Parameter Reduction in Estimated Model Sets for Robust Control

Ryozo Nagamune
Department of Mechanical Engineering,
University of British Columbia,
Vancouver, BC, V6T 1Z4, Canada
e-mail: nagamune@mech.ubc.ca

Jongeun Choi
Department of Mechanical Engineering and
Department of Electrical and Computer Engineering,
Michigan State University,
East Lansing, MI 48824-1226
e-mail: jchoi@egr.msu.edu

This paper proposes two techniques for reducing the number of uncertain parameters in order to simplify robust controller design and to reduce conservatism inherent in robust controllers. The system is assumed to have a known structure with parametric uncertainties that represent plant dynamics variation. An original set of parameters is estimated by nonlinear least-squares (NLS) optimization using noisy frequency response functions. Utilizing the property of asymptotic normality for NLS estimates, the original parameter set can be reparameterized by an affine function of the smaller number of uncorrelated parameters. The correlation among uncertain parameters is detected by the principal component analysis in one technique and optimization with a bilinear matrix inequality in the other. Numerical examples illustrate the usefulness of the proposed techniques.
[DOI: 10.1115/1.4000661]

Keywords: dynamics variation, nonlinear least-squares estimates, parametric uncertainties, parameter reduction, robust control

1 Introduction

Plant dynamics variation abounds in practical control problems. Such variation is caused by, e.g., the change in operating 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 very 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 control techniques [1] are known to be powerful tools. These techniques are based on models representing dynamics variation, and various modeling and system identification methodologies to acquire such models have been developed [2–4] in the last decades.

In modeling, we always have to consider the trade-off between the accuracy and simplicity of the model. Although a complex model can capture system properties accurately, it is often not preferable for controller design purpose due to unduly high complexity, leading to high computational cost. Especially, if we employ too many parameters to represent dynamics variation, numerical controller design based on modern robust control techniques, such as μ-synthesis and linear matrix inequality (LMI)-based controller synthesis (see, e.g., Ref. [5]), often falls into computational infeasibility, as well as unacceptably conservative system performances. Therefore, model set simplification is an important step.

For a model set involving parametric variation, there are mainly two approaches to model simplification, i.e., first is the model order reduction [6] and parameter number reduction, and the latter is the topic in this paper. Parameter reduction techniques that reduce the number of parameters in a single model have been developed by using sensitivity analysis and principle component analysis together [7,8]. In these paper, a goal is to find out such parameters in an exactly known single model that does not influence so much on the output. On the other hand, the main objective of this paper is to present theoretically sound and yet practical algorithms that will reduce the complexity in interrelated uncertain parameters by detecting hidden correlations in estimated uncertain parameters. To be practical, this reduction procedure takes into account the parameter estimation error inherited from the original noisy data.

One of the most standard techniques for parameter number reduction is the principal component analysis (PCA) [9]. One application of PCA to parameter reduction in hard disk drive systems was presented in Ref. [10]. However, pure PCA does not take into account the effect of estimation errors and noise in experimental data on parameter reduction. In this paper, we will discuss in detail how the noise on frequency response function data affects the parameter reduction stage.

This paper proposes parameter reduction techniques to simplify robust controller design and to reduce the conservativeness inherent in these controllers. The system is assumed to have a known structure with parametric uncertainties caused by plant dynamics variation. An original set of parameters is estimated by nonlinear least-squares (NLS) optimization using noisy frequency response functions. By utilizing the property of asymptotic normality for NLS estimates, the original parameter set can be reparameterized by an affine function of the smaller number of uncorrelated parameters. The correlation among uncertain parameters is detected by PCA in one technique and optimization with a bilinear matrix inequality (BMI) in the other. For the former technique, we will assume that the true parameters are random variables generated by an affine function of another uncorrelated random variables and prove the asymptotic convergence of the identified parameter set to the true parameter set. The latter technique has no convergence analysis but it will be practically useful when the number of plant samples is small, and thus, it is hard to perform any statistical inference.

The paper is organized as follows. In Sec. 2, we give one example to motivate a parameter reduction problem. In Sec. 3, we review the NLS technique for parameter estimation. This section also introduces our assumptions of the plant and the data. Section 4 discusses asymptotic properties, i.e., strong consistency and asymptotic normality of the NLS estimates. Using the asymptotic properties, Sec. 5 proposes parameter reduction techniques based

Contributed by the Dynamic Systems Division of ASME for publication in the JOURNAL OF DYNAMIC SYSTEMS, MEASUREMENT, AND CONTROL. Manuscript received November 9, 2008; final manuscript received September 18, 2009; published online February 2, 2010. Assoc. Editor: Qian Wang.

Copyright © 2010 by ASME

Downloaded 02 Feb 2010 to 35.9.134.245. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

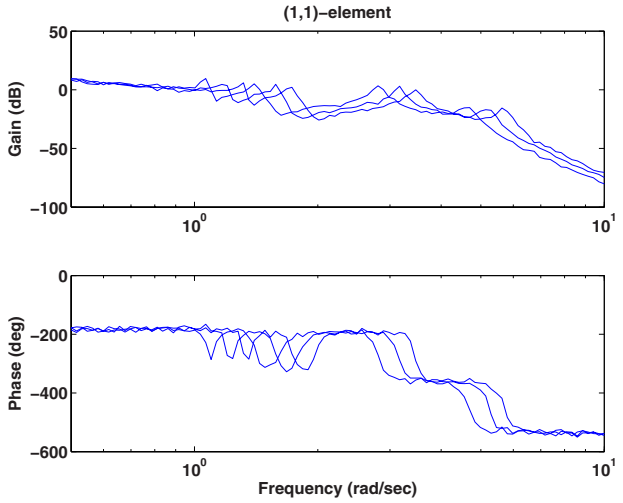

Fig. 1 Three noisy FRF data

on PCA and on optimization with a BMI. Numerical examples are given in Sec. 6 to illustrate the usefulness of the proposed techniques.

The notation used in this paper is standard. The set of positive real 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.

2 Motivating Example

Suppose that we would like to control a number of plants with a single controller. For robust control system design, we first build a mathematical model set that captures dynamics of these plants. To this end, we pick up a relatively small number of sample plants. Here, as an example, suppose that we select three plants with noisy frequency response function (FRF) data as in Fig. 1.

Since three plants are selected among a number of plants, our task is to construct a model set capturing not only these three FRFs but also the "intermediate" plant dynamics. To reduce the conservatism pertaining to robust control systems, as well as to simplify robust controller design, we also would like the set to have a small size in some sense, and to be characterized by a small number of parameters. It is very natural to assume that each of the true plant is subject to a common set of physical laws whose parameters may vary due to another process. Hence, detecting the process that generates parameter variation is essential to improve robust performance.

As far as a model set with only parametric uncertainties is concerned, the standard procedure to this model set identification problem is as follows.

1. Guess and assume the plant structure with parameters. The structure is assumed to be common to all the plants.
2. For each FRF, identify the parameters by some parameter estimation method. That is, for many FRFs, obtain a set of parameters.
3. Find correlations between the set of parameters, and reduce the dimensionality of parameters.

