Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

# Parameter Reduction of Nonlinear Least-Squares Estimates via the Singular Value Decomposition

**Ryozo Nagamune** * **Jongeun Choi** **

* *Department of Mechanical Engineering, University of British Columbia, Vancouver, BC V6T 1Z4 Canada (e-mail: nagamune@mech.ubc.ca).*
** *Department of Mechanical Engineering, Michigan State University, East Lansing, MI 48824 USA (e-mail: jchoi@egr.msu.edu)*

**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. Based on 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 over NLS estimates from different plants is detected by the singular value decomposition. A numerical example illustrates the usefulness of the proposed technique.

## 1. 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 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 and adaptive control techniques (Zhou (1998); Åström and Wittenmark (1995)) 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 (Ljung (1999); Chen and Gu (2000)).

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 (Jolliffe (2002)), Conway et al. recently developed a parameter reduction method using the singular value decomposition (Conway et al. (2007)). A possible drawback is that they do not consider the effect of error of the estimates and 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.

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. With 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 over NLS estimates from different plants is detected by the singular value decomposition.

The paper is organized as follows. In Section 2, we review the nonlinear least-squares technique for parameter estimation. This section also introduces our assumptions of the plant and the data. Section 3 discusses asymptotic properties, i.e., strong consistency and asymptotic normality, of the NLS estimates. Based on the asymptotic properties, Section 4 proposes a parameter reduction technique based on the singular value decomposition. A practical example is given after presenting the reduction technique.

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.

## 2. PARAMETER ESTIMATION BY NONLINEAR LEAST-SQUARES OPTIMIZATION

As is written in Ljung (1999), 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.

### *2.1 Model structure*

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

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

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

### *2.2 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 Brillinger (2001); Ljung (1993)). Some of it can be viewed as the quantization and electronic noise of the data acquisition system. Such noise level $\sigma$ can be suppressed effectively by averaging sinusoidal output signals over many periods. This is the major advantage of identifying LTI systems based on FRFs.

### *2.3 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 $\theta$) of the NLS problem.

## 3. ASYMPTOTIC PROPERTIES OF NONLINEAR LEAST-SQUARES ESTIMATES

Next, we will review and discuss two important properties of the NLS estimate $\hat{\theta}^M$, i.e., *strong consistency* and *asymptotic normality* (Davidson and Mackinnon (1993), Pintelon and Schoukens (2001)).

### *3.1 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 (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 Jennrich (1969)). *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 $\theta = \theta^\star$.*

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

$$[G(\theta)](s) = \frac{K}{Ts+1}, \tag{7}$$

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

$$D_M(\theta, \theta') := \sum_{m=1}^{M} \left|\frac{K}{Tj\omega_m+1} - \frac{K'}{T'j\omega_m+1}\right|^2. \tag{8}$$

In this case, provided that the frequency points $\{\omega_m\}_{m=1}^M$ are taken at even intervals within a fixed frequency range $[\underline{\omega}, \overline{\omega}]$, we have uniform convergence in the condition **C1**:

$$\lim_{M\to\infty} \frac{D_M(\theta,\theta')}{M} = \underbrace{\frac{1}{\overline{\omega}-\underline{\omega}} \int_{\underline{\omega}}^{\overline{\omega}} \left| \frac{K}{Tj\omega+1} - \frac{K'}{T'j\omega+1} \right|^2 d\omega}_{=:D(\theta,\theta')}. \tag{9}$$

In addition, it is easy to prove that, for the function $D(\theta,\theta')$ in (9), the condition **C2** holds for $\underline{\omega} < \overline{\omega}$. Therefore, the NLS estimate $\hat{\theta}^M$ of $\theta^\star$ is strongly consistent.

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

### *3.2 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 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 3.* A model set $\mathcal{G}$ is said to be *uniformly stable for a set* $\Theta$ if all the transfer functions in the set

$$\mathcal{G}(\Theta) := \{[G(\theta)](s) : \theta \in \Theta\} \tag{10}$$

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 evaluated at $\theta^\star$.

*Theorem 4.* Assume the following.

