential algorithm is presented along with its convergence proof. The complexity of this algorithm in terms of the maximum number of Newton iterations is shown to be *linear* in the number of observed data points.

1 Problem Statement

In this paper, we consider a single input-single output discrete-time system

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

where $y_i \in R$ is the system output, $q_i \in R^m$ the measurable regressor, $\theta \in R^m$ the unknown parameter vector to be identified, and $v_i \in R$ the noise. In vector form, the system can be re-written more compactly as

$$y = q\theta + v \tag{1.2}$$

where

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \ q = \begin{pmatrix} q_1^T \\ q_2^T \\ \vdots \\ q_n^T \end{pmatrix}, v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

The purpose of parametric system identification is to find an estimate $\hat{\theta}$ of the unknown parameter vector θ from available input-output measurements y and q. Up to present date, the mainstream of parameter estimation research is to use the stochastic framework [13] and [20] where the noise is assumed to be a sequence of random variables with some known probabilistic properties. A huge body of literature exists along this direction and various theoretically sound and practically efficient schemes have been developed. A deficiency with this paradigm, however, is that to obtain a reliable estimate from a relatively small data set, much a priori statistical information of the unknown noise sequence needs to be available. If the statistics of the noise sequence are unknown or erroneously assumed, the probabilistic arguments may be invalid. An alternative method is the bounded error parameter estimation, also denoted as the membership set approach, where the a priori information is a given bound $\epsilon > 0$ on the unknown noise, i.e.,

$$|v_i| \le \epsilon, \quad i = 1, 2, ..., n. \tag{1.3}$$

Then, the membership set

$$\Omega = \{\theta \in R^m \ : \ -\epsilon \le y_i - q_i^T \theta \le \epsilon, \ i = 1, 2, ..., n\} \tag{1.4}$$

is the set of all parameter estimates that are consistent with the equation (1.1), the input-output data and the assumed noise bound (1.3). Depending on the specific application, both stochastic and bounded-error approaches have their places: If one can make qualitative assumptions about the averaging properties of the noise sequence, then a stochastic framework may be quite appropriate. If such assumptions can not be made or a hard bound on the noise sequence is readily available, e.g., in the case of quantization, then a bounded error approach may be more suitable.

There is a large body of research (see for example five special issues (1990, 1992, 1994, 1995) and the survey papers by Milanese and Vicino [14], Norton [16], Piet-Lahanier and Water [19] and the references therein) devoted to the bounded error estimation. Within the bounded error method, there are basically two approaches. The first one aims towards obtaining the exact membership set. There are several advantages to constructing the exact set. For example, the description of this set may reveal how its size can be reduced by a better experimental design or a better choice of inputs. Also, it may be desirable to know the maximal "distance" (parameter estimation error) between two points of the membership set. Unfortunately, despite substantial progress in recent years, the computational complexity involved in obtaining the membership set can be very high. This is because in general the set can become fairly complex, e.g., the number of vertices can grow exponentially as the number of observations increases. To reduce the computational burden, approaches of approximating the membership set are widely used. The idea is to obtain an approximate outer bounding of the membership set by a simple-shaped set, usually an ellipsoid [10] and [11] or an orthotope [8] or a parallelotope [24]. By minimizing the "volume" of this outer bounding set, one attempts to reduce the level of conservatism introduced by these approximations. The computations involved in the outer bounding approach are much simpler and can sometimes be made recursive. However, these approaches are often conservative if the membership set is "geometrically" very different from the shape of the outer bounding sets. Also, recursive implementation of outer bounding algorithms can introduce further conservatism [17].

The other approach in the bounded error estimation is to find just one estimate in the membership set that has some optimality properties. The most popular one is the Chebyshev θ_c center [23] of the set Ω

$$\theta_c = \arg\min_{\theta\in\Omega}\left\{\max_{\eta\in\Omega}\|\theta-\eta\|\right\}$$

where $\|\cdot\|$ represents the 2 norm throughout the paper unless otherwise stated. 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. For a given data set, the calculation of θ_c is basically a linear programming [15] problem. It should be pointed out that θ_c is in general not a smooth function of the data y_i and q_i so that the change in θ_c corresponding to a new observed data could not be easily followed. This implies that no sequential algorithm could be easily developed for computing θ_c. Another shortcoming of the Chebyshev center is that the optimality of θ_c is in terms of parameter estimation error. If the output error $y_i - q_i^T\hat{\theta}$ is of concern, nothing can be said. To this end, the projection estimate θ_p, also denoted as the constrained least squares estimate, is proposed [9]

$$\theta_p = \arg\min_{\hat{\theta}\in\Omega}\|y-q\hat{\theta}\|^2 = \arg\min_{\hat{\theta}\in\Omega}\sum_{i=1}^{n}(y_i-q_i^T\hat{\theta})^2.$$

This is the best estimate in the membership set Ω that minimizes the "average" output error. We now make the following observations concerning θ_c and θ_p:

- Notice that θ_p is the constrained least squares estimate. It is well-known [7] that the unconstrained least squares estimate is in general not in the membership set. Thus, by the convexity of the set Ω and the quadratic form of the cost function, θ_p is very often at the boundary of the membership set Ω. On the other hand, θ_c is usually an interior point of Ω.

- The Chebyshev center θ_c depends only on the membership set Ω and not on the constraints $-\epsilon \le y_i - q_i^T\theta \le \epsilon$ that are redundant in forming the set Ω. On contrary, the projection estimate θ_p depends on all the constraints $-\epsilon \le y_i - q_i^T\theta \le \epsilon$ $(i =$

$1, 2, ..., n)$ even when some of them are redundant. To be more specific, consider the example

$$y_i = q_i\theta + v_i, \quad y_i, q_i, \theta, v_i \in R, \ \ |v_i| \leq 1, \ \theta = 1. \tag{1.5}$$

Let $\{q_i\}_1^n = \{1, 1, 1, 1, ..., 1\}$, $\{v_i\}_1^n = \{-1, 0, -3/4, -3/4, -3/4, ..., -3/4\}$. Then, $\{y_i\}_1^n = \{0, 1, 1/4, 1/4, 1/4, ..., 1/4\}$ and the membership set

$$\Omega = [0, 1]$$

for $n > 1$. By simple calculations, we obtain

$$\theta_c = 0.5 \ \text{ and } \ \theta_p = (n + 2)/4n$$

for $n > 2$. We see that θ_c does not depend on the constraint $-1 \leq y_i - q_i\theta \leq 1$ for $2 < i \leq n$ while θ_p does even though these constraints are redundant in forming Ω.

