System Identification Using Balanced Parameterizations

Chun Tung Chou and Jan M. Maciejowski, *Senior Member, IEEE*

***Abstract*—Some general issues in the "black-box" identification of multivariable systems are first discussed. It is then suggested that balanced parameterizations can be used to give identifiable forms. A particular advantage is that balanced parameterizations are known for several useful classes of linear dynamic models, including stable minimal models, minimum-phase models, positive-real models, and normalized coprime factor models. Before optimizing the parameters of balanced parameterizations, an initial model must be found. We use realization-based methods and so-called "subspace" methods for this purpose. These methods are very effective at finding accurate initial models without preliminary estimation of various structural indexes. The paper ends with two simulation examples, which compare the use of balanced parameterizations with more traditional ones, and three "real" examples based on practical problems: a distillation column, an industrial dryer, and the (irrational) spectrum of sea waves.**

***Index Terms*—Balanced parameterizations, identifiable forms, multivariable systems, parameter estimation, system identification.**

I. Introduction

THE identification of multivariable systems is of extreme importance in practice. The majority of systems being manipulated or controlled, whether they are industrial processes, vehicles, aircraft, consumer products or, in a different arena, markets and economies, have more than one input and more than one output and usually exhibit significant interactions between each other. The very widespread tendency to control each variable separately, pretending that the others do not exist, is a compromise which to some extent has been driven by the difficulty of obtaining multivariable models.

Currently, commercial pressures in the process industries put a considerable premium on the achievement of maximal efficiency and quality, and a key to achieving this is to implement multivariable control, which exploits the interactions between variables. It is now not uncommon, for example in the petrochemical sector, to find multivariable controllers handling 30 or 40 variables simultaneously. An essential ingredient of such a control scheme is a multivariable model of the process being controlled.

An important requirement in some applications is that the identification algorithm should, as far as possible, work routinely, automatically, and reliably, with no human intervention. Current methods cannot yet deliver this requirement, but the methods presented in this paper seem to be more promising in this respect than others, at least for multivariable problems. One can expect that it will always be necessary to monitor the results of an identification algorithm, either by a human or automatically. But, it should be possible to avoid the need for human decision making as part of the identification algorithm.

Conceptually, the identification problem for multivariable systems is very similar to that for single-variable systems. But, it is known from system theory that, whereas in the single-variable case there is only one "structural" decision to be made, namely the order of the model, in the multivariable case a relatively large number of other "structural" decisions is required. Making these decisions has always been a stumbling block for multivariable identification and has certainly made it difficult to move toward methods which could be applied "routinely and automatically." A major advance of the past ten years has been the development of so-called "subspace methods" for system identification. These appear to be very successful at finding multivariable models without making such decisions explicitly. As will be seen later in the paper, we use such methods to obtain initial models, and these methods undoubtedly make a great contribution to the success of the complete identification scheme which we present here.

In the next three sections we motivate estimation of parameters in "balanced" parameterizations. This is essentially the same as conventional parametric identification, except that we advocate an unusual choice of parameters. Then, we discuss briefly the methods we use to obtain initial models. Estimation of parameters involves the optimization of the parameters of these initial models. We then show how to compute gradients, which are needed for parameter estimation, when balanced parameterizations are used. Finally, we present some examples. These are chosen to illustrate the use of balanced parameterizations for various classes of models. Some simulation examples are also included, which compare the use of a balanced parameterization with the use of a more conventional observable parameterization. It is assumed that the reader is familiar with the basics of conventional

Manuscript received June 23, 1995. Recommended by Associate Editor, V. Solo. This work was supported by the European Research Network on System Identification (ERNSI), under European Commission Contract SC1*CT920779, and by Trinity College, Cambridge. The data for the first example was provided by SAST, Ltd. and for the second example by Cambridge Control, Ltd.

C. T. Chou is with the Department of Electrical Engineering, Systems and Control Laboratory, Delft University of Technology, 2600 GA Delft, The Netherlands.

J. M. Maciejowski is with the Engineering Department, Cambridge University, Cambridge, CB2 1PZ, U.K. (e-mail: jmm@eng.cam.ac.uk).

Publisher Item Identifier S 0018-9286(97)05064-2.

parametric identification, as described in books such as [8], [27], [38], and [46].

It should be noted that an identification scheme for multivariable systems which is very similar to ours in outline was already proposed by Akaike [1]. There, initial models were obtained by a kind of "subspace" method, and they were improved upon by using nonlinear optimization to obtain maximum likelihood estimates, though without using balanced parameterizations.

II. Problem Setting

Our objective is to find linear time-invariant (LTI) dynamic models for multivariate data. Mostly, we assume that the data is available in the form of an input vector time series $\{u(t)\colon u(t) \in \mathbf{R}^m, t = 1, \cdots, N\}$ and an output vector time series $\{y(t)\colon y(t) \in \mathbf{R}^p, t = 1, \cdots, N\}$. This corresponds to measurements taken from some process which is to be identified. Sometimes only the output measurements are available, and the objective is to find a model which produces an output with the same spectrum, or second-order statistics, as the observed output, when the model is driven by white noise. A variation of this problem occurs when the output data itself is not available, but its spectrum or autocovariance (either exact or approximate) is given; this is called the "stochastic realization" problem.

A relatively novel use of identification is as a way of obtaining approximate linear models from complex, detailed, nonlinear models. Such nonlinear models are being built increasingly frequently for industrial processes and have long been used in macroeconomics. The idea is to apply identification methods to data generated by the nonlinear model, as if it were data generated by some industrial plant. A very important difference, however, is that the experimental conditions can be chosen freely; for example, the input signal can be chosen to be an impulse, which is not usually possible with a real system. Approximate linearization of a macroeconomic model using identification techniques was treated in [28]. There only an "initial model," in the terminology of this paper, was found, with no subsequent optimization of parameters.

In this paper we shall represent LTI multivariable systems as state-space models

$$x(t+1) = A_d x(t) + B_d u(t) + K_d e(t) \tag{1}$$

$$y(t) = C_d x(t) + D_d u(t) + e(t). \tag{2}$$

It will be assumed that nothing is known about the internal structure of the model so that the state vector cannot usually be given a physical interpretation. We have used the subscript "d" to emphasize that the model description assumes discrete time. Later in the paper we shall be jumping between discrete and continuous-time representations, and this notation will help to avoid confusion there. The term $e(t)$ is a multivariable white noise process.

We will usually assume that we are only interested in *asymptotically stable* models. This assumption of stability is usually required either because we know *a priori* that the system being modeled is stable or because we make some further assumptions which imply stability (usually that $\{u(t)\}$ and $\{y(t)\}$ are stationary stochastic processes and that the effects of any initial conditions are negligible).

There are several reasons for using state-space models. Transfer function models are very natural and convenient in the single-input/single-output (SISO) case, namely when $m = p = 1$. But in the multivariable case, they have the drawback that each element of the transfer function matrix is independent of all the others. This means that it is difficult to exploit any common behavior that may be present in the real system—which is usually the case—when identifying a model. Each element has to be identified separately from the others, in essence. ARMAX models do not suffer from this problem, since they can represent dynamic behavior which is common to the various input–output paths, as well as any dynamics common to the "input–output" and the "noise-output" behavior. But in the multivariable case, the use of ARMAX models becomes quite complicated. For example, enforcement of the "strictly proper" condition, or of causality, is much more complicated than in the SISO case.

