2008 American Control Conference
Westin Seattle Hotel, Seattle, Washington, USA
June 11-13, 2008

Parameter reduction of nonlinear least-squares estimates via nonconvex optimization

Ryozo Nagamune and Jongeun Choi

Abstract— **This paper proposes a technique for reducing the number of uncertain parameters in order to simplify robust and adaptive controller design. The system is assumed to have a known structure with parametric uncertainties that represent plant dynamics variation. An original set of parameters is identified by nonlinear least-squares (NLS) optimization using noisy frequency response functions. Using the property of asymptotic normality for NLS estimates, the original parameter set is re-parameterized by an affine function of the smaller number of uncorrelated parameters. The correlation among uncertain parameters is detected by optimization with a bilinear matrix inequality. A numerical example illustrates the usefulness of the proposed technique.**

I. INTRODUCTION

Plant dynamics variation abounds in practical control problems. Such variation is caused by, e.g., the change of operational points and conditions, time-varying properties, and limited manufacturing tolerance for cheap and massive production. For instance, in the mass-spring-damper system, spring and/or damper coefficients may vary depending on the position of the mass due to nonlinearity. Also, in batch fabrication, it is costly to try to produce millions of products with exactly same dynamics. Taking into consideration plant dynamics variation is crucial to achieve satisfactory control systems for any conceivable situation.

In order to deal with plant dynamics variation, robust and adaptive control techniques [16], [17] are known to be powerful tools. These techniques are based on models representing dynamics variation, and various modeling and system identification methodologies for such models have been developed [12], [2] by the last decades.

In modeling, we always have to consider the trade-off between accuracy and simplicity of the model. Although a complex model can capture system properties in detail, it is often not preferable for controller design purpose due to unduly high computational cost. Especially, if we employ too many parameters to represent dynamics variation, numerical controller design based on modern robust control techniques often falls into computational infeasibility. Therefore, model set simplification is an important step.

For a model set involving parametric variation, there are mainly two ways of model simplification, i.e., model order reduction and parameter number reduction, and the latter is the main topic in this paper. Based on the idea of the principal component analysis [8], Conway et al. recently developed a parameter reduction method using the singular value decomposition [5]. A possible drawback is that they do not consider the effect of noise in experimental data on parameter reduction, while this paper discusses in detail how the noise on frequency response function data affects the parameter reduction stage.

R. Nagamune is with Department of Mechanical Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada nagamune@mech.ubc.ca

J. Choi is with Department of Mechanical Engineering, Michigan State University, East Lansing, MI 48824-1226, USA jchoi@mgr.msu.edu

This paper proposes a parameter reduction technique to simplify robust and adaptive controller design. The system is assumed to have a known structure with parametric uncertainties caused by plant dynamics variation. An original set of parameters is identified by nonlinear least-squares (NLS) optimization using noisy frequency response functions. Using the property of asymptotic normality for NLS estimates, the original parameter set is re-parameterized by an affine function of the smaller number of uncorrelated parameters. The correlation among uncertain parameters is detected by optimization with a bilinear matrix inequality.

Notation used in this paper is standard. The set of positive numbers and positive integers are denoted by $\mathbb{R}_+$ and $\mathbb{Z}_+$, respectively. The set of p dimensional real vector is $\mathbb{R}^p$, and the set of $p \times q$ complex matrices is $\mathbb{C}^{p\times q}$. (If $p = q = 1$, these indices are omitted.) For a complex matrix M, $\mathrm{Re}(M)$ and $\mathrm{Im}(M)$ respectively mean the real and the imaginary part of M, and M^T and M^* are respectively the transpose and the complex conjugate transpose of M. Other notation will be explained in due course.

II. PARAMETER ESTIMATION BY NONLINEAR LEAST-SQUARES OPTIMIZATION

As is written in [12, page 13], the model construction requires three basic entities, that is, the model structure, the data, and the optimality criterion. In the following, we will explain what these entities are in this paper. Throughout this paper, we assume that the system to be modeled is a scalar system, but extensions of the results in this paper to multivariable cases are straightforward.

A. Model structure

It is assumed that we have *a priori* information on the structure of a continuous-time linear time-invariant (LTI) true system:

$$[G(\theta)]\,(s), \quad \theta \in \Theta \subset \mathbb{R}^p, \qquad (1)$$

where θ is a parameter vector and Θ is a set determined by *a priori* knowledge of parameters. (For example, we may know that some parameters in θ must be positive.) The structure of G may come from either physical laws or experimental data.

978-1-4244-2079-7/08/$25.00 ©2008 AACC.

Simple examples are standard first and second order transfer functions:

$$\begin{aligned} [G(\theta)]\,(s) &:= \frac{K}{Ts+1}, & \theta &:= [K, T]^T, \\ [G(\theta)]\,(s) &:= \frac{K\omega^2}{s^2+2\zeta\omega s+\omega^2}, & \theta &:= [K, \zeta, \omega]^T. \end{aligned} \tag{2}$$

In what follows, we suppose that the true system is represented as

$$[G(\theta^\star)]\,(s), \tag{3}$$

with the true parameter vector $\theta^\star \in \Theta$.