- Although θ_p is a continuous function of the data y_i and q_i and also for a given data set θ_p can be determined by convex programming, there is no sequential algorithm available in the literature for computing θ_p. By sequential algorithm, we mean that when the new data y_i and q_i become available, the calculation of θ_p at ith time should use the previous θ_p at $(i-1)$th time.

- The projection estimate θ_p is the optimal point within Ω that minimizes the "average" output error. Since the search is carried out within the set Ω, $\epsilon^2 - (y_i - q_i^T\hat{\theta})^2 \geq 0$ for every $\hat{\theta} \in \Omega$. In other words, the output error $(y_i - q_i^T\hat{\theta})^2$ is minimized within the set Ω if and only if the complementary distance $\epsilon^2 - (y_i - q_i^T\hat{\theta})^2$ is maximized. Thus, the projection estimate θ_p can be re-formulated as

$$\theta_p = \arg\min_{\hat{\theta}\in\Omega} \sum_{i=1}^{n} (y_i - q_i^T\hat{\theta})^2 = \arg\max_{\hat{\theta}\in\Omega} \sum_{i=1}^{n} (\epsilon^2 - (y_i - q_i^T\hat{\theta})^2).$$

- The projection estimate θ_p is the one that minimizes the arithmetic-average output error. Generally, the "average" output error does not have to be in the arithmetic form. For instance,

$$\max_{\hat{\theta}\in\Omega} \Pi_{i=1}^{n} (\epsilon^2 - (y_i - q_i^T\hat{\theta})^2) \quad \text{or} \quad \max_{\hat{\theta}\in\Omega} \sum_{i=1}^{n} \ln(\epsilon^2 - (y_i - q_i^T\hat{\theta})^2)$$

implies that the "geometric-average" output error is minimized.

In this paper, our goal is to obtain a sequential and easy to implement algorithm which produces a single parameter estimate within the membership set. Although not optimal in terms of the worst-case parameter estimation error like the Chebyshev center θ_c, this estimate is interpolatory and therefore from Information-Based Complexity [23] the worst-case parameter error is guaranteed to be within a factor of two of the parameter error provided by the Chebyshev center θ_c. Moreover, the estimate provided by our algorithm is optimal in terms of some "average" output error. The main results of the paper are briefly described below:

1. We propose to use the analytic center θ_a of Ω as the estimate of unknown θ. We show in the paper that θ_a maximizes the complementary "average" output error or equivalently minimizes the logarithm weighted arithmetic average output error

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

   The computation involved in the evaluation of θ_a is shown to be relatively easy. Moreover, with little extra computation, two ellipsoids E_{in} and E_{out} centered at the analytic center θ_a can be obtained inscribing and outscribing the membership set Ω respectively, i.e.

   $$E_{in} \subseteq \Omega \subseteq E_{out}.$$

   These two ellipsoids can be useful in bounded error estimation. They may reveal the geometric structure of the membership set Ω and also allow to detect some of those constraints which do not shape the membership set. It is important to point out here that the idea of analytic center is not new. It was first presented in [21] and since then it has been widely studied in the literature of linear programming; see, e.g., [12] and [15]. It has also been mentioned in [22] as a potential tool for solving model matching problems in control. The use of analytic-center approach in identification, particularly in bounded error identification, is however new and provides a vital and easy to calculate alternative to traditional Chebyshev center and projection estimate approaches.

2. While well studied in the linear programming literature, most existing algorithms for calculating the analytic center are batch type. Due to the nature of identification, the input-output data is observed sequentially in time i. In this paper, we present a sequential algorithm for calculating the analytic center. Let $\theta_a(k)$ denote the analytic center based on the observed data $y_i, q_i, 1 \leq i \leq k$. The proposed algorithm calculates the analytic center $\theta_a(k)$ based on the previous analytic center $\theta_a(k-1)$. In other words, earlier computations of $\theta_a(k-1)$ would not be wasted in evaluating the current analytic center $\theta_a(k)$. The convergence of the algorithm is achieved. More significantly, it is shown that the number of Newton iterations [18] required to compute *a sequence* of analytic center $\theta_a(i)$ up to time $i = k$ is linear in k, comparable to the complexity of the batch algorithm for computing *a single* analytic center $\theta_a(k)$ at time $i = k$. This result is new in the literature of linear programming as well.

We end this section by giving an outline of the paper. Section 2 defines the analytic center and derives some properties of the analytic center. Sections 3-4 present algorithms, along with their convergence and complexity analysis, for computing the analytic center non-sequentially and sequentially, respectively. Some remarks are given in Section 5 and Section 6 contains the proofs.

2 Analytic Center

Consider the system (1.1) and the noise bound (1.3). Define

$$A = (a_1, a_2, ..., a_{2n}) := (q_1, -q_1, q_2, -q_2, ..., q_n, -q_n) \in R^{m \times 2n},$$

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

Then, the membership set Ω of (1.4) can be re-written as

$$\Omega = \{\theta \in R^m \ : \ -\epsilon \leq y_i - q_i^T \theta \leq \epsilon, \ i = 1, 2, ..., n\}$$

$$= \{\theta \in R^m \ : \ A^T \theta \leq c\} \tag{2.1}$$

where the vector inequality applies element-wise.

In the paper, we assume that

(**A1**) The rank of the matrix A is m for $n \geq m$. This is necessary otherwise the membership set Ω is unbounded.

(**A2**) The membership set Ω has a non-empty interior, i.e., there exists some $\rho > 0$ and some $\hat{\theta} \in \Omega$ so that

$$-\epsilon + \rho \leq y_i - q_i^T \hat{\theta} \leq \epsilon - \rho, \quad i = 1, 2, ..., n. \tag{2.2}$$

Let $\phi_d(\theta)$ denote the (dual) potential function

$$\phi_d(\theta) := \sum_{i=1}^{2n} \ln(c_i - a_i^T \theta). \tag{2.3}$$

The analytic center θ_a of Ω is an interior point of Ω maximizing the (dual) potential function $\phi_d(\theta)$

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

To derive the optimality condition, note

$$0 = \frac{d\phi_d(\theta)}{d\theta}|_{\theta=\theta_a} = \sum_{i=1}^{2n} \frac{a_i}{c_i - a_i^T \theta_a}$$

$$= (a_1, a_2, ..., a_{2n}) \begin{pmatrix} \frac{1}{c_1 - a_1^T \theta_a} & 0 & \cdots & 0 \\ 0 & \frac{1}{c_2 - a_2^T \theta_a} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{c_{2n} - a_{2n}^T \theta_a} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = A S_a^{-1} e = A x_a$$

where

$$0 < s_a = c - A^T \theta_a \in R^{2n}, S_a = \mathrm{diag}(s_a) \in R^{2n \times 2n},$$

$$e = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = S_a x_a = X_a s_a \in R^{2n}, 0 < x_a = S_a^{-1} e \in R^{2n}, X_a = \mathrm{diag}(x_a) \in R^{2n \times 2n}.$$

Furthermore,

$$2n = e^T e = e^T X_a s_a = x_a^T (c - A^T \theta_a) = x_a^T c = c^T x_a.$$

Since $\phi_d(\theta)$ is a strictly concave function in Ω and Ω is convex and bounded, the analytic center θ_a exists and is unique. Moreover, the first-order necessary optimality condition must also be sufficient. Therefore, $\theta \in \Omega$ is the analytic center θ_a if and only if there exists some

$$x > 0,\ Ax = 0,\ s = c - A^T \theta > 0,\ Xs - e = 0 \tag{2.4}$$

with $X = \mathrm{diag}(x)$ and $S = \mathrm{diag}(s)$ non-singular.