Another problem which appears with ARMAX models is that of needing an "identifiable parameterization." The observed input–output behavior of a system can only determine the transfer function of a model, not a particular representation of that transfer function. But there are infinitely many different ARMAX models which correspond to the same transfer function. It is usually considered essential to fix a representation in such a way that the identification algorithm is prevented from wandering over a family of ARMAX models, all of which correspond to the same transfer function. This is certainly necessary for practical reasons; otherwise, the algorithm may wander into representations involving enormous numerical values for the coefficients, for instance, and run into numerical difficulties caused by the limitations of finite precision. An alternative way of dealing with this problem is by "regularization" [34]; the conditioning of the estimation is improved, for example, by penalizing excessive norms of the parameter vector. It is more controversial whether an identifiable parameterization is needed for statistical reasons. Either identifiability, or some alternative way of obtaining convergence (such as regularization) seems to be necessary if the statistical properties of parameter estimates (namely, the coefficients in the model) are to be analyzed. But it is questionable whether such properties are of interest. Statistical properties of the transfer function estimates, and of estimates of system invariants such as pole locations, are certainly important, but there are results which show that these properties do not depend on the particular parameterization or even on whether a parameterization is identifiable or not [27], [44].

State-space models share with ARMAX models the benefit of representing common dynamics and the problem of requiring an identifiable parameterization. They are no better than ARMAX models in this respect, but also no worse in the multivariable case. (Whereas in the SISO case the problems with ARMAX models can be dealt with more easily.) State-space models have some mild advantages: 1) causality is built into the model structure; 2) the "strictly proper"

property can be determined by inspection; 3) in general the best and most reliable numerical algorithms for system transformation, simplification, and analysis, as well as the exploitation of models in the synthesis of filters and control systems, assume that the model is given in state-space form. A much stronger advantage of state-space models, as we shall show later in this paper, is that there exist very nice identifiable parameterizations of them. In particular, it is known how to parameterize some specific classes of models which are of interest in applications. But perhaps the strongest argument in favor of state-space models is the empirical fact that we can assemble identification procedures for them which work well and reliably.

It is usually desirable to find as simple a model as possible for the data. This means first that the final state-space model should be minimal. It also means that the minimal state dimension (McMillan degree), n, should be as small as possible, given the data. Actually, there is an implicit (sometimes explicit) tradeoff between the acceptable model complexity and how well it matches the data.

It is worth repeating that it is assumed in this paper that we are concerned with "black-box" identification, namely that we have no information on the internal structure of the real system which generated the data. If such information is available, then the methods presented here are not appropriate because they do not preserve structural information of this kind.

III. Identifiable Parameterizations

A parameterization of a model is *identifiable* if only one parameter (vector) corresponds to each transfer function, that is, if the map from parameters to input–output behavior is injective. Consider the "purely deterministic" state-space model (1), (2) when there is no noise input $e(t)$

$$x(t+1) = A_d x(t) + B_d u(t) \tag{3}$$

$$y(t) = C_d x(t) + D_d u(t). \tag{4}$$

It is known that the set of all possible input–output behaviors of this system is a manifold of dimension $(m+p)n+mp$ [7]. However, the number of elements in the A_d, B_d, C_d, and D_d matrices is $(n+m+p)n+mp$, so it is clear that n^2 degrees of freedom must be removed if parameter identifiability is to be achieved. It is most often suggested that this restriction be achieved by setting most of the parameters to zero or one and using one of the standard reachable or observable forms [17], [23].

As the parameters of such a form range over the real numbers, both unstable and nonminimal systems will be encountered. Most commonly, the system being identified is known to be stable, so it is clearly desirable to avoid estimating an unstable model. Avoiding unstable models is not usually a problem with algorithms which rely on minimizing some cost function because the value of the cost function usually becomes very large for an unstable model. Avoiding nonminimal models involves the avoidance of an algebraic set in the $(m+p)n$-dimensional parameter space associated with the matrices A_d, B_d, and C_d. Nonminimal models are locally unidentifiable in the sense that at each such model there is a manifold (embedded in the algebraic set mentioned above) in the parameter space on which the transfer function remains invariant. As mentioned earlier, such models must either be avoided, or the estimation algorithm must take measures to deal with the consequent ill-conditioning.

The issue of which identifiable form to use, and even of whether to use one at all, is far from settled. True canonical forms (in the algebraic sense) partition the set of systems so that each system can be represented by exactly one set of parameter values, but several (many!) structural forms are needed to cover the set of all systems [10], [45]. Glover and Willems [14] have argued that this is an avoidable problem; one can use "overlapping" or "pseudo-canonical" forms which have the advantage that each form can represent the majority of systems—typically, each form can represent "almost all" systems, in the measure-theoretic sense—but now each system can be represented in several such forms. Each system still has a unique representation in each form, however, which is all that matters for identification. Hanzon and Ober [18], [19] have proposed an overlapping form which bears some resemblance to the balanced form, which will be introduced in the next section, which is designed to represent systems in the neighborhood of those at which the structure of the balanced form changes. McKelvey [34] has argued that it is unnecessary to use an explicit parameterization at all; it is enough to "regularize" the parameter estimation problem. In fact, one of the schemes he proposes estimates a system which is close to being balanced. A nice discussion and resumé of much of the relevant literature is given by Veres [49].

It is worth mentioning that much of the discussion which appears in the literature about parameterizations of systems is concerned with the difficulties of estimating the correct "structure," namely which particular form of a parameterization is most appropriate, given a set of data. We believe that currently available methods for obtaining initial models, such as those described later in Section VI-B, help to reduce the severity of these problems. Evidence which we have for this is that if we deliberately generate data using a model which does not have the generic structure and then apply our identification methods to the data under the assumption that the structure *is* generic, then the algorithms operate successfully and exhibit no difficulties [6]. The probable explanation for this is that in the multivariable case the space of all stable systems of given input and output dimensions and a given McMillan degree are connected (in contrast to the SISO case) [4], [15], [40] so that the algorithms have no problem with working in the immediate neighborhood of systems with nongeneric structure, since the parameterization remains reasonably well-conditioned in nearly all directions in the parameter space. (This should be equally true for all algorithms.) When there are several generic structures to choose from, as is the case with balanced parameterizations in the SISO case, then the correct choice of such a structure is important, since the algorithm has no possibility of reaching the neighborhood of the correct system if an incorrect choice of structure is made. It appears that this can be done successfully by examining the structure of initial models which are obtained using the

methods discussed in Section VI-B. Some evidence for this is mentioned in Section VII.

IV. Balanced Parameterizations

The concept of balanced state-space realizations was introduced by Moore [36]. He showed that given any continuous-time stable state-space realization of an LTI system

$$\dot{x}(t) = A_c x(t) + B_c u(t) \tag{5}$$
$$y(t) = C_c x(t) + D_c u(t) \tag{6}$$

there exists a state coordinate transformation T which brings this state-space realization to the balanced form, namely a state-space realization which has equal and diagonal reachability and observability Gramians. The diagonal elements of the balanced Gramians are known as Hankel singular values and are input–output invariants.

One can define balanced realizations for discrete-time stable systems in a similar way. From a result of Mullis and Roberts [37], it follows that a digital filter implementation based on balanced realizations has minimum round-off noise due to finite word-length effects. Also, Thiele [47] has shown that balanced realizations minimize a particular sensitivity measure of the transfer function due to variations in the entries of the state-space matrices.

A state-space system which is close to being nonminimal gives rise to an ill-conditioned parameter estimation problem, in the sense that perturbations of the parameter estimates in certain directions (in the parameter space) have very little effect on the input–output behavior of the estimated model; in a nonminimal model they would have no effect. The consequence of this is that the accuracy of parameter estimation (as measured by the variance of the parameter estimates, for example) is low. In a sense, the use of a balanced realization gives an estimation problem which is as well-conditioned as possible *for a given transfer function*. We see that this occurs from the following result [29], [36]: if $\underline{\sigma}[G_r]$ denotes the smallest singular value of the reachability Gramian, and $\underline{\sigma}[G_o]$ denotes the smallest singular value of the observability Gramian of a system, then the balanced realization of a system is the one which maximizes the value of $\min(\underline{\sigma}[G_r], \underline{\sigma}[G_o])$. In this sense the balanced realization of a given system is the one which is "furthest away" from nonminimality.

