Proceedings of the 35th
Conference on Decision and Control
Kobe, Japan • December 1996

Multivariable Least Squares Frequency Domain Identification using Polynomial Matrix Fraction Descriptions

R.A. de Callafon‡ D. de Roover§ P.M.J. Van den Hof

Mechanical Engineering Systems and Control Group
Delft University of Technology
Mekelweg 2, 2628 CD Delft, The Netherlands
E-mail: r.a.decallafon@wbmt.tudelft.nl

Abstract

In this paper an approach is presented to estimate a linear multivariable model on the basis of (noisy) frequency domain data via a curve fitting procedure. The multivariable model is parametrized in either a left or a right polynomial matrix fraction description and the parameters are computed by using a two-norm minimization of a multivariable output error. Additionally, input-output or element-wise based multivariable frequency weightings can be specified to tune the curve fitting error in a flexible way. The procedure is demonstrated on experimental data obtained from a 3 input 3 output Wafer Stepper system.

1 Introduction

Formulating a procedure that is able to estimate a model on the basis of frequency domain data has gained considerable attention in the research on system identification. Although the clear distinction between time and frequency domain data is generally overestimated [12], estimation of models by fitting complex frequency domain data has several advantages compared to time domain approaches. Firstly, representing data in the frequency domain domain can yield substantial data reduction [14]. Secondly, compressing a huge amount of time domain data into a finite number of frequency points facilitates noise reduction directly. Both aspects are used extensively in commercially available sophisticated test equipment for spectral analysis.
Based on Least Squares (LS) estimation techniques, as used by Levi in [10] and further refined by Sanathanan and Koerner in [15], multivariable frequency domain curve fitters have been formulated in the literature. One is referred to [11], [4] and the more recently introduced procedure in [1]. Basically, the procedures differ in the way the multivariable model is parametrized and whether or not the procedure allows for a specification of the model order and a (multivariable) weighting on the curve fit error. As such, in [11] a multivariable model is found by the composition of scalar subsystems, while the order of the subsequent transfer functions is determined by testing the residuals. A similar approach can be found in [4], wherein a Chebyshev polynomial basis is used to improve numerical conditioning of the LS-problem. In [1] the model is parametrized directly by means of a matrix numerator polynomial and a scalar common denominator polynomial, whereas only a scalar frequency dependent weighting on the curve fit error is allowed.
Several alternatives to a LS-approach can also be found in the literature. In [13] a subspace based algorithm in the frequency domain is presented that allows the user to specify an additional frequency weighting. In [9] a frequency domain curve fitter has been developed in which a maximum amplitude of a (weighted) curve fit error is being considered. Furthermore, so-called $\mathcal{H}_\infty$-identification procedures, currently applicable to scalar frequency domain data, can guarantee an upper bound on the additive error, see e.g. [8] and the references therein. Unfortunately, a maximum amplitude criterion can be highly sensitive to noise, whereas the available $\mathcal{H}_\infty$-identification procedures might yield high order models for moderately damped processes [5].
Based on the LS-approach, this paper presents a multivariable frequency domain curve fitter in which the aim is to minimize the two-norm on a (weighted) curve fit error for a model having a limited McMillan degree. The multivariable model is parametrized by either a left or right polynomial Matrix Fraction Description (MFD). By use of Kronecker calculus it will be shown that both a pre, post or element-wise multivariable frequency weighting on the curve fit error can handled relatively easily. Furthermore, it will be shown that the iteration described by [15], denoted by SK-iteration, can be generalized to estimate a polynomial MFD. Due to the subsequent convex optimization steps in the SK-iteration, this approach supports the estimation of models with many parameters. Similar to the

‡The work of Raymond de Callafon is financially supported by the Dutch Systems and Control Theory Network

§Research is sponsored by Philips' Research Laboratories, Eindhoven, the Netherlands

approach followed by [1] and supported by the work of [17], the resulting estimate can be used as an initial value for a Gauss-Newton optimization.
Although cumbersome iterations can be avoided by the use of a realization based algorithm as reported in [13], the possibility to prespecify the McMillan degree of the model and to introduce a flexible element-wise frequency weighting on the multivariable data is quite helpful from a practical point of view. The procedure will be illustrated by fitting a multivariable model on the frequency response obtained from the positioning mechanism present in a wafer stepper.