- $\hat{\theta}^M$ is a strongly consistent LS estimate of $\theta^\star$.
- For a given compact parameter set $\Theta$, the model set $\mathcal{G}(\Theta)$ is uniformly stable.
- $G(\theta)$ is smooth in $\Theta$.
- The true parameter $\theta^\star$ is in the interior of $\Theta$ in (10).
- Frequency points $\{\omega_m\,;\, m = 1,\ldots,M\}$ are distributed uniformly over a frequency range $[\underline{\omega}, \overline{\omega}]$ such that

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

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

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

Then, we have

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

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

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

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

The proof is omitted due to the space limitation, and will be presented in our companion paper (Nagamune and Choi (2008)). The error covariance matrix $W(\theta^\star)$ in (15) will play an important role in the parameter reduction step.

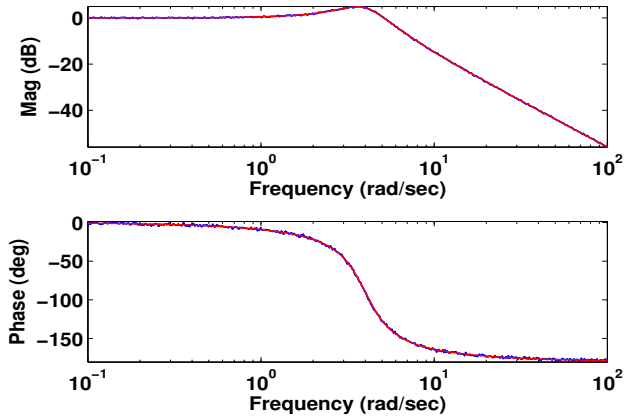


Fig. 1. Computer generated noisy FRF data for a second order system (blue-wiggly line) and its identified system (red-solid line).

*Remark 5.* If $\Sigma(\theta^\star)$ becomes singular, then the true parameter $\theta^\star$ is not identifiable, and we have to modify the parameterization.

*Remark 6.* The Fisher information matrix $\mathbb{I}(\theta^\star)$ (Kay (1993)) 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 (Kay (1993); Emery and Nenarokomov (1998)), 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 $\mathbb{I}(\theta^\star)$:

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

where $\mathbb{E}$ is an expected operator. Notice that this CRLB approaches to $W(\theta^\star)$ as $M$ increases.

*Remark 7.* 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 Emery and Nenarokomov (1998)). In practice, $[\underline{\omega}, \overline{\omega}]$ should contain all the modes of the dynamical system.

### *3.3 A numerical example on asymptotic normality*

Consider a second order dynamical system:

$$[G(\theta)](s) := \frac{K\omega^2}{s^2+2\zeta\omega s+\omega^2},\ \theta := [K,\zeta,\omega]^T. \tag{17}$$

Let the true parameters be $\theta^\star := [1,\ 0.3,\ 4]^T$.

FRFs are contaminated by complex-valued white noise $e_m$ resulting the property in (5) with $\sigma = 0.02$, and shown in Fig. 1. The number of frequency points is $M = 500$, and they are evenly distributed in log-scale over the fixed frequency interval $[0.1, 100]$.

An NLS algorithm with an initial point $\theta_0 = [20,\ 20,\ 20]^T$ converges to $\hat{\theta}^M = [0.99897,\ 0.30014,\ 3.99999]^T$ for computer generated FRFs. By approximating $\Sigma(\theta^\star)$ with

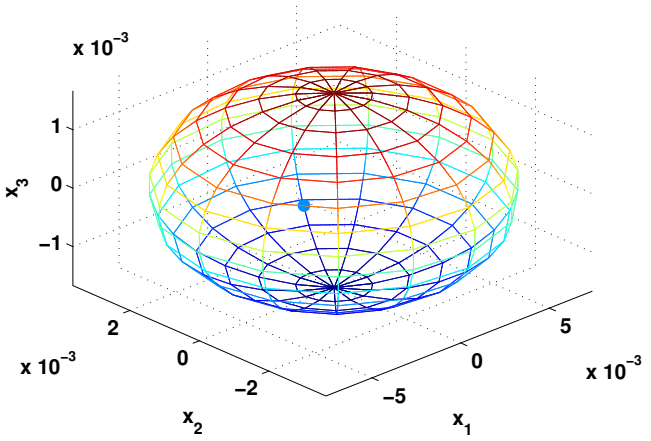


Fig. 2. A confidence region (ellipsoid) contains the error random variable (black-dot) with probability 0.971.

