Proceedings of the American Control Conference
Arlington, VA June 25-27, 2001

Quadratically Constrained Least Squares Identification

Tobin H. Van Pelt
Hunter System Technologies, LLC
PO Box 975 • Sausalito, CA 94966
vanpelt@hst.cncdsl.com

Dennis S. Bernstein[1]
Dept. of Aerospace Engineering
Univ. of Michigan • Ann Arbor, MI 48109
dsbaero@engin.umich.edu

Abstract

In this paper we investigate the consistency of parameter estimates obtained from least squares identification with a quadratic parameter constraint. For generality, we consider infinite impulse response systems with arbitrarily colored output noise. In the case of finite data, we show that there always exists a generally indefinite quadratic constraint that yields the true parameters of the system when a persistency condition is satisfied. When the autocorrelation matrix of the output noise is known to within a scalar multiple, we show that the QCLS estimator is consistent. Furthermore, we develop a heuristic iterative method for applying QCLS identification when the noise statistics are unknown. Finally, we give an example comparing this method to standard least squares identification, as well as an instrumental variable technique.

1 Introduction

A critical issue associated with least squares identification is the effect of noise on the parameter estimates [1,2]. A desirable property of any system identification procedure is that the estimator be consistent, that is, the parameter estimates converge with probability one to the true parameters as the number of data points increases. In the case of standard least squares it is well known that the estimator is not generally consistent. As a result, there exist many variants of the least squares method that attempt to remedy this lack of consistency. One example is generalized least squares which is consistent, but the cost is nonquadratic in the parameters and requires numerical optimization procedures. Another example is the instrumental variable method which is generically consistent. For finite data the accuracy of this method may be poor [3], and, consequently, the choice of good instruments depends on the system and the noise in question [1,2,4].

For least squares identification, the transfer function parameterization has an inherent ambiguity that results from simultaneous scaling of the numerator and denominator polynomials. It is traditional to normalize the leading denominator coefficient of the transfer function to be unity ($a_0 = 1$), and most treatments of least squares identification proceed with the tacit assumption that this normalization entails no loss of generality with respect to the consistency of parameter estimates.

An alternative normalization, considered in [5–7], has the form of a quadratic constraint on the transfer function coefficients. In [5,6] the authors consider the problem of under-modeling in a deterministic, noise-free setting, where the true system to be identified may not belong to the model set. In more recent work involving quadratic constraints in deterministic, noise-free identification [7], the authors consider a positive-definite Euclidean parameter constraint and show that this leads to statistically biased pole estimates. This case, which invokes constraint (20) in [6], corresponds to $N = \gamma I_{2n+2}$ in the present paper.

The least squares problem with a quadratic constraint was considered in [8]. This problem has the form

$$\min_x \|Ax - b\|_2 \quad \text{subject to} \quad \|Cx - d\|_2 = \alpha, \tag{1}$$

where $A \in \mathbb{R}^{m\times q}$, $x \in \mathbb{R}^q$, $C \in \mathbb{R}^{p\times q}$, $b \in \mathbb{R}^m$, $d \in \mathbb{R}^p$, and $\alpha > 0$. In the present paper we consider the quadratically constrained least squares (QCLS) problem

$$\min_\theta \theta^{\mathrm{T}} M \theta \quad \text{subject to} \quad \theta^{\mathrm{T}} N \theta = \gamma, \tag{2}$$

where $M \in \mathbb{R}^{(2n+2)\times(2n+2)}$ is nonnegative definite, $N \in \mathbb{R}^{(2n+2)\times(2n+2)}$ is symmetric, $\theta \in \mathbb{R}^{2n+2}$ and $\gamma > 0$. It can be seen that (1) with $b = 0$ and $d = 0$ has the same form as (2) with $M = A^{\mathrm{T}}A$ and $N = C^{\mathrm{T}}C$. However, for the case of system identification, N may be indefinite, which is not allowed in the framework of [8].

For the case of finite data, we show that, when a persistency condition is satisfied, there always exists a generally indefinite quadratic constraint matrix N such that (2) yields the true parameters of the system as a solution. However, since the appropriate constraint matrix N depends on the noise realization, this solution cannot be implemented in practice. Nevertheless, when the output error noise statistics are known to within a scalar multiple and a persistency condition is satisfied, we construct a definite quadratic constraint matrix N, such that the QCLS estimate is consistent. Note that the assumption that the output error noise statistics are known is weaker than the widely invoked assumption that the equation error noise statistics are known. For

[1]This research was supported in part by the Air Force Office of Scientific Research under grants F49620-98-1-0037 and F49620-97-1-0406.

practical implementation when the output error noise statistics are unknown, we develop a heuristic iterative method which involves an indefinite constraint matrix. This constraint matrix approximates the constraint matrix that gives the true parameter values. Proofs of the results in this paper appear in [9].

