Proceedings of the American Control Conference Baltimore, Maryland • June 1994

Least-Squares Parameter Set Estimation for Robust Control Design

Robert L. Kosut*
Integrated Systems, Inc.
3260 Jay St.
Santa Clara, CA 95054, U.S.A.

Brian D. O. Anderson†
Dept. Of Systems Engineering &
Cooperative Research Center for Robust and Adaptive Systems
The Australian National University
Canberra, ACT 2601, Australia

Abstract Two least-squares based methods are presented for obtaining ARX model sets. The first is obtained using properties of high-order ARX models and the second uses a stochastic embedding scheme on the residuals from an ARX model of any order. Either of the ARX model sets is useful for robust control of systems with uncertain parameters. Using the high order ARX model approach, the parameter uncertainty lies in a confidence ellipsoid. Using the stochastic embedding approach, the parameter uncertainty is ia confidence box. For scalar plants, both cases can be handled using convex programming to obtain the exact stability robustness margin for a particular controller. However, because the uncertainty description is probabilistic, the robustness property has to be associated with a confidence level, *i.e.*, a probability of stability.

1 Introduction

Suppose it is desired to control the single-input-single-output stable *discrete-time* system,

$$y = Gu + He, \quad e \in \mathbf{E}(\lambda) \tag{1}$$

where:

- G and H are known only to be causal linear-time-invariant (LTI) stable systems with unknown transfer functions $G(z)$ and $H(z)$, respectively.
- The sequences y and u are, respectively, the sensed output, and the applied control input. All that is available is the finite data record
$$\{y_t, u_t \mid t = 1, \ldots, N\} \tag{2}$$
where y_t and u_t are the values of the sequences y and u, respectively, at the sample time t.
- The sequence e is unpredictable, but is known to be a member of the set $\mathbf{E}(\lambda)$ where λ is unknown: likely candidiates for $\mathbf{E}(\lambda)$ are $\mathbf{POW}(\lambda)$, the set of sequences with power bound λ, or $\mathbf{IID}(\lambda)$, i.i.d. zero-mean sequences with variance λ.

*Supported by AFOSR, Directorate of Mathematical and Computer Sciences, Contract No. F49620-90-C-0064, NSF U.S./Australia Cooperative Research Program, Travel Grant INT-9014152, and the Australian National University as a Visiting Fellow.

†Supported by the Cooperative Rersearch Center Program of the Australian Government

The problem addressed here, referred to as *set estimation*, is to use the above information to determine a set description of the plant *useful for robust control design.* For example, consider the plant set[1]

$$\mathbf{M} : \left\{ y = (\widehat{G} + \Delta\widehat{W})u + \widehat{H}e \;\middle|\; \|\Delta\|_{\mathcal{H}_\infty} \le \widehat{\gamma},\ e \in \mathbf{E}(\widehat{\lambda}) \right\}$$

If $\mathbf{E}(\widehat{\lambda}) = \mathbf{POW}(\widehat{\lambda})$, then $\mathbf{M}$ is typical starting point for $\mathcal{H}_\infty$ control design. If $\mathbf{E}(\widehat{\lambda}) = \mathbf{IID}(\widehat{\lambda})$, then mixed $\mathcal{H}_2/\mathcal{H}_\infty$ control design methods apply. There are many combinations possible. However, in all the above cases, the quantities with "hats" are available *a priori* to the designer, which is not the case here. These quantities are to be estimated from the data and *a priori* information.

In the remainder of the paper, we show how least-squares estimation can be used to obtain a plant set estimate. We examine two cases: (i) high order model set estimation, and (ii) model set estimation via stochastic embedding. Related techniques can be found in [5] and the references therein.