All minimal, stable, balanced continuous-time systems can be parameterized explicitly. Furthermore, this parameterization gives a true canonical form, in the sense that to each transfer function there corresponds a unique parameter vector and that every minimal stable system of given dimensions can be represented by some parameter vector [22], [39], [52]. A complete explicit parameterization was first derived by Ober, and the particular form we use here is taken from [42].

There are two types of parameters within this parameterization—structural parameters and real-valued parameters. The structural parameters are integer-valued, being the number of distinct Hankel singular values, k, the multiplicity of the jth Hankel singular value, n_j ($j = 1, \cdots, k$), and the rank r_j of blocks B_j and C_j ($j = 1, \cdots, k$), which comprise the "B_c" and "C_c" matrices. Each set of structural parameters specifies a certain "structure" of a system, whereas each set of admissible real-valued parameters specifies a particular (stable, minimal and balanced) state-space realization, given such a structure. Note that the number of real-valued parameters required to specify a system varies from structure to structure. We shall use the term *generic structure* to refer to the structure which occurs when all the Hankel singular values are distinct ($k = n$, $n_j = 1$, $r_j = 1$). This structure is generic in the sense that almost all of the $p \times m$ systems of McMillan degree n can be represented in this form. The number of independent real-valued parameters needed to specify a system within the generic structure is $n(m + p) + mp$, which is in fact the dimension of the manifold of all $p \times m$ stable systems of McMillan degree n.

For system identification using balanced parameterizations, the generic structure is usually sufficient, since almost all systems can be represented in this form. The generic structure has a much simpler parameterization compared with the nongeneric ones; since we need the generic structure, we give details here only for this case.

In this case, the parameterization is given in terms of the parameters as follows:

$$\begin{array}{ccccccc} \sigma_1 & > \cdots > & \sigma_j & > \cdots > & \sigma_n & \sigma_j \in \mathbf{R} \\ U_1 & \cdots & U_j & \cdots & U_n & U_j \in \mathbf{R}^{p\times 1} \\ \tilde{B}_1 & \cdots & \tilde{B}_j & \cdots & \tilde{B}_n & \tilde{B}_j \in \mathbf{R}^{1\times m} \\ \tilde{D} & & & & & \tilde{D} \in \mathbf{R}^{p\times m} \end{array}$$

where $\sigma_n > 0$, $U_j^T U_j = 1$, $\tilde{B}_j \neq 0$, and the first nonzero element of $\tilde{B}_j$ is positive.

Every stable continuous-time transfer function of McMillan degree n, with distinct Hankel singular values, can be realized by (A_c, B_c, C_c, D_c) as follows:

$$A_c = [a_{ij}]$$

where

$$a_{ij} = \begin{cases} -\dfrac{1}{2\sigma_j}\tilde{B}_j\tilde{B}_j^T, & \text{for } i = j \\ \dfrac{1}{\sigma_i^2 - \sigma_j^2}(\sigma_j\tilde{B}_i\tilde{B}_j^T - \sigma_i C_i^T C_j), & \text{for } i \neq j \end{cases} \tag{7}$$

$$B_c = \begin{bmatrix} \tilde{B}_1 \\ \vdots \\ \tilde{B}_n \end{bmatrix} \tag{8}$$

$$C_c = [C_1 \quad \cdots \quad C_j \quad \cdots \quad C_n]$$

where

$$C_j = U_j\sqrt{\tilde{B}_j\tilde{B}_j^T} \tag{9}$$

$$D_c = \tilde{D}. \tag{10}$$

It appears from this parameterization that once one has decided to use the generic structure, then there are no more integer parameters to be specified. However, this is only true when there is more than one output. In the single-output case ($p = 1$), the parameters U_j "degenerate" into integer parameters; they become scalars, and under the constraints that $U_j^T U_j = 1$, they can only take the values of $s_j =$

± 1. This results in 2^n different structures for the single-output case (whose union is generic in the sense that distinct singular values occur generically). This is in contrast with the observable form, for which there is only one structure in the single-output case and $\binom{n-1}{p-1}$ generic structures in the multioutput case [27].

With the generic structure the parameters U_j (which are orthogonal matrices in the nongeneric case) become unit vectors. For the multioutput ($p > 1$) case, if one parameterized them directly, i.e., $U_j^T = [u_{j,1} \ \cdots \ u_{j,p}]^T$, then one would also have to introduce the extra nonlinear equality constraints $u_{j,1}^2 + \cdots + u_{j,p}^2 = 1$ for $1 \leq j \leq n$. However, this can be avoided by parameterizing these unit vectors using polar coordinates, as in (11), shown at the bottom of the page, and the following bounds on the parameters: $-\pi/2 < \phi_{j,1} < 3\pi/2$ and $-(\pi/2) < \phi_{j,i} < \pi/2$ for $2 \leq i < p$. Now it is known that the set of all $p \times 1$ unit vectors (which is geometrically equivalent to the "hyper-sphere" S^{p-1}) cannot be covered by one coordinate chart, and the polar coordinates are no exception. In other words, some of the unit vectors cannot be described by the above parameterization; these are vectors of the form $(0, u_{j,2}, \cdots, u_{j,p})$ where $u_{j,2} > 0$. So, some nonlinear equality constraints have been removed at the expense of neglecting part of the parameter space. However, if one really wants to cover the whole parameter space, then an alternative idea is to use an adaptive type of local coordinates. This can be described as follows: if U_j^0 is the current estimate of U_j, one can view U_j^0 as the point with coordinates $(u_{j,1}^0, \cdots, u_{j,p}^0)$ on the sphere S^{p-1}, and the idea is to parameterize the "hemisphere" on S^{p-1} whose pole is U_j^0. All the points $\overline{U}_j$ on this "hemisphere" can be described by

$$\overline{U}_j = MU_j \tag{12}$$

where U_j is as defined in (11) but with the constraint on the first parameter changed to $-\pi/2 < \phi_{j,1} < \pi/2$, and M a $p \times p$ unitary matrix whose first column is U_j^0. In order to show that $\overline{U}_j$ parameterizes the said "hemisphere," we need to verify that the angle between the vectors $\overline{U}_j$ and U_j^0 is always less than $\pi/2$. Let α be this angle, then

$$\begin{aligned} \cos(\alpha) &= (U_j^0)^T \overline{U}_j && (13)\\ &= (Me_1)^T MU_j \\ &\qquad \text{(by definition of } M \\ &\qquad\quad \text{and since } e_1 = [1, 0, \cdots, 0]^T) && (14)\\ &= \cos(\phi_{j,p-1}) \cos(\phi_{j,p-2}) \cos(\phi_{j,p-3}) \\ &\quad \cdots \cos(\phi_{j,2}) \cos(\phi_{j,1}) \\ &\qquad \text{(since } M \text{ is unitary)}. && (15) \end{aligned}$$

Hence, $\cos(\alpha)$ is strictly positive and the result follows. Finally, we remark that the matrix M can easily be found by applying QR factorization to the vector U_j^0.

An advantage of using a balanced parameterization for system identification is now apparent. One only has to check a set of simple linear inequality constraints—that $\sigma_1 > \cdots > \sigma_n > 0$, that the first nonzero element of each row of B is positive, and that the bounds on $\phi_{i,j}$ hold—in order to see whether the model is minimal and stable. Since each of these parameters can vary over a connected open subset of the real line, it is possible to replace it by another unconstrained parameter. For example, the inequality constraints between the Hankel singular values (σ_i) can be avoided by using the alternative parameters $\tau_i = \sigma_i - \sigma_{i+1}, (i = 1, \cdots, n-1), \tau_n = \sigma_n$, and even the positivity constraint could be removed by using $\log \tau_i$.