2 Null Space Condition

Consider the single-input single-output system

$$y_0(k) = G(\mathbf{q}^{-1};\vartheta)u(k), \tag{3}$$

where $u(k)$ is the *system input*, $y_0(k)$ is the *system output*, and $G(\mathbf{q}^{-1};\vartheta)$ is the nth-order proper transfer function

$$G(\mathbf{q}^{-1};\vartheta) \triangleq \frac{\beta_0+\beta_1\mathbf{q}^{-1}+\cdots+\beta_n\mathbf{q}^{-n}}{\alpha_0+\alpha_1\mathbf{q}^{-1}+\cdots+\alpha_n\mathbf{q}^{-n}}, \tag{4}$$

where $\mathbf{q}^{-1}$ is the backward-shift operator and the *system parameter vector* $\vartheta \in \mathbb{R}^{2n+2}$ is defined by

$$\vartheta \triangleq [\ \alpha_0\ \cdots\ \alpha_n\ \ \beta_0\ \cdots\ \beta_n\ ]^{\mathrm{T}}. \tag{5}$$

We assume that $\alpha_0 \neq 0$, which is equivalent to the assumption that $G(\mathbf{q}^{-1};\vartheta)$ is causal. Furthermore, we define $\operatorname{cone}(\mathcal{T}) \triangleq \{\alpha x : \alpha > 0, x \in \mathcal{T}\}$ and note the following.

Remark 2.1 $G(\mathbf{q}^{-1};\vartheta) = G(\mathbf{q}^{-1};\eta\vartheta)$ *for all nonzero* $\eta \in \mathbb{R}$*, or, equivalently,* $G(\mathbf{q}^{-1};\vartheta) = G(\mathbf{q}^{-1};\theta)$ *for all* $\theta \in \operatorname{cone}(\vartheta)\cup\operatorname{cone}(-\vartheta)$*. This nonuniqueness of the system parameterization can be removed by choosing* $\theta \in \operatorname{cone}(\vartheta)\cup\operatorname{cone}(-\vartheta)$ *such that the first component of* θ *is unity, that is,* $G(\mathbf{q}^{-1};\vartheta) = G(\mathbf{q}^{-1};\theta)$*, where* $\theta = (1/\alpha_0)\vartheta$*. Although this normalization yields a unique system parameterization, we choose not to do this in order to facilitate the following analysis.*

We assume that the system output $y_0(k)$ is corrupted by $w(k)$ so that the *measured output* $y(k)$ is given by

$$y(k) = y_0(k) + w(k) = G(\mathbf{q}^{-1};\vartheta)u(k) + w(k), \tag{6}$$

which corresponds to the output error case in [1]. Throughout this paper, we assume that the noise sequence $w(k)$ satisfies

$$w(k) = H(\mathbf{q}^{-1})v(k), \tag{7}$$

where $H(\mathbf{q}^{-1})$ is a stable transfer function and $v(k)$ is a sequence of independent random variables with zero mean, variance σ_v^2, and bounded fourth moments. Moreover, for later use, we give the following definition.

Definition 2.1 *The input sequence* $\{u(k)\}_{k=0}^{\infty}$ *is* quasi-stationary *if it is bounded and*

$$\lim_{l\to\infty}\frac{1}{l}\sum_{k=1}^{l}u(k)u(k-\tau) \tag{8}$$

exists and is bounded for all $\tau = 0,1,2,\ldots$.

Next we write (6) in regression form as

$$\sum_{i=0}^{n}\alpha_i y(k-i) - \sum_{i=0}^{n}\beta_i u(k-i) = \sum_{i=0}^{n}\alpha_i w(k-i) \tag{9}$$

which can further be written as

$$\phi^{\mathrm{T}}(k)\vartheta = \psi^{\mathrm{T}}(k)\vartheta, \tag{10}$$

where the *regression vector* $\phi(k) \in \mathbb{R}^{2n+2}$ is defined as

$$\phi(k) \triangleq \left[y(k)\ \cdots\ y(k-n)\ \ -u(k)\ \cdots\ -u(k-n)\right]^{\mathrm{T}} \tag{11}$$

and the *noise vector* $\psi(k) \in \mathbb{R}^{2n+2}$ is defined as

$$\psi(k) \triangleq [\ w(k)\ \cdots\ w(k-n)\ \ 0\ \cdots\ 0\ ]^{\mathrm{T}}. \tag{12}$$

Henceforth, we consider a finite measured output sequence $\{y(k)\}_{k=0}^{l}$ generated by (6) with system input sequence $\{u(k)\}_{k=0}^{l}$ and noise sequence $\{w(k)\}_{k=0}^{l}$. Assuming $l \geq n$, we define the *regression matrix* $\Phi \in \mathbb{R}^{(l-n+1)\times(2n+2)}$ by

$$\Phi \triangleq \begin{bmatrix}\phi^{\mathrm{T}}(n)\\ \vdots\\ \phi^{\mathrm{T}}(l)\end{bmatrix} \tag{13}$$

and the *noise matrix* $\Psi \in \mathbb{R}^{(l-n+1)\times(2n+2)}$ by

$$\Psi \triangleq \begin{bmatrix}\psi^{\mathrm{T}}(n)\\ \vdots\\ \psi^{\mathrm{T}}(l)\end{bmatrix}. \tag{14}$$