$\Sigma_M(\hat{\theta}^M)$, the estimation error covariance matrix $W$ is approximated as

$$\begin{aligned} W(\theta^\star) &\approx \frac{\sigma^2}{500}\Sigma_M(\hat{\theta}^M)^{-1}, \\ &= 10^{-5}\begin{bmatrix} 0.1383 & 0.0443 & 0.0369 \\ 0.0443 & 0.0525 & 0.0590 \\ 0.0369 & 0.0590 & 0.6392 \end{bmatrix}, \\ &=: VSV^T, \end{aligned} \tag{18}$$

which is a positive definite matrix. The matrix $V$ denotes the orthonormal coordinate transformation that contains eigenvectors of $W$ in (18). From the estimation error covariance in (18), an approximate confidence region with some probability is the ellipsoid defined by

$$(\hat{\theta} - \theta^\star)^T W^{-1}(\hat{\theta} - \theta^\star) \le \chi^2, \tag{19}$$

where $\chi^2$ is a value for the chi-squared distribution with three degrees of freedom. The 97.1% confidence region with $\chi^2 = 3^2$ is shown with respect to the orthonormal basis in Fig. 2. In this simulation experiment, the estimation error is

$$\hat{\theta} - \theta^\star = 10^{-3} \times [-1.0341,\ 0.1385,\ -0.0112]^T.$$

The error point with respect to the orthonormal basis is then

$$V^T(\hat{\theta} - \theta^\star) = 10^{-3} \times [0.0801,\ 0.9109,\ -0.5024]^T,$$

which is contained in the 97.1% confidence region in Fig. 2. In practice, to obtain the confidence region, we have to use prior knowledge of $\sigma$ or estimate it from samples.

## 4. 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 $G(\theta^\star)$ with single FRF data. In this section, by considering multiple true systems $G(\theta^\star_\ell)$, $\ell = 1, 2, \dots$, 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

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

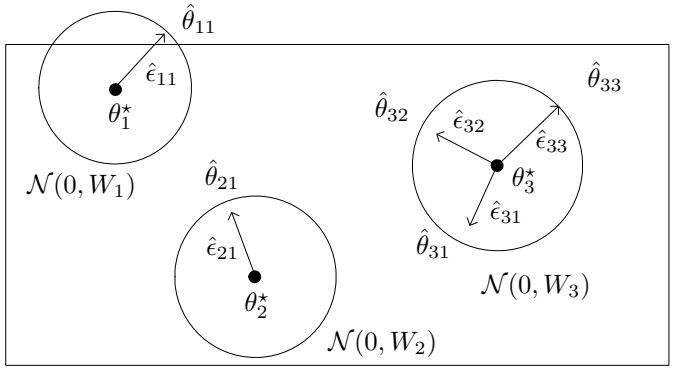


Fig. 3. Three samples of $\theta^\star_\ell$ are distributed in the square support of the probability density function for $\theta^\star_\ell$. For each sample of $\theta^\star_\ell$, there is an asymptotic normal distribution of its NLS estimates. Ellipsoids correspond to approximate confidence regions with some probability.

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 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.

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

$$\epsilon_{\ell k} := \hat{\theta}_{\ell k} - \theta^\star_\ell, \quad \ell = 1, 2, \dots,\ k = 1, 2, \dots. \tag{20}$$

By Theorem 4, for a fixed $\ell$, errors $\{\epsilon_{\ell k}; k = 1, 2, \dots\}$ are asymptotically normally distributed as $M \to \infty$:

$$\epsilon_{\ell k} \to_d \mathcal{N}(0, W_\ell),\ W_\ell := W(\theta^\star_\ell) = \frac{\sigma^2 \Sigma^{-1}(\theta^\star_\ell)}{M}. \tag{21}$$

Few samples from (21) for three true parameter vectors are illustrated in Fig. 3. For each true parameter vector $\theta^\star_\ell$, there is an asymptotic normal distribution of its NLS estimates.

Given a finite number of NLS estimates

$$\{\hat{\theta}_{\ell k} \in \mathbb{R}^p;\ \ell = 1, \dots, L,\ k = 1, \cdots, K\}, \tag{22}$$

where $p$ is the number of parameters, and the $\ell$-th asymptotic error covariances

$$\{W_\ell; \ell = 1, \dots, L\}, \tag{23}$$

the parameter reduction problem is to find a parameter set

$$\left\{\theta := \bar{\theta} + V\lambda; \lambda \in \mathbb{R}^q,\ |\lambda(i)| \le 1, i = 1, \dots, q\right\} \tag{24}$$

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 (22) in some sense. Next, we will provide a parameter reduction method based on the singular value decomposition.

