Linköping University Post Print

Nonlinear Gray-Box Identification Using Local Models Applied to Industrial Robots

Erik Wernholt and Stig Moberg

N.B.: When citing this work, cite the original article.

Original Publication:

Erik Wernholt and Stig Moberg, Nonlinear Gray-Box Identification Using Local Models Applied to Industrial Robots, 2011, Automatica, (47), 4, 650-660.

<http://dx.doi.org/10.1016/j.automatica.2011.01.021>

Copyright: Elsevier Science B.V., Amsterdam.

<http://www.elsevier.com/>

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-61670>

Nonlinear Gray-Box Identification Using Local Models Applied to Industrial Robots[★]

Erik Wernholt^a, Stig Moberg^{a,b}

^a*Division of Automatic Control, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden.*

^b*ABB AB – Robotics, SE-721 68 Västerås, Sweden.*

Abstract

In this paper, we study the problem of estimating unknown parameters in nonlinear gray-box models that may be multivariable, nonlinear, unstable, and resonant at the same time. A straightforward use of time-domain predication-error methods for this type of problem easily ends up in a large and numerically stiff optimization problem. We therefore propose an identification procedure that uses intermediate local models that allow for data compression and a less complex optimization problem. The procedure is based on the estimation of the nonparametric frequency response function (FRF) in a number of operating points. The nonlinear gray-box model is linearized in the same operating points, resulting in parametric FRFs. The optimal parameters are finally obtained by minimizing the discrepancy between the nonparametric and parametric FRFs. The procedure is illustrated by estimating elasticity parameters in a six-axes industrial robot. Different parameter estimators are compared and experimental results show the usefulness of the proposed identification procedure. The weighted logarithmic least squares estimator achieves the best result and the identified model gives a good global description of the dynamics in the frequency range of interest for robot control.

Key words: System identification; multivariable systems; nonlinear systems; closed-loop identification; frequency response methods; industrial robots.

1 Introduction

When building a mathematical model of a physical object, the user generally has two sources of information; prior knowledge and experimental data. This gives the two modeling extremes, *white-box models* that are the result of extensive physical modeling from first principles, and *black-box models*, where the model is just a vehicle to describe the experimental data without any physical interpretations of its parameters. In between comes gray-box models that are parameterizations based on various degrees of physical insights. Compared with identification methods using black-box models, gray-box models have some particular benefits. By including prior information, the model set (number of parameters) can be reduced while still providing a good approximation of

the true system. This gives a reduced mean-square error (bias and variance) (Ljung, 2008). Since the gray-box model in contrast to the black-box model has a physical interpretation, the model is also useful in many ways, such as design optimization and virtual prototyping, which are important areas in industry. Some standard references for identification of gray-box models are Ljung (1999), Bohlin (2006), and Schittkowski (2002). There are also software packages available for identifying such models, e.g., Ljung (2007), Kristensen et al. (2004), Bohlin (2006), and Schittkowski (2002).

In this paper, we will mainly consider identification of nonlinear gray-box models with the lightest shade of gray, where a white-box model contains some unknown or uncertain parameters that need to be estimated from data. To be even more specific, we will allow our systems to be *multivariable*, *nonlinear*, *unstable*, and *resonant* at the same time. Usually, in the literature, at least one of the first three properties is left out to reduce the problem complexity. Identification of such complex systems is therefore challenging, both in finding suitable model structures and efficient identification methods. A straightforward use of time-domain predication-

[★] Parts of this paper were presented at the 17th IFAC World Congress in Seoul, 2008. This work was supported by ABB AB – Robotics and the VINNOVA Industry Excellence Center LINK-SIC at Linköping University. Corresponding author: E. Wernholt. Tel. +46-706-326913. Fax +46-13-282622.

Email addresses: erikw@isy.liu.se (Erik Wernholt), stig.moberg@se.abb.com (Stig Moberg).

error methods (Ljung, 1999) for this type of system easily ends up in a large optimization problem, where each iteration in the optimization routine involves a number of simulations of a large and numerically stiff ODE for many samples. We therefore propose an identification procedure that uses intermediate local models that allow for data compression and a less complex optimization problem. The procedure is based on the estimation of the nonparametric frequency response function (FRF) in a number of operating points. The nonlinear gray-box model is linearized in the same operating points, resulting in parametric FRFs. The optimal parameters are finally obtained by minimizing a parametric criterion, measuring the discrepancy between the nonparametric and parametric FRFs.

The main motivation for our study is identification of accurate dynamic models for industrial robots, which incorporate all the four mentioned properties. We will come back to this application in Sections 5 and 6 when identification of elasticity parameters (spring-damper pairs) is used as an example of the proposed identification procedure.

Various aspects of the identification procedure will now be explained in the coming sections, starting with a more detailed outline of the procedure in Section 2. The nonparametric FRF estimation is described in Section 3, and some variants of the parametric criterion are presented in Section 4. This is followed by an experimental evaluation in Section 5 and a small comparison with time domain identification in Section 6. Finally, Section 7 concludes the paper.

2 Gray-Box Identification Using Local Models

In this paper we consider the problem of identifying parameters θ in the following nonlinear gray-box model

$$x_{t+1} = f(x_t, u_t, \theta) + g(x_t, u_t, \theta)w_t, \quad (1a)$$

$$y_t = h(x_t, u_t, \theta) + e_t, \quad (1b)$$

with state vector $x_t \in \mathbb{R}^{n_x}$, input vector $u_t \in \mathbb{R}^{n_u}$, output vector $y_t \in \mathbb{R}^{n_y}$, and $\theta \in \Theta \subset \mathbb{R}^{n_\theta}$ a vector of unknown parameters that specifies the mappings $f(\cdot)$, $g(\cdot)$, $h(\cdot)$ that may all be nonlinear. Furthermore, w_t and e_t are process- and measurement noise vectors, that are assumed to be mutually independent zero-mean white processes with covariance matrices $E(w_t w_t^T) = Q_w$ and $E(e_t e_t^T) = Q_e$.

One solution to the problem of identifying θ in (1) is to apply a nonlinear prediction error method (Ljung, 1999, pp. 146–147). The idea then is to find the parameters that will minimize the prediction errors

$$\varepsilon_{t,\theta} = y_t - \hat{y}_{t,\theta}, \quad t = 1, 2, \dots, N_{\text{tot}},$$

where $\hat{y}_{t,\theta}$ is the model's prediction of y_t given previous measurements $\{u_1, y_1, \dots, u_{t-1}, y_{t-1}, u_t\}$. For the minimization, one can choose different norms, but a common choice is a quadratic criterion

$$V(\theta) = \sum_{t=1}^{N_{\text{tot}}} \varepsilon_{t,\theta}^T \Lambda_t^{-1} \varepsilon_{t,\theta}, \quad (2)$$

with weighting matrix Λ_t and N_{tot} equal to the total number of measurement samples. For a general nonlinear system as in (1), it is often very hard to determine a predictor on formal probabilistic grounds. In most cases there is no explicit form available for the optimal solution. This implies that a predictor must be constructed either by *ad hoc* approaches, or by some approximation of the optimal solution, e.g., by using sequential Monte Carlo techniques (Doucet et al., 2001) or the Extended Kalman Filter (Anderson and Moore, 1979). We will further discuss such solutions in Section 2.4.

In this paper, we propose another solution to handle the nonlinearities in the system, where θ is identified by using intermediate local models. By performing experiments that sequentially excite the local system behavior in a number of operating points $(u_0^{(i)}, y_0^{(i)})$, $i = 1, \dots, Q$, the criterion (2) can be approximated by

$$V(\theta) = \sum_{i=1}^Q \sum_{t=1}^{N_{\text{tot},i}} [\varepsilon_{t,\theta}^{(i)}]^T [\Lambda_t^{(i)}]^{-1} \varepsilon_{t,\theta}^{(i)}, \quad (3)$$

using the notation $\varepsilon_{t,\theta}^{(i)} = y_t^{(i)} - \hat{y}_{t,\theta}^{(i)}$ where $y_t^{(i)} = y_{t+t_i} - y_0^{(i)}$ denotes the measured output around operating point i with output vector $y_0^{(i)}$. The N_{tot} samples in (2) are therefore distributed according to Fig. 1 (top part), where $N_{\text{tot},i,0}$ samples are needed to move the system to operating point i , in which $N_{\text{tot},i}$ samples are collected, such that $t_i = N_{\text{tot},i,0} + \sum_{r=1}^{i-1} (N_{\text{tot},r} + N_{\text{tot},r,0})$ and $N_{\text{tot}} = \sum_{i=1}^Q (N_{\text{tot},i} + N_{\text{tot},i,0})$.