We are now in a position to state the main theorem of this section.

Theorem 2.1 *Consider the system (1.1), the noise bound (1.3) and the membership set (1.4). Then,*

1.

$$\theta_a = arg \max_{\hat{\theta} \in \Omega} \Pi_{i=1}^n (\epsilon^2 - (y_i - q_i^T \hat{\theta})^2) = arg \max_{\hat{\theta} \in \Omega} \sum_{i=1}^n \ln(\epsilon^2 - (y_i - q_i^T \hat{\theta})^2). \tag{2.5}$$

2. Let

$$L = (\frac{q_1}{\epsilon + y_1 - q_1^T \theta_a}, \frac{-q_1}{\epsilon - y_1 + q_1^T \theta_a}, \ldots, \frac{q_n}{\epsilon + y_n - q_n^T \theta_a}, \frac{-q_n}{\epsilon - y_n + q_n^T \theta_a}),$$

and define two ellipsoids E_{out} and E_{in}

$$E_{out} = \{\theta \in R^m \ :\ (\theta - \theta_a)^T L L^T (\theta - \theta_a) \le (2n)^2\}$$

and

$$E_{in} = \{\theta \in R^m \ :\ (\theta - \theta_a)^T L L^T (\theta - \theta_a) \le 1\}. \tag{2.6}$$

Then,

$$E_{in} \subseteq \Omega \subseteq E_{out}.$$

Proof: See Section 6.

According to the above theorem, the analytic center θ_a has a clear meaning in bounded error estimation. In fact, it minimizes the "average" output error. Also with little computation, two ellipsoids E_{out} and E_{in}, that outscribes and inscribes the membership set Ω respectively, can be easily determined. Note that neither the Chebyshev center θ_c nor the projection estimate θ_p has this property.

It is also interesting to compare the output error distribution generated by θ_a with that generated by θ_p. The projection estimate θ_p minimizes the average error at a price that the maximum error can be large. On the other hand, the so called minmax estimate θ_{mm}

$$\theta_{mm} = \arg\min_{\hat{\theta}\in\Omega}\left\{\max_{1\le i\le n}(y_i - q_i^T\hat{\theta})^2\right\} = \arg\max_{\hat{\theta}\in\Omega}\left\{\min_{1\le i\le n}(\epsilon^2 - (y_i - q_i^T\hat{\theta})^2)\right\}$$

minimizes the maximum error at a price that the average error can be large. Usually, the error distribution generated by θ_{mm} is flatter than that generated by θ_p. To see the output error distribution generated by the analytic center θ_a, assume the noise bound $\epsilon = 1$ for simplicity. Note this can always be done by multiplying $1/\epsilon$ to both sides of (1.1) and re-naming y_i/ϵ, q_i/ϵ and v_i/ϵ as the new output, the new regressor and the new noise respectively. Now, because of the logarithmic weights in (2.5) and $0 \le 1 - (y_i - q_i^T\theta)^2 = \epsilon^2 - (y_i - q_i^T\theta)^2 \le 1$ for all $\theta \in \Omega$, small complementary output errors in particular the minimum complementary output error $\min_i(\epsilon^2 - (y_i - q_i^T\theta)^2) = \min_i(1 - (y_i - q_i^T\theta)^2)$ in evaluating the analytic center θ_a play a more important role than they play in evaluating the projection estimate θ_p. Thus, the minimum complementary output error generated by θ_a tends to be larger than that generated by θ_p. Equivalently, the maximum output error $\max_i(y_i - q_i^T\theta)^2$ generated by θ_a tends to be smaller than that generated by θ_p. On the other hand, the analytic center θ_a minimizes the "average" output error while the minmax estimate θ_{mm} emphasizes a single maximum error. Hence, the average output error due to θ_a is smaller than that due to θ_{mm}. In summary, in terms of the output error distribution, the analytic center estimate θ_a is in between the minmax estimate θ_{mm} and the projection estimate θ_p. To illustrate the idea, consider the scalar example of (1.5) again. Let

$$n = 2, \{q_i\}_{i=1}^2 = \{-0.5, 1\}, \{v_i\}_{i=1}^2 = \{0.5, 1\}.$$

Then,

$$\{y_i\}_{i=1}^2 = \{0, 2\}, \Omega = [1, 2] \text{ and } \theta_p = 1.6, \theta_a = 1.4557, \theta_{mm} = 1.333.$$

The following Figure shows the output errors at $i = 1, 2$ generated by θ_p ($|y_1 - q_1^T\theta_p| = 0.8$ and $|y_2 - q_2^T\theta_p| = 0.4$ denoted by "+"), by $\theta_a(x)$ ($|y_1 - q_1^T\theta_a| = 0.73$ and $|y_2 - q_2^T\theta_p| = 0.54$ denoted by "x") and by θ_{mm} ($|y_1 - q_1^T\theta_{mm}| = 0.667$, $|y_2 - q_2^T\theta_{mm}| = 0.667$ denoted by "o") respectively.

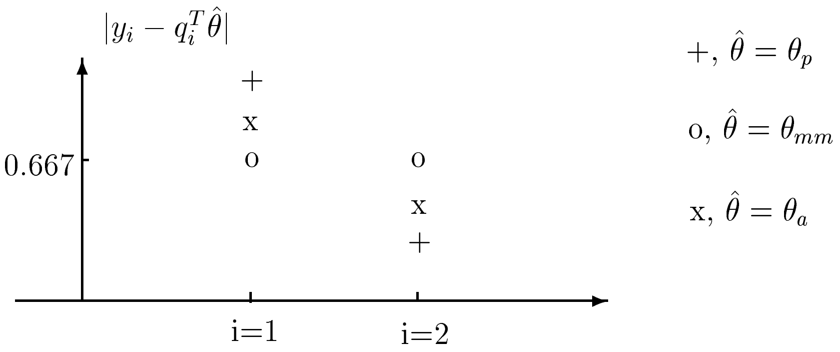

Figure 1

The error distribution generated by θ_a is flatter than that of θ_p but not as flat as that of θ_{mm}. Thus, in the case that both the maximum error and the average error are of concern, the analytic center estimate may be more suitable.