In step 3, the most standard technique in data analysis is PCA [9]. One of our objectives in this paper is to propose PCA that takes into account estimate error covariances under the assumptions that a plant structure is known in step 1 and that nonlinear least-squares estimates are used in step 2 for FRFs with a known noise level.

Another objective is to propose a technique based on numerical optimization. Here, under the same assumptions in steps 1 and 2 above, we will formulate an optimization problem with a BMI and give a solution procedure for it. Although there is no convergence analysis, the technique will be practically useful when the number of samples is small.

3 Parameter Estimation by NLS

As is written in Ref. [12] (p. 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 the extensions of the results in this paper to multivariable cases are straightforward.

3.1 Model Structure. It is assumed that we have a priori information on the structure of a continuous-time linear time-invariant true system.

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

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. Simple examples are the structures of standard first- and second-order transfer functions.

$$[G(\theta)](s) := \frac{K}{Ts+1}, \quad \theta := [K,T]^T \tag{2}$$

$$[G(\theta)](s) := \frac{K\omega^2}{s^2+2\zeta\omega s+\omega^2}, \quad \theta := [K,\zeta,\omega]^T \tag{3}$$

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

$$[G(\theta^\star)](s) \tag{4}$$

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

3.2 Frequency Domain Experimental Data. For the true system (4), we take noisy FRF data as

$$\hat{G}_m = [G(\theta^\star)](j\omega_m) + e_m, \quad m = 1, \ldots, M \tag{5}$$

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, with the following property:

$$\boldsymbol{e} := \begin{bmatrix} \begin{bmatrix} \mathrm{Re}\{e_1\} \\ \mathrm{Im}\{e_1\} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \mathrm{Re}\{e_M\} \\ \mathrm{Im}\{e_M\} \end{bmatrix} \end{bmatrix} \sim \mathcal{N}(0, \sigma^2 I_{2M}) \tag{6}$$

This means that the $2M$-dimensional real random vector $\boldsymbol{e}$ is generated by a normal distribution with zero mean and covariance $\sigma^2 I_{2M}$.

Remark 3.1. Different types of noise models could be considered (as in Refs. [3,11]) if we know more about the sources of noise. However, in this paper, we assume that FRF data are corrupted by the generic Gaussian white noise as in Eq. (5). The origin of the complex-valued white noise e_m could be from the asymptotic normal distribution of the Fourier transform of white noise in the time domain experimental data (see more details in Refs. [12,13]). The complex-valued white noise could be also

Downloaded 02 Feb 2010 to 35.9.134.245. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

viewed as the quantization and electronic noise from the imperfect measurement system.

3.3 Nonlinear Least-Squares Optimization. For the given model structure (1) and FRF data

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

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

$$\hat{\theta}^M := \arg\min_{\theta \in \Theta} \sum_{m=1}^{M} |\hat{G}_m - [G(\theta)](j\omega_m)|^2 \tag{7}$$

The minimization problem (7) is in general a nonconvex 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.

4 Asymptotic Properties of NLS Estimates

Next, we will discuss two important properties of the NLS estimate $\hat{\theta}^M$, i.e., *strong consistency* and *asymptotic normality* [11,14].

4.1 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 Eq. (7) for a large number of samples?" The precise definition is given next.

DEFINITION 4.1. *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.*

A condition for strong consistency was presented in Ref. [15].

THEOREM 4.1. *(Theorem 6 in Ref. [15]). Let D_M be a distance between two parameter vectors defined by*

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

If the following conditions hold, then the NLS estimate $\hat{\theta}^M$ of $\theta^\star$ is strongly consistent: (i) $D_M(\theta, \theta')/M$ converges uniformly to a continuous function $D(\theta, \theta')$ and (ii) $D(\theta, \theta^\star) = 0$ if and only if $\theta = \theta^\star$.

As an illustration of this theorem, let us consider a simple first-order structure

$$[G(\theta)](s) = \frac{K}{Ts + 1} \tag{8}$$

where $\theta := [K, T]^T$, $K > 0$, $T > 0$. Then by defining $\theta' := [K', T']^T$, we have

$$D_M(\theta, \theta') := \sum_{m=1}^{M} \left| \frac{K}{Tj\omega_m + 1} - \frac{K'}{T'j\omega_m + 1} \right|^2 \tag{9}$$

In this case, provided that the frequency points $\{\omega_m\}_{m=1}^M$ are taken at even intervals within a fixed frequency range $[\underline{\omega}, \bar{\omega}]$, we have the uniform convergence in the condition (i)

$$\lim_{M \to \infty} \frac{D_M(\theta, \theta')}{M} = \underbrace{\frac{1}{\bar{\omega} - \underline{\omega}} \int_{\underline{\omega}}^{\bar{\omega}} \left| \frac{K}{Tj\omega + 1} - \frac{K'}{T'j\omega + 1} \right|^2 d\omega}_{=: D(\theta, \theta')} \tag{10}$$

In addition, it is easy to prove that, for the function $D(\theta, \theta')$ in Eq. (10), the condition (ii) holds for $\underline{\omega} < \bar{\omega}$. Therefore, the NLS estimate $\hat{\theta}^M$ of $\theta^\star$ is strongly consistent in this example.

4.2 Asymptotic Normality. If a NLS estimate is strongly consistent, our next concern is to identify the distribution of the NLS estimate. It turns out that under some assumptions, 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 4.2. *A model set $\mathcal{G}$ is said to be uniformly stable for a set Θ if all the transfer functions in the set $\mathcal{G}(\Theta) := \{[G(\theta)] \times (s) : \theta \in \Theta\}$ are stable.*

In the next theorem, we use the notation

$$[\nabla G(\theta^\star)](s) := \left[\frac{\partial}{\partial \theta} G(\theta) \right]_{\theta = \theta^\star} (s) \tag{11}$$

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

THEOREM 4.2. *Assume the following: (i) $\hat{\theta}^M$ is a strongly consistent NLS estimate of $\theta^\star$. (ii) For a given compact parameter set Θ, the model set $\mathcal{G}(\Theta)$ is uniformly stable. (iii) $G(\theta)$ is smooth in Θ. (iv) The true parameter $\theta^\star$ is in the interior of Θ. (v) Frequency points $\{\omega_m; m = 1, \ldots, M\}$ are distributed uniformly over a frequency range $[\underline{\omega}, \bar{\omega}]$ such that*

$$\lim_{M \to \infty} \Sigma_M(\theta^\star) = \Sigma(\theta^\star) \tag{12}$$

where $\Sigma(\theta^\star)$ is a positive definite matrix and

$$\Sigma_M(\theta^\star) := \frac{\sum_{m=1}^{M} \mathrm{Re}\{[\nabla G(\theta^\star)](j\omega_m)[\nabla G(\theta^\star)](j\omega_m)^*\}}{M} \tag{13}$$

Then the estimate $\hat{\theta}^M$ is asymptotically normal with mean $\theta^\star$ and covariance matrix $W(\theta^\star)$.

$$\hat{\theta}^M \to_d \mathcal{N}(\theta^\star, W(\theta^\star)), \quad \text{as } M \to \infty \tag{14}$$

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

$$W(\theta^\star) := \frac{\sigma^2 \Sigma^{-1}(\theta^\star)}{M} \tag{15}$$

The proof of this theorem is given in Appendix B. The error covariance matrix $W(\theta^\star)$ in Eq. (15) will play an important role in the parameter reduction step.