B. Frequency domain experimental data

For the true system (3), we take noisy frequency response function (FRF) data as

$$\hat{G}_m = [G(\theta^\star)]\,(j\omega_m) + e_m,\ m = 1, \ldots, M. \tag{4}$$

where $\omega_m \in \mathbb{R}_+$ is the frequency of the sinusoidal input signal, $\hat{G}_m \in \mathbb{C}$ contains both gain and phase information, and $M \in \mathbb{Z}_+$ is the number of frequencies. The term e_m is a complex-valued white noise random variable resulting the following property:

$$\boldsymbol{e} := \begin{bmatrix} \begin{bmatrix} \mathrm{Re}\,\{e_1\} \\ \mathrm{Im}\,\{e_1\} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \mathrm{Re}\,\{e_M\} \\ \mathrm{Im}\,\{e_M\} \end{bmatrix} \end{bmatrix} \sim \mathcal{N}(0, \sigma^2 I_{2M}), \tag{5}$$

meaning that $\boldsymbol{e}$ is generated by a normal distribution with zero mean and covariance $\sigma^2 I_{2M}$.

The origin of the complex-valued white noise e_m is from the asymptotic normal distribution of the Fourier transform of white noise (see more details in [1], [11]). Some of it can be viewed as the quantization and electronic noise of the data acquisition system. Such noise level σ can be suppressed effectively by averaging sinusoidal output signals over many periods. This is the major advantage for identifying LTI systems based on FRFs.

C. Nonlinear least-squares optimization

For the given model structure (1) and FRF data

$$\left\{(\omega_m, \hat{G}_m); m = 1, \ldots, M\right\},$$

we consider to find the least-squares estimate $\hat{\theta}^M$ that minimizes the residual sum of squares:

$$\hat{\theta}^M := \arg\min_{\theta\in\Theta} \sum_{m=1}^{M} \left|\hat{G}_m - [G(\theta)]\,(j\omega_m)\right|^2. \tag{6}$$

The minimization problem (6) is in general a nonlinear least-squares (NLS) optimization problem with a constraint $\theta \in \Theta$, for which it is nontrivial to guarantee the existence and the uniqueness of the global solution. From now on, we *assume* the existence and the uniqueness of the global minimizer (the NLS estimate of θ) of the NLS problem. Numerically, we need to provide the optimization procedure with a reasonably good guess of $\hat{\theta}^M$, and have to be content with the suboptimal, instead of optimal, minimizer.

III. ASYMPTOTIC PROPERTIES OF NONLINEAR LEAST-SQUARES ESTIMATES

Next, we will discuss two important properties of the NLS estimate $\hat{\theta}^M$, i.e., *strong consistency* and *asymptotic normality* [6].

A. Strong consistency

Our first concern is the consistency. Roughly speaking, the consistency relates to a fundamental question: "Can we recover the true parameter $\theta^\star$ by minimizing the residual in (6) for a large number of samples?" The precise definition is given next.

Definition 1: An estimate $\hat{\theta}^M$ of $\theta^\star$ is *strongly consistent* if $\hat{\theta}^M$ converges to $\theta^\star$ almost surely (i.e., with probability one) as M (the number of data) goes to infinity.

Theorem 2 (Theorem 6 in [9]): Let D_M be a distance between two parameter vectors defined by

$$D_M(\theta, \theta') := \sum_{m=1}^{M} \left|[G(\theta)]\,(j\omega_m) - [G(\theta')]\,(j\omega_m)\right|^2.$$

If the following conditions hold, then the NLS estimate $\hat{\theta}^M$ of $\theta^\star$ is strongly consistent.

C1: *$D_M(\theta, \theta')/M$ converges uniformly to a continuous function $D(\theta, \theta')$, and*

C2: *$D(\theta, \theta^\star) = 0$ if and only if $\theta = \theta^\star$.*

Proof: See [9].

In this paper, we consider the case when an NLS estimate is strongly consistent.

B. Asymptotic normality

If an NLS estimate is strongly consistent, our next concern is to identify the distribution of the NLS estimate. It turns out that, under some conditions, the NLS estimate has asymptotically normal distribution. This property will become important later in parameter reduction. To present our result on asymptotic normality, we will introduce the following concept.

Definition 3: A model set $\mathcal{G}$ is said to be *uniformly stable for a set* Θ if all the transfer functions in the set

$$\mathcal{G}(\Theta) := \{[G(\theta)]\,(s) : \theta \in \Theta\} \tag{7}$$

are stable.

In the next theorem, we use the notation

$$[\nabla G(\theta^\star)](s) := \left[\frac{\partial}{\partial\theta} G(\theta)\right]_{\theta=\theta^\star}(s), \tag{8}$$

to denote the gradient vector evaluated at $\theta^\star$.