It then follows from (10) that

$$\Phi\vartheta = \Psi\vartheta. \tag{15}$$

Next, we define the *noise-free regression matrix* $\Phi_0 \in \mathbb{R}^{(l-n+1)\times(2n+2)}$ by

$$\Phi_0 \triangleq \begin{bmatrix}\phi_0^{\mathrm{T}}(n)\\ \vdots\\ \phi_0^{\mathrm{T}}(l)\end{bmatrix}, \tag{16}$$

where the *noise-free regression vector* $\phi_0(k) \in \mathbb{R}^{2n+2}$ is defined as

$$\phi_0(k) \triangleq \left[y_0(k)\ \cdots\ y_0(k-n)\ \ -u(k)\ \cdots\ -u(k-n)\right]^{\mathrm{T}}. \tag{17}$$

Noting $\Phi - \Psi = \Phi_0$, (15) can be written as $\Phi_0\vartheta = 0$, which yields the *null space condition*

$$\vartheta \in \mathcal{N}(\Phi_0). \tag{18}$$

Finally, define the nonnegative definite matrix $M_0 \in \mathbb{R}^{(2n+2)\times(2n+2)}$ by $M_0 \triangleq \Phi_0^{\mathrm{T}}\Phi_0$, and note that $\operatorname{rank} M_0 = \operatorname{rank}\Phi_0$ and $\mathcal{N}(M_0) = \mathcal{N}(\Phi_0)$. Therefore, the null space condition (18) can equivalently be written as $\vartheta \in \mathcal{N}(M_0)$.

We now give the following definitions.

Definition 2.2 *The input sequence* $\{u(k)\}_{k=0}^{l}$ *is* persistently exciting *for* $G(\mathbf{q}^{-1};\vartheta)$ *if* rank $\Phi_0 = 2n+1$.

Assume $\{u(k)\}_{k=0}^{l}$ is persistently exciting for $G(\mathbf{q}^{-1};\vartheta)$. It then follows that the dimension of $\mathcal{N}(\Phi_0)$ equals one, and in accordance with Remark 2.1 $G(\mathbf{q}^{-1};\vartheta)$ is uniquely determined by the null space condition (18).

Definition 2.3 *The input sequence* $\{u(k)\}_{k=0}^{l}$ *and noise sequence* $\{w(k)\}_{k=0}^{l}$ *are* jointly persistently exciting *for* $G(\mathbf{q}^{-1};\vartheta)$ *if* rank $\Phi = 2n+2$.

The assumption that $\{u(k)\}_{k=0}^{l}$ is persistently exciting for $G(\mathbf{q}^{-1};\vartheta)$ and $\{u(k)\}_{k=0}^{l}$ and $\{w(k)\}_{k=0}^{l}$ are jointly persistently exciting for $G(\mathbf{q}^{-1};\vartheta)$ are independent, that is, one does not in general imply the other.

3 Standard Least Squares

In this section we review the standard least squares problem in order to provide a framework for developing QCLS identification. First, define the *output vector* $Y \in \mathbb{R}^{l-n+1}$ by

$$Y \triangleq \begin{bmatrix} y(n) & \cdots & y(l) \end{bmatrix}^{\mathrm{T}} \tag{19}$$

and the *standard least squares regression matrix* $\Phi_{\mathrm{LS}} \in \mathbb{R}^{(l-n+1)\times(2n+1)}$ by

$$\Phi_{\mathrm{LS}} \triangleq \begin{bmatrix} \phi_{\mathrm{LS}}^{\mathrm{T}}(n) \\ \vdots \\ \phi_{\mathrm{LS}}^{\mathrm{T}}(l) \end{bmatrix}, \tag{20}$$

where the *standard least squares regression vector* $\phi_{\mathrm{LS}}(k) \in \mathbb{R}^{2n+1}$ is defined by

$$\phi_{\mathrm{LS}}(k) \triangleq \begin{bmatrix} -y(k-1) & \cdots & -y(k-n) & u(k) & \cdots & u(k-n) \end{bmatrix}^{\mathrm{T}}. \tag{21}$$

Partitioning Φ we note that

$$\Phi = \begin{bmatrix} \Phi_1 & \Phi_2 \end{bmatrix} = \begin{bmatrix} Y & -\Phi_{\mathrm{LS}} \end{bmatrix}. \tag{22}$$

Now consider the model $G(\mathbf{q}^{-1};\theta)$ of $G(\mathbf{q}^{-1};\vartheta)$, where the *model parameter vector* $\theta \in \mathbb{R}^{2n+2}$ is defined by

$$\theta \triangleq \begin{bmatrix} a_0 & \cdots & a_n & b_0 & \cdots & b_n \end{bmatrix}^{\mathrm{T}}, \tag{23}$$

and where Remark 2.1 is also valid for θ. Note that the definition of θ assumes that the system order n is known. We partition θ as $\theta = \begin{bmatrix} a_0 & \tilde{\theta}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$, where $\tilde{\theta} \in \mathbb{R}^{2n+1}$.