2 Problem formulation

To formulate the multivariable frequency domain identification problem, consider the following set $\mathcal{G}$ of noisy complex frequency response data observations $G(\omega_j)$, evaluated at N frequency points ω_j.

$$\mathcal{G} := \{G(\omega_j) \mid G(\omega_j) \in \mathbb{C}^{p\times m}, \text{ for } j \in 1,\ldots,N\} \quad (1)$$

The aim of the identification problem discussed in this paper is to find a linear time invariant multivariable model P of *limited complexity*, having m inputs and p outputs, that *approximates* the data $\mathcal{G}$ in (1).
To address the limited complexity, the model $P(\theta)$ is parametrized by a either a left or right polynomial MFD that depends on a real valued parameter θ of limited dimension. The specific parametrization of the polynomial MFD of $P(\theta)$ is discussed in the next section. The approximation of the data $\mathcal{G}$ by the model $P(\theta)$ is addressed by considering the following additive error.

$$E_a(\omega_j,\theta) := [G(\omega_j) - P(\xi(\omega_j),\theta)] \text{ for } j \in 1,\ldots,N \quad (2)$$

The complex variable $\xi(\cdot)$ in (2) is used to denote the frequency dependency of the model $P(\theta)$. In this way, $\xi(\omega_j) = i\omega_j$ to represent a continuous time model, whereas $\xi(\omega_j) = e^{i\omega_j T}$ (shift operator) or $\xi(\omega_j) = (e^{i\omega_j} - 1)/T$ (δ operator) to represent a discrete time model with sampling time T.
To tune the additive error E_a in (2), both an input-output frequency weighted curve fit error E_w with

$$E_w(\omega_j,\theta) := W_{out}(\omega_j)E_a(\omega_j,\theta)W_{in}(\omega_j) \quad (3)$$

and an element-wise frequency weighted curve fit error E_s with

$$E_s(\omega_j,\theta) := S(\omega_j).*E_a(\omega_j,\theta) \quad (4)$$

will be considered in this paper. In (4) .* is used to denote the Schur product; an element-by-element multiplication.
Using the notation E to denote the frequency weighted curve fit error E_w in (3) and E_s in (4), the deviation of the data $\mathcal{G}$ is characterized by following the norm function $J(\theta)$.

$$J(\theta) := \sum_{i=1}^{N} \text{tr}\{E(\omega_j,\theta)E^*(\omega_j,\theta)\} = \|E(\theta)\|_F^2 \quad (5)$$

In (5) * is used to denote the complex conjugate transpose, $\text{tr}\{\cdot\}$ is the trace operator and $\|E(\theta)\|_F$ denotes the Frobenius norm operating on the matrix $E(\theta) = [E(\omega_1,\theta) \cdots E(\omega_N,\theta)]$. Consequently, the goal of the procedure described in this paper is to find a real valued parameter $\hat{\theta}$ of limited complexity that can be formulated by the following minimization.

$$\hat{\theta} := \arg \min_{\theta \in \mathbb{R}} J(\theta) \quad (6)$$

3 Parametrization

3.1 Polynomial matrix fraction descriptions

The multivariable model is represented by either a left or right polynomial MFD, respectively given by

$$P(\xi,\theta) = A(\xi^{-1},\theta)^{-1}B(\xi^{-1},\theta) \quad (7)$$

$$P(\xi,\theta) = B(\xi^{-1},\theta)A(\xi^{-1},\theta)^{-1} \quad (8)$$

where A and B denote parametrized polynomial matrices in the indeterminate ξ^{-1}.
For a model having m inputs and p outputs, the the polynomial matrix $B(\xi^{-1},\theta)$ is parametrized by

$$B(\xi^{-1},\theta) = \sum_{k=d}^{d+b-1} B_k\,\xi^{-k} \quad (9)$$

where $B_k \in \mathbb{R}^{p\times m}$, d denotes the number of leading zero matrix coefficients and b the number of non-zero matrix coefficients in $B(\xi^{-1},\theta)$. For the left MFD in (7), $A(\xi^{-1},\theta)$ is parametrized by

$$A(\xi^{-1},\theta) = I_{p\times p} + \xi^{-1}\sum_{k=1}^{a} A_k\,\xi^{-k+1} \quad (10)$$

where $A_k \in \mathbb{R}^{p\times p}$ and a denotes the number of non-zero matrix coefficients in the monic polynomial $A(\xi^{-1},\theta)$. The parameter θ is determined by the corresponding unknown matrix coefficients in the polynomials. Hence,

$$\theta = \begin{bmatrix} B_d & \cdots & B_{d+b-1} & A_1 & \cdots & A_a \end{bmatrix} \quad (11)$$

and $\theta \in \mathbb{R}^{p\times(mb+pa)}$ for the left MFD in (7). Dual results can be formulated for the right MFD in (8).
Additionally to the full polynomial parametrization presented here, so-called structural parameters d_{ij}, b_{ij} and a_{ij} with $d := \min\{d_{ij}\}$, $b := \max\{b_{ij}\}$, and $a := \max\{a_{ij}\}$ can be used to specify a none-full polynomial parametrization. In this way, the parameter θ as given in (11) may contain prespecified zero entries at specific locations. This may occur in a discrete time model with $\xi^{-1} = z^{-1}$ where the value of d_{ij} has a direct connection with the number of time delays from the jth input to the ith output.