Theorem 4: Assume the following.

- $\hat{\theta}^M$ is a strongly consistent least-squares estimate of $\theta^\star$.
- For a given compact parameter set Θ, the model set $\mathcal{G}(\Theta)$ is uniformly stable.
- $G(\theta)$ is smooth in Θ.
- The true parameter $\theta^\star$ is in the interior of Θ in (7).
- Frequency points $\{\omega_m\,;\, m = 1, \ldots, M\}$ are distributed uniformly over a frequency range $[\underline{\omega}, \overline{\omega}]$ such that

$$\lim_{M\to\infty} \Sigma_M(\theta^\star) = \Sigma(\theta^\star), \tag{9}$$

where $\Sigma(\theta^\star)$ is a positive definite matrix, and Σ_M is defined by

$$\Sigma_M(\theta^\star) := \frac{\sum_{m=1}^{M} \mathrm{Re}\left\{[\nabla G(\theta^\star)](j\omega_m)[\nabla G(\theta^\star)](j\omega_m)^*\right\}}{M}. \tag{10}$$

Then, we have

$$\hat{\theta}^M \to_d \mathcal{N}(\theta^\star, W(\theta^\star)), \text{ as } M \to \infty, \tag{11}$$

where $\to_d$ denotes "converges in distribution" and

$$W(\theta^\star) := \frac{\sigma^2 \Sigma^{-1}(\theta^\star)}{M}. \tag{12}$$

(In words, $\hat{\theta}^M$ is asymptotically normal with mean $\theta^\star$ and covariance matrix $W(\theta^\star)$.)

Proof: See Appendix A.

The error covariance matrix $W(\theta^\star)$ in (12) will play an important role in the parameter reduction step.

Remark 5: The Fisher information matrix $\mathbb{I}(\theta^\star)$ [10] of the model (4) can be easily computed by

$$\mathbb{I}(\theta^\star) = \frac{M\Sigma_M(\theta^\star)}{\sigma^2}.$$

By the Cramér-Rao theorem [10], [7], the covariance matrix of any unbiased estimator $\hat{\theta}$ is lower bounded by the Cramér-Rao Lower Bound (CRLB), or the inverse of the Fisher information matrix $\mathbb{I}(\theta^\star)$:

$$\mathbb{E}\left\{(\hat{\theta}-\theta^\star)(\hat{\theta}-\theta^\star)^T\right\} \succeq \mathbb{I}(\theta^\star)^{-1} = \frac{\sigma^2\Sigma_M^{-1}(\theta^\star)}{M}. \tag{13}$$

Notice that this CRLB approaches to $W(\theta^\star)$ as M increases.

Remark 6: The choice of $[\underline{\omega}, \overline{\omega}]$ can significantly affect the error covariance matrix W. We want to select $[\underline{\omega}, \overline{\omega}]$ to minimize the "size" of the covariance matrix W. The optimization is usually considered in terms of the determinant or the trace of W or $\mathbb{I}$ (see more details in [7]). In general, $[\underline{\omega}, \overline{\omega}]$ should contain all the modes of the dynamical system.

IV. PARAMETER REDUCTION

So far, we have derived the asymptotic error covariance matrix $W(\theta^\star)$ of the nonlinear least-squares estimate $\hat{\theta}$[1] for a single true system $G(\theta^\star)$. In this section, by considering multiple true systems $G(\theta_\ell^\star)$, $\ell = 1, 2, \ldots$, with the same model structure, and a corresponding set of NLS estimates and error covariances, we will re-parameterize the set with a fewer number of uncorrelated parameters. This step is called *parameter reduction*. Such multiple true systems represents the dynamics variation caused by manufacturing tolerance, change of operating points, and/or time varying nature of the plant. A time varying correlation on parameters can be represented by a collection of time invariant correlations with given short time intervals.

[1]Hereafter, we omit the superscript M of $\hat{\theta}^M$ for notational simplicity.

For the ℓ-th dynamical system, we denote the true parameter by $\theta_\ell^\star$, and its NLS estimate based on FRF data by $\hat{\theta}_\ell$. Then, the estimation error is

$$\epsilon_\ell := \hat{\theta}_\ell - \theta_\ell^\star, \quad \ell = 1, 2, \ldots. \tag{14}$$

Under the conditions in Theorem 4, ϵ_ℓ is asymptotically normally distributed as M goes to infinity:

$$\epsilon_{\ell k} \to_d \mathcal{N}(0, W_\ell), \; W_\ell := W(\theta_\ell^\star) = \frac{\sigma^2\Sigma^{-1}(\theta_\ell^\star)}{M}. \tag{15}$$

Given a finite number of NLS estimates

$$\{\hat{\theta}_\ell \in \mathbb{R}^p;\, \ell = 1, \ldots, L\}, \tag{16}$$

where p is the number of parameters, and the ℓ-th asymptotic error covariances

$$\{W_\ell; \ell = 1, \ldots, L\}, \tag{17}$$

the parameter reduction problem is to find a parameter set

$$\left\{\theta := \bar{\theta} + V\lambda; \lambda \in \mathbb{R}^q, \; \|\lambda\|_\infty \le 1\right\} \tag{18}$$

with $q < p$, or equivalently $\bar{\theta} \in \mathbb{R}^p$ and $V \in \mathbb{R}^{p\times q}$, so that the set approximates all the given estimates in (16) in some sense. Next, we will provide a parameter reduction method based on a bilinear matrix inequality (BMI).