Next, we follow the standard approach in which a_0 is fixed (typically set to unity), and we define the *standard least squares cost*

$$J(\tilde{\theta}) \triangleq \|a_0 Y - \Phi_{\mathrm{LS}}\tilde{\theta}\|_2^2. \tag{24}$$

Noting (22), it follows that (24) can be written as

$$J(\tilde{\theta}) = \|a_0\Phi_1 + \Phi_2\tilde{\theta}\|_2^2. \tag{25}$$

The standard least squares problem is then given by

$$\min_{\tilde{\theta}\in\mathbb{R}^{2n+1}} J(\tilde{\theta}). \tag{26}$$

Solutions to (26) are given by the following result.

Proposition 3.1 *Let* $\{u(k)\}_{k=0}^{l}$ *and* $\{y(k)\}_{k=0}^{l}$ *satisfy (6), and assume* $\{u(k)\}_{k=0}^{l}$ *and* $\{w(k)\}_{k=0}^{l}$ *are jointly persistently exciting for* $G(\mathbf{q}^{-1};\vartheta)$. *Then*

$$\hat{\tilde{\theta}}_l = -a_0\left(\Phi_2^{\mathrm{T}}\Phi_2\right)^{-1}\Phi_2^{\mathrm{T}}\Phi_1 \tag{27}$$

is the unique solution of (26).

The resulting model using (27) is then given by $G(\mathbf{q}^{-1};\hat{\theta}_l)$, where $\hat{\theta}_l$ is defined by $\hat{\theta}_l \triangleq \begin{bmatrix} a_0 & \hat{\tilde{\theta}}_l^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}$.

Next we define the *parameter estimation error* by

$$\Delta\theta_l \triangleq (a_0/\alpha_0)\,\vartheta - \hat{\theta}_l, \tag{28}$$

and the *standard least squares bias* by

$$\Delta\tilde{\theta}_l \triangleq (a_0/\alpha_0)\,\tilde{\vartheta} - \hat{\tilde{\theta}}_l. \tag{29}$$

Furthermore, assuming $G(\mathbf{q}^{-1};\vartheta)$ is stable and $\{u(k)\}_{k=0}^{\infty}$ is quasi-stationary, we define the *asymptotic standard least squares bias* by

$$\Delta\tilde{\theta}_\infty = \lim_{l\to\infty}\Delta\tilde{\theta}_l. \tag{30}$$

If $\Delta\tilde{\theta}_\infty = 0$ with probability one, then $\hat{\tilde{\theta}}_l$ is *consistent*. In general, the standard least squares estimator (27) is not consistent.

4 Quadratically Constrained Least Squares

In this section we develop an alternative approach to determining ϑ. Instead of fixing a_0, we minimize the least squares cost (24) with a quadratic constraint on θ. Hence, we consider the *generalized least squares cost*

$$\mathcal{J}(\theta) \triangleq \|\Phi\theta\|_2^2 = \|a_0 Y - \Phi_{\mathrm{LS}}\tilde{\theta}\|_2^2. \tag{31}$$

By defining $M \in \mathbb{R}^{(2n+2)\times(2n+2)}$ as $M \triangleq \Phi^{\mathrm{T}}\Phi$, it then follows that (31) can be rewritten as

$$\mathcal{J}(\theta) = \theta^{\mathrm{T}} M \theta. \tag{32}$$

Next, let $\gamma > 0$, let $N \in \mathbb{R}^{(2n+2)\times(2n+2)}$ be symmetric, and define the *parameter constraint set* $\mathcal{D}_\gamma(N)$ by

$$\mathcal{D}_\gamma(N) \triangleq \{\theta \in \mathbb{R}^{2n+2} : \theta^{\mathrm{T}} N \theta = \gamma\}. \tag{33}$$

The *quadratically constrained least squares* (QCLS) *problem* is then given by

$$\min_{\theta \in \mathcal{D}_\gamma(N)} \mathcal{J}(\theta). \tag{34}$$

The existence of solutions to (34) is treated in [9].