Note that the parameterization described above only holds for continuous-time systems. We have used the subscript c in this section to emphasize that the parameterization is valid for continuous-time systems only. One can obtain an implicit parameterization for discrete-time systems by means of the usual bilinear transformation between the s and z domains

$$z = \frac{1+s}{1-s} \tag{16}$$

$$s = \frac{z-1}{z+1}. \tag{17}$$

This transformation preserves stability, minimality, and balancing.

We finish this section by showing what the balanced parameterization looks like in a given case. We consider $n = 4$, $m = 2$, $p = 3$, and we assume the generic case. Notice that although this canonical form looks much more complicated than the more usual observable form, it has exactly the same number of free parameters (20, with 6 more in the "D_c" matrix); see (18)–(20), shown at the bottom of the next page. Here B_i is the ith row of B_c, and C_i is the ith column of C_c.

V. Useful Classes of Models

The parameterization presented in the previous section is useful when it is known *a priori* that the process being identified is stable and when an output-error criterion is used, namely that the model is

$$x(t+1) = A_d x(t) + B_d u(t) \tag{21}$$

$$y(t) = C_d x(t) + D_d u(t) + \nu(t). \tag{22}$$

Depending on the nature of the process being identified and on how the model is to be used, some other classes of models

$$U_j = \begin{bmatrix} u_{j,1} \\ u_{j,2} \\ u_{j,3} \\ \vdots \\ u_{j,p-1} \\ u_{j,p} \end{bmatrix} = \begin{bmatrix} \cos(\phi_{j,p-1}) \cos(\phi_{j,p-2}) \cos(\phi_{j,p-3}) \cdots \cos(\phi_{j,2}) \cos(\phi_{j,1}) \\ \cos(\phi_{j,p-1}) \cos(\phi_{j,p-2}) \cos(\phi_{j,p-3}) \cdots \cos(\phi_{j,2}) \sin(\phi_{j,1}) \\ \cos(\phi_{j,p-1}) \cos(\phi_{j,p-2}) \cos(\phi_{j,p-3}) \cdots \sin(\phi_{j,2}) \\ \vdots \\ \cos(\phi_{j,p-1}) \sin(\phi_{j,p-2}) \\ \sin(\phi_{j,p-1}) \end{bmatrix} \tag{11}$$

are also useful in system identification. In this section, we shall examine three such classes of linear models: minimum-phase, positive-real, and coprime factor models. We shall discuss how the concept of "balancing" can be extended to these classes of models and present a balanced parameterization for each one of them.

A. Minimum-Phase Models

One of the uses of models is to predict future system responses. If the prediction is based entirely on past inputs, then a stable model is appropriate. However, if it is based on both past inputs and past outputs, then a stable and minimum-phase model is required. A particular way to write down this model, if the system equations are given by (1) and (2), is the one-step-ahead predictor in innovation form [17], [27]

$$\hat{x}(t+1) = A_d\hat{x}(t) + B_d u(t) + K_d e(t) \tag{23}$$

$$\hat{y}(t) = C_d\hat{x}(t) + D_d u(t) \tag{24}$$

$$y(t) = \hat{y}(t) + e(t). \tag{25}$$

In this case, one requires the eigenvalues of both A_d and $A_d - K_dC_d$ to lie inside the unit circle, or equivalently, the system with realization (A_d, K_d, C_d, I) to be stable and minimum-phase.

One can define minimum-phase balanced realizations in a way analogous to Moore's notion of balanced realizations for stable systems. There are several different ways of defining minimum-phase balanced realizations [33], [42], and the following definition is taken from the latter reference: let $\mathcal{S}$ denote the realization of a continuous-time stable and minimum-phase system and $\mathcal{S}^{-1}$ denote the inverse realization. Then $\mathcal{S}$ is said to be continuous-time minimum-phase balanced if the reachability Gramian of $\mathcal{S}$ and the observability Gramian of $\mathcal{S}^{-1}$ are equal and diagonal. Discrete-time minimum-phase balanced realizations are defined analogously.

McGinnie [33] has derived a minimum-phase balanced parameterization based on the above definition. Interestingly, the parameterization so obtained can be described by the same set of parameters used in Section IV. Moreover, the state-space matrices of this minimum-phase parameterization have the same structures as those described in Section IV (in the generic case).

Using the same notation as in Section IV, except that we relabel all the B's by K's and replace m by p, all $p \times p$ continuous-time stable and minimum-phase transfer functions of McMillan degree n have a minimal state-space realization (A_c, K_c, C_c, D_c) whose generic form is given by

$$A_c = [a_{ij}] \tag{26}$$

where

$$a_{ij} = \begin{cases} -\dfrac{1}{2\sigma_i}\tilde{K}_i\tilde{K}_i^T, & \text{if } i = j \\ \dfrac{1}{\sigma_i^2 - \sigma_j^2}\Big[\sigma_j(1+\sigma_i^2)\tilde{K}_i\tilde{K}_j^T \\ \quad -\sigma_i\sqrt{1+\sigma_i^2}\sqrt{1+\sigma_j^2} \\ \quad \cdot\sqrt{\tilde{K}_i\tilde{K}_i^T}\sqrt{\tilde{K}_j\tilde{K}_j^T}U_iU_j^T\Big], & \text{if } i \neq j \end{cases} \tag{27}$$

$$K_c = \begin{bmatrix} \tilde{K}_1 \\ \vdots \\ \tilde{K}_n \end{bmatrix} \tag{28}$$

$$C_c = \tilde{D}\tilde{C}_c$$

where

$$\tilde{C}_c = [\tilde{C}_1 \quad \cdots \quad \tilde{C}_j \quad \cdots \quad \tilde{C}_n]$$

and

$$\tilde{C}_j = \left(\sigma_j\tilde{K}_j^T + \sqrt{1+\sigma_j^2}U_j\sqrt{\tilde{K}_j\tilde{K}_j^T}\right) \tag{29}$$

$$D_c = \tilde{D} \text{ where } \tilde{D} \text{ is invertible.} \tag{30}$$

Note that the constraints stated in Section IV also apply here, and in addition $\tilde{D}$ is required to be invertible. This parameterization is valid for continuous-time systems, but

$$A_c = \begin{bmatrix} -\dfrac{\|B_1\|^2}{2\sigma_1} & \dfrac{\sigma_2B_1B_2^T - \sigma_1C_1^TC_2}{\sigma_1^2-\sigma_2^2} & \dfrac{\sigma_3B_1B_3^T - \sigma_1C_1^TC_3}{\sigma_1^2-\sigma_3^2} & \dfrac{\sigma_4B_1B_4^T - \sigma_1C_1^TC_4}{\sigma_1^2-\sigma_4^2} \\ \dfrac{\sigma_1B_2B_1^T - \sigma_2C_2^TC_1}{\sigma_2^2-\sigma_1^2} & -\dfrac{\|B_2\|^2}{2\sigma_2} & \dfrac{\sigma_3B_2B_3^T - \sigma_2C_2^TC_3}{\sigma_2^2-\sigma_3^2} & \dfrac{\sigma_4B_2B_4^T - \sigma_2C_2^TC_4}{\sigma_2^2-\sigma_4^2} \\ \dfrac{\sigma_1B_3B_1^T - \sigma_3C_3^TC_1}{\sigma_3^2-\sigma_1^2} & \dfrac{\sigma_2B_3B_2^T - \sigma_3C_3^TC_2}{\sigma_3^2-\sigma_2^2} & -\dfrac{\|B_3\|^2}{2\sigma_3} & \dfrac{\sigma_4B_3B_4^T - \sigma_3C_3^TC_4}{\sigma_3^2-\sigma_4^2} \\ \dfrac{\sigma_1B_4B_1^T - \sigma_4C_4^TC_1}{\sigma_4^2-\sigma_1^2} & \dfrac{\sigma_2B_4B_2^T - \sigma_4C_4^TC_2}{\sigma_4^2-\sigma_2^2} & \cdots\dfrac{\sigma_3B_4B_3^T - \sigma_4C_4^TC_3}{\sigma_4^2-\sigma_3^2} & -\dfrac{\|B_4\|^2}{2\sigma_4} \end{bmatrix} \tag{18}$$