*Remark 8.* We only consider an affine mapping from $\lambda$ to $\theta$ in (24). Such parameterization occurs in many robust control and adaptive control applications. The parameter reduction to fewer number of uncorrelated uncertainties in the solution (24) will significantly reduce the numerical burdens and complexity of the robust controller synthesis based on convex optimization. In the reduction, the NLS estimation covariance matrix $W_\ell$ will play a key role for

detecting the correlation among $\{\theta_\ell^\star\}$. In particular, obtaining uncorrelated parameter $\lambda$ from the identified systems with $\hat{\theta}_{\ell k}$ is a necessary pre-process for the synthesis of multiple robust controllers which involves a large number of iterations of convex optimization (Choi et al. (2006)).

In this paper, 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 $\mathbb{E}_\ell\left\{\lambda_\ell \lambda_\ell^T\right\} = \Lambda$ [3] as

$$\theta_\ell^\star = \bar{\theta} + V\lambda_\ell, \quad \ell = 1, 2, \ldots, \tag{25}$$

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.

Next, we will explain how to obtain the unknown parameters from the estimates and covariances, for the infinite sample case ($\ell = \infty$, $k = \infty$) and for the finite sample case ($\ell < \infty$, $k < \infty$).

### *4.1 Infinite sample case*

Although infinite samples are impossible in practice, the following theorem justifies the parameter reduction method which will be proposed in the next subsection for finite sample cases.

*Theorem 9.* In the case of infinite samples, the unknown parameters $\bar{\theta}$, $q$ and $V$ are obtained by

$$\bar{\theta} = \mathbb{E}_\ell\mathbb{E}_k\left\{\hat{\theta}_{\ell k}\right\} = \lim_{L\to\infty}\lim_{K\to\infty}\frac{1}{L}\frac{1}{K}\sum_{\ell=1}^{L}\sum_{k=1}^{K}\hat{\theta}_{\ell k}, \tag{26}$$

$$q = \mathrm{rank}(P - W), \tag{27}$$

$$P := \mathbb{E}_\ell\mathbb{E}_k\left\{(\hat{\theta}_{\ell k} - \bar{\theta})(\hat{\theta}_{\ell k} - \bar{\theta})^T\right\}, \tag{28}$$

$$W := \mathbb{E}_\ell\{W_\ell\}, \tag{29}$$

$$V = U(:, 1:q)\Sigma(1:q, 1:q)^{1/2}\Lambda^{-1/2} \in \mathbb{R}^{p\times q}, \tag{30}$$

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 $\Sigma$, and $P^{1/2}$ denotes a matrix square root of a positive definite matrix $P$.

We will prove this theorem. The estimation error (20) becomes

$$\epsilon_{\ell k} = \hat{\theta}_{\ell k} - (\bar{\theta} + V\lambda_\ell). \tag{31}$$

The nominal parameter $\bar{\theta} \in \mathbb{R}^p$ can be obtained by averaging both sides of (31), 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 (21),

$$\begin{aligned} P =& \mathbb{E}_\ell\mathbb{E}_k\left\{(V\lambda_\ell + \epsilon_{\ell k})(V\lambda_\ell + \epsilon_{\ell k})^T\right\}, \\ =& \mathbb{E}_\ell\left\{(V\lambda_\ell)(V\lambda_\ell)^T\right\} + \mathbb{E}_\ell\mathbb{E}_k\left\{\epsilon_{\ell k}\epsilon_{\ell k}^T\right\}, \\ =& V\Lambda V^T + W, \end{aligned} \tag{32}$$

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

$$V\Lambda V^T = P - W = U\Sigma U^T, \tag{33}$$

we can determine the reduced number of parameters $q$ as

$$q = \mathrm{rank}\Sigma = \mathrm{rank}(P - W), \tag{34}$$

and the matrix $V$ as in (30).

### *4.2 Finite sample case*

In practice, we only have a finite number of samples. For sample sets $\{\hat{\theta}_{\ell k} : \ell = 1, \ldots, L,\ k = 1, \ldots, K\}$ and $\{W_\ell : \ell = 1, \ldots, L\}$, the matrices $\bar{\theta}$, $P$ and $W$ in Theorem 9 can be approximated respectively by

$$\bar{\theta}^s := \frac{1}{L}\frac{1}{K}\sum_{\ell=1}^{L}\sum_{k=1}^{K}\hat{\theta}_{\ell k}, \tag{35}$$

$$P^s := \frac{1}{L}\frac{1}{K}\sum_{\ell=1}^{L}\sum_{k=1}^{K}(\hat{\theta}_{\ell k} - \bar{\theta}^s)(\hat{\theta}_{\ell k} - \bar{\theta}^s)^T, \tag{36}$$

$$W^s := \frac{1}{L}\sum_{\ell=1}^{L}W_\ell. \tag{37}$$

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 9, the approximations become better as the numbers of samples $L$ and $K$ increases.