3.2 Model order

Due to the indeterminate ξ^{-1}, it can be verified that the MFD of (7) or (8) gives rise to a (strictly) proper transfer function matrix $P(\xi,\theta)$, regardless of the value of the integers $d_{i,j}$, $b_{i,j}$ or $a_{i,j}$. Hence, there are no restrictions on the size of the structural parameters, other than a limitation on the McMillan degree of the resulting model $P(\xi,\hat{\theta})$. For the connection between the structural parameters and the McMillan degree of $P(\xi,\theta)$, the following result can be given.

Lemma 3.1 *Consider a parameter $\hat{\theta}$ such that $A_a \neq 0$ and $B_{d+b-1} \neq 0$. Define*

$$\eta := \max\{a, d+b-1\} \tag{12}$$

and $\bar{A}(\xi,\hat{\theta}) := \xi^{\eta}A(\xi^{-1},\hat{\theta})$, $\bar{B}(\xi,\hat{\theta}) := \xi^{\eta}B(\xi^{-1},\hat{\theta})$. Let n be used to denote the McMillan degree of the multivariable transfer function model $P(\xi,\hat{\theta})$ obtained by (7) or (8), then

$$n = deg\ det\{\bar{A}(\xi,\hat{\theta})\}$$

if and only if $\bar{A}(\xi,\hat{\theta})$ and $\bar{B}(\xi,\hat{\theta})$ are left coprime over $\mathbb{R}[\xi]$ in case of (7) and right coprime over $\mathbb{R}[\xi]$ in case of (8).

Proof: The proof is given for (8). With the condition $A_a \neq 0$, $B_{d+b-1} \neq 0$, it follows that $\bar{A}(\xi) := \xi^{\eta}A(\xi^{-1})$ and $\bar{B}(\xi) := \xi^{\eta}B(\xi^{-1})$ are polynomial matrices in ξ. In case of (8), $P(\xi) = \bar{B}(\xi)\bar{A}(\xi)^{-1}$ and a state space realization [A,B,C,D] for $P(\xi)$ can be obtained, such that $\dim \mathsf{A} = \deg\ \det\{\bar{A}(\xi)\}$ and {A,B} controllable, see e.g [3]. Furthermore, {C,A} is observable if and only if $\bar{A}(\xi)$ and $\bar{B}(\xi)$ are right coprime over $\mathbb{R}[\xi]$, see theorem 6.1 in [3]. Dually, the result can be shown for (7). ∎

Under some mild condition on the polynomials $A(\xi^{-1},\hat{\theta})$ and $B(\xi^{-1},\hat{\theta})$ being estimated, lemma 3.1 gives a direct relation between the deg det$\{\bar{A}(\xi,\hat{\theta})\}$ and the McMillan degree of the resulting estimate $P(\xi,\hat{\theta})$. In case of the left MFD (7), deg det$\{\bar{A}(\xi,\hat{\theta})\}$ generally will be equal to ηp. Hence, the structural parameters give rise to (an upper bound) on the McMillan degree of the model being estimated. For a more detailed discussion on the exact relation between the McMillan degree, the row degree of the polynomial matrices $A(\xi^{-1},\theta)$, $B(\xi^{-1},\theta)$ and the observability indices of a model computed by a left polynomial MFD, one is referred to [6] or [16].

Compared to a parametrization of the multivariable model $P(\xi,\theta)$ using a scalar common denominator polynomial $d(\xi^{-1},\theta)$ as presented in [1], the parametrization using a (left) MFD is more flexible, as a scalar common denominator restricts $A(\xi^{-1},\theta)$ to be $I_{p\times p}d(\xi^{-1},\theta)$. A model with one output that is parametrized by the left MFD of (7), constitutes a scalar common denominator polynomial $A(\xi^{-1},\theta)$.