2 Least-Squares Estimation

Parameter estimation via least-squares with an ARX[2] model structure is perhaps the most widely used approach to system identification. The standard (*e.g.*, [7]) parametric ARX model set is:

$$\mathbf{M}_{\mathbf{ARX}} : \left\{ A_\theta y = B_\theta u + e \;\middle|\; \theta \in \mathbb{R}^p,\ e \in \mathbf{IID}(\lambda) \right\} \tag{3}$$

where

$$\begin{aligned} A_\theta &= 1 + \sum_{i=1}^{n} a_i z^{-i}, \quad B_\theta = \sum_{i=1}^{m} b_i z^{-i} \\ \theta &= [a_1 \cdots a_n \; b_1 \cdots b_m]^T \in \mathbb{R}^p, \quad p = n + m \end{aligned}$$

Thus, the ARX model output at tome t is given by,

$$\begin{aligned} y_t &= \phi_t^T \theta + e_t \\ \phi_t^T &= [-y_{t-1} \cdots - y_{t-n} \; u_{t-1} \cdots u_{t-m}] \end{aligned}$$

[1] If Δ is stable, $\|\Delta\|_{\mathcal{H}_\infty} = \sup_\omega |\Delta(e^{j\omega})|$, otherwize, $\|\Delta\|_{\mathcal{H}_\infty} = \infty$.

[2] ARX is the standard acronymn for system models with an AutoRegressive term (Ay) and an eXternal input (Bu).

The least-squares parameter estimate, based on a finite data record, is found from:

$$\widehat{\theta} = \arg\min_{\theta} \frac{1}{N}\sum_{t=1}^{N}(y_t - \theta^T\phi_t)^2 \quad (4)$$

Limiting Properties It is well known [7] that

$$\widehat{\theta} \to \overline{\theta}, \text{ as } N \to \infty, \text{ w.p. } 1$$

where

$$\overline{\theta} = \arg\min_{\theta} \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{err}(\omega,\theta)d\omega$$

with the "error" spectrum given by,

$$\begin{aligned} S_{err}(\omega,\theta) &= |A_\theta(e^{j\omega})G(e^{j\omega}) - B_\theta(e^{j\omega})|^2 S_{uu}(\omega) \\ &+ \lambda|A_\theta(e^{j\omega})H(e^{j\omega})|^2 \end{aligned}$$

Much more can be said when the system (1) is in the ARX model set (3), *i.e.*, when there exists a parameter $\theta_0 \in \mathbb{R}^p$ such that, $A_{\theta_0}G - B_{\theta_0} = 0$ and $A_{\theta_0}H = 1$. In this case, $\overline{\theta} = \theta_0$ and the parameter error $\widehat{\theta} - \theta_0$ is asymptotically normally distributed *i.e.*, as $N \to \infty$,

$$\sqrt{N}(\widehat{\theta} - \theta_0) \to \mathcal{N}(0, \lambda\mathcal{E}(\phi_t\phi_t^T)^{-1}) \quad (5)$$

where $\mathcal{E}(\cdot)$ denotes expectation. In addition, consistent estimates of $\mathcal{E}(\phi_t\phi_t^T)$ and λ, are given by

$$\widehat{R} = \frac{1}{N}\sum_{t=1}^{N}\phi_t\phi_t^T \quad \widehat{\lambda} = \frac{1}{N-p}\sum_{t=1}^{N}\widehat{e}_t^2$$

with

$$\widehat{e}_t = y_t - \phi_t^T\widehat{\theta}$$

the estimated prediction error.

Large N Properties To simplify notation set

$$\widehat{A} = A_{\widehat{\theta}}, \quad \widehat{B} = B_{\widehat{\theta}}$$

Define $\delta \in \mathbb{R}^p$ as the normalized (unknown) parameter error,

$$\delta = \widehat{R}^{1/2}(\widehat{\theta} - \theta_0) \quad (6)$$

Then, the true system (1) can be expressed as:

$$\widehat{A}y = \widehat{B}u - \delta^T\widehat{R}^{-1/2}\phi + e \quad (7)$$

It follows that for sufficiently large N [7]:

$$\delta \in \mathcal{N}(0, \frac{\lambda}{N}) \text{ and } \frac{N}{p}\frac{\delta^T\delta}{\widehat{\lambda}} \in \mathcal{F}(p, N-p) \quad (8)$$

where $\mathcal{F}(p, N-p)$ is the F-distribution with degrees of freedom p and $N-p$. Hence,