Remark 4.1. In most practical applications, a feasible range of $\theta^\star$ can be obtained from a priori knowledge of the plant such as its geometrical dimensions and material properties. Hence, we can formulate identification problems so that assumptions (ii)–(iv) are satisfied. Assumption (i) can be verified by Theorem 4.1 as in the presented first-order example. However, for general structures, it may be hard to verify the conditions. For a given parameterization $G(\theta)$, assumption (v) can be approximately verified by checking the positivity of Σ_M at an estimate of $\theta^\star$. If Σ_M cannot be positive definite, this implies that the model structure is not sound, and therefore, has to be modified.

Remark 4.2. The Fisher information matrix $\mathbb{I}(\theta^\star)$ [16] of the model (5) can be easily computed by

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

By the Cramér–Rao theorem [16,17], 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 $\mathbb{I}(\theta^\star)$.

$$\mathbb{E}\{(\hat{\theta} - \theta^\star)(\hat{\theta} - \theta^\star)^T\} \geqslant \mathbb{I}(\theta^\star)^{-1} = \frac{\sigma^2 \Sigma_M^{-1}(\theta^\star)}{M} \tag{16}$$

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

Remark 4.3. The choice of $[\underline{\omega}, \bar{\omega}]$ can significantly affect the

Downloaded 02 Feb 2010 to 35.9.134.245. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

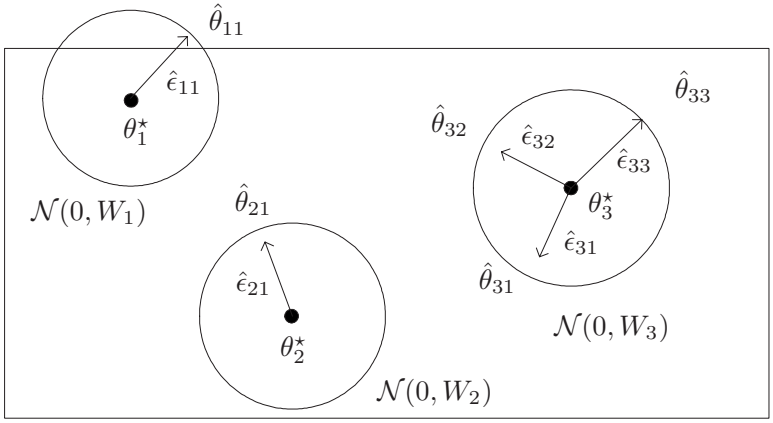

Fig. 2 Three samples of $\theta_\ell^\star$ are distributed in the square support of the probability density function for $\theta_\ell^\star$. For each sample of $\theta_\ell^\star$, there is an asymptotic normal distribution of its NLS estimates. Ellipsoids correspond to approximate confidence regions with some probability.

error covariance matrix W. We want to select $[\underline{\omega}, \bar{\omega}]$ 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 Ref. [17]). In general, $[\underline{\omega}, \bar{\omega}]$ should contain all the significant modes of the dynamical system.

5 Parameter Reduction

So far, we have derived the asymptotic error covariance matrix $W(\theta^\star)$ of the NLS estimate $\hat{\theta}$ [1] for a single true system $G(\theta^\star)$ with single FRF data. 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 reparameterize the set with a fewer number of uncorrelated parameters. This step is called *parameter reduction*. Such multiple true systems represent dynamics variation caused by manufacturing tolerance, change in operating points, and time-varying nature of the plant.

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

$$\epsilon_{\ell k} := \hat{\theta}_{\ell k} - \theta_\ell^\star, \quad \ell = 1,2, \ldots, \ k = 1,2, \ldots \tag{17}$$

By Theorem 4.2, for a fixed ℓ, $\{\epsilon_{\ell k}\}$ is a random process with an asymptotic normal distribution as $M \to \infty$.

$$\epsilon_{\ell k} \to_d \mathcal{N}(0, W_\ell), \quad W_\ell := W(\theta_\ell^\star) = \frac{\sigma^2 \Sigma^{-1}(\theta_\ell^\star)}{M} \tag{18}$$

Few samples from Eq. (18) for three true parameter vectors $\theta_\ell^\star$ are illustrated in Fig. 2. For each $\theta_\ell^\star$, there is an asymptotic normal distribution of its NLS estimates.

Given a finite number of NLS estimates

$$\{\hat{\theta}_{\ell k} \in \mathbb{R}^p; \ \ell = 1, \ldots, L, \ k = 1, \ldots, K\} \tag{19}$$

where p is the number of parameters, and the ℓth asymptotic error covariances

$$\{W_\ell; \ \ell = 1, \ldots, L\} \tag{20}$$

the parameter reduction problem is to find a set

$$\{\theta := \bar{\theta} + V\lambda; \quad \lambda \in \mathbb{R}^q, \ |\lambda(i)| \le 1, \quad i = 1, \ldots, q\} \tag{21}$$

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 Eq. (19) in some sense. Next, we will propose two parameter reduction techniques; one is based on PCA and the other uses a BMI.

5.1 Parameter Reduction Via PCA. PCA can reduce the dimensionality of an empirical data set of correlated variables. The eigenvectors of the covariance matrix of the data are the principal components. These vectors form a basis often called the Karhunen–Loeve transform (KLT) and decorrelate the data in their new coordinates. By observing the associated variance levels, most important principal components can be chosen to represent the original data set.

However, direct blind application of PCA or KLT to parameter reduction can be problematic. We should decide on the threshold of the variance level of the estimates for reduction by PCA. But how can we say if a certain level of variance is negligible? One could choose the threshold level based on the noise level, and this is the approach that we take.

In our problem, the data set contains hidden features of a manufacturing process, which can be represented in terms of a small number of variables and the NLS estimation error noise. Hence the objective of this subsection is to provide a theoretically sound PCA like algorithm, which removes the noise components first and then apply PCA to possibly reduce the dimensionality of the original variables. In this subsection, the assumption on the process of generating the true parameters $\theta_\ell^\star$, $\ell = 1, 2, \ldots$ is as follows.

ASSUMPTION. *The true parameters $\theta_\ell^\star$, $\ell = 1, 2, \ldots$ are generated by means of a stationary random process $\{\lambda_\ell\} \subset \mathbb{R}^q$ with zero mean $\mathbb{E}_\ell\{\lambda_\ell\} = 0$ [2] and some covariance $\mathbb{E}_\ell\{\lambda_\ell \lambda_\ell^T\} = \Lambda$ [3] as*

$$\theta_\ell^\star = \bar{\theta} + V\lambda_\ell, \quad \ell = 1,2, \ldots \tag{22}$$

where $\bar{\theta} \in \mathbb{R}^p$, $V \in \mathbb{R}^{p \times q}$, and $q \in \mathbb{Z}_+$ are unknown and to be determined.

Under this assumption, we will explain how to obtain the un-

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

[2] E_ℓ is the expectation operator over ℓ.

[3] The covariance matrix $\Lambda = \mathbb{E}_\ell\{\lambda_\ell \lambda_\ell^T\}$ is a user's choice. An example of a random process $\{\lambda_\ell\}$ that appears in robust control applications is the uniform distribution, with each vector element $\lambda(i)$, $i = 1, \ldots, q$, having the probability density function $f_{\lambda(i)}$ and the covariance matrix Λ as $f_{\lambda(i)} = 1/2$, $\lambda(i) \in [-1, 1]$, $\Lambda = I_q/3$.