Geometrically, the parameter set (18) is a q-dimensional hyperrectangle in $\mathbb{R}^p$ ($q < p$), and the NLS estimates are points in $\mathbb{R}^p$. In order to find a hyperrectangle that passes close to all the points, we take the following two steps:

1) Find a q-dimensional *hyperplane* that passes close to all the NLS estimates.
2) Find a hyperrectangle in the obtained hyperplane so that the size is minimized while maintaining closeness to all the NLS estimates.

The minimization of the hyperrectangle size is important for less conservative robust controller design.

The problem in step 1) can be written mathematically as

$$\begin{aligned} &\min_{\bar{\theta}\in\mathbb{R}^p, V\in\mathbb{R}^{p\times q}, \lambda_\ell\in\mathbb{R}^q, \ell=1,\ldots,L} \gamma \\ &\text{subj. to } \max_{1\le\ell\le L} \left\| W_\ell^{-1/2}\left(\hat{\theta}_\ell - (\bar{\theta} + V\lambda_\ell)\right)\right\|^2 < \gamma. \end{aligned} \tag{19}$$

Here, to measure the "distance" between an NLS estimate and the hyperplane, we take into account the error covariance matrix. In terms of matrix inequalities, we can express the inequality constraint in (19) as

$$\begin{bmatrix} \gamma I_{LK} & (\hat{\Theta} - (\bar{\Theta} + (I_{LK}\otimes V)\Lambda))^T \\ \star & \boldsymbol{W} \end{bmatrix} > 0, \tag{20}$$

where $\otimes$ denotes the Kronecker product, and

$$\begin{aligned} \hat{\Theta} &:= \mathrm{diag}\left[\hat{\theta}_{11}, \cdots, \hat{\theta}_{1K}, \cdots, \hat{\theta}_{L1}, \cdots, \hat{\theta}_{LK}\right] \\ &\quad \in \mathbb{R}^{pLK\times LK}, \\ \bar{\Theta} &:= I_{LK}\otimes\bar{\theta} \in \mathbb{R}^{pLK\times LK}, \\ \Lambda &:= \mathrm{diag}\left[I_K\otimes\lambda_1, \cdots, I_K\otimes\lambda_L\right] \in \mathbb{R}^{qLK\times LK}, \\ \boldsymbol{W} &:= \mathrm{diag}\left[I_K\otimes W_1, \cdots, I_K\otimes W_L\right] \in \mathbb{R}^{pLK\times pLK}. \end{aligned}$$

This is a BMI with unknowns γ, $\bar{\Theta}$, V and Λ. To find a suboptimal solution via LMIs, we utilize a standard technique of alternating two LMI optimization problems:

- Fix $\bar{\Theta}$ and V, and solve LMI with respect to γ and Λ.
- Fix Λ, and solve LMI with respect to γ, $\bar{\Theta}$ and V.

After finding a q-dimensional hyperplane, in step 2), to minimize the "size" of the parameter set (18) for robust control purpose, as well as to satisfy the constraint $\|\lambda\|_\infty \le 1$, we should adjust the nominal parameter $\bar{\theta}$ and the matrix V. The problem is to find a hyperrectangle

$$\mathcal{H} := \left\{ \lambda := \bar{\lambda} + T\tilde{\lambda},\ \left\|\tilde{\lambda}\right\|_\infty \le 1 \right\} \tag{21}$$

with shortest sides (without rotation, i.e., T is a diagonal matrix) that contains the suboptimal solutions $\{\lambda_\ell\}_{\ell=1}^{L}$. This problem has an explicit solution:

$$\bar{\lambda}(i) := \frac{\min\limits_{\ell=1,\ldots,L} \lambda_\ell(i) + \max\limits_{\ell=1,\ldots,L} \lambda_\ell(i)}{2}, \quad i = 1, \ldots, q,$$

$$T := \operatorname{diag}\left[\max_\ell \left|\lambda_\ell(1) - \bar{\lambda}(1)\right|, \cdots, \max_\ell \left|\lambda_\ell(q) - \bar{\lambda}(q)\right| \right].$$

The following inclusion relation holds:

$$\begin{aligned} &\left\{\theta := \bar{\theta} + V\lambda_\ell,\ \ell = 1, \ldots, L\right\}, \\ &\subset \left\{\theta := \bar{\theta} + V\lambda,\ \lambda \in \mathcal{H}\right\}, \\ &= \left\{\theta := \bar{\theta}_{\text{new}} + V_{\text{new}}\tilde{\lambda},\ \left\|\tilde{\lambda}\right\|_\infty \le 1\right\}, \end{aligned} \tag{22}$$

where $\bar{\theta}_{\text{new}} = \bar{\theta} + V\bar{\lambda}$ and $V_{\text{new}} = VT$. In this way, we have obtained the set (18) that approximates all the NLS estimates.