$$\text{Prob}\{\delta^T\delta \le \frac{p}{N}\alpha(\eta)\widehat{\lambda}\} = \eta \quad (9)$$

can be determined from an F-distribution table. To be safe, suppose we set η very high, say, $\eta = .999$. Then for typical numbers such as $N \ge 1000$ and $n = m = 10$, we get $\alpha(.999) = 2.27$. For large n, say $n = m = 60$, and large $N \gg n$, we get $\alpha(.999) \approx 1.45$, and so on. In addition, for large N, $e \in \mathbf{E}(\widehat{\lambda})$ with high probability. Hence, for large n, m, and N, the system (1) is in the model set

$$\mathbf{M}_{\text{ARX}} : \begin{cases} \widehat{A}y = \widehat{B}u - \delta^T\widehat{R}^{-1/2}\phi + e \\ \delta^T\delta \le \frac{p}{N}\alpha(\eta)\widehat{\lambda} \\ e \in \mathbf{IID}(\widehat{\lambda}) \end{cases} \quad (10)$$

with probability of at least η. Observe that the model uncertainty in this set is represented as a *norm bound on the normalized parameter error* with a probability tag which can be selected close to one. Putting aside for the moment the issue of whether or not the true system is in the model set, it is therefore logical to pursue robust control with this type of parametric uncertainty.

3 Robust Control with ARX Set

In this section we discuss the issue of robust control design retaining still the assumption that the true system is in the ARX model set $\mathbf{M}_{\text{ARX}}$ of (10). Suppose we apply the LTI feedback controller

$$u = -\widehat{K}y \quad (11)$$

where $\widehat{K}$ stabilizes the "nominal" ARX system ($\delta = 0$),

$$\widehat{A}y = \widehat{B}u + e$$

Applying the control to the actual system model (7), gives the closed-loop system

$$\begin{bmatrix} y \\ u \end{bmatrix} = \begin{bmatrix} T_\delta \\ -Q_\delta \end{bmatrix} e = \frac{1}{1-\delta^T h}\begin{bmatrix} \widehat{T} \\ -\widehat{Q} \end{bmatrix} e$$

where

$$\begin{aligned} \widehat{T} &= \frac{1}{\widehat{A}+\widehat{B}\widehat{K}}, \quad \widehat{Q} = \frac{\widehat{K}}{\widehat{A}+\widehat{B}\widehat{K}} \\ \widehat{h} &= \widehat{R}^{-1/2}\begin{bmatrix} D\widehat{T} \\ D\widehat{Q} \end{bmatrix}, \quad D = \begin{bmatrix} z^{-1} \\ \vdots \\ z^{-n} \end{bmatrix} \end{aligned}$$

Because $\widehat{K}$ stabilizes the nominal system, $\widehat{T}$, $\widehat{Q}$ and $\widehat{h}$ are all stable.

Recall from the Nyquist theorem that since $\widehat{h}$ is stable, the closed-loop system is stable if and only if, for all ω,

$$|1 - \delta^T\widehat{h}(e^{j\omega})| \ne 0, \quad \forall\delta^T\delta \le r^2 < r_{stab}^2$$

where r_{stab}, the "real" stability margin is given by,

$$r_{stab} = \inf_{\omega}\inf_{r(\omega)}\left\{\delta^T\delta = r(\omega)^2 \;\middle|\; \delta^T\widehat{h}(e^{j\omega}) = 1\right\}$$

Calculating $r(\omega)$ involves finding the minimum norm (least-squares) solution to the over-determined set of equations $\delta^T\widehat{h}(e^{j\omega}) = 1$ at each frequency. As shown in [10], an easy calculation yields,

$$r(\omega)^2 = \begin{cases} 1/[\|a\|^2 - (a^Tb)^2/\|b\|^2], & b \ne 0 \\ 1/\|a\|^2, & b = 0 \end{cases} \quad (12)$$

where

$$a = \text{Re } h(e^{j\omega}), \; b = \text{Im } h(e^{j\omega})$$

Hence, a "probability of stability" can be stated as follows. Since

$$\text{Prob}\{\delta^T\delta \le \frac{p}{N}\alpha(\eta)\widehat{\lambda}\} = \eta$$

it follows that

$$\frac{p}{N}\alpha(\eta)\widehat{\lambda} < r_{stab}^2 \Rightarrow \text{Prob}\{(1 - \delta^T h)^{-1} \text{ stable}\} \ge \eta$$

In the more general case, calculating the stability margin, r_{stab}, for other than the two-norm of the parameter error, can be cast as a a convex programming problem, *e.g.*, [10]. However, no closed form or convex programming solution is known in the MIMO case.