Downloaded 02 Feb 2010 to 35.9.134.245. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

known parameters from the estimates and covariances for the case of an infinite number of samples ($\ell=\infty, \quad k=\infty$) and for the case of a finite number of samples ($\ell<\infty, \quad k<\infty$).

5.1.1 The Case of an Infinite Number of Samples. Although an infinite number of samples is impossible in practice, the following theorem justifies the parameter reduction method, which will be proposed for finite sample cases later.

THEOREM 5.1. *In the case of infinite samples, the unknown parameters* $\bar{\theta}$, q, *and* V are obtained by

$$\bar{\theta} = \mathbb{E}_\ell \mathbb{E}_k\{\hat{\theta}_{\ell k}\} = \lim_{L,K\to\infty} \frac{1}{L}\frac{1}{K}\sum_{\ell=1}^{L}\sum_{k=1}^{K} \hat{\theta}_{\ell k} \tag{23}$$

$$q = \operatorname{rank}(P - W) \tag{24}$$

$$V = U(:,1:q)\Sigma(1:q,1:q)^{1/2}\Lambda^{-1/2} \in \mathbb{R}^{p\times q} \tag{25}$$

where $U(:,1:q)$ is a matrix consisting of the first q *columns of* U, $\Sigma(1:q,1:q)$ *is a matrix consisting of the first* q *rows and first* q *columns of* Σ, *and* $\Lambda^{1/2}$ *denotes a matrix square root of a positive definite matrix* Λ *and*

$$P := \mathbb{E}_\ell \mathbb{E}_k\{(\hat{\theta}_{\ell k} - \bar{\theta})(\hat{\theta}_{\ell k} - \bar{\theta})^T\}$$

$$W := \mathbb{E}_\ell\{W_\ell\}$$

We will prove this theorem. Due to Eq. (22), the estimation error (17) can be written as

$$\epsilon_{\ell k} = \hat{\theta}_{\ell k} - (\bar{\theta} + V\lambda_\ell) \tag{26}$$

The nominal parameter $\bar{\theta} \in \mathbb{R}^p$ in Eq. (23) can be obtained by averaging both sides of Eq. (26) by letting M go to infinity, and by using assumptions $\mathbb{E}_\ell\{\lambda_\ell\}=0$ and $\mathbb{E}_k\{\epsilon_{\ell k}\}=0$.

For the nominal parameter vector $\bar{\theta}$, the error covariance matrix P is, as M goes to infinity in Eq. (18)

$$P = \mathbb{E}_\ell \mathbb{E}_k\{(V\lambda_\ell + \epsilon_{\ell k})(V\lambda_\ell + \epsilon_{\ell k})^T\} = \mathbb{E}_\ell\{(V\lambda_\ell)(V\lambda_\ell)^T\} + E_\ell \mathbb{E}_k\{\epsilon_{\ell k}\epsilon_{\ell k}^T\} = V\Lambda V^T + W \tag{27}$$

Here, we have used $\mathbb{E}_k\{\epsilon_{\ell k}\}=0$ and $\mathbb{E}_k\{\epsilon_{\ell k}\epsilon_{\ell k}^T\}=W_\ell$. By taking the singular value decomposition (SVD) of the matrix $P-W$,

$$V\Lambda V^T = P - W = U\Sigma U^T \tag{28}$$

We can determine the reduced number of parameters q as

$$q = \operatorname{rank}\Sigma = \operatorname{rank}(P - W) \tag{29}$$

and the matrix V as in Eq. (25).

5.1.2 The Case of a Finite Number of Samples. In practice, we have only a finite number of samples. For sample sets $\{\hat{\theta}_{\ell k}: \ell = 1,\ldots,L, \quad k=1,\ldots,K\}$ and $\{W_\ell: \ell=1,\ldots,L\}$, the matrices $\bar{\theta}$, P, and W in Theorem 5.1 can be approximated, respectively, by

$$\bar{\theta}^s := \frac{1}{L}\frac{1}{K}\sum_{\ell=1}^{L}\sum_{k=1}^{K} \hat{\theta}_{\ell k} \tag{30}$$

$$P^s := \frac{1}{LK-1}\sum_{\ell=1}^{L}\sum_{k=1}^{K} (\hat{\theta}_{\ell k} - \bar{\theta}^s)(\hat{\theta}_{\ell k} - \bar{\theta}^s)^T \tag{31}$$

$$W^s := \frac{1}{L}\sum_{\ell=1}^{L} W_\ell \tag{32}$$

In finite sample cases, the reduced number q of parameters must be determined by truncating relatively small singular values of $P^s - W^s$. Due to Theorem 5.1, the approximations become better as the numbers of samples L and K increases.

Remark 5.1. We only consider an affine mapping from λ to θ in (21). Such parameterization occurs in many control applications. However, this "affinity" assumption may be limited for some manufacturing processes. In this case, our approach can be generalized by a nonlinear version of PCA, called *kernel PCA* [18].

5.2 Parameter Reduction Via BMI. In the parameter reduction method via PCA, we assumed that true parameter vectors are generated by an affine transformation of a random vector sequence. In contrast, in what follows, we will not assume anything about the process of true parameter vectors. This assumption is more practical than the one in the previous subsection. Under this assumption, we will obtain, for the prespecified reduced number of uncorrelated parameters q, the parameter set (21) based on only the NLS estimates and their error covariances.

Geometrically, the parameter set (21) is a q-dimensional hyperrectangle in $\mathbb{R}^p$ ($q<p$) and the NLS estimates are points in $\mathbb{R}^p$. To find a hyperrectangle that passes close to all these points, we take the following two steps.

Step 1. Find a q-dimensional *hyperplane* that passes close to all the NLS estimates.

Step 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 in step 2 is important for less conservative robust controller design.

The problem in step 1 can be written mathematically as

$$\min_{\theta_{\text{opt}}\in\mathbb{R}^p, V_{\text{opt}}\in\mathbb{R}^{p\times q}, \lambda_\ell\in\mathbb{R}^q, \ell=1,\ldots,L} \gamma$$

$$\text{subject to } \|W_\ell^{-1/2}(\hat{\theta}_{\ell k} - (\theta_{\text{opt}} + V_{\text{opt}}\lambda_\ell))\|^2 < \gamma$$

$$k = 1,\ldots,K, \quad \ell = 1,\ldots,L \tag{33}$$

Here, in measuring the "distance" between an NLS estimate and the hyperplane, we take into account the error covariance matrix, indicating how much we can trust the estimate. In terms of matrix inequalities, we can express the inequality constraint in Eq. (33) for each ℓ and k as

$$\begin{bmatrix} \gamma & (\hat{\theta}_{\ell k} - (\theta_{\text{opt}} + V_{\text{opt}}\lambda_\ell))^T \\ \star & W_\ell \end{bmatrix} > 0 \tag{34}$$

where $\star$ denotes entries that follow from symmetry and $P>0$ means that a symmetric matrix P is positive definite. By gathering all ℓ and k, this condition is equivalent to