Example 7: We illustrate the proposed parameter reduction method with an example, taken from the book [3, Ch. 11]. Consider the following set of true system dynamics:

$$\mathcal{S} := \left\{ G(s) = \prod_{m=1}^{5} \left[G_m(\delta)\right](s) : \delta \in [-0.2, 0.2] \right\}, \tag{23}$$

$$\begin{aligned} \left[G_1(\delta)\right](s) &= \frac{0.64013}{s^2}, \\ \left[G_2(\delta)\right](s) &= \frac{0.912s^2 + 0.4574s + 1.433(1+\delta)}{s^2 + 0.3592s + 1.433(1+\delta)}, \\ \left[G_3(\delta)\right](s) &= \frac{0.7586s^2 + 0.9622s + 2.491(1+\delta)}{s^2 + 0.7891s + 2.491(1+\delta)}, \\ \left[G_4(\delta)\right](s) &= \frac{9.917(1+\delta)}{s^2 + 0.1575s + 9.917(1+\delta)}, \\ \left[G_5(\delta)\right](s) &= \frac{2.731(1+\delta)}{s^2 + 0.2613s + 2.731(1+\delta)}. \end{aligned}$$

(Frequency is scaled by 10^{-4}; see [3, eq.(11.4)].) For each of five δ-values, $\delta = 0, 0.05, 0.1, \pm 0.2$, we took noisy FRF data with noise variance $\sigma^2 = 0.01$.

By regarding eight parameters as components of uncertain θ, the NLS estimates were obtained as

$$\left\{ \hat{\theta}_\ell \in \mathbb{R}^8 : \ell = 1, \ldots, 5 \right\}. \tag{24}$$

We also computed the approximated asymptotic error covariances:

$$\left\{ W_\ell : \ell = 1, \ldots, 5 \right\}. \tag{25}$$

Here, we selected $q = 1$ and performed parameter reduction. In Figure 1, it is shown the noisy FRF data (blue lines), and Bode plots of transfer functions obtained by optimally perturbing one uncertain parameter λ (red lines). As can be seen in the figure, a model set with one parameter can capture the FRF data quite well, which indicates that the original 8 parameters were redundant to represent the uncertain system. This parameter reduction will lead to the reduction of the burden and the conservativeness in robust controller design.

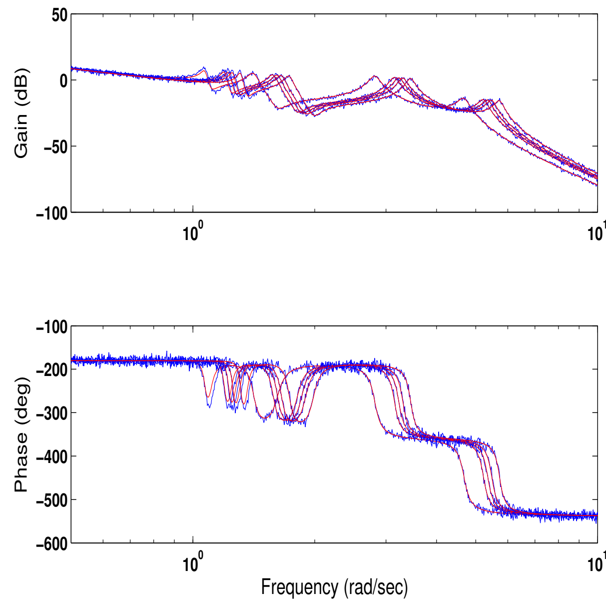

Fig. 1. The noisy FRF data (blue lines), and Bode plots of transfer functions obtained by optimally perturbing one uncertain parameter λ (red lines).

V. CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed a new parameter reduction technique for robust and adaptive control. The technique has been developed based on asymptotic properties of nonlinear least squares estimates, that is, strong consistency and asymptotic normality, and utilized optimization involving a bilinear matrix inequality to detect the correlation of original parameters.

The essential necessary assumption in this paper is that *we know the structure of the true system*, which is not realistic in many applications. Important future work is automatic detection of the structure of the true system from the combination of *a priori* information and experimental frequency response function data.

APPENDIX

A. *Proof of Theorem 4*

To prove Theorem 4, we first review the known result, presented in [15], on asymptotic normality of NLS estimates. Then, we apply this result to our problem formulation.