2.1 Linearized Gray-Box Model

The predictor $\hat{y}_{t,\theta}^{(i)}$ for the local model can be simplified compared to $\hat{y}_{t,\theta}$ since only the local (linear) behavior needs to be captured. Assume now that the behavior of (1) around an operating point $(u_0^{(i)}, y_0^{(i)})$ can be described by the linearized model

$$x_{t+1}^{(i)} = A_\theta^{(i)} x_t^{(i)} + B_\theta^{(i)} u_t^{(i)} + B_{w,\theta}^{(i)} w_t^{(i)}, \quad (4a)$$

$$y_t^{(i)} = C_\theta^{(i)} x_t^{(i)} + D_\theta^{(i)} u_t^{(i)} + e_t^{(i)}, \quad (4b)$$

where $x_t^{(i)} = x_{t+t_i} - x_0^{(i)}$, $u_t^{(i)} = u_{t+t_i} - u_0^{(i)}$, $y_t^{(i)} = y_{t+t_i} - y_0^{(i)}$, and $x_0^{(i)}$ is obtained as, e.g., the solution to

Fig. 1. Illustration how the N_{tot} samples are distributed into Q operating points, where $N_{\text{tot},i,0}$ samples are “wasted” to get to operating point i where the $N_{\text{tot},i}$ samples are further distributed into n_u parts to estimate a nonparametric FRF by using the n_u steady-state responses, $u_t^{(i,1)}, y_t^{(i,1)}, \dots, u_t^{(i,n_u)}, y_t^{(i,n_u)}$, each of length N_i .

$x_0^{(i)} = f(x_0^{(i)}, u_0^{(i)}, \theta)$, $y_0^{(i)} = h(x_0^{(i)}, u_0^{(i)}, \theta)$. Note that $x_0^{(i)}$ therefore, in general, will be θ -dependent. The matrices $A_\theta^{(i)}, B_\theta^{(i)}, C_\theta^{(i)}, D_\theta^{(i)}$ are the partial derivatives of $f(\cdot)$ and $h(\cdot)$ w.r.t. x_t and u_t and evaluated in $(x_0^{(i)}, u_0^{(i)})$, and $B_{w,\theta}^{(i)} = g(x_0^{(i)}, u_0^{(i)}, \theta)$.

By assuming w_t and e_t to be Gaussian and neglecting transients, the optimal one-step ahead predictor $\hat{y}_{t,\theta}^{(i)}$ for (4) is given by the steady-state Kalman filter

$$\hat{x}_{t+1}^{(i)} = A_\theta^{(i)} \hat{x}_t^{(i)} + B_\theta^{(i)} u_t^{(i)} + K_\theta^{(i)} (y_t^{(i)} - \hat{y}_{t,\theta}^{(i)}), \quad (5a)$$

$$\hat{y}_{t,\theta}^{(i)} = C_\theta^{(i)} \hat{x}_t^{(i)} + D_\theta^{(i)} u_t^{(i)}. \quad (5b)$$

Here, $K_\theta^{(i)} = A_\theta^{(i)} P_\theta^{(i)} [C_\theta^{(i)}]^T [C_\theta^{(i)} P_\theta^{(i)} [C_\theta^{(i)}]^T + Q_e]^{-1}$ where $P_\theta^{(i)}$ is the positive semi-definite solution of the stationary Riccati equation (omitting indices)

$$P = A P A^T - A P C^T (C P C^T + Q_e)^{-1} C P A^T + B_w Q_w B_w^T.$$

By using (5), the prediction error $\varepsilon_{t,\theta}^{(i)}$ can be written as

$$\varepsilon_{t,\theta}^{(i)} = [H_\theta^{(i)}(q)]^{-1} [y_t^{(i)} - G_\theta^{(i)}(q) u_t^{(i)}], \quad (6)$$

where q is the difference operator, $u_{t+1} = q u_t$, and

$$G_\theta^{(i)}(q) = C_\theta^{(i)} [qI - A_\theta^{(i)}]^{-1} B_\theta^{(i)} + D_\theta^{(i)}, \quad (7a)$$

$$H_\theta^{(i)}(q) = C_\theta^{(i)} [qI - A_\theta^{(i)}]^{-1} K_\theta^{(i)} + I. \quad (7b)$$

In case the measurement data around operating point i can be described by (4) with $\theta = \theta_0$, the prediction error $\varepsilon_{t,\theta_0}^{(i)}$ is the innovation with covariance matrix $\Lambda_0^{(i)} = C_{\theta_0}^{(i)} P_{\theta_0}^{(i)} [C_{\theta_0}^{(i)}]^T + Q_e$. The weights $\Lambda_t^{(i)}$ in (3) should therefore be close to $\Lambda_0^{(i)}$.

2.2 Intermediate Local Models

In this paper, we will not use (3) and (6) directly for the parameter estimation, but instead capture and compress the local behavior by an intermediate estimation of the nonparametric (FRF) in each operating point. There are

several reasons for this, which will be outlined in Section 2.4. The criterion (3) will now be approximated and rewritten to show how the nonparametric FRF can be estimated and used for the parameter estimation (c.f. (11) and (12)). For simplicity, we will assume steady-state response data to avoid transients and initial conditions.

To handle multivariable systems, the $N_{\text{tot},i}$ samples in operating point i are further split into n_u parts in a similar way as in Ljung (2003). Each part contains $N_{i,0}$ transient samples followed by N_i samples of the steady-state response (assuming a periodic excitation), such that $N_{\text{tot},i} = n_u(N_i + N_{i,0})$. For details, see Fig. 1. By neglecting the transient samples and using constant weights $\Lambda_t^{(i)} = \Lambda^{(i)}$, (3) is approximated by

$$V(\theta) = \sum_{i=1}^Q \sum_{l=1}^{n_u} \sum_{t=1}^{N_i} [\varepsilon_{t,\theta}^{(i,l)}]^T [\Lambda^{(i)}]^{-1} \varepsilon_{t,\theta}^{(i,l)}, \quad (8)$$

where $\varepsilon_{t,\theta}^{(i,l)} = [H_\theta^{(i)}(q)]^{-1} [y_t^{(i,l)} - G_\theta^{(i)}(q) u_t^{(i,l)}]$ is the prediction error in operating point i for data part l using the model (6). By using Parseval's relationship and assuming periodic data, (8) can be rewritten as

$$V(\theta) = \sum_{i=1}^Q \sum_{l=1}^{n_u} \sum_{k=1}^{N_i} [X_{k,\theta}^{(i,l)}]^H [\Phi_{k,\theta}^{(i)}]^{-1} X_{k,\theta}^{(i,l)}, \quad (9)$$

with $(\cdot)^H$ denoting complex conjugate transpose,

$$X_{k,\theta}^{(i,l)} = Y_k^{(i,l)} - G_{k,\theta}^{(i)} U_k^{(i,l)}, \quad \Phi_{k,\theta}^{(i)} = H_{k,\theta}^{(i)} \Lambda^{(i)} [H_{k,\theta}^{(i)}]^H,$$

and using the short notation

$$G_{k,\theta}^{(i)} = G_\theta^{(i)}(e^{j\omega_k}), \quad H_{k,\theta}^{(i)} = H_\theta^{(i)}(e^{j\omega_k}).$$

$Y_k^{(i,l)}$ and $U_k^{(i,l)}$ are the discrete Fourier transform (DFT) of the output and input signals,

$$Y_k^{(i,l)} = N_i^{-1/2} \sum_{t=1}^{N_i} y_t^{(i,l)} e^{j\omega_k t}, \quad \omega_k = \frac{2\pi k}{N_i},$$

and similarly for $U_k^{(i,l)}$. The criterion (9) can be more

compactly written by using the matrix trace as

$$V(\theta) = \sum_{i=1}^Q \sum_{k=1}^{N_i} \text{tr}([\mathbf{X}_{k,\theta}^{(i)}]^H [\Phi_{k,\theta}^{(i)}]^{-1} \mathbf{X}_{k,\theta}^{(i)}) \quad (10)$$

where $\mathbf{X}_{k,\theta}^{(i)} = \mathbf{Y}_k^{(i)} - G_{k,\theta}^{(i)} \mathbf{U}_k^{(i)}$ and the n_u DFT vectors are collected into matrices

$$\begin{aligned} \mathbf{U}_k^{(i)} &= [U_k^{(i,1)} \ U_k^{(i,2)} \ \dots \ U_k^{(i,n_u)}], \\ \mathbf{Y}_k^{(i)} &= [Y_k^{(i,1)} \ Y_k^{(i,2)} \ \dots \ Y_k^{(i,n_u)}]. \end{aligned}$$