$$B_c = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \\ b_{41} & b_{42} \end{bmatrix} \tag{19}$$

$$C_c = \begin{bmatrix} \cos(\phi_{12})\cos(\phi_{11})\|B_1\| & \cos(\phi_{22})\cos(\phi_{21})\|B_2\| & \cos(\phi_{32})\cos(\phi_{31})\|B_3\| & \cos(\phi_{42})\cos(\phi_{41})\|B_4\| \\ \cos(\phi_{12})\sin(\phi_{11})\|B_1\| & \cos(\phi_{22})\sin(\phi_{21})\|B_2\| & \cos(\phi_{32})\sin(\phi_{31})\|B_3\| & \cos(\phi_{42})\sin(\phi_{41})\|B_4\| \\ \sin(\phi_{12})\|B_1\| & \sin(\phi_{22})\|B_2\| & \sin(\phi_{32})\|B_3\| & \sin(\phi_{42})\|B_4\| \end{bmatrix} \tag{20}$$

since the bilinear transformation preserves the minimum-phase property, it can be used to obtain an implicit parameterization of discrete-time minimum-phase systems. If, instead of allowing $\tilde{D}$ to be any invertible matrix, we choose

$$\tilde{D} = I - \tilde{C}_c(I - A_c + K_c\tilde{C}_c)^{-1}K_c \tag{31}$$

then the feed-through term of the corresponding discrete-time system will always be the identity matrix. In other words, this choice of $\tilde{D}$ gives us an implicit parameterization of all discrete-time innovation systems. Note that the parametric constraints given in Section IV also ensure that the above choice of $\tilde{D}$ is invertible.

The matrix B_d which appears in (23) can now be allowed to have arbitrary real entries. This allows nonminimum-phase transfer function matrices relating $u(t)$ to $y(t)$ to be identified, the minimum-phase "restriction" applying only to the transfer function between $e(t)$ and $y(t)$. This also allows the possibility that the pair (A_d, B_d) may be uncontrollable, which is reasonable if the transfer function from $u(t)$ to $y(t)$ is of McMillan degree smaller than n.

The possibility of parameterizing minimum-phase systems with almost no constraints on the parameters is a significant advantage for identification.

B. Positive-Real Models

A different situation occurs when we wish to obtain an innovation model

$$x(t+1) = A_dx(t) + K_de(t) \tag{32}$$
$$y(t) = C_dx(t) + e(t) \tag{33}$$

but instead of the measurements $y(t)$ we have only estimates of the *autocovariance* function

$$R(\tau) = Ey(t+\tau)y(t)^T \qquad (\tau = 0, 1, 2, \cdots, N) \tag{34}$$

or of the spectral density of $y(t)$ (which is the discrete Fourier transform of the autocovariance). We still want an innovations representation of $y(t)$, but now the requirement is that its autocovariance should match the given one (approximately).

Under the assumption that the model (32), (33) is stable, and letting $P = E\{x(t)x(t)^T\}$ and $Q = E\{e(t)e(t)^T\}$, it can be shown that the autocovariance sequence $\{R(0), R(1), \cdots\}$ is equal to the impulse response of the following system:

$$x(t+1) = A_dx(t) + B_du(t) \tag{35}$$
$$y(t) = C_dx(t) + R(0)u(t) \tag{36}$$

where

$$B_d = A_dPC_d^T + K_dQ \tag{37}$$
$$P = A_dPA_d^T + K_dQK_d^T \tag{38}$$
$$R(0) = C_dPC_d^T + Q. \tag{39}$$

Hence the standard approach [3], [13] to the "stochastic realization" problem is as follows.

Step 1) Find a system (A_d, B_d, C_d, D_d) which is stable and has the sequence $\{R(\tau)\colon \tau \geq 0\}$ as its impulse response.

Step 2) Solve the following positive-real Riccati equation:

$$P = A_dPA_d^T + (B_d - A_dPC_d^T) \cdot [R(0) - C_dPC_d^T]^{-1}(B_d - A_dPC_d^T)^T \tag{40}$$

[which arises by substituting (37) and (39) into (38)] for P and then find K_d from (37).

The solution of (40) is not unique, but it can be shown that if the smallest nonnegative solution is taken, then the resulting triple (A_d, K_d, C_d) gives an innovation representation [3], [13]. Note that $D_d = R(0)$.

Of course, in the context of system identification we expect to find, in Step 1), a system whose impulse response matches the given covariance sequence only approximately. But, the complication is that if the impulse response of (A_d, B_d, C_d, D_d) is not a valid autocovariance sequence, then the rest of the theory will not apply—we shall not be able to find a solution of (40). Now a sequence $\{R(\tau)\colon \tau \geq 0\}$ is a valid autocovariance sequence if and only if the Toeplitz matrix built from it is positive definite [5]. A stable system which has a positive definite impulse response is called *positive real*. To solve the approximate stochastic realization problem, we need to identify a positive-real system in Step 1).

There are several different ways of defining a balanced positive-real system [11], [16], and one of the possibilities is to assume that the minimal positive definite solutions to both (40) and its dual are equal and diagonal. Based on the continuous-time counterpart of this definition, a balanced parameterization for continuous-time positive-real systems has been derived [33], [42]. The fact that this parameterization is of continuous-time systems again does not present any limitation, since the bilinear transformation preserves the positive-real property; we thus have an implicit parameterization of discrete-time positive-real systems. As was the case with the two parameterizations already introduced above, this parameterization can also be described by the same set of parameters given in Section IV. Using the same notation as that given in Section IV with $m = p$, nearly all $p \times p$ positive-real systems (A_c, B_c, C_c, D_c) of McMillan degree n can be parameterized as follows [33]:

$$A = [a_{ij}] \tag{41}$$

where

$$a_{ij} = \begin{cases} \tilde{B}_iD_s^{-1}C_i - \dfrac{1+\sigma_i^2}{2\sigma_i}\tilde{B}_iD_s^{-1}\tilde{B}_i^T, & \text{if } i = j \\ \tilde{B}_iD_s^{-1}C_j - \dfrac{1}{\sigma_j^2 - \sigma_i^2} \cdot [\sigma_j(1-\sigma_i^2)\tilde{B}_iD_s^{-1}\tilde{B}_j^T \\ \quad -\sigma_i(1-\sigma_j^2)C_i^TD_s^{-1}C_j, & \text{if } i \neq j \end{cases} \tag{42}$$

$$D_s = \tilde{D} + \tilde{D}^T$$
$$B_c = \begin{bmatrix} \tilde{B}_1 \\ \vdots \\ \tilde{B}_n \end{bmatrix} \tag{43}$$
$$C_c = [C_1 \quad \cdots \quad C_j \quad \cdots \quad C_n]$$

where

$$C_j = D_s^{1/2} U_j \sqrt{\tilde{B}_j D_s^{-1} \tilde{B}_j^T} \tag{44}$$

$$D_c = \tilde{D}. \tag{45}$$

In addition to the constraints stated in Section IV, the following two constraints are now required in order to guarantee that the representation is positive-real: $1 > \sigma_1$ and $\tilde{D} + \tilde{D}^T > 0$.