$$\begin{bmatrix} \gamma I_{LK} & (\hat{\Theta} - (\theta_{\text{opt}} + (I_{LK}\otimes V_{\text{opt}})\Lambda))^T \\ \star & \boldsymbol{W} \end{bmatrix} > 0 \tag{35}$$

where $\otimes$ denotes the Kronecker product and

$$\hat{\Theta} := \operatorname{diag}[\hat{\theta}_{11},\ldots,\hat{\theta}_{1K},\ldots,\hat{\theta}_{L1},\ldots,\hat{\theta}_{LK}] \in \mathbb{R}^{pLK\times LK}$$

$$\theta_{\text{opt}} := I_{LK}\otimes\theta_{\text{opt}} \in \mathbb{R}^{pLK\times LK}$$

$$\Lambda := \operatorname{diag}[I_K\otimes\lambda_1,\ldots,I_K\otimes\lambda_L] \in \mathbb{R}^{qLK\times LK}$$

$$\boldsymbol{W} := \operatorname{diag}[I_K\otimes W_1,\ldots,I_K\otimes W_L] \in \mathbb{R}^{pLK\times pLK}$$

This is a BMI with unknowns γ, θ_{opt}, V_{opt}, and Λ. To find a suboptimal solution via LMIs, we alternate the following two LMI optimization problems: (i) Fix $(\theta_{\text{opt}}, V_{\text{opt}})$, and solve LMI with respect to (γ,Λ). (ii) Fix Λ, and solve LMI with respect to $(\gamma,\theta_{\text{opt}},V_{\text{opt}})$. The initial points for θ_{opt} and V_{opt} can be, for example, the ones corresponding to the sample mean value $\bar{\theta}^s$ in (30) and a matrix V obtained from its error covariance matrix.

After finding a q-dimensional hyperplane in step 2 to minimize the "size" of the parameter set (21)) for robust control purpose, as

Downloaded 02 Feb 2010 to 35.9.134.245. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

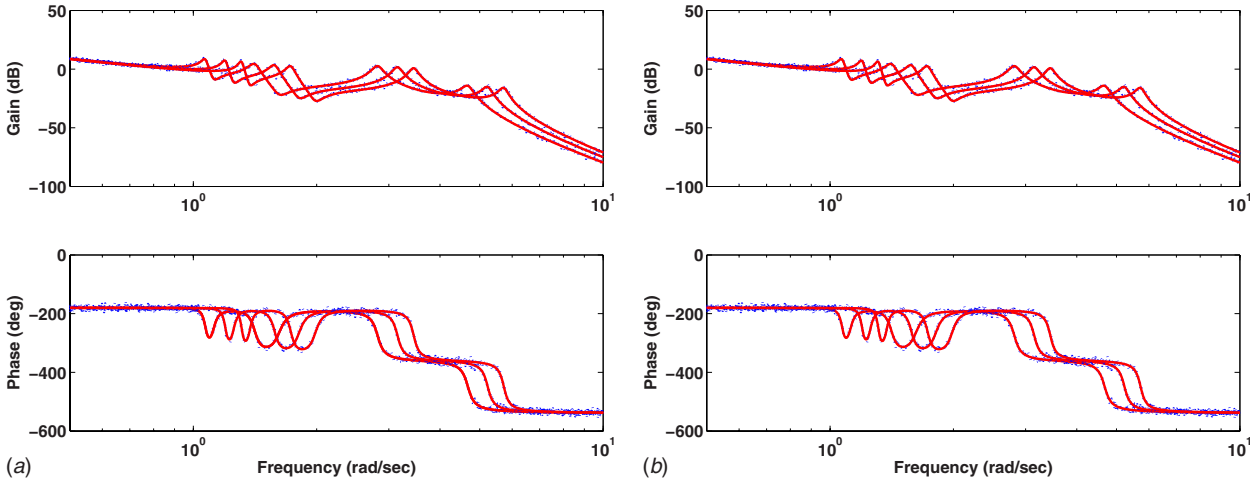

Fig. 3 The noisy FRF data (dotted lines) and Bode plots of transfer functions obtained by optimally perturbing one uncertain parameter λ (solid lines). The left two and right two figures, respectively, correspond to parameter reduction based on PCA and BMI.

well as to satisfy the constraint $|\lambda(i)| \le 1$, we need to adjust the nominal parameter θ_{opt} and the matrix V_{opt}. The problem is to find a hyperrectangle

$$\mathcal{H} := \{\lambda := \bar{\lambda} + T\delta\lambda, \quad |\delta\lambda(i)| \le 1, \ i = 1, \ldots, q\} \tag{36}$$

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

$$\bar{\lambda}(i) := \frac{1}{2}\left(\min_{\ell=1,\ldots,L} \lambda_\ell(i) + \max_{\ell=1,\ldots,L} \lambda_\ell(i)\right), \ i = 1, \ldots, q$$

$$T := \text{diag}[\max_{\ell}|\lambda_\ell(1) - \bar{\lambda}(1)|, \ldots, \max_{\ell}|\lambda_\ell(q) - \bar{\lambda}(q)|]$$

Since $\{\lambda_\ell\}_{\ell=1}^{L} \subset \mathcal{H}$, the following relation holds

$$\{\theta := \theta_{\text{opt}} + V_{\text{opt}}\lambda_\ell, \ \ell = 1, \ldots, L\}$$

$$\subset \{\theta := \theta_{\text{opt}} + V_{\text{opt}}\lambda, \quad \lambda \in \mathcal{H}\}$$

$$= \{\theta := \bar{\theta} + V\delta\lambda, \quad |\delta\lambda(i)| \le 1, \ i = 1, \ldots, q\} \tag{37}$$

where $\bar{\theta} := \theta_{\text{opt}} + V_{\text{opt}}\bar{\lambda}$, $\quad V := V_{\text{opt}}T$. In this way, we have obtained set (21) that approximates all the NLS estimates.

6 Practical Examples

We illustrate the proposed parameter reduction methods with two examples. One is a single-input-single-output (SISO) example from a hard disk drive application, and the other is a multiple-input-multiple-output (MIMO) example from a machine tool application.

6.1 SISO Example. This example was taken from the book [19] (ch. 11). Consider the following set of true system dynamics:

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

$$[G_1(\delta)](s) := \frac{0.64013}{s^2}$$

$$[G_2(\delta)](s) := \frac{0.912s^2 + 0.4574s + 1.433(1+\delta)}{s^2 + 0.3592s + 1.433(1+\delta)}$$

$$[G_3(\delta)](s) := \frac{0.7586s^2 + 0.9622s + 2.491(1+\delta)}{s^2 + 0.7891s + 2.491(1+\delta)}$$

$$[G_4(\delta)](s) := \frac{9.917(1+\delta)}{s^2 + 0.1575s + 9.917(1+\delta)}$$

$$[G_5(\delta)](s) := \frac{27.31(1+\delta)}{s^2 + 0.2613s + 27.31(1+\delta)}$$

We have considered the following case: For uniformly distributed $\delta = 0, \pm 0.2$, we took $K=3$ noisy FRF data with noise variance $\sigma^2 = 0.01$. We will use the parameter reduction technique via PCA and BMI. By regarding the eight parameters as components of uncertain θ, the NLS estimates $\{\hat{\theta}_{\ell k} \in \mathbb{R}^8\}$ and the approximated asymptotic error covariances $\{W_\ell\}$ were obtained. Based on these estimates and covariances, we reduced the number of uncertain parameters.