4 High-Order ARX Model Set

The analysis in the previous sections depended on the true system being a member of the parametric ARX model set. Although this is never true, it is a fact (*e.g.*, [7]) that the system (1) is equivalent to the *infinite order* ARX model:

$$Ay = Bu + e \quad (13)$$

where

$$A = H^{-1} = 1 + \sum_{k=1}^{\infty} a_k z^{-k} \quad (14)$$

$$B = H^{-1}G = \sum_{k=1}^{\infty} b_k z^{-k} \quad (15)$$

Because H^{-1} and $H^{-1}G$ are stable, there exists positive constants M_a, M_b, and $\rho < 1$ such that

$$|a_k| \leq M_a \rho^k, \quad |b_k| \leq M_b \rho^k$$

It therefore follows that the n-th order ARX model,

$$A_n y = B_n u + e \quad (16)$$

with

$$A_n = 1 + \sum_{k=1}^{n} a_k z^{-k} \quad (17)$$

$$B_n = \sum_{k=1}^{n} b_k z^{-k} \quad (18)$$

can, for sufficiently large n,arbitrarily well approximate the infinite order ARX model, and hence, the original system. Specifically, it is easily shown that $\|A_nG - B_n\|_{\mathcal{H}_\infty}$ and $\|A_nH - 1\|_{\mathcal{H}_\infty}$ approach zero exponentially as n goes to infinity. As a result, it is very tempting to use the previous result with very high order models, particularly since it is very easy to compute such models. The following example illustrates this approach.

Example The following example is presented without figures – no room! Take the true $G(z)$ as the zero-order-hold equivalent of $\Omega^2/(s^2 + 2\zeta\Omega s + \Omega^2)$ with $\Omega = 2\pi(.1)$ rad/sec, $\zeta = .15$, and a sampling frequency of 2 hz. Take the noise model as $H = 1$ with $e \in \mathbf{IID}(.2^2)$ and with input $u \in \mathbf{IID}(1)$. Hence, the rms noise to signal ratio at the output is rms(e)/rms(Gu) = .28. The system was simulated with u and e drawn from the previous distributions. Least-squares identification was performed for time samples $t = 1 : 512$ for model orders $n = m = 1 : 40$. All the models were validated on a different data set for $t = 513 : 1024$. From the validation set, the minimum rms value occurs at $n = 14$. Because larger orders show an increase in rms, any models of order higher than 14 must be trying to fit the particular noise realization during identification. The fits in magnitude are very similar, but get more noisy at high frequencies. Since the true noise has a constant spectrum, we see that the higher orders are being used primarily to flatten the estimated spectrum. One would therefore expect a significant number of pole-zero cancelations in the estimated transfer function. This phenomena is dramtically visible in a pole zero plot of the estimated transfer function. Moreover, it is interesting to examine the Hankel singular values as well.

An $\mathcal{H}_2$ controller was designed using the 14th order model. (We know this controller stabilizes the true system, because in this case the true system is being simulated. Normally, of course, the true system is unknown to the designer.) Although the closed-loop system is stable and seems to behave well, the robust analysis suggests testing the worst-case plant possible from the ARX model set. By examing the paramaetric stability margins *vs.* frequency (12) for probability .95 and .995, respectively, it can be seen that at about .5 hz, the closed-loop system is just stable for a .95 probability and just unstable for .995. The interpretion is that there exists a plant in the ARX set which *could* have produced the data with the same statistics. Since this plant would cause considerable trouble for this controller, the controller should be redesigned. , *e.g.*, the loop gain should be reduced in the .5 hz range.