3 Non-sequential Algorithm

Optimality properties of the analytic center θ_a in bounded error identification have been given in the previous section. In this section, we discuss the calculation of θ_a by means of a non-sequential algorithm. This subject has been studied in the literature of linear programming [15]. However, for easy exposition of the sequential algorithm to be presented in the next section, we will provide some discussion on calculating θ_a from the given data set.

For a given data set and an arbitrary small error tolerance $\delta > 0$, we aim to calculate an approximate analytic center θ_δ which is δ-close to the analytic center θ_a. The closeness can be expressed in many different ways. The standard way in linear programming is as follows: $(x_\delta, \theta_\delta, s_\delta)$ is said to be a δ-approximate center pair if $s_\delta = c - A^T\theta_\delta > 0$, $x_\delta > 0$, $Ax_\delta = 0$, $c^T x_\delta = 2n$, and

$$\|X_\delta(c - A^T\theta_\delta) - e\| = \|X_\delta s_\delta - e\| \leq \delta < 1.$$

Clearly, θ_δ does not coincide exactly with θ_a. However, the following results indicate the quality of a δ-approximate center; see [12] for proof and further details.

Lemma 3.1 *Let $(x_\delta, \theta_\delta, s_\delta)$ be a δ-approximate center pair and let (x_a, θ_a, s_a) be the analytic center pair of Ω. Then,*

$$0 \leq \phi_d(\theta_a) - \phi_d(\theta_\delta) \leq \frac{\delta^2}{2(1-\delta)}$$

and

$$\|S_\delta^{-1} s_a - e\| \leq \frac{\delta}{1-\delta}.$$

Therefore, for small δ, $\phi_d(\theta_a)$ and $\phi_d(\theta_\delta)$ are “close”. This implies that the output errors generated by θ_δ and by θ_a are small.

Non-sequential algorithm for computing θ_δ:

Consider the system (1.1). Suppose the data set y_i, q_i $(i = 1, 2, ..., n)$, the noise bound ϵ and the tolerance $\delta > 0$ are given. The goal is to compute a δ-approximate center pair $(x_\delta, \theta_\delta, s_\delta)$ of Ω.

Initial: Set $j = 0$ and $x(0) = \frac{1}{\epsilon} e = \begin{pmatrix} \frac{1}{\epsilon} \\ \vdots \\ \frac{1}{\epsilon} \end{pmatrix} \in R^{2n}$.

Phase I:

Step 1: For any j, define $X(j) = \mathrm{diag}(x(j))$,

$$\theta(j) = (AX(j)^2A^T)^{-1}AX(j)^2(c - X(j)^{-1}e),$$

$$s(j) = c - A^T\theta(j),$$

$$S(j) = \mathrm{diag}(s(j)),$$

$$p(j) = X(j)(c - A^T\theta(j)) - e = X(j)s(j) - e.$$

If $\|p(j)\| \leq \delta \leq 1/2$, set $\theta_\delta = \theta(j), x_\delta = x(j), s_\delta = s(j)$ and stop; else if $\delta < \|p(j)\| \leq 1/2$, then $(\theta(j), x(j), s(j))$ is a 1/2-approximate center and go to Step 3 in Phase II; otherwise $\|p(j)\| > 1/2$ and go to Step 2.

Step 2: Define

$$\hat{x}(j+1) = x(j) - \frac{1}{4}X(j)p(j)/\|p(j)\|,$$

$$x(j+1) = 2n\hat{x}(j+1)/(c^T\hat{x}(j+1)).$$

Set $j = j + 1$ and go to Step 1.

Phase II: After Phase I, we have a 1/2-approximate center pair.

Step 3: Solve for $d_x(j), d_\theta(j)$ and $d_s(j)$ from the system of linear equations (Newton's method)

$$S(j)d_x(j) + X(j)d_s(j) = e - X(j)s(j),$$

$$Ad_x(j) = 0,$$

$$A^T d_\theta(j) + d_s(j) = 0$$

and define

$$x(j+1) = x(j) + d_x(j),\ \theta(j+1) = \theta(j) + d_\theta(j),\ s(j+1) = s(j) + d_s(j).$$

Go to Step 4.

Step 4: If $\|X(j+1)(c - A^T\theta(j+1)) - e\| \leq \delta$, set $\theta_\delta = \theta(j+1), x_\delta = x(j+1), s_\delta = s(j+1)$ and stop; otherwise, set $j = j + 1$ and go to Step 3.

The algorithm has two phases. The purpose of Phase I is to find a 1/2-approximate center from the initial $x(0)$ and that of Phase II is to find a δ-approximate center from a 1/2-approximate center. For Phase II, we have the following results from [15]:

Lemma 3.2 *Let (x, θ, s) be a $\delta(\le 1/2)$-approximate center pair of Ω. Apply the Newton method to solve for d_x, d_θ and d_s from the system of linear equations*

$$\begin{aligned} S d_x + X d_s &= e - Xs, \\ A d_x &= 0, \\ A^T d_\theta + d_s &= 0, \end{aligned}$$

and assign

$$x^+ = x + d_x, \ \theta^+ = \theta + d_\theta, \ s^+ = s + d_s.$$

Then, we have

$$x^+ > 0, \quad Ax^+ = 0, \quad c^T x^+ = 2n, \quad s^+ > 0$$

and

$$\|X^+(c - A^T\theta^+) - e\| \le \frac{\sqrt{2}\delta^2}{4(1-\delta)} \le \delta^2.$$

In other words, the error $\|X(c - A^T\theta) - e\|$ converges to zero quadratically.