The three largest singular values of $P^s - W^s$ are 0.1881, 0.0790, and 0.0058. We need to select q in Eq. (29) by neglecting small singular values. In practical problems with finite samples, one may need some trial and error to select an appropriate q. Theoretical results in this paper guarantees that as M, L, and K go to infinity, we are able to recover the true value of q, which is one in this example.

Here, we selected the number of reduced parameters $q=1$ and the performed parameter reduction. In Fig. 3, it is shown the noisy FRF data (dotted lines) and Bode plots of transfer functions obtained by optimally perturbing one uncertain parameter λ (solid lines). As can be seen in these figures, a model set with one parameter can capture the FRF data quite well, which indicates that the original eight parameters were redundant to represent the uncertain system.

Finally, we compared the two proposed techniques with the standard PCA without taking into account estimation error covariance matrices. For the case $L=K=3$, we tried various number M of frequency gridding and computed the average with over ten different noise realizations of the norm of nominal estimation error.

$$\|[\|\theta_\ell^* - \hat{\theta}_{\ell,k}^*\|]_{\ell,k}\|$$

where $\hat{\theta}_{\ell,k}^*$ is a point in the reduced parameter set, which is closest to the NLS estimate $\hat{\theta}_{\ell,k}$, and $[m_{\ell,k}]_{\ell,k}$ is a matrix whose (ℓ,k) entry is $m_{\ell,k}$. Table 1 shows the improvement in percentage of the

Downloaded 02 Feb 2010 to 35.9.134.245. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

Table 1 Nominal error improvement in percentage

M	70	80	90	100
PCA with W	9.07	7.99	11.4	11.8
BMI	11.5	17.3	18.0	10.4

proposed methods over the standard PCA without considering estimation error covariance matrices W. As can be seen in the table, both of the proposed methods outperform the standard PCA around 10% in average, even though this is not theoretically guaranteed. In fact, through a number of simulations, we found that PCA with W almost always perform better than PCA without W, while the performance of BMI varies under different conditions such as initial points, number of samples, and different noise realizations.

6.2 MIMO Example. In this example, the FRF data for a machine tool with a ball screw drive are used. The system has two inputs and two outputs. This system can be regarded as a simple mass-spring-damper system. However, the system parameters change depending on the mass (or table) location over the possibly long ball screw. This is illustrated in Fig. 4, where FRF data (dotted lines) were taken at three different locations of the mass.

From a priori knowledge about the plant, we assumed the model structure as

$$[G(\theta)](s) := \frac{\theta_1}{s^2}\begin{bmatrix} M_1(s) & M_2(s) \\ M_2(s) & M_3(s) \end{bmatrix}$$

$$M_1(s) := K_1 \frac{s^2 + 2\theta_4\theta_5 s + \theta_5^2}{s^2 + 2\theta_2\theta_3 s + \theta_3^2}$$

$$M_2(s) := \frac{K_2}{s^2 + 2\theta_2\theta_3 s + \theta_3^2} \cdot \frac{-s + \theta_8}{s + \theta_8}$$

$$M_3(s) := K_3 \frac{s^2 + 2\theta_6\theta_7 s + \theta_7^2}{s^2 + 2\theta_2\theta_3 s + \theta_3^2} \cdot \frac{-s + \theta_8}{s + \theta_8}$$

Here, the first-order term in M_2 and M_3 can be interpreted as the Padé approximation of a time delay and K_i, $i=1,2,3$ are fixed constants: $K_1=0.001$, $K_2=0.0004$, $K_3=0.0005$. There are eight parameters $\theta := [\theta_1, \ldots, \theta_8]^T$ for NLS optimization.

In this example, we used the parameter reduction method based on BMI. When we applied the method to the FRF data, we figured out by trial and error that two parameters are sufficient to represent the main characteristics of the data. In fact, this parameter reduction into two parameters did not deteriorate the FRF fitting at all in this example. Therefore, we set $q=2$.

The frequency responses of three models in the model set with reduced number of parameters are shown in Fig. 4 as solid lines with experimental FRF data as dotted lines. As can be seen in the