1) A result in [15]: Consider a general NLS optimization problem to find a true parameter $\theta^\star \in \Theta$:

$$\hat{\theta}^N := \arg\min_{\theta\in\Theta} \sum_{n=1}^{N} \left(y_n - f_n(\theta)\right)^2, \qquad (26)$$

where $y_n := f_n(\theta^\star) + e_n \in \mathbb{R}$ and $f_n : \Theta \to \mathbb{R}$ are given, and $e_n \in \mathcal{N}(0, \sigma^2)$ is a real random process with a normal distribution. For a functional f_n, we denote its gradient and the Hessian by $\nabla f_n : \Theta \to \mathbb{R}^p$ and $\nabla^2 f_n : \Theta \to \mathbb{R}^{p\times p}$, respectively. In addition, let $\tilde{\Sigma}_N : \Theta \times \mathbb{R} \to \mathbb{R}^{p\times p}$ be an operator defined by

$$\tilde{\Sigma}_N(\theta, \tau_N) := \frac{1}{\tau_N} \sum_{n=1}^{N} \nabla f_n(\theta) \nabla f_n(\theta)^T, \qquad (27)$$

where $\{\tau_N\}_{N=1}^{\infty}$ is a positive sequence that satisfies

$$\lim_{N\to\infty} \tau_N = \infty. \qquad (28)$$

The following result [15] is available for asymptotic normality of NLS estimates.

Theorem 8 (Theorem 5 in [15]): Let $\hat{\theta}^N$ be a strongly consistent least-squares estimate of $\theta^\star$. Under the regularity conditions **A1-A5** below, we have

$$\sqrt{\tau_N}(\hat{\theta}^N - \theta^\star) \to_d \mathcal{N}(0, \sigma^2 \tilde{\Sigma}^{-1}), \qquad (29)$$

where $\tilde{\Sigma} := \lim_{N\to\infty} \tilde{\Sigma}_N(\theta^\star, \tau_N)$.

2) Regularity conditions:

A1: $\nabla f_n(\theta)$ and $\nabla^2 f_n(\theta)$ exist for all θ in the neighborhood of $\theta^\star$ which is in the interior of Θ. There exists $\{\tau_N\}_{N=1}^{\infty}$ satisfying (28) and

$$\tilde{\Sigma}_N(\theta^\star, \tau_N) \to \tilde{\Sigma}, \text{ as } N \to \infty, \qquad (30)$$

where $\tilde{\Sigma}$ is positive definite.

A2: As N goes to infinity,

$$\max_{1\le n\le N} \frac{1}{\tau_N} \nabla f_n(\theta^\star)^T \tilde{\Sigma}^{-1} \nabla f_n(\theta^\star) \to 0.$$

A3: As N goes to infinity, and $||\theta - \theta^\star|| \to 0$,

$$\tilde{\Sigma}_N(\theta, \tau_N) \tilde{\Sigma}_N^{-1}(\theta^\star, \tau_N)$$

converges to the identity matrix uniformly.

A4: There exists a $\delta > 0$ such that, for any (j,k)-entry of the Hessian $\nabla^2 f_n(\theta)$, denoted by $\left[\nabla^2 f_n(\theta)\right]_{j,k}$,

$$\limsup_{N\to\infty} \frac{1}{\tau_N} \sum_{n=1}^{N} \sup_{\theta\in\mathcal{B}_\delta(\theta^\star)} \left(\left[\nabla^2 f_n(\theta)\right]_{j,k}\right)^2 < \infty, \qquad (31)$$

where $\mathcal{B}_\delta(\theta^\star) := \{\theta \in \Theta : \|\theta - \theta^\star\| \le \delta\}$ is a δ-neighborhood of $\theta^\star$.

A5: Take δ that satisfies (31). If, for a pair (j,k), the following holds:

$$\sum_{n=1}^{\infty} \sup_{||\theta-\theta^\star||\ge\delta} \left|\left[\nabla^2 f_n(\theta)\right]_{j,k}\right|^2 = \infty, \qquad (32)$$

then there exists a K, independent of n, such that

$$\begin{aligned} \sup_{\substack{s\neq t \\ s,t\in\mathcal{B}_\delta(\theta^\star)}} & \frac{\left|\left[\nabla^2 f_n(s) - \nabla^2 f_n(t)\right]_{j,k}\right|}{\|s-t\|} \\ & \le K \sup_{\theta\in\mathcal{B}_\delta(\theta^*)} \left|\left[\nabla^2 f_n(\theta)\right]_{j,k}\right|, \end{aligned} \qquad (33)$$

for all n.