For Phase I, let any $x > 0$, with $Ax = 0$, and any $\theta \in \Omega$, i.e., $s = c - A^T\theta > 0$, consider the (primal) homogeneous potential function

$$\phi_p(x) := 2n \ln(c^T x) - \sum_{k=1}^{2n} \ln x_k.$$

Note that $\phi_p(x)$ is homogeneous of degree 0, i.e., $\phi_p(\alpha x) = \phi_p(x)$ for any $\alpha > 0$. Now, consider a primal-dual potential function

$$\phi(x, \theta) := \phi_p(x) - \phi_d(\theta) = 2n \ln(c^T x) - \sum_{k=1}^{2n} \ln(x_k s_k)$$

$$= 2n \ln((c - A^T\theta)^T x) - \sum_{k=1}^{2n} \ln(x_k s_k)$$

$$= 2n \ln(s^T x) - \sum_{k=1}^{2n} \ln(x_k s_k) = 2n \ln(\sum_{k=1}^{2n} s_k x_k) - \ln(\Pi_{k=1}^{2n}(x_k s_k)).$$

Note that for $y_i > 0, \ i = 1, 2, ..., k$, $(\frac{\sum_{i=1}^k y_i}{k})^k \ge \Pi_{i=1}^k y_i$ and this implies $k \ln(\sum_{i=1}^k y_i) - \ln(\Pi_{i=1}^k y_i) \ge k \ln k$. Thus, it follows that $\phi(x, \theta) \ge 2n \ln(2n)$, i.e., for any $x > 0$ with $Ax = 0$ and any $\theta \in \Omega$,

$$\phi_p(x) \ge \phi_d(\theta) + 2n \ln(2n).$$

In particular, let θ be the one satisfying (2.2), then

$$\phi_p(x) \geq 2n \ln \rho + 2n \ln(2n).$$

Now, let $x > 0$, $Ax = 0$ and $c^T x = 2n$. Define

$$x^+ = x - \frac{1}{4} \frac{Xp}{\|p\|},$$

with

$$\theta = (AX^2A^T)^{-1}AX^2(c - X^{-1}e), \quad p = X(c - A^T\theta) - e.$$

Then,

$$Ax^+ = A(x - \frac{1}{4}\frac{Xp}{\|p\|)}) = -\frac{1}{4}AX^2[c - A^T(AX^2A^T)^{-1}AX^2(c - X^{-1}e)]/\|p\| = 0$$

and

$$x^+ = X(e - \frac{1}{4}\frac{p}{\|p\|}) > 0.$$

Moreover, if $x^+ = x + d_x$ with $d_x = -\frac{1}{4}\frac{Xp}{\|p\|}$, we have from [12] that

$$\log(c^T x^+) - \log(c^T x) \leq \frac{1}{c^T x} c^T (x^+ - x) = \frac{1}{c^T x} c^T d_x,$$

$$-\sum \log x_j^+ + \sum \log x_j \leq -e^T X^{-1} d_x + \frac{\|X^{-1}d_x\|^2}{2(1 - \|X^{-1}d_x\|_\infty)},$$

$$\begin{array}{rcl} \phi_p(x^+) - \phi_p(x) & \leq & \frac{2n}{c^T x} c^T d_x - e^T X^{-1} d_x + \frac{\|X^{-1}d_x\|^2}{2(1 - \|X^{-1}d_x\|_\infty)} \\ & = & (c - X^{-1}e)^T d_x + \frac{\|X^{-1}d_x\|^2}{2(1 - \|X^{-1}d_x\|_\infty)} \\ & \leq & -\frac{1}{4}\|p\| + 1/24. \end{array}$$

Thus, as long as $\|p\| \geq 1/2 > 0$, it is guaranteed that

$$\phi_p(x^+) - \phi_p(x) \leq -1/12.$$

Since $\phi_p(x_a)$ is bounded from below by $2n \ln \rho + 2n \ln(2n)$ and each iteration reduces the primal potential function $\phi_p(x)$ by at least $1/12$, we shall generate a pair (x, θ), in finite steps, such that

$$\|X(c - A^T\theta) - e\| \leq 1/2.$$

It is important to point out that if the above inequality holds, then $c - A^T\theta \geq 0$ and this implies $\theta \in \Omega$ and it is a 1/2-approximate center. Once a 1/2-approximate center is obtained, we may use Phase II to obtain a δ-approximate center.

Combining Lemma 3.2 and the above discussion and also noting that fact that the amount of work in Phase II is one Newton iteration and the amount of work in Phase I is basically solving a linear least squares problem, identical to one Newton iteration, the following theorem is in place.

Theorem 3.1 *The non-sequential algorithm for calculating a δ-approximate center terminates in a finite number of iterations and the maximum number of Newton iterations required is bounded by $O(n \cdot \ln \frac{\epsilon}{\rho} + \ln \ln \frac{2}{\delta})$ with ϵ and ρ defined in (1.3) and (2.2) respectively. Moreover, if $1/2 \leq \delta$, the maximum number of Newton iterations is bounded by $O(n \cdot \ln \frac{\epsilon}{\rho})$.*

Proof: See Section 6.

4 Sequential Algorithm

Due to the nature of identification, the input-output data y_i, q_i are observed sequentially in time i. In the remaining of the paper, with a slight abuse of notation, we use the upper i to emphasize the dependence on time i. At each i, the membership set is given by

$$\begin{aligned} \Omega^i &= \{\theta \in R^m \ : \ -\epsilon \leq y_k - q_k^T\theta \leq \epsilon, \ k = 1, 2, ..., i\} \\ &= \{\theta \in R^m \ : \ (A^i)^T\theta \leq c^i\} \end{aligned} \tag{4.1}$$

where

$$A^i = (a_1, a_2, ..., a_{2i}) := (q_1, -q_1, q_2, -q_2, ..., q_i, -q_i) \in R^{m\times 2i},$$

$$\text{and} \quad c^i = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_{2i-1} \\ c_{2i} \end{pmatrix} := \begin{pmatrix} \epsilon + y_1 \\ \epsilon - y_1 \\ \vdots \\ \epsilon + y_i \\ \epsilon - y_i \end{pmatrix} \in R^{2i}.$$

For identification purposes, we need to compute the analytic center θ_a^i or a δ-approximate center of Ω^i for each $m \le i \le n$ (note that for $i < m$, rank $A^i < m$ and therefore Ω^i is unbounded). One way is to directly apply the Non-sequential algorithm presented in Section 3 to find a sequence of δ-approximate centers for $m \le i \le n$. However, the total computation involved is high. Here, we present a sequential algorithm in which the calculation of $\theta_a^i(\theta_\delta^i)$ is based on θ_a^{i-1} (θ_δ^{i-1}). In other words, earlier computations would not be wasted in computing the current analytic center or the δ-approximate center. It will be shown that the total number of Newton iterations required to compute a sequence of δ-approximate centers θ_δ^m, θ_δ^{m+1},...,θ_δ^n is bounded by $O(n \cdot (\ln \frac{\epsilon}{\rho} + \ln \ln \frac{2}{\delta}))$, linearly in n and thus comparable to the complexity of the Non-sequential algorithm for computing *a single* δ-approximate center θ_δ^n.

Sequential Algorithm:

Consider the system (1.1), the noise bound ϵ and the tolerance $\delta > 0$. Suppose the data set is given sequentially, i.e., at time i, only $y_k, q_k, k = 1, 2, ..., i$ are available. The goal is to compute a sequence of δ-approximate center pair $(x_\delta^i, \theta_\delta^i, s_\delta^i)$, $i = m, m+1, ..., n$ sequentially.