If the input DFT matrix $\mathbf{U}_k^{(i)}$ has full rank, the *nonparametric FRF estimate* can be calculated as

$$\hat{G}_k^{(i)} = \mathbf{Y}_k^{(i)} [\mathbf{U}_k^{(i)}]^{-1}. \quad (11)$$

The criterion (10) can now finally be rewritten once more to show how (11) can be used for identification:

$$\begin{aligned} V(\theta) &= \sum_{i=1}^Q \sum_{k=1}^{N_i} \text{tr}([\tilde{G}_{k,\theta}^{(i)} \mathbf{U}_k^{(i)}]^H [\Phi_{k,\theta}^{(i)}]^{-1} \tilde{G}_{k,\theta}^{(i)} \mathbf{U}_k^{(i)}) \\ &= \sum_{i=1}^Q \sum_{k=1}^{N_i} \text{vec}(\tilde{G}_{k,\theta}^{(i)})^H [W_{k,\theta}^{(i)}]^{-1} \text{vec}(\tilde{G}_{k,\theta}^{(i)}), \end{aligned} \quad (12)$$

with

$$\tilde{G}_{k,\theta}^{(i)} = \hat{G}_k^{(i)} - G_{k,\theta}^{(i)}, \quad (13)$$

and

$$W_{k,\theta}^{(i)} = [\mathbf{U}_k^{(i)} [\mathbf{U}_k^{(i)}]^H]^{-T} \otimes \Phi_{k,\theta}^{(i)}. \quad (14)$$

Here, $\otimes$ is the Kronecker product and (12) is obtained by using $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$, $\text{tr}(AB) = \text{tr}(BA)$, and $\text{tr}(ABCD) = \text{vec}(D^T)^T (C^T \otimes A) \text{vec}(B)$. The operator $\text{vec}(B) = [b_1^T b_2^T \dots b_n^T]^T$ stacks the columns b_i of a matrix B into a long column vector.

The new criterion (12) now measures the (weighted) discrepancy (13) between the nonparametric FRF estimate $\hat{G}_k^{(i)}$ and the parametric FRF $G_{k,\theta}^{(i)}$, obtained by linearizing (1), for all the Q different operating points. To be noted is that the weighting matrix (14) can be seen as the covariance expression for the nonparametric FRF estimate (11), under certain assumptions on the measurements. See Section 3 for more details.

2.3 Proposed Identification Procedure

To summarize, the proposed identification procedure is given by the following three steps:

- (1) Collect experimental data from Q operating points.
- (2) Estimate intermediate local models for each operating point, in this paper nonparametric FRFs $\hat{G}_k^{(i)}$, $i = 1, \dots, Q$.

- (3) Obtain the optimal parameter vector $\hat{\theta}$ by minimizing the discrepancy $\tilde{G}_{k,\theta}^{(i)} = \hat{G}_k^{(i)} - G_{k,\theta}^{(i)}$ between $\hat{G}_k^{(i)}$ and the parametric FRF, $G_{k,\theta}^{(i)}$, for all excited frequencies and operating points.

This identification procedure was first suggested in Öhr et al. (2006) and various aspects of the procedure are further analyzed in Wernholt (2007) and in the cited references by Wernholt and co-authors in this paper.

Remark 1 Note that the parametric FRF, $G_{k,\theta}^{(i)}$, is a function of the nonlinear gray-box model (1) such that for each parameter vector θ during the minimization, (1) is linearized according to Section 2.1 before calculating the FRF $G_{k,\theta}^{(i)} = G_\theta^{(i)}(e^{j\omega_k})$ with $G_\theta^{(i)}(q)$ from (7a).

2.4 Discussion

As was previously mentioned, it is common to use prediction-error methods for system identification. One simple (and often used) way of deriving an *ad hoc* predictor for a nonlinear system is to disregard the process noise w_t in (1). The predictor is then actually a simulation of the noise-free model. However, this only works if the model is stable, since a stable predictor always is required. Another option is to apply the ideas in Forsell and Ljung (2000), where the predictor is modified such that an output error model can be used also for unstable systems. This is appealing since the user often is mainly interested in modeling the dynamics of the plant, the noise properties are less interesting. These ideas are also exploited in Larsson et al. (2009), where the nonlinear dynamics of an unstable fighter aircraft is identified, using the Extended Kalman Filter (Anderson and Moore, 1979). See also the work done on stochastic gray-box identification by, e.g., Bohlin (2006) and Kristensen et al. (2004). However, a drawback with all these time-domain solutions is that they are numerically intensive and sensitive for our setup due to a resonant, numerically stiff and large multivariable system (many states and parameters). In each iteration of the optimization routine, a large number of simulations of this stiff system must be carried out to obtain the gradient, and many samples are needed in order to capture the resonant behavior.

One important feature of using intermediate models therefore is that they enable data compression, which is also emphasized in, e.g., Gawthrop and Wang (2005). Thousands of data points can be compressed into an FRF with a few excited frequencies, which reduces computations in the nonlinear gray-box identification substantially (see Section 6 for an example). It is important to acknowledge that the idea of estimating intermediate models is not new and is commonly used in, e.g., modal analysis where usually a linear parametric model is

fitted to an estimated FRF (Behi and Tesar, 1991; Verboven et al., 2005). Another related result is presented in Johansen and Foss (1995), where local black-box models are combined into a global model through interpolation. However, such a model will have very limited extrapolation features since it is mainly a way of obtaining an accurate nonlinear black-box model in the manifold spanned by the local models. The difference in our setup is that the prior information incorporated into the nonlinear gray-box model (1) makes it probable that, once the parameters have been properly estimated, the model will give an accurate description of the behavior in a fairly large range of the state space. The parameters of this global model are however estimated by only using a few intermediate local models.

Other features of the proposed identification procedure, besides data compression, is that unstable output-error models works fine in the frequency domain as long as the input and output signals are bounded (which is handled by closed-loop experiments). In the time domain, this is much more involved (Forssell and Ljung, 2000). In the frequency domain, it is also easy to validate the model such that all important resonances are captured, and model requirements in the frequency domain are easily handled.

The proposed procedure also has some possible problems. Local minima in the optimization give problems here as well as for prediction error methods. In addition comes some difficulties with biased nonparametric FRF estimates due to closed-loop data and nonlinearities. The bias due to closed-loop data is sometimes negligible and can also be reduced by proper experiment design to increase the signal-to-noise ratio, such as periodic excitation (see Section 3). Another possible problem is if a linear model is unable to describe the dynamics in an operating point even for small perturbations due to the nonlinearities. This can be the case for certain nonlinearities such as friction or backlash, but the problem can often be reduced by the choice of excitation (e.g., avoid zero velocity to reduce the effects due to Coulomb friction). Identifiability can also be lost due to the linearization.

3 FRF Estimation

As an intermediate step toward the parameter identification, estimates of the nonparametric FRF (11) in a number of operating points are needed. In the literature (see, e.g., Guillaume, 1998; Pintelon and Schoukens, 2001; Wernholt and Moberg, 2008a) various estimation methods exist that differ in requirements on the measurement setup and signal-to-noise ratio, as well as in bias and variance properties. Here we will mainly consider the simplest version (11) due to its connection to the prediction error criterion.

3.1 Properties of Nonparametric FRF Estimates

Assume now that N samples of data are generated from a noise-corrupted linear system $G(q)$ in open loop

$$y_t = G(q)u_t + v_t.$$

For a scalar input u_t , the nonparametric FRF estimate $\hat{G}_k = Y_k/U_k$ (also called Empirical Transfer Function Estimate, ETFE), with U_k and Y_k the DFTs of the input and output signals, has the following statistical properties (see Lemma 6.1 in Ljung, 1999):

$$E(\hat{G}_k) = G_k + \frac{\rho_1}{\sqrt{N}U_k}, \quad \text{Cov}(\hat{G}_k) = \frac{\Phi_k}{|U_k|^2} + \frac{\rho_2}{N|U_k|^2}.$$

Here, $G_k = G(e^{j\omega_k})$, Φ_k is the spectrum of the additive noise v_t , and ρ_i are constant bounds that depend on the system impulse response, maximum input amplitude, and the covariance function of the noise. Moreover, it can be shown that the FRF estimates at different frequencies are asymptotically ($N \rightarrow \infty$) uncorrelated. In general, the FRF estimate will therefore be asymptotically unbiased with a variance equal to a noise-to-signal ratio $\Phi_k/|U_k|^2$, usually resulting in a very crude estimate that must be smoothed over neighboring frequencies to reduce the random fluctuations. However, these properties are drastically different if the data is *periodic*, for example $N = N_p P$ with P an integer number of periods and N_p samples in each period, such that $u_t = u_{t+N_p}$. In that case, $\rho_1 = 0$ and $|U_k|^2$ increases linearly with P for the excited frequencies $\omega_k = 2\pi k/N_p$. This gives an unbiased estimate and the variance decays like $1/P$ (assuming that v_t contains no pure sinusoids at the excited frequencies such that Φ_k is approximately constant when P is increased).