3) Reducing our problem formulation to the form in (26): Since the formulation in (4) and (6) contains complex numbers, by dividing the square term into real and imaginary parts, we can rewrite the cost function in (6) as

$$\begin{aligned} &\sum_{m=1}^{M} \left(\mathrm{Re}\left\{\hat{G}_m\right\} - \mathrm{Re}\left\{\left[G(\theta)\right](j\omega_m)\right\}\right)^2, \\ &+ \sum_{m=1}^{M} \left(\mathrm{Im}\left\{\hat{G}_m\right\} - \mathrm{Im}\left\{\left[G(\theta)\right](j\omega_m)\right\}\right)^2. \end{aligned}$$

By comparing this equation with (26), we define $N := 2M$, and for $m = 1, \dots, M$,

$$\begin{aligned} y_n &:= \begin{cases} \mathrm{Re}\left\{\hat{G}_m\right\}, & \text{if } n = 2m-1, \\ \mathrm{Im}\left\{\hat{G}_m\right\}, & \text{if } n = 2m, \end{cases} \\ f_n(\theta) &:= \begin{cases} \mathrm{Re}\left\{\left[G(\theta)\right](j\omega_m)\right\}, & \text{if } n = 2m-1, \\ \mathrm{Im}\left\{\left[G(\theta)\right](j\omega_m)\right\}, & \text{if } n = 2m. \end{cases} \end{aligned} \qquad (34)$$

4) Checking the regularity conditions in our problem formulation: In what follows, we verify the aforementioned regularity conditions **A1-A5** for the function $f_n(\theta)$ defined in (34).

A1: Let $\tau_N = N/2$. Then,

$$\tilde{\Sigma}_N(\theta^\star, \tau_N) = \frac{2}{N} \sum_{n=1}^{N} \nabla f_n(\theta^\star) \nabla f_n(\theta^\star)^T = \Sigma_M(\theta^\star),$$

which, due to the assumption (9), converges to $\Sigma(\theta^\star)$ that is positive definite as N (or M) goes to infinity.

A2: Since τ_N satisfies $\lim_{N\to\infty} \tau_{N-1}/\tau_N = 1$, it can be shown (see [15]) that **A2** is implied by **A1**.

A3: Using the matrix norm $||\cdot||$, we have

$$\begin{aligned} &\left\|\tilde{\Sigma}_N(\theta, \tau_N)\tilde{\Sigma}_N^{-1}(\theta^\star, \tau_N) - I\right\| \\ &\le \left\|\tilde{\Sigma}_N(\theta, \tau_N) - \tilde{\Sigma}_N(\theta^\star, \tau_N)\right\| \left\|\tilde{\Sigma}_N^{-1}(\theta^\star, \tau_N)\right\|. \end{aligned}$$

By using the smoothness assumption of G, the series expansion around $\theta = \theta^\star$ gives us

$$\begin{aligned}&\left\|\tilde{\Sigma}_N(\theta,\tau_N) - \tilde{\Sigma}_N(\theta^\star,\tau_N)\right\| \\ &\le K_N \|\theta - \theta^\star\| + O(\|\theta-\theta^\star\|^2),\end{aligned}$$

where K_N is bounded. Therefore, for any $\epsilon > 0$, there exist a positive integer $\bar{N}$ and a neighborhood of $\theta^\star$ such that $\left\|\tilde{\Sigma}_N^{-1}(\theta^\star,\tau_N)\right\|$ is bounded by some constant for all $N \ge \bar{N}$ (due to (30)), and

$$\left\|\tilde{\Sigma}_N(\theta,\tau_N)\tilde{\Sigma}_N^{-1}(\theta^\star,\tau_N) - I\right\| < \epsilon, \forall N \ge \bar{N}.$$

A4: For each (j,k), we have, as M goes to infinity,

$$\begin{aligned}&\frac{1}{\tau_N}\sum_{n=1}^{N}\sup_{\theta\in\mathcal{B}_\delta(\theta^\star)}\left(\left[\nabla^2 f_n(\theta)\right]_{j,k}\right)^2 \\ &= \frac{1}{M}\sum_{m=1}^{M}\left\{\sup_{\theta\in\mathcal{B}_\delta(\theta^\star)}\left(\left[\nabla^2 \mathrm{Re} G_m(\theta)\right]_{j,k}\right)^2\right. \\ &\qquad\left. + \sup_{\theta\in\mathcal{B}_\delta(\theta^\star)}\left(\left[\nabla^2 \mathrm{Im} G_m(\theta)\right]_{j,k}\right)^2\right\}, \\ &\to \int_{\underline{\omega}}^{\overline{\omega}}\left\{\sup_{\theta\in\mathcal{B}_\delta(\theta^\star)}\left(\left[\nabla^2 \mathrm{Re}\left[G(\theta)\right](j\omega)\right]_{j,k}\right)^2\right. \\ &\qquad\left. + \sup_{\theta\in\mathcal{B}_\delta(\theta^\star)}\left(\left[\nabla^2 \mathrm{Im}\left[G(\theta)\right](j\omega)\right]_{j,k}\right)^2\right\} d\omega.\end{aligned}$$

Since the function G is smooth around $\theta^\star$, and since the integral is taken within a finite interval, the last term becomes finite for some small $\delta > 0$.