Next, define $N_{\rm LS} \in \mathbb{R}^{(2n+2)\times(2n+2)}$ by

$$N_{\rm LS} \triangleq \begin{bmatrix} \gamma & 0 & \cdots & 0 \\ 0 & 0 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix}. \tag{35}$$

Setting $N = N_{\rm LS}$ it can be seen that the QCLS problem (34) is equivalent to the standard least squares problem (26). Specifically, note that the parameter constraint set $\mathcal{D}_\gamma(N_{\rm LS})$ is equivalent to fixing $a_0 = \pm 1$. Other choices of N lead to alternative parameter constraints and yield different solutions to (34).

Next we define $\Delta M \in \mathbb{R}^{(2n+2)\times(2n+2)}$ by $\Delta M \triangleq M - M_0$ and consider the QCLS problem with $N = \Delta M$. In the noise-free case $\Delta M = 0$, and thus $\mathcal{D}_\gamma(N) = \varnothing$ for $N = \Delta M$. In the noisy case, however, the following result gives conditions under which the QCLS problem correctly identifies $G(\mathbf{q}^{-1};\vartheta)$.

Theorem 4.1 *Let* $\{u(k)\}_{k=0}^{l}$ *be persistently exciting for* $G(\mathbf{q}^{-1};\vartheta)$*, assume* $\mathcal{D}_\gamma(\Delta M) \neq \varnothing$*, and assume* $\mathcal{N}(M) \cap \mathcal{N}(\Delta M) = \{0\}$*. Then* $G(\mathbf{q}^{-1};\vartheta) = G(\mathbf{q}^{-1};\theta)$ *for all* θ *that solves (34).*

Note that if $\{u(k)\}_{k=0}^{l}$ and $\{w(k)\}_{k=0}^{l}$ are jointly persistently exciting for $G(\mathbf{q}^{-1};\vartheta)$, it follows that $\mathcal{N}(M) \cap \mathcal{N}(\Delta M) = \{0\}$. Thus, Theorem 4.1 holds in this case and $G(\mathbf{q}^{-1};\vartheta) = G(\mathbf{q}^{-1};\theta)$ for all θ that solve (34).

To graphically illustrate the QCLS problem, consider $\{u(k)\}_{k=0}^{l}$ and $\{y(k)\}_{k=0}^{l}$ generated from the zero-order (static) system $G(\mathbf{q}^{-1};\vartheta) = \beta_0/\alpha_0$, where $\vartheta = [\ \alpha_0 \ \ \beta_0\ ]^{\rm T}$, so that $y(k) = \beta_0/\alpha_0 u(k) + w(k)$. Hence, the model parameter vector $\theta \in \mathbb{R}^2$ is given by $\theta = [\ a_0 \ \ b_0\ ]^{\rm T}$. Furthermore, we assume that $\{u(k)\}_{k=0}^{l}$ and $\{w(k)\}_{k=0}^{l}$ are jointly persistently exciting for $G(\mathbf{q}^{-1};\vartheta)$, so that $M > 0$. Next, consider the solution $\hat{\theta}_{\rm LS}$ to the QCLS problem using $N = N_{\rm LS}$ (standard least squares). Figure 1 shows the θ-plane with level sets of $\mathcal{J}(\theta)$, cone(ϑ), the parameter constraint set $\mathcal{D}_\gamma(N_{\rm LS})$, and cone$(\hat{\theta}_{\rm LS})$, all corresponding to a specific input-output data set. The angle between cone(ϑ) and cone$(\hat{\theta}_{\rm LS})$, represented by $\angle\theta_{\rm LS}$, illustrates the bias in the estimate of ϑ. Next, by choosing $N = \Delta M = M - M_0$ it turns out that the constraint set $\mathcal{D}_\gamma(\Delta M)$ is given by the hyperbolas shown in Figure 1, which are tangent to the level set $\mathcal{J}(\theta) = \gamma$. In this case, the solution to the QCLS problem is an element of cone(ϑ). In other words, the set $\mathcal{D}_\gamma(\Delta M)$ quadratically constrains θ such

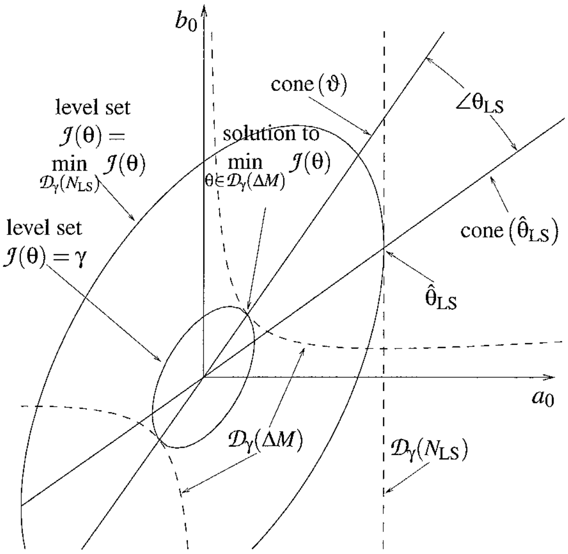

Figure 1: θ-plane of the QCLS problem.

that the minimizer of the least squares cost $\mathcal{J}(\theta)$ is a scalar multiple of ϑ.