Similar results can be obtained for the multi-input case, stating that $\hat{G}_k = \mathbf{Y}_k \mathbf{U}_k^{-1}$ is unbiased for periodic data and with a covariance matrix given by (cf. (14))

$$\begin{aligned} \text{Cov}(\hat{G}_k) &= E(\text{vec}(\hat{G}_k - G_k)[\text{vec}(\hat{G}_k - G_k)]^H) \\ &= [\mathbf{U}_k \mathbf{U}_k^H]^{-T} \otimes \Phi_k, \end{aligned}$$

where we have neglected a term similar to $\rho_2/(N|U_k|^2)$ above. For details, see for example Theorem 8.2.5 in Brillinger (1981) or Wernholt and Gunnarsson (2007).

If the open-loop system to be identified is unstable, it is necessary to collect data in closed loop. Consider there-

Fig. 2. Closed-loop measurement setup.

fore the setup in Fig. 2, where the controller takes as input the difference between the reference signal r and the measured and sampled output y , and u is the input. The additive noise v contains various sources of noise and disturbances. The nonparametric FRF estimate then usually gets biased due to the correlation between u and v . Assuming a linear plant and controller,

$$y_t = G(q)u_t + v_t, \quad u_t = F(q)(r_t - y_t),$$

and the same number of inputs and outputs $n_y = n_u = n$, the estimate is given by

$$\begin{aligned} \hat{G}_k &= \mathbf{Y}_k \mathbf{U}_k^{-1} = G_k + \mathbf{V}_k \mathbf{U}_k^{-1} \\ &= G_k + \mathbf{V}_k (\mathbf{R}_k - \mathbf{V}_k)^{-1} (G_k + F_k^{-1}), \end{aligned}$$

where $\mathbf{R}_k$ and $\mathbf{V}_k$ are the DFT matrices of the reference and measurement noise signals, similarly as for $\mathbf{Y}_k$ and $\mathbf{U}_k$. It is then hard to calculate the bias due to the matrix inverse, but it is approximately proportional to the noise-to-signal ratio ($\Phi_k/|R_k|^2$ for $n = 1$). Fortunately, the signal-to-noise ratio can be made arbitrarily large at the excited frequencies by using a periodic excitation. The FRF estimate will therefore be asymptotically unbiased (as the number of periods, $P \rightarrow \infty$) at the excited frequencies also in closed loop. For details, see Wernholt and Gunnarsson (2007).

3.2 Excitation Signals

As has already been stressed, the quality of the nonparametric FRF estimate is highly dependent on the excitation where $\mathbf{U}_k$ must be invertible and periodic data is desirable. We will therefore assume periodic data with $N_i = N_p P$, where P is an integer number of periods (important to avoid leakage effects in the DFTs) and N_p is the number of samples in each period.

For multivariable systems, the *orthogonal random phase multisine* signal has been proposed by Dobrowiecki and Schoukens (2007) to minimize the FRF uncertainty in open loop, given certain input amplitude constraints. Here, this signal will be used in closed loop, which corresponds to an optimal experiment design with output amplitude constraints. The excitation signal in operating point i is then given by $\mathbf{R}_k^{(i)} = \mathbf{R}_{d,k}^{(i)} \mathbf{T}$, where $\mathbf{T}$ is an orthogonal matrix, $\mathbf{T}_{il} = e^{\frac{2\pi j}{n}(i-1)(l-1)}$, with $\mathbf{T}\mathbf{T}^H = n\mathbf{I}$. $\mathbf{R}_{d,k}$ is a diagonal matrix where each diagonal element is the DFT of a random phase multisine signal, which in the time domain can be written as

$$r_t = \sum_{k=1}^{N_f} A_k \cos(\omega_k t + \phi_k),$$

with N_f the number of excited frequencies (limited by $N_f \leq N_p/2$ assuming periodic excitation and N_p even),

amplitudes A_k , frequencies ω_k chosen from the grid $\{\omega_k = 2\pi k/N_p, k = 1, \dots, N_p/2\}$, and random phases ϕ_k uniformly distributed between 0 and 2π . The selection of frequencies as well as the amplitude spectrum will affect the parameter estimation in the next step. Using too many frequencies for a given total excitation power will give a low signal-to-noise ratio, which increases both the bias and the variance in the nonparametric FRF estimate. The amplitude spectrum should also reflect the parameter sensitivity (cf. $\Psi_{0,k}^{(i)}$ in (16)), such that θ influence $G_{k,\theta}^{(i)}$ for the selected frequencies.

The nonparametric FRF estimate can be improved by collecting more experimental data and average over multiple blocks (by performing $n_u M$ experiments in each operating point, with M an integer number of blocks) and/or periods. The covariance matrix can then also be estimated. For a linear system, averaging over different periods is sufficient, whereas for a nonlinear system it can be essential to average over blocks where $\mathbf{R}_k^{(i)}$ in each block have different realizations of the random phases. The reason is that nonlinearities otherwise might distort the estimate and give a too low uncertainty estimate. See Schoukens et al. (2005) for details and Wernholt and Gunnarsson (2008) for experimental results.

3.3 Optimal Operating Points

Given a nonlinear gray-box model (1), the information about the unknown parameters will vary depending on the operating point. Therefore, given a limited total measurement time, one should perform experiments around the operating point(s) that contribute the most to the information about the unknown parameters. In Wernholt and Löffberg (2007), this problem is formulated as a convex optimization problem by using a set of candidate operating points, obtained by gridding the workspace. It is also shown that the experiment design is efficiently solved by considering the dual problem. Given thousands of candidates, only a few operating points are selected in the optimum. See Wernholt and Löffberg (2007) for details and examples.

4 Parameter Estimation

When the nonparametric FRFs have been estimated from data, the last step is to linearize (1) in the same operating points, calculate the parametric FRFs, $G_{k,\theta}^{(i)}$, $i = 1, \dots, Q$, and estimate the unknown parameters by minimizing a cost function $V(\theta)$, for example (12). First, two different parameter estimators will be analyzed and compared. Next, the selection of weights and the solution of the optimization problem are discussed.

4.1 Weighted Nonlinear Least Squares

Minimizing (12) for the N_f excited frequencies gives the *weighted nonlinear least squares (NLS) estimator*:

$$\hat{\theta}_{N_f}^{\text{NLS}} = \arg \min_{\theta \in \Theta} V_{N_f}^{\text{NLS}}(\theta), \quad (15a)$$

$$V_{N_f}^{\text{NLS}}(\theta) = \sum_{i=1}^Q \sum_{k=1}^{N_f} [\mathcal{E}_{k,\theta}^{(i)}]^H [W_k^{(i)}]^{-1} \mathcal{E}_{k,\theta}^{(i)}, \quad (15b)$$

$$\mathcal{E}_{k,\theta}^{(i)} = \text{vec}(\hat{G}_k^{(i)}) - \text{vec}(G_{k,\theta}^{(i)}), \quad (15c)$$

where $W_k^{(i)}$ is assumed to be a Hermitian ($W = W^H$) weighting matrix. The asymptotic properties ($N_f \rightarrow \infty$) of this estimator are given in the following result under certain assumptions:

- (1) The nonparametric FRF is unbiased and given by $\hat{G}_k^{(i)} = G_{0,k}^{(i)} + \eta_k^{(i)}$ with $\text{vec}(\eta_k^{(i)})$ a zero mean circular complex random vector, independent over i and k , with finite covariance matrix $W_{0,k}^{(i)}$.
- (2) The true system is in the model class: $G_{0,k}^{(i)} = G_{k,\theta_0}^{(i)}$.
- (3) Θ is a compact set where $V_{N_f}^{\text{NLS}}(\theta)$ and its first- and second-order derivatives are continuous for any value of N_f .
- (4) For N_f large enough, the expected value of $V_{N_f}^{\text{NLS}}(\theta)$ has a unique global minimum θ_0 in Θ .