4 Computational procedure

4.1 Iterative minimization

In this section, the minimization of (6) will be discussed by means of an iterative procedure of convex optimization steps similar to the SK-iteration of [15]. The attention will be restricted to a parametrization of $P(\xi,\theta)$ based on the left MFD (7) as dual results can be obtained for a right MFD. To extend the SK-iteration to the multivariable case, first consider the (unweighted) additive curve fit error of (2).

For a model $P(\xi,\theta)$ parametrized by left MFD, (2) can be written as

$$E_a(\omega_j,\theta) = A(\xi(\omega_j)^{-1},\theta)^{-1}\tilde{E}(\omega_j,\theta) \tag{13}$$

where $\tilde{E}(\omega_j,\theta)$ is the equation error defined by

$$\tilde{E}(\omega_j,\theta) := A(\xi(\omega_j)^{-1},\theta)G(\omega_j) - B(\xi(\omega_j)^{-1},\theta). \tag{14}$$

Substituting the parametrization (7) for the polynomials A, B, the equation error in (14) can be represented by

$$\tilde{E}(\omega_j,\theta) = G(\omega_j) - \theta\Phi(\omega_j) \tag{15}$$

where θ is given in (11) and

$$\Phi(\omega_j) = \begin{bmatrix} I_{m\times m}\xi(\omega_j)^{-d} \\ \vdots \\ I_{m\times m}\xi(\omega_j)^{-(d+b-1)} \\ G(\omega_j)\xi(\omega_j)^{-1} \\ \vdots \\ G(\omega_j)\xi(\omega_j)^{-a} \end{bmatrix} \tag{16}$$

with $\Phi(\omega_j) \in \mathbb{C}^{(mb+pa)\times m}$.

A matrix $\tilde{E}(\theta)$ can be formed by stacking $\tilde{E}(\omega_j,\theta)$ column-wise for $j \in 1,\ldots,N$ and this yields

$$\arg\min_{\theta\in\mathbb{R}} \|\tilde{E}(\theta)\|_F^2 = \arg\min_{\theta\in\mathbb{R}} \|\mathsf{G} - \theta\mathsf{P}\|_F^2 \tag{17}$$

where G and P are found by stacking the real and imaginary part of respectively $G(\omega_j)$ and $\Phi(\omega_j)$ for $j \in 1,\ldots,N$. Due to the linear appearance of the parameter θ, (17) corresponds a standard least squares problem that can be solved by numerical reliable tools as e.g a QR-factorization with (partial) pivoting [7].

Due to the fact that $A(\xi^{-1},\theta)$ in (13) also depends on the parameter θ, the linear appearance of the parameter θ in (13) is violated. In order to facilitate the convexity in minimizing the two-norm on the equation error in (17), an iterative procedure similar as in [15] can be used. An estimate $\hat{\theta}_t$ in step t is computed by replacing $A(\xi(\omega_j)^{-1},\theta)$ in (13) by a *fixed* $A(\xi(\omega_j)^{-1},\hat{\theta}_{t-1})$ based on an estimate $\hat{\theta}_{t-1}$ obtained from the previous step $t-1$. In this way the Frobenius norm of an output weighted equation error $\tilde{E}_w(\omega_j,\hat{\theta}_{t-1},\theta) = A(\xi(\omega_j)^{-1},\hat{\theta}_{t-1})^{-1}\tilde{E}(\omega_j,\theta)$ needs to be minimized repeatedly according to

$$\hat{\theta}_t = \arg\min_{\theta\in\mathbb{R}} \|\tilde{E}_w(\hat{\theta}_{t-1},\theta)\|_F^2.$$

This generalizes the SK-iteration to multivariable models parametrized by a left polynomial MFD. A dual approach can be formulated for a right polynomial MFD.

The estimate obtained from the SK-iteration is not optimal in the sense of (6) in presence of noise and/or incorrect model order, but it does provide a tool to find an initial estimate for a GN-optimization [17]. Furthermore, the convex optimization to be solved in each step of the multivariable SK-iteration supports the estimation of models with many parameters. The computational procedure to obtain the parameter $\hat{\theta}$ in case of the (weighted) curve fit errors of (3) and (4) is presented in the subsequent sections.

4.2 Input-output weighting

The input-output weighted curve fit error of (3) can be rewritten into

$$E_w(\omega_j,\theta) = \tilde{W}_{out}(\omega_j,\theta)\tilde{E}(\omega_j,\theta)W_{in}(\omega_j) \qquad (18)$$

where $\tilde{W}_{out}(\omega_j,\theta) := W_{out}(\omega_j)A(\xi(\omega_j)^{-1},\theta)^{-1}$ and $\tilde{E}(\omega_j,\theta)$ is given in (14).