Caveat Emptor There are several impediments to using the high-order ARX estimation and robust control design approach just described. First, if $\mathcal{H}_2$ or $\mathcal{H}_\infty$ methods are useed for design, the controller (11) will also be of high order – although in most cases it is easy to reduce the controller because of the many stable near pole zero cancelations in the estimated model. Secondly, a determination of what is meant precisely by high order is dependent on *a priori* knowledge about the true system. Thirdly, the statistical properties are based on very large data lengths, and a precise value of "large" depends on the true system properties.

To offset the high order problem, an alternative is to use a more parsimonious model parametrization. For example, use of Laguerre or Kautz expansions, as proposed in [11], can result in considerably fewer parameters to obtain the same level of approximation as a model expanded in the backward shift operator z^{-1}. However, the efficacy of this approach depends on prior information regarding the accuracy of some dominant pole locations.

5 Residual Analysis and Stochastic Embedding

The stochastic embedding principle developed by Goodwin *et al.*[3] provides a completely stochastic framework for model error estimation. No assumptions on model order or data length are required. The basic idea is to view model error as a realization of a random variable with zero mean, whose variance is described by a few parameters which captures the structure of the model error. In other words, the complicated problem of relating the bias to the data and the mismatch in structure of the true system and the model is avoided. By modeling the error in this seemingly rudimentary form, an error model set parametrization is obtained which is described by a small number of parameters, yet is capable of representing a large set of error models. In [3] a Maximum Likelihood (ML) approach is used to define the estimates of these model error parameters. In this section, we use the stochastic embedding principle, but applied in a slightly different way, avoiding some of the approximations involved in solving for the ML estimate. Although no prior assumptions about high model order are required, the results presented here – unlike those in [3] – are asymptotic in data length.

To apply the stochastic embedding principle of [3], we proceed as follows: suppose we have obtained ARX estimates $\widehat{A}, \widehat{B}$ from some finite data record. Since the true system (1) is $y = Gu + He$, it follows that the prediction error is:

$$\widehat{e} = \widehat{A}y - \widehat{B}u = \Delta u + w \quad (19)$$

with

$$\Delta = \widehat{A}G - \widehat{B} \quad (20)$$

$$w = \widehat{A}He \quad (21)$$

As before, our goal is to make some kind of estimate covering Δ. Towards this end, form the (asymptotic) prediction error correlation coefficients:

$$c_k = \mathcal{E}(\widehat{e}_t\widehat{e}_{t+k}), \quad \forall k \geq 0 \quad (22)$$

If $u \in \mathbf{IID}(\lambda_u)$, then:

$$c_k = \begin{cases} \lambda_w + \lambda_u \sum_{i=1}^{\infty} \delta_i^2, & k = 0 \\ \lambda_u \sum_{i=1}^{\infty} \delta_i \delta_{i+k}, & k \geq 1 \end{cases} \quad (23)$$

where

$$\lambda_w = \mathcal{E}(w_t)^2 \quad (24)$$

and $\{\delta_1, \delta_2, \ldots\}$ are the impulse response coefficients of Δ, *i.e.*,

$$\Delta(z) = \sum_{i=1}^{\infty} \delta_i z^{-i} \tag{25}$$

Since Δ is stable, there exist positive constants M_δ and $\rho < 1$ such that

$$|\delta_i| \leq M_\delta \rho^{i-1} \tag{26}$$

Hence,

$$|c_k| \leq \begin{cases} \lambda_w + \lambda_u \dfrac{M_\delta^2}{1-\rho^2}, & k = 0 \\ \lambda_u \dfrac{\rho^k M_\delta^2}{1-\rho^2} & k \geq 1 \end{cases} \tag{27}$$

Observe that for $k \geq 1$, the correlation coefficients decay exponentially with k. Hence, information is contained in the correlation coefficients about the model error Δ in the form of bounds on the impulse reponse coefficients $\{\delta_1, \delta_2, \ldots\}$. The problem with (27) is that only bounds are provided and thus it is difficult to infer estimates of M_δ and ρ from c_k or estimates of c_k. The key is to utilize the stochastic embedding put forth in [3].