Result 2 *The NLS estimator (15) is consistent ($\lim_{N_f \rightarrow \infty} \hat{\theta}_{N_f}^{\text{NLS}} = \theta_0$), and $\sqrt{N_f}(\hat{\theta}_{N_f}^{\text{NLS}} - \theta_0)$ is asymptotically Normal distributed with covariance matrix P_θ ,*

$$P_\theta = \frac{1}{2} \left[\frac{1}{N_f} \sum_{i=1}^Q \sum_{k=1}^{N_f} \Re \left\{ \overline{\Psi_{0,k}^{(i)}} \Xi_k^{(i)} [\Psi_{0,k}^{(i)}]^T \right\} \right]^{-1} \times \left[\frac{1}{N_f} \sum_{i=1}^Q \sum_{k=1}^{N_f} \Re \left\{ \overline{\Psi_{0,k}^{(i)}} \Sigma_k^{(i)} [\Psi_{0,k}^{(i)}]^T \right\} \right] \times \left[\frac{1}{N_f} \sum_{i=1}^Q \sum_{k=1}^{N_f} \Re \left\{ \overline{\Psi_{0,k}^{(i)}} \Xi_k^{(i)} [\Psi_{0,k}^{(i)}]^T \right\} \right]^{-1}, \quad (16)$$

with the Jacobian matrix $[\Psi_{0,k}^{(i)}]^T = \frac{\partial \text{vec}(G_{k,\theta}^{(i)})}{\partial \theta} \Big|_{\theta=\theta_0}$,

$\overline{(\cdot)}$ denoting complex conjugate, and

$$\Xi_k^{(i)} = [W_k^{(i)}]^{-1}, \quad \Sigma_k^{(i)} = [W_k^{(i)}]^{-1} W_{0,k}^{(i)} [W_k^{(i)}]^{-1}.$$

The covariance is minimized by using the weights

$$W_k^{(i)} = W_{0,k}^{(i)}, \quad (17)$$

which also simplifies (16) to

$$P_\theta = \frac{1}{2} \left[\frac{1}{N_f} \sum_{i=1}^Q \sum_{k=1}^{N_f} \Re \left\{ \overline{\Psi_{0,k}^{(i)}} [W_{0,k}^{(i)}]^{-1} [\Psi_{0,k}^{(i)}]^T \right\} \right]^{-1}.$$

This result is a variant of classical results on static non-linear regression, such as Jennrich (1969). See also (Pintelon and Schoukens, 2001, Thm. 7.21) and (Ljung, 1999, Thm. 9.1) for similar ideas.

Under the given assumptions, the NLS estimator therefore results in unbiased parameters and $\text{Cov}(\hat{\theta}_{N_f}^{\text{NLS}}) \sim \frac{1}{N_f} P_\theta$, with P_θ from (16). The first assumption is true asymptotically, also in a closed-loop setup, given that we use a periodic excitation signal with infinitely many periods. The second assumption, that the true system is in the model class, is necessary when talking about consistency, but is in reality of course violated. Therefore, we will also have a bias term that depends on the chosen weights, see Section 4.3 for additional comments.

4.2 Weighted Logarithmic Least Squares

For systems with a large dynamic range, the NLS estimator may become ill-conditioned. Therefore, the *weighted logarithmic least squares (LLS) estimator* has been suggested (Pintelon and Schoukens, 2001, pp. 206–207)

$$\hat{\theta}_{N_f}^{\text{LLS}} = \arg \min_{\theta \in \Theta} V_{N_f}^{\text{LLS}}(\theta), \quad (18a)$$

$$V_{N_f}^{\text{LLS}}(\theta) = \sum_{i=1}^Q \sum_{k=1}^{N_f} [\mathcal{E}_{k,\theta}^{(i)}]^H [W_k^{(i)}]^{-1} \mathcal{E}_{k,\theta}^{(i)}, \quad (18b)$$

$$\mathcal{E}_{k,\theta}^{(i)} = \log \text{vec}(\hat{G}_k^{(i)}) - \log \text{vec}(G_{k,\theta}^{(i)}), \quad (18c)$$

where $\log G = \log |G| + j \arg G$. According to Pintelon and Schoukens (2001), this estimator has improved numerical stability and is particularly robust to outliers in the measurements. However, from a theoretical point of view, the estimator is slightly biased. The bias can be neglected if the FRF uncertainty is fairly small, see Pintelon and Schoukens (2001) for details. By neglecting the bias, and using $\frac{\partial}{\partial \theta} \log G_{k,\theta} = \frac{1}{G_{k,\theta}} \frac{\partial}{\partial \theta} G_{k,\theta}$ and $\mathcal{E}_{k,\theta_0} = \log \hat{G}_k - \log G_{k,\theta_0} \approx \eta_k / G_{k,\theta_0}$, one can show a similar result as Result 2 for the LLS estimator.

Result 3 *The asymptotic covariance matrix for the LLS estimator (18) is approximately given by (16) with*

$$\Xi_k^{(i)} = [G_{d,k,\theta_0}^{(i)} W_k^{(i)} [G_{d,k,\theta_0}^{(i)}]^H]^{-1}, \quad \Sigma_k^{(i)} = \Xi_k^{(i)} W_{0,k}^{(i)} \Xi_k^{(i)},$$

and $G_{d,k,\theta_0}^{(i)} = \text{diag}(\text{vec}(G_{k,\theta_0}^{(i)}))$. Using the weights

$$W_k^{(i)} = \left[G_{d,k,\theta_0}^{(i)} \right]^{-1} W_{0,k}^{(i)} \left[G_{d,k,\theta_0}^{(i)} \right]^{-H}, \quad (19)$$

gives approximately the same covariance as for the NLS estimator (when using (17)).

4.3 Selection of Weights

Even if the covariance is minimized by using the weights (17) and (19), respectively, the chosen weights will in general be different for a number of reasons. First, the true covariance matrix $W_{0,k}^{(i)}$ is usually not known so the user must instead be content with an estimated covariance matrix $\widehat{W}_{0,k}^{(i)}$. Second, the weights also reflect where the user requires the best model fit. This is important in case the model is unable to describe every detail in the measurements. The weights should then primarily be selected to distribute the bias, not to get minimum variance.

For a resonant system, it is often easier to use the LLS estimator in the way that even constant weights will make sure that both resonances and anti-resonances are matched by the model. This is due to the fact that the logarithm in the LLS estimator inherently gives the relative error, compared to the absolute error when using the NLS estimator. With the NLS estimator, the anti-resonances are easily missed if not choosing $[W_k^{(i)}]^{-1}$ large at those frequencies.

4.4 Solving the Optimization Problem

The minimization problem to be solved, (15) or (18), is unfortunately non-convex. In the example in Section 5, the problem is solved using *fminunc* in MATLAB, which is a gradient-based method which only returns a local optimum. Due to the existence of local minima, a good initial parameter vector, θ_{init} , is important. The problem can be partly solved by random perturbations of the initial parameters in order to reduce the risk falling into local minima. Or, alternatively stated, to obtain a local minimum which is good enough for the purpose of the model. The quality of the resulting model, as well as problems with local minima and identifiability properties, depend on the choices of estimator, weights, and operating points for the experiments. This will be illustrated in Sections 5 and 6.

4.5 Final Identification Procedure

The identification procedure outlined in the previous sections is now summarized:

- (1) Select a number of operating points, e.g., by solving an optimal experiment design problem according to Wernholt and Löffberg (2007).
- (2) Design the excitation signal and collect experimental data from each operating point. Preferably by using an orthogonal random phase multisine signal with multiple periods and/or realizations, see Section 3.2.
- (3) Estimate the nonparametric FRF (11) in each operating point, including uncertainty estimates by using multiple periods and/or realizations.
- (4) Select a parameter estimator, e.g., NLS (15) or LLS (18) (preferred for resonant systems) with suitable weights according to Section 4.3.
- (5) Solve the corresponding minimization problem, (15) or (18), by using optimization software.

5 Application Example

In this section, the described identification procedure, see Section 4.5, is applied to an industrial robot from the ABB IRB6600 series, see Fig. 3. Similar results can be found in Wernholt and Moberg (2008b) and some early results in Öhr et al. (2006).

5.1 Modeling of Robot Manipulators

Most publications concerning robot manipulators only consider elasticity in the rotational actuated joint direction. If gear elasticity is considered we then get the *flexible joint model*, and if link deformation restricted to a plane perpendicular to the preceding joint is included we get the *flexible link model*. These models are described in, e.g., De Luca (2000). However, these models cannot accurately describe a modern industrial robot with weight- and cost optimized components, and thus more elasticity, which results in more complicated vibration modes inside the controller bandwidth. A more general model, called the *extended flexible joint model*, was suggested in Moberg and Hanssen (2007). The model is a lumped parameter model, which consists of a serial kinematic chain of rigid bodies. The rigid bodies are connected by elastic elements, representing torsional deflection. The elastic elements consist of one or more elastic

Fig. 3. The IRB6640 robot from ABB.