### *4.3 A numerical example*

We illustrate the proposed parameter reduction method with an example, taken from the book (Chen et al., 2006, Ch. 11). Consider the following set of true system dynamics:

$$\mathcal{S} := \left\{G(s) = \prod_{m=1}^{5}\left[G_m(\delta)\right](s) : \delta \in [-0.2, 0.2]\right\}, \tag{38}$$

where

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

(Frequency is scaled by $10^{-4}$; see (Chen et al., 2006, eq.(11.4)).) For each of three $\delta = 0, \pm 0.2$, we took $K = 3$ noisy FRF data with noise variance $\sigma^2 = 0.01$.

[2] $\mathbb{E}_\ell$ is the expectation operator over $\ell$.

[3] The covariance matrix $\Lambda = \mathbb{E}_\ell\left\{\lambda_\ell\lambda_\ell^T\right\}$ 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 $\Lambda$ as

$$f_{\lambda(i)} = \frac{1}{2},\ \lambda(i) \in [-1, 1], \quad \Lambda = \frac{1}{3}I_q.$$

By regarding eight parameters as components of uncertain $\theta$, the NLS estimates were obtained as

$$\left\{\hat{\theta}_{\ell k} \in \mathbb{R} : \ell = 1, \ldots, 3,\ k = 1, \ldots, 3\right\}. \tag{39}$$

We also computed the approximated asymptotic error covariances:

$$\{W_\ell : \ell = 1, \ldots, 3\}. \tag{40}$$

Based on these estimates and covariances, we reduced the number of uncertain parameters $\lambda$ to one. The three largest singular values of $P^s - W^s$ are

$$0.1886, 0.0100, 0.0002.$$

We need to select $q$ in (34) 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 obtain the true value of $q$, which is one in this example.

Here, we selected $q = 1$ and performed parameter reduction. In Figure 4, it is shown the noisy FRF data (blue lines), and Bode plots of transfer functions obtained by optimally perturbing one uncertain parameter $\lambda$ (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.

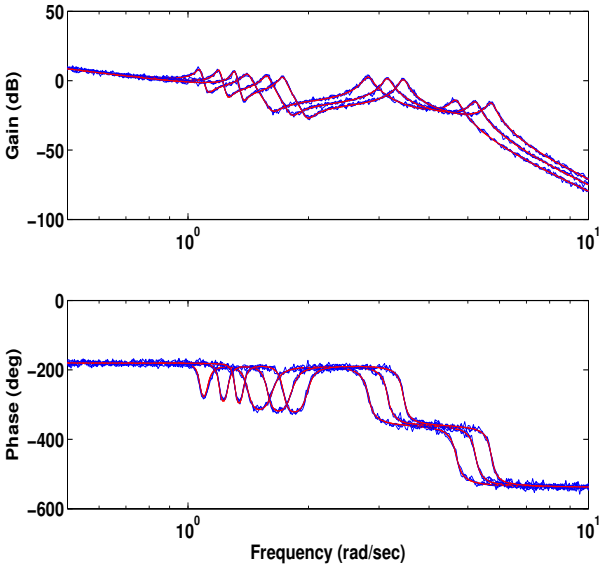


Fig. 4. The noisy FRF data (blue lines), and Bode plots of transfer functions obtained by optimally perturbing one uncertain parameter $\lambda$ (red lines).

## 5. CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed a 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 the singular value decomposition to detect the correlation of original parameters. The basic idea of the proposed reduction method is based on the principal component analysis (PCA). Our main contribution is the proposed parameter reduction technique based on PCA and its asymptotic convergence results based on the statistical properties of NLS estimates.

The essential necessary assumption in this paper is that *we know the structure of the true system*, which is not realistic 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.

## REFERENCES

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

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

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

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

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*.

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

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

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

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

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

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

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

R. Nagamune and J. Choi, "Parameter reduction of nonlinear least-square estimates via nonconvex optimization," to appear in *2008 American Control Conference*.

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

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

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.

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

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

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