So again, we can estimate parameters, with only mild constraints upon them, in the knowledge that we are not straying outside the class of systems which is of interest. In this case the parameters σ_i are the *canonical correlation coefficients* [1], [11]. They are a little more constrained than the corresponding parameters in the earlier parameterizations, since they are bounded above as well as below, but each of them still ranges over a connected open subset of $\mathbf{R}$, so this is still an easy constraint to handle. Remarkably, when a positive-real system is parameterized in this way, the solution to the continuous-time positive-real Riccati equation which corresponds to (40) is diag $(\sigma_1, \cdots, \sigma_n)$, and *this solution is also the solution to the discrete-time Riccati equation* (40), since solutions to these two equations are left invariant by the bilinear transformation. So, if we identify a positive-real system in this parameterization, then we can obtain K_d immediately from (37) since we already know P.

C. Normalized Coprime Factor Models

So far we have assumed that the process being modeled is stable. In the case of unstable processes, one can perform identification in closed loop, with the process stabilized by a controller. In this case the relevant class of models is that of minimal systems which are not necessarily stable. So a possible model is

$$x(t+1) = A_d x(t) + B_d u(t) \tag{46}$$

$$y(t) = C_d x(t) + D_d u(t) + \nu(t) \tag{47}$$

with (A_d, B_d, C_d, D_d) a minimal realization. One can define a minimal balanced state-space realization by assuming that the solutions to two particular algebraic Riccati equations are equal and diagonal [21], [41], and this leads to a parameterization of all continuous-time minimal systems [41], [42]. Unfortunately, one *cannot* obtain an implicit parameterization of discrete-time minimal systems using the bilinear transformation, as we have done for other model classes, because the transformation is not defined if the continuous-time system has poles at $+1$.

An alternative is to model the system using coprime factors, for example

$$\overline{y}(z) = \tilde{M}(z)^{-1} \tilde{N}(z) \overline{u}(z) + \overline{\nu}(z) \tag{48}$$

where $\tilde{N}(z)$ and $\tilde{M}(z)$ are normalized left coprime factors of the transfer function matrix, and $\overline{u}(z)$, $\overline{y}(z)$, and $\overline{\nu}(z)$ are z transforms of the signals $u(t)$, $y(t)$, and $\nu(t)$. We shall show later how an implicit parameterization of these factors can be obtained from their continuous-time counterparts.

Let $G_c(s)$ be a $p \times m$ continuous-time transfer function matrix of McMillan degree n and $\tilde{N}_c(s)$ and $\tilde{M}_c(s)$ be its normalized left coprime factors such that $G_c(s) = \tilde{M}_c(s)^{-1} \tilde{N}_c(s)$. Define the transfer function $F_c(s) = [\tilde{N}_c(s) \;\; \tilde{M}_c(s)]$, then it can be shown that $\tilde{N}_c(s)$ and $\tilde{M}_c(s)$ are normalized left coprime if and only if $F_c(s)$ is coinner. (Here "normalized" means that $\tilde{N}_c(s)\tilde{N}_c(s)^* + \tilde{M}_c(s)\tilde{M}_c(s)^* = I$ for all s, and "coinner" means that $F_c(j\omega)F_c(j\omega)^* = I$ for all ω [53].)

By using the (general form of the) stable balanced canonical form given in Section IV, Ober and McFarlane [41] derived a canonical form for all $p \times (m+p)$ coinner functions of McMillan degree n and hence obtained a canonical form for continuous-time normalized left coprime factors. Again, only the generic structure of the canonical form will be given here. Note that the canonical form given here is slightly different from that given in [41]; we have modified the expressions so that they can be given in terms of the same set of parameters used in Section IV. So, almost all $p \times (m+p)$ coinner functions (A_c, B_c, C_c, D_c) of McMillan degree n can be parameterized as follows:

$$A_c = [a_{ij}]$$

where

$$a_{ij} = \begin{cases} \dfrac{-1}{2\sigma_i(1-\sigma_i^2)} \tilde{B}_i \tilde{B}_i^T, & \text{if } i = j \\ \dfrac{-1}{\sigma_i^2 - \sigma_j^2} \cdot [\sigma_j \tilde{B}_i \tilde{B}_j^T - \sigma_i (1-\sigma_j^2) C_i^T C_j], & \text{if } i \neq j \end{cases} \tag{49}$$

$$B_c = [B_{c1} \quad B_{c2}]$$

where

$$B_{c1} \in \mathbf{R}^{n \times m}, \; B_{c2} \in \mathbf{R}^{n \times p} \tag{50}$$

$$B_{c1} = \begin{bmatrix} B_{c1,1} \\ \vdots \\ B_{c1,n} \end{bmatrix}$$

$$B_{c2} = \begin{bmatrix} B_{c2,1} \\ \vdots \\ B_{c2,n} \end{bmatrix} \tag{51}$$

$$B_{c1,i} = \tilde{B}_i S^{1/2} - \sigma_i C_i^T R^{-1/2} \tilde{D} \tag{52}$$

$$B_{c2,i} = -\tilde{B}_i S^{1/2} \tilde{D}^T - \sigma_i C_i^T R^{-1/2} \tag{53}$$

$$R = I + \tilde{D}\tilde{D}^T$$

$$S = I + \tilde{D}^T \tilde{D} \tag{54}$$

$$C_c = [C_1 \quad \cdots \quad C_n]$$

where

$$C_i = \frac{1}{\sqrt{1-\sigma_i^2}} \sqrt{\tilde{B}_i \tilde{B}_i^T} U_i \tag{55}$$

$$D_c = [D_{c1} \quad D_{c2}]$$

where

$$D_{c1} = R^{-1/2} \tilde{D} \in \mathbf{R}^{p \times m}$$

and

$$D_{c2} = R^{-1/2} \in \mathbf{R}^{p \times p}. \tag{56}$$

Since the Hankel singular values of a coinner function are less than unity, we need the constraint $1 > \sigma_1$ in addition to those stated in Section IV. Note that the state-space realizations of the coprime factors $\tilde{N}_c(s)$ and $\tilde{M}_c(s)$ are, respectively, given by $(A_c, B_{c1}, C_c, D_{c1})$ and $(A_c, B_{c2}, C_c, D_{c2})$. Note also that (A_c, B_c, C_c, D_c) is in fact balanced in the sense of Moore with balanced Gramian diag $(\sigma_1, \cdots, \sigma_n)$.

The discrete-time coprime factors can now be calculated as follows. The key point is that if one applies the bilinear transformation to a continuous-time coinner function $G_c(s)$, then the resulting discrete-time function $G_d(z)$ is also coinner. Hence if $(A_d, [B_{d1}\ \ B_{d2}], C_d, [D_{d1}\ \ D_{d2}])$ is the discrete-time realization obtained under this transformation, with both "B" and "D" matrices partitioned in the same way as their continuous-time counterparts, then $(A_d, B_{d1}, C_d, D_{d1})$ and $(A_d, B_{d2}, C_d, D_{d2})$ are realizations of the numerator and denominator, respectively, of the discrete-time normalized left coprime factors.

VI. Parameter Estimation

A. Outline of Algorithm

We propose that system identification can be carried out in the following manner. Choose a discrete-time model class which is of interest. If this is the class of asymptotically stable minimal systems, or of minimum-phase systems, or of positive-real systems, or of coprime fractions, then parameterize it using a balanced parameterization, together with the bilinear transformation. Let θ denote the vector of real-valued parameters which appear in this parameterization.

Next, choose a criterion function $V(\theta)$ which measures the difference between the available data and the performance of the model in some way.

Now estimate the parameters as follows.

Step 1) Obtain an initial model somehow.

Step 2) Set $k := 1$. From the initial model obtain the initial value of the parameter vector, θ_1, and if the number of outputs is one, the sign parameters s_j are as well.

Step 3) Set $k := k + 1$. Use an optimization algorithm to find a parameter vector θ_k such that $V(\theta_k) < V(\theta_{k-1})$. If possible, increase the efficiency of this step by computing the gradient $\nabla_\theta V(\theta)$. Section VI-C shows how this can be done if $V(\theta)$ is differentiable.