Fig. 4. An extended flexible joint dynamic model with three rigid bodies (RB), three joints, two motors, and thus five degrees-of-freedom.

joints. An actuated joint consists of a motor and a gear-box, and the corresponding spring-damper pair models the torsional deflection of the gearbox. A non-actuated joint models elasticities in, e.g., bearings and links. One example of this model is illustrated in Fig. 4, consisting of two motors, three rigid bodies, and three rotational spring-damper pairs. Thus, the model has two actuated joints, one non-actuated joint, and five degrees-of-freedom (DOF). Generally, if the number of actuated and non-actuated joints are Υ_a and Υ_{na} , respectively, the system has $2\Upsilon_a + \Upsilon_{na}$ DOF. The model equations can then be described as

$$\begin{aligned} M(q_g, q_e) \begin{bmatrix} \ddot{q}_g \\ \ddot{q}_e \end{bmatrix} + c(q_g, q_e, \dot{q}_g, \dot{q}_e) + g(q_g, q_e) &= \begin{bmatrix} \tau_g \\ \tau_e \end{bmatrix}, \\ K_g(\eta^{-1}q_m - q_g) + D_g(\eta^{-1}\dot{q}_m - \dot{q}_g) &= \tau_g, \\ -K_e q_e - D_e \dot{q}_e &= \tau_e, \\ M_m \ddot{q}_m + f_m(\dot{q}_m) + \eta^{-T} \tau_g &= \tau_m, \end{aligned}$$

where the angular positions are described by $q_g \in \mathbb{R}^{\Upsilon_a}$ (actuated joints), $q_e \in \mathbb{R}^{\Upsilon_{na}}$ (non-actuated joints), and $q_m \in \mathbb{R}^{\Upsilon_a}$ (motor). The torques are $\tau_g \in \mathbb{R}^{\Upsilon_a}$ (actuated joints), $\tau_e \in \mathbb{R}^{\Upsilon_{na}}$ (non-actuated joints), and $\tau_m \in \mathbb{R}^{\Upsilon_a}$ (motors). The inertia matrices for the motors and joints are M_m and $M(\cdot)$, respectively. The Coriolis and centrifugal torques are described by $c(\cdot)$, and $g(\cdot)$ is the gravity torque. The nonlinear friction in motor bearings and gearboxes is $f_m(\cdot)$ and η is the gear ratio matrix. Finally, K_g , K_e , D_g , and D_e are the diagonal stiffness and damping matrices for the actuated and non-actuated directions. As an example, the complete equations for the model in Fig. 4 are given in Moberg and Hanssen (2009). Since the inertia matrices are invertible (Spong et al., 2006), the system can be formulated in state-space form

$$\dot{x}(t) = f_c(x(t), u(t), \theta), \quad (21a)$$

$$y(t) = h(x(t), u(t), \theta), \quad (21b)$$

where the states, x , can be chosen as the angular position and speed of the motors and joints, the input, u , is the actuator torque τ_m , and the measured variable is the motor angular position, i.e., $y = q_m$. The rigid body parameters of the model (mass, length, inertia, and center

of mass) are assumed to be known from a CAD model or a rigid body identification, see, e.g., Kozłowski (1998). Here, the main objective is identification of elasticity parameters (spring-damper pairs), which are denoted θ .

In the identification procedure described in the previous sections, the parametric FRF $G_{k,\theta}^{(i)}$ is obtained by linearizing the discrete-time model (1) according to Section 2.1. A problem now is that the dynamics in our gray-box model (21) is expressed in continuous time. Of course it is possible to discretize (21a), with assumptions on the inter-sample behavior of u , to obtain $f(x_t, u_t, \theta)$. However, since this must be carried out for each new θ -value during the optimization (see Remark 1), this will be very time consuming. For simplicity, we will instead first linearize (21) and then discretize the linearized model by using zero-order hold to obtain $G_{k,\theta}^{(i)}$. This gives a negligible error (since the considered frequency range is much smaller than the Nyquist frequency) but simplifies the identification procedure substantially. For more details about the linearization and discretization, see, e.g., Wernholt (2007).

5.2 Experimental Setup

The robot used in the experiments is equipped with an experimental feedback controller to allow for offline generated excitation signals. The excitation signal is the angular position reference, and its derivative is the sum of a low-frequency square wave and an orthogonal random phase multisine signal with constant amplitude spectrum for 100 log-spaced frequencies between 2–60 Hz. The square wave is used to reduce the number of zero velocity crossings, and thus reduce the influence of Coulomb friction. For the identification, only the standard sensors installed in the robot are used, i.e., the output y is the motor angular position q_m . The input u is the torque reference signal from the controller. However, in the figures, the angular acceleration $\ddot{q}_m$ is considered as output to avoid the effects of double integrators.

The identification is carried out both on a simulated robot, where the true value θ_0 is known, and on the real robot. Simulations and robot experiments were performed in 15 optimal operating points (see Wernholt and Löfberg, 2007), with 24 sub-experiments in each operating point to get $M = 4$ blocks with $n_u = 6$ parts (the robot has 6 axes). The sampling period is $T_s = 0.5$ ms and every sub-experiment has $N_i = N_p = 2 \cdot 10^4$ samples (i.e. $P = 1$). Each block has a different realization of the multisine signal in order to reduce the influence of both deterministic and stochastic disturbances, and also to enable the computation of the FRF uncertainty. Based on these measurements, the parameters are then estimated using the following estimators:

LLSUW: LLS with user-defined weights.

LLSCW: LLS with minimal-covariance weights (19).
 NLSUW: NLS with user-defined weights.
 NLSCW: NLS with minimal-covariance weights (17).

For simplicity, only diagonal weights $W_k^{(i)}$ are considered (i.e., covariance between different elements in the FRF is neglected). The user-defined weights have only a few levels and are chosen to, roughly, consider the FRF uncertainty. The minimal-covariance weights, (17) and (19), are computed using the estimated covariance matrix¹ $\widehat{W}_{0,k}^{(i)}$ and $G_{k,\theta_0}^{(i)} \approx \widehat{G}_k^{(i)}$. To assess the sensitivity to the initial parameter vector, θ_{init} , 100 optimizations are performed for each of the four estimators, using randomly perturbed initial parameters, $\theta_{\text{init}}^{[l]}$, $l = 1, \dots, 100$ (the same for all estimators). Each element in $\theta_{\text{init}}^{[l]}$ is obtained by multiplying the corresponding element in θ_0 by 10^φ , where φ is a random number from a uniform distribution on the interval $[-1, 1]$. For the simulated robot, θ_0 is known, but for the real robot, a reasonable guess of θ_0 is used. The nonlinear gray-box model for the identification is an *extended flexible joint model* with $\Upsilon_a = 6$, $\Upsilon_{na} = 6$, and thus 18 DOF or 36 states. A linear friction model is used and the parameter vector θ consists of 12 spring constants, 12 dampers, and 6 viscous friction parameters. The linear friction model improves the estimate of the unknown springs and dampers, but the identified friction is not physically correct. For an accurate friction identification, the friction model needs to be nonlinear and other excitation signals must be used for dedicated friction estimation. Note also that it is important to design the experiments (e.g. avoid zero velocity) in such a way that unmodelled properties (e.g. Coulomb friction) does not disturb the identification of the gray-box model.

5.3 Identification of Simulated Robot

The simulation model is based on the gray-box model with some modifications in order to act as a substitute to the real robot. These modifications are:

- Deterministic torque ripple disturbances (Gutt et al., 1996) are added to the torque input in order to simulate ripple disturbances in the motors, drive electronics, and gearboxes.
- Deterministic and stochastic disturbances are added to the measured motor angular position to simulate the real position sensor of the robot (Hanselman, 1990).
- A nonlinear friction model, as described in Moberg et al. (2008), is used.

The same controller and excitation signals were used as for the real robot. The model cost for all 400 identified

¹ Actually only the diagonal, since more blocks of data are needed in order to estimate all elements in $\widehat{W}_{0,k}^{(i)}$.

Fig. 5. Simulated robot: The normalized cost (computed using the LLSUW cost function) for all identified models. Each optimization is carried out using different initial parameter values.

Table 1

Simulated robot: Max and mean (arithmetic) errors of the 12 springs (K) and the 12 dampers (D) for the best model. The error is computed as $100(10^{|\log_{10}(\hat{\Theta}/\Theta_0)|} - 1)$. The motivation for this is that a stiffness parameter of $2\Theta_0$ is about as bad as $1/2\Theta_0$.

	NLSCW	NLSUW	LLSCW	LLSUW
Max Err K	1745 %	3.1 %	11 %	3.1 %
Mean Err K	224 %	1.5 %	2.8 %	1.5 %
Max Err D	1293 %	135 %	412 %	125 %
Mean Err D	436 %	55 %	115 %	41 %

models (4 estimators $\times$ 100 different initial parameter vectors) is shown in Fig. 5. To get comparable numbers, all costs are computed using the LLSUW cost function (18). Clearly, the NLS criteria is very sensitive with respect to initial parameter values, and the minimal-covariance weight is more sensitive than the user-defined weight. Table 1 shows the errors in estimated springs and dampers for the best model, i.e., the model with minimum cost function. Both NLSUW and LLSUW give quite accurate values for the spring constants but the estimated dampers are not that accurate. The NLSCW estimator fails completely. To be noted is that despite the good result for the best NLSUW model in Table 1, approximately only 5 out of 100 optimizations for NLSUW give a reasonable model, compared to approximately 60 for LLSUW. This is evident in Fig. 5, as well as in Fig. 6 where the spring constant accuracy for the 60 best models is shown. Fig. 6 also shows that the LLSUW estimator works really well from a stiffness identification point of view.