Stochastic Embedding Make the further assumption that the model error is a realization of a stochastic process such that:

$$\frac{\delta_i}{\rho^{i-1}} \in \text{IID}(\lambda_\delta) \tag{28}$$

As a result, there holds

$$\mathcal{E}_\delta c_k = \begin{cases} \lambda_w + \lambda_u \dfrac{M_\delta^2}{1-\rho^2}, & k = 0 \\ 0, & k \geq 1 \end{cases} \tag{29}$$

Since there is no information in the mean regarding the pulse reponse decay, it is necessary to compute the variance. Thus,

$$\mathcal{E}_\delta c_k^2 = \lambda_u^2 \frac{\lambda_\delta^2 \rho^{2k}}{1-\rho^4}, \quad \forall k \geq 1 \tag{30}$$

We restrict attention now to $k \geq 1$ becasue the information we seek appears only in those coefficients. Observe that the model error has been captured by a parametric stochastic description with just two free parameters λ_δ and ρ.

There are now several ways to fit the sample variance to this expression and find values for λ_δ and ρ which achieve the best fit. One approach to estimating these parameters is to fit a gaussian distribution to the correlation coefficients $c_k, k \geq 1$. That is, assume that $\{c_k \mid k = 1 : L\}$ is an independent gaussian sequence with zero mean and variance $\alpha\beta^k$ where $0 < \beta < 1$. Thus,

$$\beta = \rho^2, \quad \alpha = \lambda_u^2 \frac{\lambda_\delta^2}{1-\rho^4}, \tag{31}$$

and estimates for λ_δ and ρ can be obtained from estimates of α and β. That is, given α and β,

$$\rho = \sqrt{\beta}, \quad M_\delta = \sqrt{\lambda_\delta} = \left(\frac{\alpha(1-\beta^2)}{\lambda_u^2}\right)^{1/4} \tag{32}$$

Taking a max-likelihood approach, the negative log likelihood function is then:

$$\mathcal{L} = \frac{1}{2\alpha} \sum_{k=1}^{L} c_k^2 \beta^{-k} + \frac{L}{2} \log\alpha + \frac{L(L+1)}{4} \log\beta + \text{ constant} \tag{33}$$

The minimizing values $\widehat{\alpha}$ and $\widehat{\beta}$ satisfy:

$$\widehat{\alpha} = \frac{1}{L} \sum_{k=1}^{L} c_k^2 \widehat{\beta}^{-k} \tag{34}$$

$$0 = \sum_{k=1}^{L} \left(k - \frac{L+1}{2}\right) c_k^2 \widehat{\beta}^{-k} \tag{35}$$

Clearly a solution exists iff there is a real root $0 < \widehat{\beta} < 1$. By replacing c_k^2 with $\mathcal{E}_\delta(c_k^2)$ we get the polynomial in $\beta/\widehat{\beta}$:

$$0 = \sum_{k=1}^{L} \left(k - \frac{L+1}{2}\right) \left(\frac{\beta}{\widehat{\beta}}\right)^k$$

Clearly $\beta = \widehat{\beta}$ is a solution, and certainly the desired one in that $\widehat{\alpha} = \alpha$. In fact, it can be shown that, on average, the only real solution is $\beta = \widehat{\beta}$. Even though L can be quite large, *e.g.*, $L = 1024$ is typical, it is actually very easy to solve for $\widehat{\beta}$ by a simple bisection search over $\widehat{\beta}$. Although this method is similar to that developed in [3], here an exact solution to the max-likelihood problem is possible because the data is taken to be the asymptotic prediction error correlation coefficents, *i.e.*, *infinite* data. Clearly errors will be introduced in computing the coefficients using *finite length* data. This requires further analysis.