However, this approach is not possible in practice since ΔM is not known. Thus, we consider the solution to the QCLS problem (34) with arbitrary constraint matrix N.

Theorem 4.2 *Assume that the QCLS problem (34) has a solution, and consider* $\{u(k)\}_{k=0}^{l}$ *and* $\{y(k)\}_{k=0}^{l}$ *satisfying (6). Furthermore, consider the generalized eigenvalue problem*

$$(M - \lambda N)\theta = 0. \tag{36}$$

Then there exists a smallest positive generalized eigenvalue $\lambda_{\min}^{+}$*. Furthermore, let* $\theta \in \mathbb{R}^{2n+2}$ *be an element of the generalized eigenspace corresponding to* $\lambda_{\min}^{+}$*. Then the QCLS estimator*

$$\hat{\theta}_l = \pm\sqrt{\frac{\gamma}{\theta^{\rm T} N \theta}}\,\theta \tag{37}$$

solves (34), and $\mathcal{J}(\hat{\theta}_l) = \gamma\lambda_{\min}^{+}$.

Corollary 4.1 *Assume* $\mathcal{D}_\gamma(N) \neq \varnothing$*, consider* $\{u(k)\}_{k=0}^{l}$ *and* $\{y(k)\}_{k=0}^{l}$ *satisfying (6), and let* $\{u(k)\}_{k=0}^{l}$ *and* $\{w(k)\}_{k=0}^{l}$ *be jointly persistently exciting for* $G(\mathbf{q}^{-1};\vartheta)$*. Furthermore, let* $\lambda_{\max}$ *be the largest (positive) eigenvalue of* $M^{-1/2}NM^{-1/2}$*, where* $q \in \mathbb{R}^{2n+2}$ *is an associated eigenvector. Then*

$$\hat{\theta}_l = \pm\sqrt{\frac{\gamma}{q^{\rm T}M^{-1}q}}\,M^{-1/2}q \tag{38}$$

solves (34), and $\mathcal{J}(\hat{\theta}_l) = \gamma/\lambda_{\max}$.

We note that it was shown in [9] that $G(\mathbf{q}^{-1};\hat{\theta}_l)$, where $\hat{\theta}_l$ is given by (37), is independent of γ. Moreover, the solution of the QCLS problem is unaffected by positive scaling of M. Therefore, for numerical accuracy, it may be desirable to scale M such that $||M|| \approx 1$ while choosing $||N|| \approx 1$.

5 Consistency of QCLS

In this section we establish conditions for the consistency of the QCLS estimator (37).

Noting Theorem 4.1, it is desirable to choose N close to ΔM. Thus, we consider $N = Q$, where $Q \in \mathbb{R}^{(2n+2)\times(2n+2)}$ is defined by $Q \triangleq \mathbb{E}[\Delta M]$. Furthermore, assuming that $\{u(k)\}_{k=0}^{l}$ and $\{w(k)\}_{k=0}^{l}$ are uncorrelated,

$$Q = \mathbb{E}[\Psi^{\mathrm{T}}\Psi] = \begin{bmatrix} R_{ww} & 0_{n+1} \\ 0_{n+1} & 0_{n+1} \end{bmatrix}, \qquad (39)$$

where $R_{ww} \in \mathbb{R}^{(n+1)\times(n+1)}$ is defined by

$$R_{ww} \triangleq \mathbb{E}\begin{bmatrix} w^2(k) & w(k)w(k-1) & \cdots & w(k)w(k-n) \\ w(k)w(k-1) & w^2(k) & \cdots & w(k)w(k-n+1) \\ \vdots & \vdots & \ddots & \vdots \\ w(k)w(k-n) & w(k)w(k-n+1) & \cdots & w^2(k) \end{bmatrix}. \qquad (40)$$

Next, we examine the error in the QCLS estimator $\hat{\theta}_l$ as the number of measurements l gets large. The QCLS estimator $\hat{\theta}_l$ of ϑ is *consistent* if, with probability one,

$$\hat{\theta}_l \longrightarrow \mathrm{cone}(\vartheta) \cup \mathrm{cone}(-\vartheta) \quad \text{as} \quad l \longrightarrow \infty. \qquad (41)$$

We then have the following consistency result.