The parametric FRF for the best LLSUW model is compared to the nonparametric FRF in Fig. 7 (only element $[1, 1]$ in operating point 8). The parametric FRF of the true linearized model (TLM), $G_{k,\theta_0}^{(8)}$, is also shown as a comparison. This model has a low value of the viscous friction, and the nonlinear friction terms (e.g., Coulomb friction) are omitted during the linearization. Note that the identified gray-box model (IDM) obtains a higher value of the viscous friction to better describe the damping caused by the nonlinear friction, which is clearly vis-

Fig. 6. Simulated robot: Error in estimated stiffness for the 60 best models.

Fig. 7. Simulated robot: The magnitude (from motor torque to motor acceleration) of element $[1, 1]$ for the nonparametric FRF, $\hat{G}_k^{(8)}$, the identified parametric FRF $G_{k,\theta}^{(8)}$ (IDM), and the true linearized parametric FRF $G_{k,\theta_0}^{(8)}$ (TLM).

ible in the nonparametric FRF. Some standard deviations for the estimated springs and dampers are listed in Table 2 where the theoretical standard deviation of the best LLSUW model according to (16) is compared to the empirical standard deviation for the parameters in the 28 best LLSUW models. Note that the empirical variation of parameter values is due to the model parameter sensitivity of the cost function in relation to the stop criteria of the optimization, and to the existence of local minima in the cost function. The 28 models used are the models with cost functions within 5 % from the best model.

The best estimator with respect to accuracy and sensitivity to initial values is clearly the LLSUW estimator. The LLS estimator is, as previously explained, known to be a good choice for resonant systems. The failure of the minimal-covariance weight can be explained by the high uncertainty around the resonance and anti-resonance frequencies, and the fact that most information about stiffness and damping can be found at those frequencies. The stiffness parameters are accurately es-

Table 2

Simulated robot: max, min, and (geometric) mean standard deviation for 12 springs (K) and 12 dampers (D) estimated with the LLSUW estimator. Theoretical standard deviation according to (16), σ_{th} , and based on the 28 best models, σ_{28} .

	$\sigma_{28}^{\min}$	$\sigma_{th}^{\min}$	$\sigma_{28}^{\max}$	$\sigma_{th}^{\max}$	$\sigma_{28}^{\text{mean}}$	$\sigma_{th}^{\text{mean}}$
K	0.4 %	0.4 %	4.9 %	7.5 %	1.3 %	1.7 %
D	11 %	8.4 %	32 %	108 %	27 %	32 %

Table 3

Real robot: max, min, and (geometric) mean standard deviation for 12 springs (K) and 12 dampers (D) estimated with the LLSUW estimator. Theoretical standard deviation according to (16), σ_{th} , and based on the 54 best models, σ_{54} .

	$\sigma_{54}^{\min}$	$\sigma_{th}^{\min}$	$\sigma_{54}^{\max}$	$\sigma_{th}^{\max}$	$\sigma_{54}^{\text{mean}}$	$\sigma_{th}^{\text{mean}}$
K	1.3 %	0.3 %	41.5 %	2.7 %	3.7 %	0.9 %
D	30 %	4.1 %	54 %	47 %	47 %	10 %

Fig. 8. Real robot: The normalized cost (computed using the LLSUW cost function) for all identified models.

timated for the simulated robot (max error of 3%, see Table 1). The stiffness parameter uncertainty estimate based on the result for different initial parameter values is in agreement with the theoretical uncertainty based on the FRF uncertainty (see Table 2). The dampers are harder to estimate and this can partly be explained by the FRF uncertainty around the resonance frequencies. The nonlinear friction also increases the FRF damping in a way that cannot be fully modeled by a linear viscous friction.

5.4 Identification of Real Robot

The same procedure as in Section 5.3 is now applied to the real robot. The model cost for all identified models is shown in Fig. 8, and Table 3 shows the standard deviations for the LLSUW estimator. The parametric FRF for the best LLSUW model in one operating point is compared to the nonparametric FRF in Fig. 9, and element $[1, 1]$ is shown in Fig. 10.

The result is the same as for the simulated robot, i.e., the best estimator with respect to accuracy and sensitivity to initial values is the LLSUW estimator. See also Fig. 11 where the max and mean error of the estimated stiffness parameters are shown for the LLSUW estimator

Fig. 9. Real robot: All 36 elements of nonparametric (dashed) and parametric (solid) FRF in operating point 8.

Fig. 10. Real robot: Element [1, 1] for nonparametric and parametric FRF in operating point 8.

Fig. 11. Real robot: Error in estimated stiffness for the 60 best LLSUW models, compared to the best LLSUW model.

(the other estimators give worse results and are not displayed). The estimated stiffness for the real robot has a larger uncertainty and the model error, i.e., the cost function, is larger for the real robot if compared to the simulated robot. This is no surprise since the 18 DOF model is a low-order approximation of the real robot. Moreover, the placement of the spring-damper pairs in the 18 DOF model is probably not optimal and linear spring-dampers are assumed. Despite of this, if the re-

sulting parametric FRFs are inspected (see e.g., Fig. 9), the model requirements in the frequency domain are satisfied. More specifically, the lower resonance frequencies are estimated with an error within 0.1–1 Hz. The obtained model should therefore be useful for many purposes.

6 A Small Comparison with Time Domain Identification

This section further motivates the use of frequency domain identification for resonant systems. In a small simulation study, time domain identification is compared to frequency domain identification. The simulated system is a model of an elastic robot arm, as described in Moberg et al. (2009), and is illustrated in Fig. 12. Compared to Section 5 this is a simplified scalar system ($n_y = n_u = 1$) with stable dynamics (except integrators). The same torque ripple and measurement disturbances as described in Section 5.3 are used. The friction, springs, and dampers are linear. The identification task is to identify the three unknown spring-damper pairs, all other parameters are assumed to be known. The identification is performed in closed-loop with motor torque τ_m as input and motor position q_m as output. The evaluation is performed in the same way as described in Section 5.2, identification is performed for 100 sets of initial parameters in the range $0.1\theta_0$ to $10\theta_0$. The reference speed is a 10 s swept sinusoid (chirp) with an added low frequency sinusoid. The signals are measured for 11 s to ensure that the system is at rest when the measurement ends (to avoid frequency leakage for the non-periodic input). The reason for using chirp is that multisine excitation is tailored for frequency domain identification and yields a worse result than chirp when used for time-domain identification. Periodic multisine is used in two cases as a comparison. For the time domain identification, the *System Identification Toolbox* (Ljung, 2007) is used. The following cases are evaluated:

- FD NLS: Frequency domain, NLS estimator.
- FD LLS 1: Frequency domain, LLS estimator.
- FD LLS 2: Frequency domain, LLS estimator, multisine reference.
- FD LLS 3: Frequency domain, LLS estimator, multisine reference, average FRF from 4 different experiments using different realizations.
- TD PEM 1: Time domain, prediction error method, no disturbance model, motor speed is used as output.
- TD PEM 2: Time domain, prediction error method, disturbance model used.

Fig. 12. The benchmark system.

Table 4

Result for 100 identifications (optimizations) with random initial parameters. AM is the number of acceptable models. Max errors of springs (K) and dampers (D) are given for the best obtained model. Time is the normalized execution (optimization) time for the identification method. Note that the experiment time is the same for all cases (11 s) except for FD LLS 3 (44 s).

Method	AM	Err K	Err D	Time
FD NLS	0 %	1.1 %	3.9 %	3.3
FD LLS 1	85 %	0.3 %	7.1 %	2.6
FD LLS 2	70 %	0.3 %	1.3 %	1.0
FD LLS 3	66 %	0.1 %	0.2 %	1.0
TD PEM 1	46 %	0.7 %	2.1 %	5.7
TD PEM 2	19 %	0.1 %	0.9 %	5.9

Note that a disturbance model is required to get a stable predictor if motor position is used as output since we have a pure integrator in the system. When motor speed is used, no disturbance model is required. The results for the different cases are shown in Table 4. An acceptable model is defined as a model with a stiffness error within $\pm 1\%$ and a damping error within $\pm 10\%$. The conclusion is once again that for frequency domain identification, the LLS estimator is superior to the NLS estimator which is extremely sensitive to initial parameter values. Time domain identification is also quite sensitive to initial parameter values. For the time domain methods, the number of acceptable models can be slightly increased by decreasing the tolerance parameter of the optimizer, however the penalty in increased execution time is large. Numerical problems were indicated for the time domain methods, especially when disturbance models were used. However, the result also shows a potential to increase the accuracy by use of disturbance models. However, the best results are obtained for the frequency domain method, when multiple experiments are performed and periodic multisine excitation is used.