Using a similar approach of iterative minimization steps as used in section 4.1, the parameter θ in $\tilde{W}_{out}(\omega_j,\theta)$ in (18) is fixed to an estimate $\hat{\theta}_{t-1}$ obtained from the previous step $t-1$. Consequently, the weighted equation error $\tilde{E}_w$ defined by

$$\tilde{E}_w(\omega_j,\hat{\theta}_{t-1},\theta) := \tilde{W}_{out}(\omega_j,\theta_t)\tilde{E}(\omega_j,\theta)W_{in}(\omega_j) \qquad (19)$$

again indicates that the parameter θ to be estimated appears linearly in (19).

Although the free parameter θ appears linearly in (19), writing down a matrix representation for the weighted equation error $\tilde{E}_w$ similar to (17) would inevitably lead to additional (large) sparse matrices that need to be stored in order to compute the least squares solution. The sparse matrices arise from the frequency dependent output (and input) weighting that need to be incorporated [1]. Furthermore, the parameter θ might have a structure containing zero entries at prespecified locations if a none-full polynomial parametrization is being used.

To avoid the computational and memory storage issues that arise from dealing with (large) sparse matrices and to be able to take into account the specific structure that might be present in the parameter θ, a fairly simple and straightforward computational procedure based on Kronecker calculus is presented here. For this purpose consider the following definition.

Definition 4.1 *Consider two matrices $X \in \mathbb{C}^{n_1\times n_2}$ and $Y \in \mathbb{C}^{m_1\times m_2}$, then the Kronecker vector* $\mathrm{vec}(X) \in \mathbb{C}^{n_1n_2\times 1}$ *and the Kronecker product $X\otimes Y \in \mathbb{C}^{n_1m_1\times n_2m_2}$ are respectively defined by* $\mathrm{vec}(X) := [x_1 \ \cdots \ x_{n_2}]^T$ *and*

$$X\otimes Y := \begin{bmatrix} x_{1,1}Y & \cdots & x_{1,n_2}Y \\ \vdots & \cdots & \vdots \\ x_{n_1,1}Y & \cdots & x_{n_1,n_2}Y \end{bmatrix}$$

where $x_{i,j}$ and x_j for $i \in 1,\ldots,n_1$ and $j \in 1,\ldots,n_2$ are used to denote respectively the (i,j)th entry in X and the jth column in X.

The Kronecker product is a well known concept [2] and by stacking the columns of a matrix to obtain the corresponding Kronecker vector as mentioned in definition 4.1, the following result can be obtained.

Proposition 4.2 *Consider (complex) matrices X, Y and Z with appropriate dimensions, such that the matrix product $C := XYZ$ is well defined. Then* $\mathrm{vec}(C)$ *satisfies*

$$\mathrm{vec}(C) = [Z^T\otimes X]\mathrm{vec}(Y).$$

Proof: The proof can be found in [2]. ∎

On the basis of proposition 4.2, the Kronecker vector of the input/output weighted equation error $\tilde{E}_w(\omega_j,\hat{\theta}_{t-1},\theta)$ in (19) can be written as

$$\mathrm{vec}(\tilde{E}_w) = \mathrm{vec}(\tilde{W}_{out}GW_{in}) - [[\Phi W_{in}]^T\otimes \tilde{W}_{out}]\mathrm{vec}(\theta)$$

wherein the arguments ω_j, $\hat{\theta}_{t-1}$ and θ are left out, to avoid notational issues. As the Frobenius-norm satisfies $\|X\|_F^2 = \|\mathrm{vec}(X)\|_F^2$ for an arbitrary matrix X, the Frobenius-norm on $\tilde{E}_w$ can be characterized by a matrix representation formed by stacking $\mathrm{vec}(\tilde{E}_w(\omega_j,\hat{\theta}_{t-1},\theta))$ row-wise for $j \in 1,\ldots,N$. This yields the following estimate

$$\begin{aligned} \hat{\theta} &= \arg\min_{\theta\in\mathbb{R}} \|\mathrm{vec}(\tilde{E}_w(\hat{\theta}_{t-1},\theta))\|_F^2 \\ &= \arg\min_{\bar{\theta}\in\mathbb{R}} \|\mathsf{G}_w - \mathsf{P}_w\bar{\theta}\|_F^2 \end{aligned} \qquad (20)$$