Step 1: Start at $i = m$. Let

$$x^m(0) = \begin{pmatrix} 1/\epsilon \\ \vdots \\ 1/\epsilon \end{pmatrix} \in R^{2m}.$$

Apply the Non-sequential algorithm of the previous section with $x(0)$ replaced by $x^m(0)$ to calculate a 1/2-approximate center and a δ-approximate center θ_δ^m for Ω^m. Let T^m denotes the number of Newton iterations to find the 1/2-approximate center pair

$$(x^m(T^m), \theta^m(T^m), s^m(T^m)) = (x_{\delta=1/2}^m, \theta_{\delta=1/2}^m, s_{\delta=1/2}^m).$$

Go to Step 2.

Step 2: When new data y_{i+1}, q_{i+1} become available for $m \le i < n$, let

$$x^{i+1}(0) = \begin{pmatrix} x^i(T^i) \\ 1/\epsilon \\ 1/\epsilon \end{pmatrix} \in R^{2(i+1)}.$$

Apply the Non-sequential algorithm, with $x(0)$ replaced by $x^{i+1}(0)$, to calculate a 1/2-approximate center and a δ-approximate center θ_δ^{i+1} for Ω^{i+1}. Let T^{i+1} denotes the number of Newton iterations to find the 1/2-approximate center pair

$$(x^{i+1}(T^{i+1}), \theta^{i+1}(T^{i+1}), s^{i+1}(T^{i+1})) = (x^{i+1}_{\delta=1/2}, \theta^{i+1}_{\delta=1/2}, s^{i+1}_{\delta=1/2})$$

at $(i+1)th$ iteration. If $i = n-1$, stop. Otherwise, set $i = i+1$ and repeat Step 2.

Theorem 4.1 *The sequential algorithm terminates in a finite number of iterations and the maximum number of Newton iterations required to compute a sequence of δ-approximate centers θ_δ^m, θ_δ^{m+1},...,θ_δ^n of $\Omega^m, \Omega^{m+1}, ..., \Omega^n$ is bounded by $O(n \cdot (\ln \frac{\epsilon}{\rho} + \ln\ln \frac{2}{\delta}))$, linearly in n.*

Proof: See Section 6.

The idea of both the Non-sequential algorithm and the Sequential algorithm is to reduce the prime potential function $\phi_p(x)$. The difference is that at each i, the Non-sequential algorithm starts at $x^{i+1}(0) = \frac{1}{\epsilon} e_{2(i+1)} = \frac{1}{\epsilon}(1, 1, ..., 1)^T \in R^{2(i+1)}$ and the Sequential algorithm starts at $x^{i+1}(0) = \begin{pmatrix} x^i(T^i) \\ 1/\epsilon \\ 1/\epsilon \end{pmatrix}$. It can be easily verified that

$$\phi_p^{i+1}(\begin{pmatrix} x^i(T^i) \\ 1/\epsilon \\ 1/\epsilon \end{pmatrix}) - \phi_p^{i+1}(\frac{1}{\epsilon} e_{2(i+1)}) \leq -(1/12) \cdot T^i.$$

This implies that by using the Sequential algorithm, ϕ_p^{i+1} is simultaneously decreased by the same amount when reducing ϕ_p^i. Thus, the Sequential algorithm needs much less computation than repeatedly apply the Non-sequential algorithm.

We next illustrate the performance of the algorithm using a simulation example. The system considered is a second order system

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

with unknown $\theta = \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 5 \\ 5 \end{pmatrix}$. For simulation, the input sequence applied was $\{u_i\} = \{1, 0, 1, 0, 1, 0, ...\}$ and v_i were chosen as independent random variable uniformly distributed in the interval $[-1,\ 1]$ $(i.e., \epsilon = 1)$. Figure 2 shows the membership sets Ω^i (denoted by solid boxes) and the approximate analytic centers θ_δ^i (denoted by x's) for $\delta = 0.01$ at $i = 10, 30$ and 70 respectively.

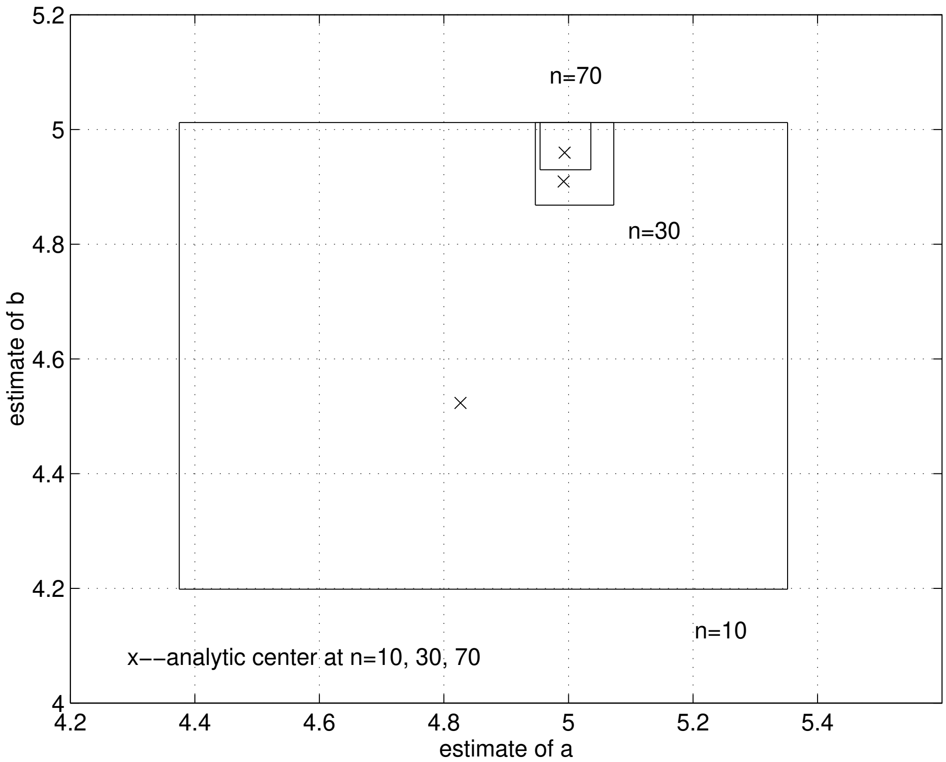

Figure 2:

5 Concluding Remarks

In this paper, an analytic-center approach is proposed for bounded error parameter estimation. The analytic center θ_a minimizes the logarithm weighted average output error and allows an easy to compute sequential algorithm. This easy-to-implement property distinguishes the analytic-center approach from the Chebyshev center θ_c, the projection estimate θ_p, and the minmax estimate θ_{mm} methods. To the best knowledge of the authors, there is no sequential algorithm for computing θ_c or θ_p or θ_{mm}. This leads to the following interesting question: Is it possible to find a sequential algorithm to calculate θ_c or θ_p or θ_{mm}? Conceptually, it is hard to have a sequential algorithm for θ_c and θ_{mm} because they are not smooth functions of the data y_i and q_i and the changes on θ_c and θ_{mm} corresponding to the new observation y_i and q_i can not be easily followed. On the other hand, θ_p is a smooth function of the data. Taking the previous projection estimate at time $(i-1)$ as the starting point for computing the projection estimate at i would provide some advantage. This topic is currently under our investigation.