Step 4) If successfully converged, stop. Otherwise, find another initial model and return to Step 2).

This outline is a rather general one, and it leaves a number of options for the user. Since the criterion function $V(\theta)$ is a nonlinear function of its parameters, it is important that a good initial estimate is obtained so as to improve the convergence speed and to avoid convergence to a local minimum. There are many possibilities here, as outlined in the next subsection.

There is a slight restriction which arises when estimating minimum-phase models using balanced parameterizations. This is that the pair (A_c, K_c) in (26) and (28), and hence the pair (A_d, K_d) in (23), must be controllable. This implies that, in the estimated model, all poles of the "deterministic" part of the model are also poles of the "stochastic" part (though not necessarily the other way round). If the data is generated by a linear system which does not satisfy this condition, then it cannot be represented by the balanced parameterization we employ. This appears to be a considerable restriction at first. But in practice it is not a significant restriction, because the "true" model is invariably nonlinear, or at least of very high McMillan degree, and is therefore in any case not a member of the set of candidate models. In this situation one of the roles of the stochastic part of the model is to account for some of the approximation error of the deterministic part. For this purpose the stochastic part can always take advantage of all the deterministic dynamics—in effect it provides some "noise" at the deterministic inputs—and consequently the estimated models always have (A_d, K_d) controllable. This situation holds irrespective of the parameterization being employed.

If there is only one output, it is important that the sign parameters (which determine the structure) are estimated correctly. Our experience with simulation examples shows that the optimization algorithm diverges if the sign parameters are incorrect. McGinnie has shown that it is in principle possible to estimate these sign parameters directly from the data [33], though this is unlikely to be useful in practice because of the very large data sets required. However, our experience of using the "N4SID" subspace algorithm to find initial estimates for about 200 artificially generated data sets (as described in Section VII) is that it has *always* given us a correct structure to date.

B. Finding an Initial Model

A real issue, which is particularly acute in the multivariable case, is how to get started, how to find a useful initial model. Possible methods for finding initial models include the instrumental variables method "IV" [27], model reduction-based methods [54], realization-based methods [24], [55], or the recently developed "subspace" methods such as MOESP [50], [51], N4SID [48], or CVA [25], [26]. The latter two classes of methods seem to be particularly effective for finding initial models of multivariable systems. These are the methods which we use in preference to others, and we outline them very briefly.

If estimates of a (discrete-time) system's impulse response are available, then the approximate realization algorithm developed in [24], [55], which obtains a low-rank approximate factorization of the Hankel matrix built up from impulse response estimates, can be used. It is in effect a numerically robust version of the Ho–Kalman algorithm [20]. If the available data is in the form of an estimated autocovariance sequence of a discrete-time stochastic process, then a closely related algorithm is available which produces an approximate stochastic realization of that process [12].

More commonly, available data is in the form of actual input and output sequences, obtained either from experiments or gathered during normal plant operation. For this case a class of methods known as "subspace algorithms" has been developed [25], [35], [48], [50]. These work by estimating the range space of an observability matrix from a low-rank approximate factorization of a Hankel matrix built up from the input–output data. These algorithms provide models of

both deterministic and stochastic systems, namely of all the matrices in the model (1), (2).

Both the "realization" and the "subspace" algorithms have several attractive features: 1) they can be guaranteed to produce a model of the required class, and stability can be enforced [32], as can the minimum-phase property of any stochastic part of the model [48]; 2) they usually produce accurate models of the data, using relatively low state dimensions; 3) they seem to be particularly easy to use.

C. Gradient Calculations

Parameter estimation algorithms can be made more efficient, both in terms of computation speed and of accuracy, if the gradient of the objective function $V(\theta)$ with respect to the parameter vector θ is supplied to the algorithm. To compute the gradient, three calculations are required.

1) The derivatives of the objective function must be obtained in terms of the derivatives of the discrete-time model matrices.
2) The derivatives of the discrete-time model matrices must be obtained in terms of the derivatives of the continuous-time model matrices, which appear in the parameterizations.
3) The derivatives of the continuous-time matrices, with respect to the parameters, must be computed.

The first of these depends on the particular identification problem being solved, since it depends on the objective function being used. We will therefore not address this here, but confine ourselves to showing how the other calculations can be performed.

In terms of state-space transformations, the bilinear transformation can be expressed as

$$A_d(\theta) = (I + A_c(\theta))\,(I - A_c(\theta))^{-1} \tag{57}$$

$$B_d(\theta) = \sqrt{2}(I - A_c(\theta))^{-1} B_c(\theta) \tag{58}$$

$$C_d(\theta) = \sqrt{2} C_c(\theta)(I - A_c(\theta))^{-1} \tag{59}$$

$$D_d(\theta) = D_c(\theta) + C_c(\theta)(I - A_c(\theta))^{-1} B_c(\theta). \tag{60}$$

By using this, the derivatives of the discrete-time state-space matrices can be related to the derivatives of the continuous-time state-space matrices as follows:

$$\frac{\partial(A_d(\theta))}{\partial\theta_i} = (I + A_d(\theta))\,\frac{\partial(A_c(\theta))}{\partial\theta_i}\,(I - A_c(\theta))^{-1} \tag{61}$$

$$\frac{\partial(B_d(\theta))}{\partial\theta_i} = (I - A_c(\theta))^{-1}\,\frac{\partial(A_c(\theta))}{\partial\theta_i}\,B_d(\theta) + \sqrt{2}\,(I - A_c(\theta))^{-1}\,\frac{\partial(B_d(\theta))}{\partial\theta_i} \tag{62}$$

$$\frac{\partial(C_d(\theta))}{\partial\theta_i} = \sqrt{2}\,\frac{\partial(C_c(\theta))}{\partial\theta_i}\,(I - A_c(\theta))^{-1} + C_d(\theta)\,\frac{\partial(A_c(\theta))}{\partial\theta_i}\,(I - A_c(\theta))^{-1} \tag{63}$$

$$\frac{\partial(D_d(\theta))}{\partial\theta_i} = \frac{\partial(D_c(\theta))}{\partial\theta_i} + \frac{1}{\sqrt{2}}\,\frac{\partial(C_c(\theta))}{\partial\theta_i} \cdot B_d(\theta) + \frac{1}{2}\,C_d(\theta)\,\frac{\partial(A_c(\theta))}{\partial\theta_i}\,B_d(\theta) + \frac{1}{\sqrt{2}}\,C_d(\theta)\,\frac{\partial(B_c(\theta))}{\partial\theta_i}. \tag{64}$$

For the parameterization of stable systems given in Section IV, the formulas for the derivatives in the SISO case were given in [30]. A difference between the SISO and multivariable cases is that in the multi-input/multioutput (MIMO) case closed-form formulas for the derivatives are not always available; it may be necessary to solve some linear equations and some Lyapunov equations in order to obtain the derivatives [31]. However, even in the MIMO case, closed-form formulas are available when all the Hankel singular values are distinct, namely in the generic case. In the appendix we give explicit formulas for the derivatives of MIMO balanced state-space matrices given in Section IV. Derivatives for balanced parameterizations of the other classes of models can be obtained similarly.

D. Statistical Properties

For the case of estimating innovations models of the form (23)–(25) in a stochastic environment, it is appropriate to consider the statistical properties of the algorithm. Asymptotically (with the number of data points) the statistical properties of the estimated transfer function do not depend on the particular parameterization used, but do depend on the cost function $V(\theta)$ and on the statistical properties of the data [27], [44] (or, equivalently, on the properties of the input and noise signals and on the true system generating the data). If $V(\theta)$ is a "prediction-error" criterion, then under rather general conditions, as the length of data increases, the estimated parameter vector converges almost surely to the value which minimizes, over the set of models in the search space, the expected value of V. In particular, this occurs even if the true system generating the data is not in the model set, which is of relevance in view of the remarks made in Section VI-A above.