where $\bar{\theta} = \mathrm{vec}(\theta) \in \mathbb{R}^{p(mb+pa)\times 1}$ according to (11). Furthermore, $\mathsf{G}_w \in \mathbb{R}^{2pmN\times 1}$ and $\mathsf{P}_w \in \mathbb{R}^{2pmN\times p(mb+pa)}$ are matrices that can be found by row-wise stacking of the real and imaginary part of respectively $\mathrm{vec}(\tilde{W}_{out}(\omega_j,\hat{\theta}_{t-1})G(\omega_j)W_{in}(\omega_j))$ and $\mathrm{vec}([\Phi(\omega_j)W_{in}(\omega_j)]^T\otimes\tilde{W}_{out}(\omega_j,\hat{\theta}_{t-1}))$ for $j \in 1,\ldots,N$.

The regression matrix P_w in (20) does not exhibit any sparse matrix structure as occurs e.g. in the method of [1]. In fact, $2pmN\times p(mb+pa)$ entries is the smallest dimension of the regression matrix P_w in order to compute a least squares parameter $\hat{\theta}$ that has $p(mb+pa)$ unknown entries (for a a left full polynomial parametrization) on the basis of N complex frequency domain points of a $p\times m$ multivariable system. In this way memory storage problems are avoided directly as much as possible.

As the parameter θ is converted into a column parameter $\bar{\theta} = \mathrm{vec}(\theta)$, any prespecified zero entries in $\bar{\theta}$ can be incorporated in the estimation of the parameter relatively easy. This can be done by omitting the columns in the regression matrix P_w that correspond to the zero entries in $\bar{\theta}$ and thereby reducing the size of the parameter to be estimated directly.

4.3 Schur weighting

Consider the Schur or element-wise frequency weighted curve fit error in (4) and rewrite this into

$$E_s(\omega_j,\theta) = S(\omega_j).*[A(\xi(\omega_j)^{-1},\theta)^{-1}\tilde{E}(\omega_j,\theta)] \qquad (21)$$

where the equation error $\tilde{E}(\omega_j, \theta)$ was defined in (14). Using a similar approach of iterative minimization steps as used in section 4.1, the parameter θ in $A(\xi(\omega_j)^{-1}, \theta)^{-1}$ in (21) is fixed to an estimate $\hat{\theta}_{t-1}$ obtained from the previous step $t-1$. Consequently, the weighted equation error $\tilde{E}_s$ defined by

$$\tilde{E}_s(\omega_j, \hat{\theta}_{t-1}, \theta) := S(\omega_j). * [A(\xi(\omega_j)^{-1}, \hat{\theta}_{t-1})^{-1} \tilde{E}(\omega_j, \theta)]$$

again indicates that the parameter θ to be estimated appears linearly. Finally, it can be verified (leaving out the arguments ω_j, $\xi(\omega_j)^{-1}$, $\hat{\theta}_{t-1}$ and θ) that $\text{vec}(\tilde{E}_s)$ can be rewritten into

$$\text{vec}(S. * [A^{-1}G]) - \text{diag}(\text{vec}(S))[\Phi^T \otimes A^{-1}]\text{vec}(\theta) \quad (22)$$

by using the result of proposition 4.2. Hence, stacking $\text{vec}(\tilde{E}_s(\omega_j, \hat{\theta}_{t-1}, \theta))$ row wise for each $j \in 1, \ldots, N$ will yield a similar expression for the minimizing argument $\hat{\theta}$ as given in (20). However, the matrix G_w in (20) now contains real and imaginary part of $\text{vec}(S(\omega_j). * [A(\xi(\omega_j)^{-1}, \hat{\theta}_{t-1})G(\omega_j)])$, whereas P_w in (20) will consist of the real and imaginary part of $\text{diag}(\text{vec}(S(\omega_j)))[\Phi(\omega_j)^T \otimes A^{-1}(\xi(\omega_j)^{-1}, \hat{\theta}_{t-1})]$ for $j \in 1, \ldots, N$. Hence, the same computational procedure can be used to incorporate an element-by-element weighted curve fit error (4) by a slight modification of the matrices in (20).

5 Application to experimental data

5.1 Description of the wafer stepper system

The multivariable curve fit procedure discussed in this paper is illustrated by curve fitting experimental data obtained from a positioning system of a wafer stepper.