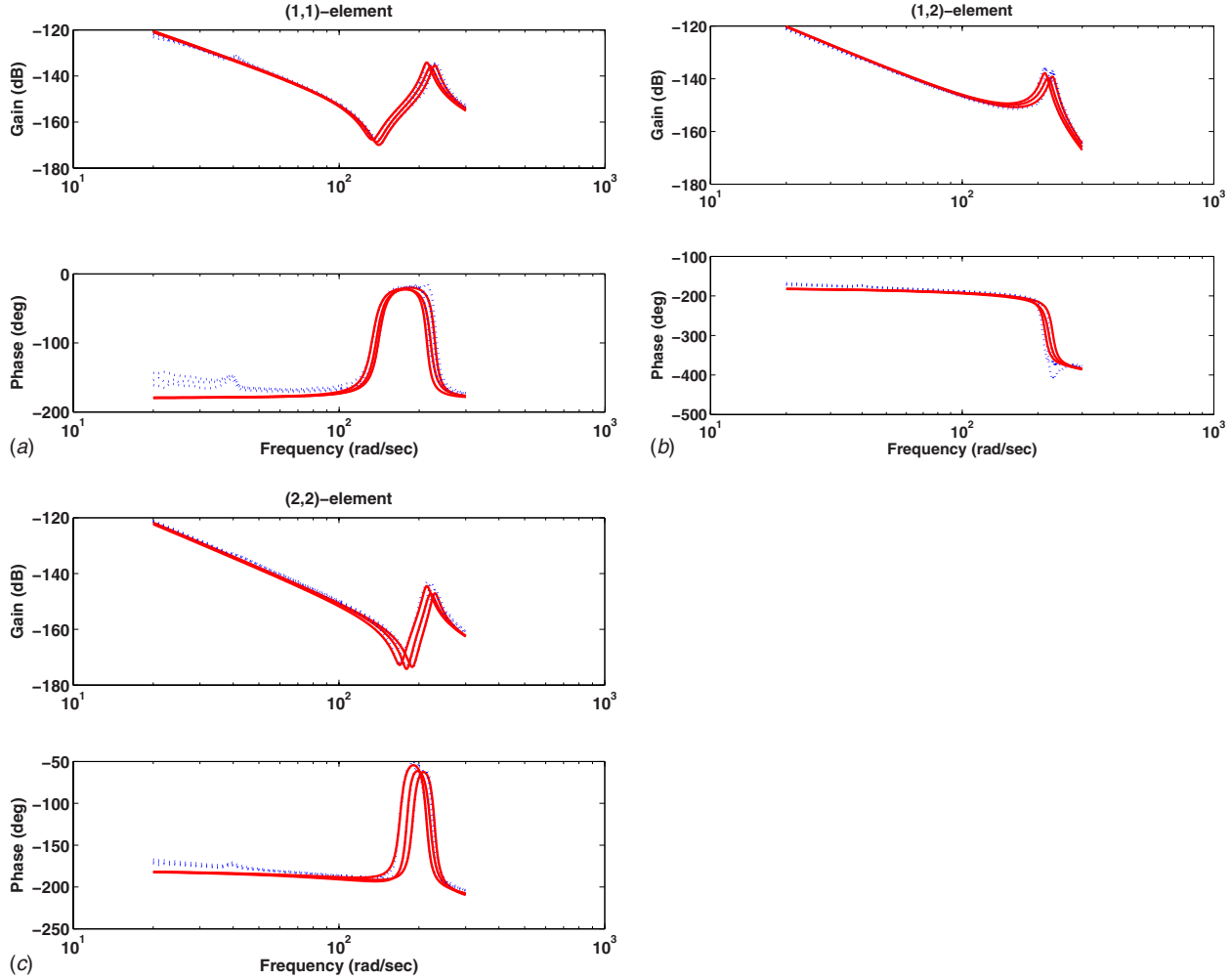

Fig. 4 The noisy FRF data (dotted lines), and Bode plots of transfer functions obtained by optimally perturbing an uncertain parameter vector λ (solid lines). The first resonance mode around 40 rad/s was not considered in the model structure for simplicity.

Downloaded 02 Feb 2010 to 35.9.134.245. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

figures, the model set captures main features of the dynamic responses of the perturbed plant. This model set will be useful in designing, e.g., a gain-scheduling controller [20,21].

7 Conclusions

In this paper, we have proposed two parameter reduction techniques for robust control. The techniques have been developed based on asymptotic properties of nonlinear least-squares estimates, that are strong consistency and asymptotic normality. We have utilized the principal component analysis in one technique and optimization involving a bilinear matrix inequality in the other 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 practical in some 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.

Acknowledgment

The authors are grateful to Professor Y. Altintas and Dr. C. Okwudire at the Manufacturing Automation Laboratory, University of British Columbia, for providing them with experimental FRF data for a ball screw machine tool. This work was supported by the National Science and Engineering Research Council (NSERC) of Canada.

Appendix A: A Gradient Computation

Here, we provide a formula to compute the gradient of a transfer function $G(\theta)$ with respect to a parameter vector $\theta \in \mathbb{R}^p$. For generality, the function $G(\theta)$ is assumed to be represented in a linear fractional transformation (LFT) form:

$$[G(\theta)] := M_{22} + M_{21}\Delta(\theta)[I - M_{11}\Delta(\theta)]^{-1}M_{12} \tag{A1}$$

(We omitted the argument s for simplicity.) Here, the matrix M is partitioned with appropriate dimensions as

$$M(s) = \begin{bmatrix} M_{11}(s) & M_{12}(s) \\ M_{21}(s) & M_{22}(s) \end{bmatrix} \tag{A2}$$

For a given set of parameters $\{\theta(i): i \in \mathcal{I}\}$, $\mathcal{I} := \{1, \ldots, p\}$, consider a matrix $\Delta(\theta) \in \mathbb{R}^{w\times w}$ having the following structure, which is linear with respect to parameters $\theta(i)$:

$$\Delta(\theta) = \sum_{i\in\mathcal{I}} \theta(i)E_i \tag{A3}$$

for some matrix $E_i \in \mathbb{R}^{w\times w}$.

Now, we will prove that, for the transfer function $G(\theta)$ in Eq. (A1) with $\Delta(\theta)$ in Eq. (A3), the ith entry of $\nabla G(\theta)$ can be computed as

$$[\nabla G(\theta)]_i = M_{21}(I - \Delta(\theta)M_{11})^{-1}E_i(I - M_{11}\Delta(\theta))^{-1}M_{12} \tag{A4}$$

The formula (A4) with an LFT format is found to be an efficient way to obtain gradients with respect to θ for calculating Σ_M in Eq. (13).

The transfer function $G(\theta)$ is given by

$$G(\theta) = M_{22} + M_{21}\sum_{i\in\mathcal{I}} \theta(i)E_i\left[I - M_{11}\sum_{i\in\mathcal{I}} \theta(i)E_i\right]^{-1}M_{12}$$

The partial derivative of $G(\theta)$ with respect to $\theta(i)$ is obtained as follows:

$$[\nabla G(\theta)]_i = M_{21}E_i[I - M_{11}\Delta(\theta)]^{-1}M_{12} + M_{21}\Delta(\theta)\frac{\partial}{\partial\theta_i}[I - M_{11}\Delta(\theta)]^{-1}M_{12} = M_{21}E_i[I - M_{11}\Delta(\theta)]^{-1}M_{12} + M_{21}\Delta(\theta)[I - M_{11}\Delta(\theta)]^{-1} \times (M_{11}E_i)[I - M_{11}\Delta(\theta)]^{-1}M_{12} = M_{21}[I - \Delta(\theta)M_{11}]^{-1}E_i[I - M_{11}\Delta(\theta)]^{-1}M_{12}$$

Appendix B: Proof of THEOREM 4.2

To prove THEOREM 4.2, we first review the known result in Ref. [22], on asymptotic normality of NLS estimates. Then, we apply this result to our problem formulation.

B.1 A Result in Ref. [22]. Consider a general NLS optimization problem to find a true parameter $\theta^\star \in \Theta$:

$$\hat{\theta}^N := \arg\min_{\theta\in\Theta}\sum_{n=1}^{N}(y_n - f_n(\theta))^2 \tag{B1}$$

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

$$\tilde{\Sigma}_N(\theta, \tau_N) := \frac{1}{\tau_N}\sum_{n=1}^{N}\nabla f_n(\theta)\,\nabla f_n(\theta)^T \tag{B2}$$

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

$$\lim_{N\to\infty} \tau_N = \infty. \tag{B3}$$

The following result [22] is available for asymptotic normality of NLS estimates.

THEOREM B.1. *(Theorem 5 in Ref. [22]). Let* $\hat{\theta}^N$ be a strongly consistent least-squares estimate of $\theta^\star$. *Under the regularity conditions A1–A5 below, we have*

$$\sqrt{\tau_N}(\hat{\theta}^N - \theta^\star) \to_d \mathcal{N}(0, \sigma^2\tilde{\Sigma}^{-1}) \tag{B4}$$

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

B.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 (44) and

$$\tilde{\Sigma}_N(\theta^\star, \tau_N) \to \tilde{\Sigma} \quad \text{as } N \to \infty \tag{B5}$$

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 $[\nabla^2 f_n(\theta)]_{j,k}$

$$\limsup_{N\to\infty}\frac{1}{\tau_N}\sum_{n=1}^{N}\sup_{\theta\in\mathcal{B}_\delta(\theta^\star)}([\nabla^2 f_n(\theta)]_{j,k})^2 < \infty \tag{B6}$$

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 Eq. (47). If for a pair (j,k), the following holds:

$$\sum_{n=1}^{\infty}\sup_{\|\theta-\theta^\star\|\ge\delta}|[\nabla^2 f_n(\theta)]_{j,k}|^2 = \infty \tag{B7}$$

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

Downloaded 02 Feb 2010 to 35.9.134.245. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

$$\sup_{\substack{s\neq t\\ s,t\in\mathcal{B}_\delta(\theta^\star)}} \frac{|[\nabla^2 f_n(s)-\nabla^2 f_n(t)]_{j,k}|}{\|s-t\|} \le K \sup_{\theta\in\mathcal{B}_\delta(\theta^\star)} |[\nabla^2 f_n(\theta)]_{j,k}| \tag{B8}$$

for all n.