Having found the estimate $\widehat{\rho}, \widehat{\lambda}_\delta$, an estimate for λ_w follows from the expression for c_0, that is:

$$\widehat{\lambda}_w = c_0 - \lambda_u \frac{\widehat{\lambda}_\delta}{1-\widehat{\rho}^2} \tag{36}$$

Hence, the end result of the max-likelihood/stochastic embedding approach is the following ARX model set:

$$\mathbf{M}'_{\text{ARX}} : \begin{cases} \widehat{A}y = \widehat{B}u + \Delta u + w \\ \Delta = \sum_{k=1}^{\infty} q_k \widehat{\rho}^{k-1} z^{-k} \\ q_k \in \text{IID}(\widehat{\lambda}_\delta) \\ \mathcal{E}(w_t)^2 = \widehat{\lambda}_w \end{cases} \tag{37}$$

6 Robust Control Analysis

In this section we discuss the issue of robust control design under the assumption that the true system is in the ARX model set $\mathbf{M}'_{\text{ARX}}$ of (37). Following the discussion in section , the control $u = -\widehat{K}y$ which stabilizes the nominal ARX system $\widehat{A}y = \widehat{B}u + e$, will stabilize plants in $\mathbf{M}'_{\text{ARX}}$ iff the closed-loop transfer function

$$\frac{1}{1 - \sum_{k=1}^{\infty} \widehat{h}_k(z) q_k} \tag{38}$$

is stable where $\widehat{Q} = \widehat{K}/(1 + \widehat{G}\widehat{K})$ as before and now

$$\widehat{h}_k(z) = \widehat{Q}(z) \widehat{\rho}^{k-1} z^{-k} \tag{39}$$

Because $q_k \in \text{IID}(\widehat{\lambda}_\delta)$, stability is meant again as a "probability of stability." In this case, the uncertain parameters q_k are bounded in magnitude, rather than in an ellipsoid as before. Specifically, since

$$\text{Prob}\{\sup_k q_k^2 \leq \alpha(\eta)\widehat{\lambda}\} = \eta$$

it follows that

$$\alpha(\eta)\widehat{\lambda} < r_{\text{stab}}^2 \Rightarrow \text{Prob}\{(1 - \sum_{k=1}^{\infty} \widehat{h}_k(z) q_k)^{-1} \text{ stable}\} \geq \eta$$

In this case the stability margin is defined as:

$$r_{\text{stab}} = \inf_{\omega} \inf_{r(\omega)} \left\{ q_k^2 = r(\omega)^2 \;\middle|\; \sum_{k=1}^{\infty} \widehat{h}_k(e^{j\omega}) q_k = 1 \right\} \tag{40}$$

Although this is an infinite dimensional convex programming problem, nonetheless a solution is readily obtained *e.g.*, [10].

7 Concluding Discussion

Two least-squares based methods have been presented for obtaining ARX model sets. The first is obtained using properties of high-order ARX models and the second uses a stochastic embedding scheme on the residuals from an ARX model of any order. Various asymptotic approximations are used in obtaining either set, and their effect is under study. Either of the ARX model sets is useful for robust control of systems with uncertain paramaters. However, in this case the uncertainty is random and so a notion of probability of stability is introduced. In other unpublished work of the authors, methods are being developed for analyzing the more practical situation where the data set is finite.

In this paper only some of the many properties of least-squares parameter estimation with ARX models have been examined. There are many properties of least-squares estimation that may yet be profitable for obtaining set estimators useful for robust control design. For example:

1. The least-squares solution is unique and provides the global minimum in closed-form, which in the ideal case allows for a complete asymptotic statistical analysis.
2. By utilizing "square-root" numerical algorithms, *i.e.*, SVD or QR algorithms, the solution can be efficiently and rapidly computed even for very high orders. In particular, with the QR method:
   (a) all models are simultaneously computed up to a specified maximum order and stored in a square-root matrix whose size is determined by the maximum model order, not data length.
   (b) prediction error variance vs. model order (up to the maximum) is available as the last "column" of QR.
   (c) results of different experiments are easily joined, *i.e.*, there is no need to re-compute over both data sets.
   (d) high-order models are easily computed; very rapid computation is possible with lattice algorithms.
3. The impulse response coefficients, up to the order of the numerator polynomial, are asymptotically unbiased.
4. The estimates have orthogonality properties which are not dependent on assumptions about the character of the uncertain disturbance, *e.g.*, gausssian, worst-case, and so on.