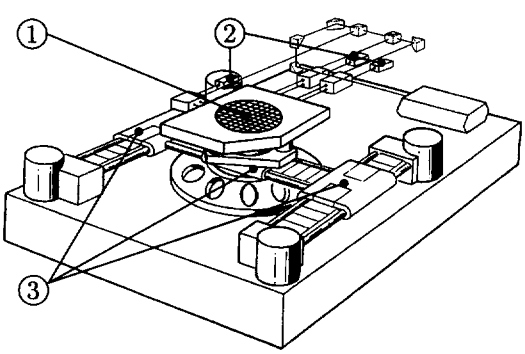

Fig. 1: Schematic view of a wafer stage; 1:wafer chuck, 2:laser interferometers, 3:linear motors.

A wafer stepper is a high accuracy positioning machine, used in chip manufacturing processes and a schematic view is depicted in figure 1. The wafer carries approximately 80 chips and is placed on a moving table, called the wafer chuck, which needs to be positioned accurately. The position of the wafer chuck on the horizontal surface of a granite block is measured by means of three laser interferometry measurements, whereas three linear motors are used to position the wafer chuck. In this way, the positioning system is considered to be a multivariable system, having three currants to the linear motors as inputs and three position measurements as outputs of the process.

5.2 Experimental results

Periodic random noise signals of 1024 points are used to excite the system. Using the resulting averaged time series, a spectral estimate is computed, resulting in a finite number of frequency domain data points that constitutes a suitable starting point for the subsequent curve fit procedure.

As the resulting model has to be used for discrete time control design purposes, the aim is to estimate a possibly *low order* discrete time multivariable model, that describes the dynamical behaviour of the positioning system in the frequency domain till approximately 400 Hz. For frequencies smaller than 100 Hz, the positioning system acts like a double integrator. To illustrate the usage of weighting functions in order to shape the curve fit error, an output weighting is used that emphasizes the frequency range between 200 and 300 Hz and starts to roll off at 300 Hz. The order of the resulting multivariable model (without the 3 double integrators) is chosen to be 12, represented by a full left polynomial matrix fraction description having 81 parameters.

The SK-iteration is started up by first estimating a high order model to compute an initial value for the modified output weighting $\tilde{W}_{out}$ in (19). After this initialization, the SK-iteration is invoked 8 times. The Bode amplitude plot and phase plot of the 18th order estimate (including the 3 double integrators) is depicted respectively in figure 2 and figure 3. It should be noted that the multivariable output weighting applied during the estimation procedure emphasizes the frequency domain area of interest.

6 Conclusions

An approach is presented to estimate a linear multivariable model on the basis of noisy frequency domain data using a two-norm minimization of a weighted curve fit error. The weighting on the curve fit error can be specified by either an input/output or an element-by-element frequency dependent multivariable weighting function. The multivariable model is parametrized in either a left or right polynomial matrix fraction description wherein structural parameters allow the specification of both full polynomial or none-full polynomial descriptions. The computational procedure is able to estimate complex models by using an iterative procedure of solving weighted multivariable least squares problems and exploits the structure of the least squares problem, thereby reducing any computation and memory requirements directly. The curve is demonstrated on experimental multivariable frequency domain data obtained from a Wafer Stepper system having 3 inputs and 3 outputs.

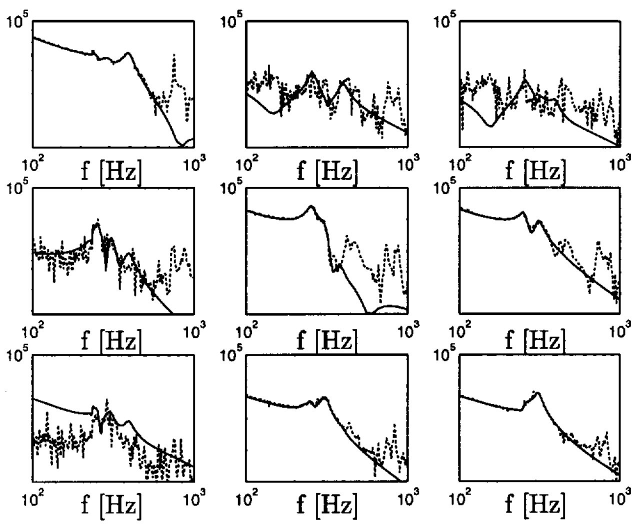

Fig. 2: Amplitude Bode plot of 18th order discrete time model $P(e^{i\omega_j},\hat{\theta})$ and the data $G(\omega_j)$.