Theorem 5.1 *Let $G(\mathbf{q}^{-1};\vartheta)$ be stable, and consider $\{u(k)\}_{k=0}^{\infty}$ and $\{y(k)\}_{k=0}^{\infty}$ satisfying (6), where $\{u(k)\}_{k=0}^{\infty}$ is quasi-stationary. Furthermore, for all $l \geq 3n$, assume $\{u(k)\}_{k=0}^{l}$ is persistently exciting for $G(\mathbf{q}^{-1};\vartheta)$, assume $\{u(k)\}_{k=0}^{l}$ and $\{w(k)\}_{k=0}^{l}$ are uncorrelated and jointly persistently exciting for $G(\mathbf{q}^{-1};\vartheta)$, let $\eta > 0$, and let $\hat{\theta}_l$ be the QCLS estimator (37) with $N = l\eta Q$. Then $\hat{\theta}_l$ is consistent.*

We note that Theorem 5.1 holds for arbitrary $\eta > 0$. This implies that Q or, equivalently, R_{ww}, need only be known to within a scalar multiple.

6 Iterative QCLS Identification and Numerical Examples

In this section we develop a heuristic iterative QCLS identification algorithm and illustrate its use with a numerical example. For the case in which the noise statistics are unknown, the matrix ΔM can be iteratively estimated and used within QCLS identification in the following manner.

First, assume that a model estimate $G(\mathbf{q}^{-1};\hat{\theta})$ exists, and define the *estimated model output* $\hat{y}_0(k)$ by $\hat{y}_0(k) \triangleq G(\mathbf{q}^{-1};\hat{\theta})u(k)$. Next, consider the sequence $\{\hat{y}_0(k)\}_{k=0}^{l}$ generated using $\{u(k)\}_{k=0}^{l}$, let $\delta > 0$, and define $\widehat{\Phi}_0 \in \mathbb{R}^{(l-n-\mu+1)\times(2n+2)}$ by

$$\widehat{\Phi}_0 \triangleq \begin{bmatrix} \hat{\phi}^{\mathrm{T}}(n+\delta) \\ \vdots \\ \hat{\phi}^{\mathrm{T}}(l) \end{bmatrix}, \qquad (42)$$

where

$$\hat{\phi}(k) \triangleq \begin{bmatrix} \hat{y}_0(k) & \cdots & \hat{y}_0(k-n) & -u(k) & \cdots & -u(k-n) \end{bmatrix}^{\mathrm{T}}. \qquad (43)$$

Furthermore, define $\widehat{M}_0 \in \mathbb{R}^{(2n+2)\times(2n+2)}$ by $\widehat{M}_0 \triangleq \widehat{\Phi}_0^{\mathrm{T}}\widehat{\Phi}_0$ and $\widehat{\Delta M} \in \mathbb{R}^{(2n+2)\times(2n+2)}$ by $\widehat{\Delta M} \triangleq M - \widehat{M}_0$, which is generally indefinite.

Now, we let $N^{(i)}$ denote the ith constraint matrix, and let $\hat{\theta}_l^{(i)}$ denote the solution to the QCLS problem with $N = N^{(i)}$. We then give the following iterative procedure for performing QCLS identification when the statistics of the noise $w(k)$ are unknown.

Iterative QCLS Algorithm

1. Set $i = 0$ and $N^{(0)} = N_{\mathrm{LS}}$.
2. Obtain the QCLS estimate $\hat{\theta}_l^{(i)}$ with $N = N^{(i)}$.
3. Compute $\widehat{\Delta M}$ using $\hat{\theta} = \hat{\theta}_l^{(i)}$.
4. Set $N^{(i+1)} = \widehat{\Delta M}$.
5. Increment i and go to 2.

Numerical results suggest that δ should be chosen larger than the time constant associated with any transients of the model $G(\mathbf{q}^{-1};\hat{\theta})$.

Next, consider the output sequence $\{y(k)\}_{k=0}^{l}$, where $l = 500$, of the 2nd-order system

$$y(k) = \frac{1}{1+.2\mathbf{q}^{-1}+.8\mathbf{q}^{-2}}u(k) + \frac{1}{1-.9\mathbf{q}^{-1}}v(k), \qquad (44)$$

where the input sequence $\{u(k)\}_{k=0}^{l}$ is given by

$$u(k) = \sin .2k + \frac{1}{2}\sin(.8k + \frac{\pi}{4}), \qquad (45)$$

and the variance of the noise $v(k)$ is given by $\sigma_v^2 = 8.39 \times 10^{-4}$ with a resulting signal-to-noise ratio SNR = 15 (std.). Simulations were repeated with 100 different realizations of the white noise sequence $\{v(k)\}_{k=0}^{l}$. Next, the QCLS identification algorithm (using $\delta = 25$ with 5 iterations) was compared to standard least squares identification ($N = N_{\mathrm{LS}}$) and the instrumental variable (IV) method.