In this small linear example, the difference in execution time between computing the frequency domain cost and simulating the system is quite small. However, for the 18 DOF nonlinear model used in Section 5, the difference is extremely large. If the execution times for time domain simulation and frequency domain cost are compared, the conclusion is that time domain identification will require between 5000 and 100000 longer optimization time. The data compression obtained by the use of the frequency domain method, is in the range 100 to 2000. The low estimates assume that the n_u experiments required for computing the multivariable FRF can be replaced by 1 experiment and that multiple experiments are not required for increasing the accuracy (disturbance model used). The high estimates assumes that the experiment times are the same for the frequency and time domain methods. If the increased sensitivity for initial parameter values cannot be handled by, e.g., pre-filtering, the

difference in optimization time will further increase. It is likely that time domain identification can reduce the required experiment time but it will certainly increase the optimization time. However, in order to apply a time domain method, the problem of finding a stable predictor for the nonlinear unstable system must be solved and numerical problems are also to be expected.

7 Conclusions and Future Work

The problem of estimating unknown parameters in a nonlinear gray-box model has been studied. It has been shown in Section 2.2 that a time-domain prediction error method can be approximated by a frequency-domain procedure that uses intermediate local models. This allows for data compression and a simplified optimization problem. The procedure is based on the estimation of the nonparametric FRF in a number of operating points. The nonlinear gray-box model is linearized in the same operating points and the optimal parameters are obtained by minimizing the discrepancy between the nonparametric FRFs and the parametric FRFs (the FRFs of the linearized gray-box model). Two different standard parameter estimators (NLS and LLS), as well as the selection of weights in the estimators, have been studied.

The proposed procedure has been successfully applied to the problem of identifying elasticity parameters (spring-damper pairs) in a six-axes industrial robot. The performance and sensitivity to varying initial parameters varies greatly between different estimators and selected weights. It was shown that the LLS estimator is superior for this type of system due to its robustness to the initial estimate. The stiffness parameters were accurately estimated, but the estimation of damping was quite hard. The minimal-covariance weights turned out to give poor results, compared to approximately constant user-defined weights. The main reason was the large FRF uncertainty around the resonances, and the fact that most information about the parameters can be found at those frequencies. The identification result of the real robot is satisfactory, the identified model gives a good global description in the frequency range of interest and the obtained model should be useful for many future purposes.

For the future, improved model accuracy, reduction of measurement time and excitation energy, and ease-of-use are desired. Nonlinear time-domain identification, gray-box model structure selection, and automatic weight selection are therefore important areas for future research. The use of additional sensors, e.g., accelerometers are also an interesting area.

References

Anderson, B. D. O. and Moore, J. B. (1979). *Optimal Filtering*. Prentice-Hall, Upper Saddle River, N.J.

- Behi, F. and Tesar, D. (1991). Parametric identification for industrial manipulators using experimental modal analysis. *IEEE Transactions on Robotics and Automation*, 7(5):642–652.
- Bohlin, T. (2006). *Practical Grey-box Process Identification: Theory and Applications*. Springer.
- Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*. McGraw-Hill, New York, expanded edition.
- De Luca, A. (2000). Feedforward/feedback laws for the control of flexible robots. In *2000 IEEE International Conference on Robotics and Automation*, pages 233–240, San Francisco, California.
- Dobrowiecki, T. and Schoukens, J. (2007). Measuring a linear approximation to weakly nonlinear MIMO systems. *Automatica*, 43(10):1737–1751.
- Doucet, A., de Freitas, N., and Gordon, N., editors (2001). *Sequential Monte Carlo Methods in Practice*. Springer Verlag.
- Forssell, U. and Ljung, L. (2000). Identification of unstable systems using output error and box-jenkins model structures. *IEEE Transactions on Automatic Control*, 45(1):137–141.
- Gawthrop, P. J. and Wang, L. (2005). Data compression for estimation of the physical parameters of stable and unstable linear systems. *Automatica*, 41(8):1313–1321.
- Guillaume, P. (1998). Frequency response measurements of multivariable systems using nonlinear averaging techniques. *IEEE Transactions on Instrumentation and Measurement*, 47(3):796–800.
- Gutt, H.-J., Scholl, F. D., and Blattner, J. (1996). High precision servo drives with DSP-based torque ripple reduction. In *IEEE AFRICON, 1996*, volume 2, pages 632–637.
- Hanselman, D. (1990). Resolver signal requirements for high accuracy resolver-to-digital conversion. *IEEE Transactions on Industrial Electronics*, 37(6):556–561.
- Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, 40(2):633–643.
- Johansen, T. A. and Foss, B. A. (1995). Identification of non-linear system structure and parameters using regime decomposition. *Automatica*, 31(2):321–326.
- Kozłowski, K. (1998). *Modelling and identification in robotics*. Advances in Industrial Control. Springer, London.
- Kristensen, N. R., Madsen, H., and Jørgensen, S. B. (2004). Parameter estimation in stochastic grey-box models. *Automatica*, 40(2):225–237.
- Larsson, R., Sjanic, Z., Enqvist, M., and Ljung, L. (2009). Direct prediction-error identification of unstable nonlinear systems applied to flight test data. In *15th IFAC Symposium on System Identification*, pages 144–149, Saint-Malo, France.
- Ljung, L. (1999). *System Identification: Theory for the User*. Prentice Hall, Upper Saddle River, New Jersey, USA, 2nd edition.
- Ljung, L. (2003). Linear system identification as curve fitting. In Rantzer, A. and Byrnes, C. I., editors, *Directions in Mathematical Systems Theory and Optimization (Lecture Notes in Control and Information Sciences)*, volume 286, pages 203–215. Springer Verlag.
- Ljung, L. (2007). *System Identification Toolbox for use with Matlab. Version 7*. The MathWorks, Inc., Natick, MA, 7 edition.
- Ljung, L. (2008). Perspectives on system identification. In *17th IFAC World Congress*, pages 7172–7184, Seoul, Korea.
- Moberg, S. and Hanssen, S. (2007). A DAE approach to feed-forward control of flexible manipulators. In *2007 IEEE International Conference on Robotics and Automation*, pages 3439–3444, Roma, Italy.
- Moberg, S. and Hanssen, S. (2009). Inverse dynamics of flexible manipulators. In *Multibody Dynamics 2009*, pages 1–20, Warsaw, Poland.
- Moberg, S., Öhr, J., and Gunnarsson, S. (2008). A benchmark problem for robust control of a multivariable nonlinear flexible manipulator. In *Proc. 17th IFAC World Congress*, Seoul, Korea.
- Moberg, S., Öhr, J., and Gunnarsson, S. (2009). A benchmark problem for robust feedback control of a flexible manipulator. *IEEE Transactions on Control Systems Technology*, 17(6):1398–1405.
- Öhr, J., Moberg, S., Wernholt, E., Hanssen, S., Pettersson, J., Persson, S., and Sander-Tavallaey, S. (2006). Identification of flexibility parameters of 6-axis industrial manipulator models. In *ISMA2006 International Conference on Noise and Vibration Engineering*, pages 3305–3314, Leuven, Belgium.
- Pintelon, R. and Schoukens, J. (2001). *System identification: a frequency domain approach*. IEEE Press, New York.
- Schittkowski, K. (2002). *Numerical data fitting in dynamical systems*. Kluwer Academic Publishers, Dordrecht.
- Schoukens, J., Pintelon, R., Dobrowiecki, T., and Rolain, Y. (2005). Identification of linear systems with nonlinear distortions. *Automatica*, 41(3):491–504.
- Spong, M. W., Hutchinson, S., and Vidyasagar, M. (2006). *Robot Modeling and Control*. Wiley.
- Verboven, P., Guillaume, P., and Caubergh, B. (2005). Multivariable frequency-response curve fitting with application to modal parameter estimation. *Automatica*, 41(10):1773–1782.
- Wernholt, E. (2007). *Multivariable Frequency-Domain Identification of Industrial Robots*. Dissertations. no. 1138, Linköping Studies in Science and Technology, Linköping University, Sweden.
- Wernholt, E. and Gunnarsson, S. (2007). Analysis of methods for multivariable frequency response function estimation in closed loop. In *46th IEEE Conference on Decision and Control*, pages 4881–4888, New Orleans, Louisiana, USA.
- Wernholt, E. and Gunnarsson, S. (2008). Estimation of nonlinear effects in frequency-domain identification of industrial robots. *IEEE Transactions on Instrumentation and Measurement*, 57(4):856–863.
- Wernholt, E. and Löfberg, J. (2007). Experiment design for identification of nonlinear gray-box models with application to industrial robots. In *46th IEEE Conference on Decision and Control*, pages 5110–5116, New Orleans, Louisiana, USA.
- Wernholt, E. and Moberg, S. (2008a). Experimental comparison of methods for multivariable frequency response function estimation. In *17th IFAC World Congress*, pages 15359–15366, Seoul, Korea.
- Wernholt, E. and Moberg, S. (2008b). Frequency-domain gray-box identification of industrial robots. In *17th IFAC World Congress*, pages 15372–15380, Seoul, Korea.