Another interesting point is that both the projection estimate θ_p and the minmax estimate θ_{mm} minimize the output error. The projection estimate θ_p minimizes the average error despite the fact that the maximum error at some points can be large while the minmax estimate θ_{mm} minimizes the worst-case error despite the fact that the average error can be large. Thus, θ_p and θ_{mm} may be viewed as two extremes. In a sense, the analytic center θ_a balances these two extremes and considers simultaneously the average and the worst-case error. It would be an interesting piece of work to work out this simple observation quantitatively.

6 Proofs

Proof of Theorem 2.1: We first prove the first part of the theorem.

$$\theta_a = \arg\max_{\hat\theta\in\Omega}\sum_{i=1}^{2n}\ln(c_i - a_i^T\hat\theta) = \arg\max_{\hat\theta\in\Omega}\sum_{i=1}^{n}\ln[(c_{2i-1} - a_{2i-1}^T\hat\theta)(c_{2i} - a_{2i}^T\hat\theta)]$$

$$= \arg\max_{\hat\theta\in\Omega}\sum_{i=1}^{n}\ln(\epsilon + y_i - q_i^T\hat\theta)(\epsilon - y_i + q_i^T\hat\theta)$$

$$= \arg\max_{\hat\theta\in\Omega}\sum_{i=1}^{n}\ln[\epsilon^2 - (y_i - q_i^T\hat\theta)^2] = \arg\max_{\hat\theta\in\Omega}\Pi_{i=1}^{n}[\epsilon^2 - (y_i - q_i^T\hat\theta)^2].$$

This completes the first part. For the second part, let $\theta\in\Omega$ which implies $s = c - A^T\theta > 0$ and

$$(\theta-\theta_a)^T LL^T(\theta-\theta_a)$$

$$
\begin{aligned}
&= \left\| \begin{pmatrix} \frac{1}{c_1 - a_1^T\theta_a} & 0 & \cdots & 0 \\ 0 & \frac{1}{c_2 - a_2^T\theta_a} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{c_{2n} - a_{2n}^T\theta_a} \end{pmatrix} \begin{pmatrix} c_1 - q_1^T\theta - c_1 + q_1^T\theta_a \\ \vdots \\ c_{2n} - q_{2n}^T\theta - c_{2n} + q_{2n}^T\theta_a \end{pmatrix} \right\|^2 \\
&= \|S_a^{-1}(s - s_a)\|^2 = \|X_a(s - s_a)\|^2 = \|X_a s - e\|^2 = \|X_a s\|^2 - 2s^T X_a e + 2n \\
&= \|X_a s\|^2 - 2(c - A^T\theta)^T x_a + 2n \\
&= \|X_a s\|^2 - 2n = \sum_{i=1}^{2n}((x_a)_i s_i)^2 - 2n \le (\sum_{i=1}^{2n}(x_a)_i s_i)^2 - 2n \\
&= (x_a^T s)^2 - 2n = (x_a^T c)^2 - 2n = (2n)^2 - 2n < (2n)^2.
\end{aligned}
$$

This implies $\theta \in \Omega \Longrightarrow \theta \in E_{out}$. Now, let $\theta \in E_{in}$, i.e.,

$$
1 \ge (\theta - \theta_a)^T LL^T(\theta - \theta_a) = (\frac{(s_1 - (s_a)_1)}{(s_a)_1})^2 + (\frac{(s_2 - (s_a)_2)}{(s_a)_2})^2 + \ldots + (\frac{(s_{2n} - (s_a)_{2n})}{(s_a)_{2n}})^2.
$$

The analytic center θ_a is an interior point of Ω and thus, $s_a = c - A^T\theta_a > 0$ and $(s_a)_i > 0$. This implies $s_i \ge 0$ for each i, otherwise $(\frac{(s_i - (s_a)_i)}{(s_a)_i})^2 > 1$ for some i. Therefore, $s = c - A^T\theta \ge 0$ for every $\theta \in E_{in}$ and hence $\theta \in \Omega$. Combining those two parts, we have $E_{in} \subseteq \Omega \subseteq E_{out}$.

Proof of Theorem 3.1: Note that $x(0) > 0$, $A^T x = 0$ and $c^T x(0) = 2n$. From the discussion before the theorem, we have that in at most $24n \ln \frac{\epsilon}{\rho}$ Newton iterations, a 1/2-approximate center is provided and the algorithm goes to Phase II. Now, according to Lemma 3.2, the convergence to a δ-approximate center from a 1/2-approximate center is quadratic and in fact is bounded by k iterations where k satisfies

$$
(\frac{1}{2})^{2^k} \le \delta
$$

or

$$
k \ge \frac{1}{\ln(2)} \ln \frac{\ln(\frac{1}{\delta})}{\ln(2)} = O(\ln\ln(\frac{2}{\delta})).
$$

Thus, the total number of Newton iterations is bounded by $O(n \ln \frac{\epsilon}{\rho} + \ln\ln(\frac{2}{\delta}))$.