For the instrumental variable method, the instruments were chosen as in [1, 2]. First, the model resulting from the standard least squares method was used with $\{u(k)\}_{k=0}^{l}$ to generate $\{\hat{y}_0(k)\}_{k=0}^{l}$. Next, the instrument was chosen as

$$\Xi = \begin{bmatrix} \hat{\phi}^{\mathrm{T}}(n) \\ \vdots \\ \hat{\phi}^{\mathrm{T}}(l) \end{bmatrix}, \tag{46}$$

where

$$\hat{\phi}(k) \triangleq \left[\hat{y}_0(k-1) \ \cdots \ \hat{y}_0(k-n) \ -u(k) \ \cdots \ -u(k-n)\right]^{\mathrm{T}}. \tag{47}$$

The instrumental variable estimate was then computed using

$$\hat{\theta}_{\mathrm{IV}_l} = -\left(\Xi^{\mathrm{T}}\Phi_2\right)^{-1}\Xi^{\mathrm{T}}\Phi_1. \tag{48}$$

Additionally, for comparison purposes, QCLS identification was performed assuming that the noise statistics were known. For this example, $R_{ww} = \sigma_v^2 \tilde{R}_{ww}$, where

$$\tilde{R}_{ww} = \begin{bmatrix} 1 & .9 & .81 \\ .9 & 1 & .9 \\ .81 & .9 & 1 \end{bmatrix}. \tag{49}$$

Thus, Q was chosen as

$$Q = \begin{bmatrix} \tilde{R}_{ww} & 0_{n+1} \\ 0_{n+1} & 0_{n+1} \end{bmatrix}. \tag{50}$$

In these examples $a_0 = \alpha_0 = 1$ when computing the parameter estimate error, so that $\Delta\theta_l = \hat{\theta}_l - \vartheta = \hat{\bar{\theta}}_l - \tilde{\vartheta}$ for each method.

Figure 2 shows the evolution of the parameter estimation error for a single run of each identification method considered. The iterative QCLS algorithm and QCLS with known statistics both exhibit less parameter error than standard least squares. Note that iterative QCLS does slightly better than QCLS with known noise statistics since the former method exploits information concerning the actual noise realization. Furthermore, the parameter estimate error of iterative QCLS is similar to that of QCLS with known statistics.

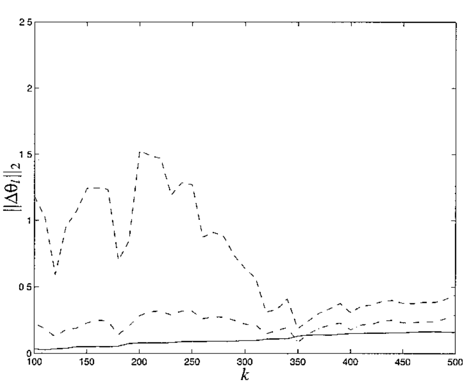

Figure 2: Parameter estimation error. (· · · · LS ; ; - - - IV ; —— iterative QCLS ; -·-·- QCLS with known statistics)

Table 1 shows the average $\Delta\theta_l$ (plus or minus one standard deviation) over the 100 runs considered, as well as the number of runs resulting in unstable models. The results show that the instrumental variable method yields highly uncertain estimates, and 37% of the identified models were unstable for this example. Moreover, the standard least squares method produced models with large bias. In contrast, the iterative QCLS algorithm produced models that had less bias than both the standard least squares and instrumental variable methods. Additionally, 2% of the models were unstable in the iterative QCLS case.

Method	$\|\Delta\theta_l\|_2$	Unstable Models (%)
LS	$2.35 \pm .0036$	0
IV	1.09 ± 2.19	37
iterative QCLS	$.281 \pm .020$	2
QCLS	$.150 \pm .0068$	0

Table 1: Average parameter estimation error.

Acknowledgement

We wish to thank Spilios Fassois for lengthy discussions and numerous helpful suggestions.

References

[1] L. Ljung. *System Identification: Theory for the User*. Prentice Hall, 2nd edition, 1999.

[2] T. Söderström and P. Stoica. *System Identification*. Prentice Hall, 1989.

[3] T. Söderström and P. Stoica. On the generic consistency of instrumental variable estimates. In *Proc. 9th IFAC World Congress*, Budapest, 1984.

[4] T. Söderström and P. Stoica. *Instrumental Variable Methods for System Identification*. Springer-Verlag, New York, 1983.

[5] M. Gevers. Estimation of Transfer Functions: Underbiased or Overbiased? In *Proc. IEEE Conf. Decision and Control*, pages 3200–3201, 1990.

[6] B. D. Moor, M. Gevers, and G. C. Goodwin. L_2-overbiased, L_2-underbiased and L_2-unbiased Estimation of Transfer Functions. *Automatica*, 30(5):893–898, 1994.

[7] P. Lemmerling and B. D. Moor. Misfit Versus Latency. In *Proc. IFAC System Identification*, Santa Barbara, CA, 2000.

[8] W. Gander. Least squares with a quadratic constraint. *Numer. Math.*, 36(291-307), 1981.

[9] T. H. Van Pelt and D. S. Bernstein. Parameter consistency and quadratically constrained least squares identification. *IEEE Trans. Autom. Cont.*, submitted November 2000.