Propery 3 is quite appealing. Specifically, if the input u is white, then the first m impulse response coefficients of the true input/output transfer function G are asymptotically unbiased. Recall that m is the order of B_θ, the numerator polynomial in the ARX model set. It is interesting that the result does not depend on n, the order of A_θ, the ARX denominator polynomial. Moreover, the true system need not be in the ARX model set!

This result was originaly proven in [9]. Other useful results following from this fact can be found in [1]. To summarize the result, let

$$\overline{\theta} = \arg\min_{\theta \in \mathbb{R}^p} \mathcal{E}(y_t - \theta^T \phi_t)^2 \qquad (41)$$

$$\overline{G} = \frac{\overline{B}}{\overline{A}} = \sum_{k=1}^{\infty} \overline{g}_k z^{-k} \qquad (42)$$

Thus $\{\overline{g}_1, \overline{g}_2, \ldots\}$ is the pulse response sequence associated with the asymptotic transfer function estimate of G in (1) which has the pulse response sequence $\{g_1, g_2, \ldots\}$. That is,

$$G = \sum_{k=1}^{\infty} g_k z^{-k}$$

The result from [9] is as follows:

$$u \in \mathrm{IID}(\lambda_u) \Longrightarrow \overline{g}_k = g_k, \ \ \forall k = 1:m \qquad (43)$$

This result together with the use of an appropriate affine model expansion (*e.g.*, Laguerre or Kautz) may prove most beneficial. An affine model set, *e.g.*, a Laguerre expansion for G, can offset the issue of determining what is meant by a large data length. Moreover, with this model, it is possible to precisely compute statistical properties for any given model order or data length – no asymptotic assumptions are required, *e.g.*, [6]. The stochastic embedding scheme developed in [3] uses such an approach, and if applied as proposed here, it may be possible to eliminate all the asymptotic requirements, *i.e.*, large data length or high model order.

References

[1] H. Aling and R.L. Kosut, "Unbiased least-squares estimates with structure incompatibilities," *Proc. 1993 ECC*, Groningen, The Netherlands, 29 June-2July, 1993.

[2] B. D. O. Anderson *et al.*, *Stability of Adaptive Systems: Passivity and Averaging Analysis*, MIT Press, Cambridge, MA, 1986.

[3] G.C. Goodwin, M. Gevers, and B. Ninness, "Quantifying the error in estimated transfer functions with applications to model order selection," *IEEE Trans. on Automatic Control*, vol. 37, no.7, pp. 913-928, July 1992.

[4] Ulf Grenander and Gabor Szego, *Toeplitz Forms and Their Applications*, Chelsea Pub. Co. 1st Edition 1958, 2nd Edition 1984.

[5] "Special Issue on System Identification for Robust Control Design," edited by R.L. Kosut, G.C. Goodwin, and M.P. Polis, *IEEE Trans. on Automatic Control*, vol. 37, no.7, pp. 899-1008, July 1992.

[6] R.L. Kosut and B.D.O. Anderson, "Statistical analysis of least-squares identification for robust control design: output error case with affine parametrization," *Proc. 1993 ACC*, San Francisco, CA, June 1993.

[7] L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, Inc., 1987.

[8] M. Massoumnia and R.L. Kosut, "A family of norms for system identification problems," *Proc. 1993 ACC*, San Francisco, CA, June 1993.

[9] C.T. Mullis and R.A. Roberts, "The use of second-order information in the approximation of discrete-time linear systems, *IEEE Trans. Acoustics, Speech, and Sig. Proc.*, vol. ASSP-24, No.3, pp.226-238, June 1976.

[10] A. Rantzer, "Convex robustness specifications for real parametric uncertainty in linear systems," *Proc. 1992 ACC*, pp. 583-587, June 1992.

[11] B. Wahlberg and L.Ljung, "Hard frequency-domain model error bounds from least-squares like identification techniques", *IEEE Trans. Aut. Contr.*, vol. 37, no. 7, July, 1992.