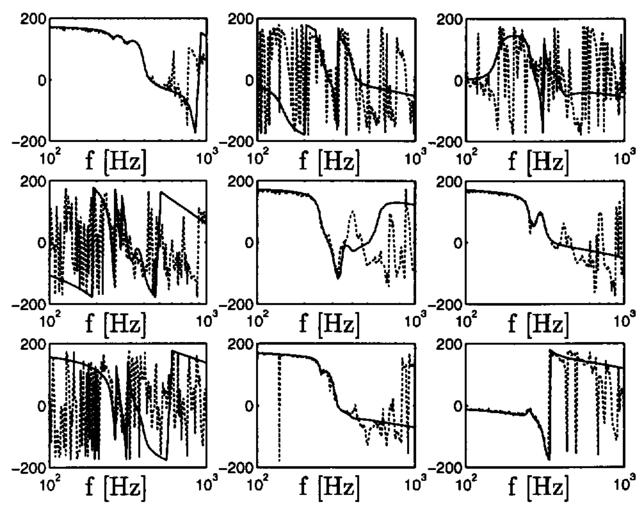

Fig. 3: Phase Bode plot of 18th order discrete time model $P(e^{i\omega_j},\hat{\theta})$ and the data $G(\omega_j)$.

7 Acknowledgements

The authors would like to acknowledge the support of a wafer stepper experimental setup by the Philips' Research Laboratory in Eindhoven, the Netherlands. Furthermore we like to thank Erwin Walgers, for his contribution to the experimental part of this paper.

References

[1] D.S. Bayard. High order multivariable transfer function curve fitting: algorithms, sparse matrix methods and experimental results. *Automatica,* Vol. *30,* No. *9,* pp. 1439–1444, 1994.

[2] R. Bellman. *Introduction to Matrix Computations.* McGraw–Hill, NY, 1970.

[3] C.T. Chen. *Linear System Theory and Design.* CBS College Publishing, NY, 1984.

[4] R.L. Dailey and M.S. Lukich. MIMO transfer function curve fitting using Chebyshev polynomials. In *SIAM 35th Anniversary Meeting,* Denver, USA, 1987.

[5] J.H. Friedman and P.P. Khargonekar. Identification of lightly damped systems using frequency domain techniques. In *Prepr. 10th IFAC Symp. on System Identification,* pp. 201–206, Copenhagen, Denmark, 1994.

[6] M.R. Gevers. ARMA models, their Kronecker indices and their McMillan degree. *Int. Journal of Control,* Vol. *43,* No. *6,* pp. 1745–1761, 1986.

[7] G.H. Golub and C.F. Van Loan. *Matrix Computations.* Johns Hopkins University Press, Baltimore, 1989.

[8] G. Gu and P.P. Khargonekar. A class of algorithms for identification in H_∞. *Automatica,* Vol. *28,* pp. 299–312, 1992.

[9] R.G. Hakvoort and P.M.J. Van den Hof. Frequency domain curve fitting with maximum amplitude criterion and guaranteed stability. *Int. Journal of Control,* Vol. *60,* No. *5,* pp. 809–825, 1994.

[10] E.C. Levi. Complex–curve fitting. *IEEE Trans. on Automatic Control, AC–4,* pp. 37–44, 1959.

[11] P.L. Lin and Y.C. Wu. Identification of multi–input multi–output linear systems from frequency response data. *Journal of Dynamic Systems, Measurements and Control,* Vol. *104,* pp. 58–64, 1982.

[12] L. Ljung. Some results on identifying linear systems using frequency domain data. In *Proc. 32nd IEEE Conference on Decision and Control,* pp. 3534–3538, San Antonio, USA, 1993.

[13] T. McKelvey. Frequency weighted subspace based system identification in the frequency domain. In *Proc. 34th IEEE Conference on Decision and Control,* pp. 1228–1233, New Orleans, USA, 1995.

[14] R. Pintelon, P. Guillaume, Y. Rolain, J. Schoukens, and H. Van hamme. Parametric identification of transfer functions in the frequency domain – a survey. *IEEE Trans. on Automatic Control, AC-39,* pp. 2245–2260, 1994.

[15] C.K. Sanathanan and J. Koerner. Transfer function synthesis as a ratio of two complex polynomials. *IEEE Trans. on Automatic Control, AC-8,* pp. 56–58, 1963.

[16] P.M.J. Van den Hof. System order and structure indices of linear systems in polynomial form. *Int. Journal of Control,* Vol. *55,* No. *6,* pp. 1471–1490, 1992.

[17] A.H. Whitfield. Asymptotic behaviour of transfer function synthesis methods. *Int. Journal of Control,* Vol. *45,* No. *3,* pp. 1083–1092, 1987.