A5: Without loss of generality, we assume that

$$\sup_{\theta\in\mathcal{B}_\delta(\theta^\star)}\left|\left[\nabla^2 f_n(\theta)\right]_{j,k}\right| > 0,\ \forall n.$$

In fact, if $\sup_{\theta\in\mathcal{B}_\delta(\theta^\star)}\left|\left[\nabla^2 f_n(\theta)\right]_{j,k}\right| = 0$ for some n, then the condition (33) holds irrespective of the choice of K for that n. Under this assumption, since the frequency range $[\underline{\omega},\overline{\omega}]$ is a closed set, and since the functions $\sup_{\theta\in\mathcal{B}_\delta(\theta^\star)}\left|\left[\nabla^2 \mathrm{Re} G(\theta)\right]_{j,k}(j\omega)\right|$ and $\sup_{\theta\in\mathcal{B}_\delta(\theta^\star)}\left|\left[\nabla^2 \mathrm{Im} G(\theta)\right]_{j,k}(j\omega)\right|$ are continuous with respect to ω, there exists a constant $\gamma > 0$ satisfying

$$\inf_n \sup_{\theta\in\mathcal{B}_\delta(\theta^\star)}\left|\left[\nabla^2 f_n(\theta)\right]_{j,k}\right| > \gamma. \qquad (35)$$

Since $\left[\nabla^3 G(\theta)\right](s)$ is uniformly stable over $\theta \in \Theta$, we have

$$\left|\left[\nabla^3 f_n(\theta)\right]_{i,j,k}\right| \le K_1 := \sup_{\theta\in\Theta}\left\|\left[\nabla^3 G(\theta)\right]_{i,j,k}\right\|_\infty,$$

where $\|\cdot\|_\infty$ denotes the H-infinity norm of a stable transfer function. Here, K_1 is finite due to the compactness of Θ. Thus, $\left[\nabla^2 f_n(\theta)\right]_{j,k}$ satisfies the global Lipschitz condition for all $s,t \in \Theta$ such that

$$\sup_{\substack{s\ne t \\ s,t\in\mathcal{B}_\delta(\theta^\star)}}\frac{\left|\left[\nabla^2 f_n(s) - \nabla^2 f_n(t)\right]_{j,k}\right|}{\|s-t\|} \le K_1.$$

We then select K by

$$K := \frac{K_1}{\inf_n \sup_{\theta\in\mathcal{B}_\delta(\theta^\star)}\left|\left[\nabla^2 f_n(\theta)\right]_{j,k}\right|},$$

which is independent of n, and positive due to (35). From this equation, (33) is obtained. (We remark that, in our problem formulation, (33) holds even without the condition (32).)

References

[1] D. R. Brillinger, *Time Series: Data Analysis and Theory*. Classics in Applied Mathematics. SIAM, 2001.

[2] J. Chen and G. Gu, *Control-Oriented System Identification: An H-infinity Approach*, Wiley, 2000.

[3] B. M. Chen, T. H. Lee, K. Peng, and V. Venkataramanan, *Hard Disk Drive Servo Systems*, 2nd ed., Springer, 2006.

[4] J. Choi, R. Nagamune, and R. Horowitz, "Multiple robust controller design based on parameter dependent Lyapunov functions," *17th Int. Symposium on Mathematical Theory of Networks and Systems (MTNS)*, Kyoto, Japan, 2006.

[5] R. Conway, S. Felix, and R. Horowitz, "Parametric uncertainty identification and model reduction for robust H_2 synthesis for track-following in dual-stage hard disk drives," *2007 American Control Conference*.

[6] R. Davidson and J. G. Mackinnon, *Estimation and Inference in Econometrics*, Oxford University Press, 1993.

[7] A. F. Emery and A. V. Nenarokomov, "Optimal experiment design," *Meas. Sci. Technol.*, Vol.9, pp.864-876, 1998.

[8] I. T. Jolliffe, *Principal Component Analysis: 2nd edition*. Springer Verlag, 2002.

[9] R.I. Jennrich, "Asymptotic properties of nonlinear least squares estimators," *The Annals of Mathematical Statistics*, Vol.40, No.2, pp.633-643, 1969.

[10] S. M. Kay, *Fundamentals of Statistical Signal Processing* Vol.1, Prentice Hall, Upper Saddle River, NJ; 1993.

[11] L. Ljung, "Some results of identifying linear systems using frequency domain," *32nd Conference on Decision and Control*, San Antonio, Texas, 1993.

[12] L. Ljung, *System Identification: Theory for the user* (2nd edition), Prentice Hall, Upper Saddle River, NJ; 1999.

[13] S. Nash and A. Sofer, *Linear and Nonlinear Programming*, McGraw-Hill, 1996.

[14] B. Wahlberg and L. Ljung, "Hard frequency-domain model error bounds from least-square like identification techniques," *IEEE Trans. Automat. Contr.*, vol. 36, no. 7, pp. 900–912, 1991.

[15] C-F. Wu, "Asymptotic theory of nonlinear least squares estimation," *The Annals of Statistics*, Vol.9, No.3, pp.501-513, 1981.

[16] K. Zhou, *Essentials of Robust Control*, Prentice-Hall, 1998.

[17] K. J. Åström and B. Wittenmark, *Adaptive Control: 2nd edition*, Addison-Wesley, 1995.