B.3 Reducing our Problem to the Form in Eq. (B1). Since the formulation in Eqs. (5) and (7) contains complex numbers by dividing the square term into real and imaginary parts, we can rewrite the cost function in Eq. (7) as

$$\sum_{m=1}^{M} (\mathrm{Re}\{\hat{G}_m\} - \mathrm{Re}\{[G(\theta)](j\omega_m)\})^2 + \sum_{m=1}^{M} (\mathrm{Im}\{\hat{G}_m\} - \mathrm{Im}\{[G(\theta)] \times (j\omega_m)\})^2$$

By comparing this equation with Eq. (B1), we define $N:=2M$ and for $m=1,\ldots,M$

$$y_n := \begin{cases} \mathrm{Re}\{\hat{G}_m\} & \text{if } n=2m-1 \\ \mathrm{Im}\{\hat{G}_m\} & \text{if } n=2m \end{cases}$$

$$f_n(\theta) := \begin{cases} \mathrm{Re}\{[G(\theta)](j\omega_m)\} & \text{if } n=2m-1 \\ \mathrm{Im}\{[G(\theta)](j\omega_m)\} & \text{if } n=2m \end{cases} \tag{B9}$$

B.4 Checking the Regularity Conditions in our Problem. In what follows, we verify the aforementioned regularity conditions A1–A5 for the function $f_n(\theta)$ in Eq. (B9).

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 (12), 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 Ref. [22]) that A2 is implied by A1.

A3. We have

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

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

$$\|\tilde{\Sigma}_N(\theta,\tau_N) - \tilde{\Sigma}_N(\theta^\star,\tau_N)\| \le K_N\|\theta-\theta^\star\| + O(\|\theta-\theta^\star\|^2)$$

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 $\|\tilde{\Sigma}_N^{-1}(\theta^\star,\tau_N)\|$ is bounded by some constant for all $N\ge\bar{N}$ and

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

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

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

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

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

In fact, if $\sup_{\theta\in\mathcal{B}_\delta(\theta^\star)}|[\nabla^2 f_n(\theta)]_{j,k}|=0$ for some n, then the condition (49) holds irrespective of the choice of K. Under this assumption, since the frequency range $[\underline{\omega},\bar{\omega}]$ is a closed set and since the functions $\sup_{\theta\in\mathcal{B}_\delta(\theta^\star)}|[\nabla^2 \mathrm{Re}\, G(\theta)]_{j,k}(j\omega)|$ and $\sup_{\theta\in\mathcal{B}_\delta(\theta^\star)}|[\nabla^2 \mathrm{Im}\, G(\theta)]_{j,k}(j\omega)|$ are continuous with respect to ω, there exists a constant $\gamma>0$ satisfying

$$\inf_n \sup_{\theta\in\mathcal{B}_\delta(\theta^\star)} |[\nabla^2 f_n(\theta)]_{j,k}| > \gamma \tag{B10}$$

Since $[\nabla^3 G(\theta)]$ [4] is uniformly stable over $\theta\in\Theta$, we have

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

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

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

We then select K by

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

which is independent of n and positive due to Eq. (B10). From this equation, Eq. (B8) is obtained. (We remark that, in our problem formulation, Eq. (B8) holds even without the condition Eq. (B7).)

References

[1] Zhou, K., 1998, *Essentials of Robust Control*, Prentice-Hall, Englewood Cliffs, NJ.

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

[3] Ljung, L., 1999, *System Identification: Theory for the User*, 2nd ed., Prentice-Hall, Upper Saddle River, NJ.

[4] Kosut, R. L., Lau, M. K., and Boyd, S. P., 1992, "Set-Membership Identification of Systems With Parametric and Nonparametric Uncertainty," IEEE Trans. Autom. Control, **37**(7), pp. 929–941.

[5] Skogestad, S., and Postlethwaite, I., 2005, *Multivariable Feedback Control: Analysis and Design*, Wiley, New York.

[6] Obinata, G., and Anderson, B. D. O., 2001, *Model Reduction for Control System Design*, Springer, New York.

[7] Degenring, D., Froemel, C., Dikta, G., and Takors, R., 2004, "Sensitivity Analysis for the Reduction of Complex Metabolism Models," J. Process Control, **14**(7), pp. 729–745.

[8] Sun, C., and Hahn, J., 2006, "Parameter Reduction for Stable Dynamical Systems Based on Hankel Singular Values and Sensitivity Analysis," Chem. Eng. Sci., **61**(16), pp. 5393–5403.

[9] Jolliffe, I. T., 2002, *Principal Component Analysis*, Springer-Verlag, New York.

[10] Conway, R., Felix, S., and Horowitz, R., 2007, "Parametric Uncertainty Identification and Model Reduction for Robust H_2 Synthesis for Track-Following in Dual-Stage Hard Disk Drives," American Control Conference.

[11] Pintelon, R., and Schoukens, J., 2001, *System Identification: A Frequency Domain Approach*, IEEE Press.

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

[13] Ljung, L., 1993, "Some Results of Identifying Linear Systems Using Frequency Domain," 32nd Conference on Decision and Control, San Antonio, TX.

[4] $[\nabla^3 G(\theta)]$ is a three-dimensional matrix (i,j,k) whose entry is given by $\partial^3 G/\partial\theta_i\partial\theta_j\partial\theta_k$.

Downloaded 02 Feb 2010 to 35.9.134.245. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm

[14] Davidson, R., and Mackinnon, J. G., 1993, *Estimation and Inference in Econometrics*, Oxford University Press, New York.
[15] Jennrich, R. I., 1969, "Asymptotic Properties of Nonlinear Least Squares Estimators," Ann. Math. Stat., **40**(2), pp. 633–643.
[16] Kay, S. M., 1993, *Fundamentals of Statistical Signal Processing*, Prentice-Hall, Upper Saddle River, NJ, Vol. 1.
[17] Emery, A. F., and Nenarokomov, A. V., 1998, "Optimal Experiment Design," Meas. Sci. Technol., **9**, pp. 864–876.
[18] Schölkopf, B., Smola, A., and Müller, K.-R., 1998, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem," Neural Comput., **10**(5), pp. 1299–1319.
[19] Chen, B. M., Lee, T. H., Peng, K., and Venkataramanan, V., 2006, *Hard Disk Drive Servo Systems*, Springer, New York.
[20] Dong, K., and Wu, F., 2007, "Robust and Gain-Scheduling Control for LFT Systems Through Duality and Conjugate Lyapunov Functions," Int. J. Control, **80**(4), pp. 555–568.
[21] Wu, F., and Dong, K., 2006, "Gain-Scheduling Control of LFT Systems Using Parameter-Dependent Lyapunov Functions," Automatica, **42**(1), pp. 39–50.
[22] Wu, C.-F., 1981, "Asymptotic Theory of Nonlinear Least Squares Estimation," Ann. Stat., **9**(3), pp. 501–513.

Downloaded 02 Feb 2010 to 35.9.134.245. Redistribution subject to ASME license or copyright; see http://www.asme.org/terms/Terms_Use.cfm