Proof of Theorem 4.1: Note $x^k(T^k) > 0, A^k x^k(T^k) = 0, (c^k)^T x^k(T^k) = 2k$ and $x^{k+1}(0) > 0, A^{k+1}x^{k+1}(0) = 0, (c^{k+1})^T x^{k+1}(0) = 2(k+1)$. According to Theorem 3.1, at each $k =$

$m+1, m+2, ..., n$, the required number of Newton iterations to find a δ-approximate center is bounded by $T^k + O(\ln\ln(2/\delta))$ and the required number of Newton iterations to find a $1/2$-approximate center is bounded by T^k. Therefore, the total number to find a sequence of δ-approximate centers $\theta_\delta^m, \theta_\delta^{m+1}, ..., \theta_\delta^n$ is bounded by $T^m + T^{m+1} + ... + T^n + O(n\ln\ln(2/\delta))$. We now show that

$$T^m + T^{m+1} + ... + T^n \leq 24n \ln \frac{\epsilon}{\rho}.$$

To this end, for each $k \geq m$, let $e_{2k} = (1,1,...,1)^T \in R^{2k}$. Note that $\phi_p^k(x^k(T^k)) \leq \phi_p^k(x^k(0)) - (1/12)\cdot T^k$ and this implies

$$\phi_p^k(\frac{1}{\epsilon}e_{2k}) - \phi_p^k(x^k(T^k)) = \phi_p^k(\frac{1}{\epsilon}e_{2k}) - \phi_p^k(x^k(0)) + \phi_p^k(x^k(0)) - \phi_p^k(x^k(T^k))$$

$$\geq 2k\ln(2k) - 2k\ln(\frac{1}{\epsilon}) - \phi_p^k\begin{pmatrix} x^{k-1}(T^{k-1}) \\ \frac{1}{\epsilon} \\ \vdots \\ \frac{1}{\epsilon}\end{pmatrix} + (1/12)\cdot T^k$$

$$= -2k\ln(\frac{1}{\epsilon}) + \sum_{j=1}^{2(k-1)} x^{k-1}(T^{k-1})_j + 2\ln(\frac{1}{\epsilon}) + (1/12)\cdot T^k$$

$$= -2(k-1)\ln(\frac{1}{\epsilon}) + \sum_{j=1}^{2(k-1)} x^{k-1}(T^{k-1})_j + (1/12)\cdot T^k$$

$$= \phi_p^{k-1}(\frac{1}{\epsilon}e_{2(k-1)}) - \phi_p^{k-1}(x^{k-1}(T^{k-1})) + (1/12)\cdot T^k.$$

Now, since $\phi_p^n(x^n(T^n)) \geq 2n\ln\rho + 2n\ln(2n)$, it follows that

$$2n\ln\frac{\epsilon}{\rho} = \phi_p^n(\frac{1}{\epsilon}e_{2n}) - 2n\ln\rho - 2n\ln(2n) \geq \phi_p^n(\frac{1}{\epsilon}e_{2n}) - \phi_p^n(x^n(T^n))$$

$$\geq \phi_p^{n-1}(\frac{1}{\epsilon}e_{2(n-1)}) - \phi_n^{n-1}(x^{n-1}(T^{n-1})) + (1/12)\cdot T^n \geq \ldots \geq$$

$$\geq \phi_p^m(\frac{1}{\epsilon}e_{2m}) - \phi_m^m(x^m(T^m)) + (1/12)\cdot(T^{m+1} + T^{m+2} + \ldots + T^n)$$

$$\geq \phi_p^m(x^m(0)) - \phi_m^m(x^m(T^m)) + (1/12)\cdot(T^{m+1} + T^{m+2} + \ldots + T^n)$$

$$\geq (1/12)\cdot(T^m + T^{m+1} + T^{m+2} + \ldots + T^n).$$

This completes the proof.

References

[1] Special issue on bounded-error estimation (1994). *Int. J. Adap. Contr. and Signal Proc.*, **8**, No.1.

[2] Special issue on bounded-error estimation (Part II) (1995). *Int. J. Adap. Contr. and Signal Proc.*, **9**, No.1.

[3] Special issue on parameter estimation with bounded error (1990). *Mathematics and Computers in Simulations*, **32**.

[4] Special issue on trends in system identification (1995). *Automatica*, **31**, 1689-1883.

[5] Special issue on system identification for robust control design (1992). *IEEE Transactions on Automatic Control*, **37**, No. 7.

[6] Bai, E.W., R. Tempo and H. Cho (1995). "Membership set estimator: size, optimal input, complexity and relations with least squares", *IEEE Trans. on CAS*, Part I, **42**, 266-277

[7] Bai, E.W., N. Nagpal and R. Tempo (1996). "Bounded error parameter estimation: noise models and recursive algorithms", *Automatic*, **32**, 985-999.

[8] Belforte, G., B. Bona and V. Cerone (1990). "Parameter estimation algorithms for a set-membership description of uncertainty," *Automatica*, **26**, 887-898.

[9] Ben-Israel, A. and T. N. E. Greville (1980). *Generalized Inverses: Theory and Applications*, Krieger Publishing Company.

[10] Dasgupta, S. and Y. Huang (1987). "Asymptotically convergent modified recursive least squares with data dependent updating and forgetting for systems with bounded noise", *IEEE Trans. on Info. Theory*, **33**, 383-392.

[11] Fogel, E. and Y. Huang (1982). "On the value of information in system identification-bounded noise case," *Automatica*, **18**, 229-238.

[12] Goffin, J. L., Z. Luo, and Y. Ye (1996). "Complexity analysis of an interior cutting plane method for convex feasibility problems,", *SIAM J. Optimization* **6**, 638–652.

[13] Ljung, L. (1987). *System Identification: Theory for the Users*, Prentice-Hall, Englewood Cliffs, NJ.

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

[15] Mizuno, S., M. Todd and Y. Ye (1993). "On adaptive-step primal-dual interior-point algorithms for linear programming", *Math. of Operations Res.*, **18**, 964-981.

[16] Norton, J.P. (1987). "Identification and application of bounded parameter models," *Automatica*, **23**, 497-597.

[17] Ninness, B.M. and G. C. Goodwin (1995). "Rapproachement between bounded-error and stochastic estimation theory", *Int. J. Adap. Contr. and Signal Proc.*, **9**, 107-132.

[18] Ortega J. M. and W.C. Rheinbolt (1970). "Iterative Solution of Nonlinear Equations in Several Variables," *Academic Press*, New York.

[19] 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.

[20] Soderstrom, S. and P. Stoica (1988). *System Identification*, Prentice-Hall, Englewood Cliffs, NJ.

[21] Sonnevend, G. (1986). "An analytic center for polyhedrons and new classes of global algorithms for linear programming", *Lecture Notes in Control and Information Science*, **84**, 866-876.

[22] Sonnevend, G. (1987), "A new method for solving a set of linear inequalities and its applications", *Dynamic Modeling and Control of National Economics*, Pergamon, Oxford-New York, 465-471.

[23] Traub, J.F., G. Wasikowski and H. Wozniakowski (1988). *Information-Based Complexity*, Academic, New York.

[24] Vicino and Zappa(1996). "Sequential approximation of feasible parameter sets for identification with set membership uncertainty", *IEEE Trans. on Auto. Contr.*, **41**, 774-785

[25] Walter, E. and H. Piet-Lahanier (1990). "Estimation of parameter bounds from bounded-error data: A survey", *Mathematics and Computers in Simulations*, **32**, 449-468.