If the true system, including the noise generator, *is* in the model set, then asymptotic efficiency of the parameter vector estimate is obtained with a maximum likelihood method (e.g., a particular prediction error method in the presence of Gaussian noise) [27], [46]. "Efficiency" here refers to the performance of various estimators of the *same* set of parameters, that is, assuming a particular parameterization. It says nothing about comparative estimator performance with various parameterizations. But, such comparisons make sense only if a parameterization is chosen because the parameters have some real independent meaning (such as values of physical constants), and the estimation accuracy of individual parameters then matters. In balanced parameterizations the parameters do not have any such meaning. The asymptotic estimation accuracy depends on the inverse of the expected value of the Hessian of the criterion $V(\theta)$ at the true parameter value [27]. It is certainly possible for the true system to be such that this Hessian is arbitrarily close to singularity, even with a balanced parameterization; consider for example the innovation model (26)–(30) in the case $p = 1$, $n = 2$, $\tilde{K}_1 = k > 0$, $\tilde{K}_2 = \varepsilon > 0$, and let ε approach zero arbitrarily closely. In this case the transfer function of the innovation model is arbitrarily close to being first-order, and $V(\theta)$ is almost independent of $\tilde{K}_2$. In this particular case the arguments advanced in [29] and repeated in Section IV

above suggest that the situation would be even worse with any other parameterization, but it is doubtful whether one can make such a statement in general. Indeed, one of our examples (see Section VII) suggests—admittedly for a small data set—that there are cases for which other parameterizations are preferable.

There is some evidence that certain versions of subspace methods are asymptotically efficient (in the sense of yielding efficient parameter estimates) [9]. If this is so, then "improving" the initial model found by such a method by subsequent parameter optimization adds nothing in a statistical sense, asymptotically. Of course, the situation may be very different in practice with a finite data set, especially if the data set is not very large.

VII. Examples

In this section we present three examples to show how balanced parameterizations can be used for estimating parameters of linear systems. The examples concern a distillation column, an industrial dryer, and the spectrum of sea waves, and they illustrate the estimation of stable systems, minimum phase systems, and positive-real systems, respectively. The first two examples are multivariable. We begin, however, with some artificially generated examples of systems in which the use of balanced and observable parameterizations is compared.

A. Simulation Examples

In both of the simulation examples, a white noise input of length N samples was applied to a discrete-time linear system, and a white noise disturbance was added to the system output $z(t)$ to create the "measured output" $y(t)$. In both cases the N4SID subspace algorithm (with no modeling of the output disturbance) was used to obtain an initial model. Parameters were then optimized using functions from the *MATLAB System Identification Toolbox*, with the number of iterations of the optimizer limited to at most 60 in each case. The rms signal-to-noise ratio, of input signal to disturbance signal, was 100:1 in each case.

The quality of each model was measured by

$$\epsilon = \sum_{t=1}^{N} [\hat{y}(t) - z(t)]^2 \tag{65}$$

where $\hat{y}(t)$ is the output prediction of the model based on the measured input and $z(t)$ is the noise-free output of the true plant. Thus, a good model should give a small value of ϵ. We remark that the N4SID algorithm gave the correct "sign parameters" in every case.

1) Random Systems: In this example the true system has two inputs, two outputs and four states; it is generated randomly (using the *MATLAB Control Toolbox* function "drmodel"). In each experiment, $N = 100$. We identify four different fourth-order state-space models of output-error form—one using the stable balanced canonical from and the other three using the three different choices of observability indexes of the observable form.

The above experiment was repeated 100 times using 100 different, randomly generated systems. Let ϵ_i^m denote the quality of the model of type m for the ith $(i = 1, \cdots, 100)$ experiment. The average value of ϵ_i^m over the 100 experiments for each model type m is summarized in the first column of Table I. It appears that the balanced canonical form did much worse than the observable forms. This is due to the fact that for one of the experiments ($i = 55$), the balanced form converged to a false local minimum. The value of ϵ_{55}^m for all of the four different models is listed in the last column of Table I. If this particular experiment is not included, then the average over the other 99 experiments, which is given in the second column of the table, shows that the performance of the four different types of models is almost identical.

TABLE I
Comparing Observable and Balanced Forms (Random Systems)

Model	[Observability indices]	$\bar{\epsilon}_i^m$ $(i = 1, .., 100)$	$\bar{\epsilon}_i^m$ $(i \neq 55)$	ϵ_{55}^m
Balanced		0.0216	0.0024	1.9170
Observable	[1 3]	0.0025	0.0025	0.0024
Observable	[2 2]	0.0031	0.0031	0.0024
Observable	[3 1]	0.0034	0.0035	0.0024

TABLE II
Comparing Observable and Balanced Forms (Repeated Poles)

Number of sig. fig.	Observable form	Balanced form
7	6000	10^{-5}
8	677	10^{-6}
9	5.6	10^{-6}

2) Repeated Poles: For this example the true system was SISO, with transfer function $1/(z - 0.98)^3$, and the data length was $N = 500$. This system is generally considered to be "hard to identify" because of the great sensitivity of its pole locations to parameter values. This time we calculated two optimized models—one using the balanced form and the other using the observable form. The entire optimization procedure was performed using double precision arithmetic, except that at the beginning of each optimization iteration the parameter values were truncated to x significant figures. Three sets of experiments were performed, with $x = 7, 8$, and 9, respectively. For each value of x, 100 sets of data were generated and two different models, based on the two different parameterizations, were calculated from each set of simulation data. Table II shows the average value of ϵ over 100 experiments for each value of x and the choice of parameterization. The balanced parameterization clearly did much better than the observable one in this case.

B. Distillation Column

This example concerns a fractional distillation column with three inputs (input cooling temperature, reboiling temperature, and pressure) and two outputs (top product flow rate, C4 concentration). A complex "high-fidelity" nonlinear simulation of the column was available, constructed from physical and thermodynamic "first-principles." In order to exploit this model for control, an automatic way of approximating it by a low-order linear model was required. The quality of the approximation was to be judged by its ability to reproduce low-amplitude step responses, although this is not the best criterion for control applications. As is to be expected, and as

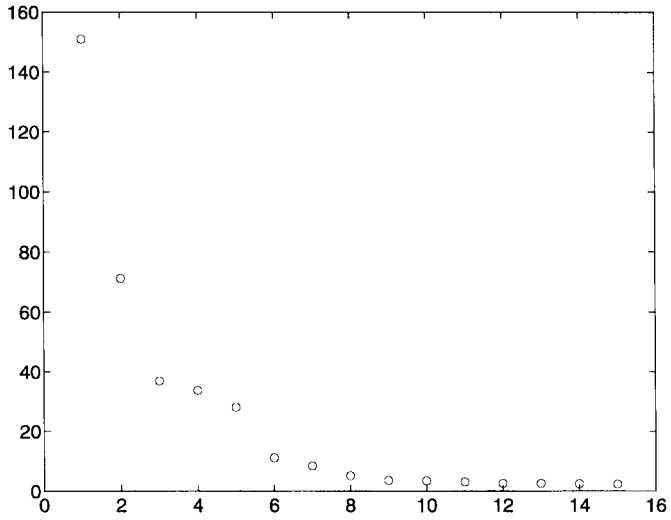

Fig. 1. The largest 15 singular values.

we shall demonstrate later, a model which reproduced the step response very well did not reproduce the frequency response particularly well.

Since a simulation model was available, it was possible to produce impulse response data from it—in fact by differencing step response data—and therefore to use an approximate realization algorithm to obtain the initial model. The only property of the model known *a priori* was that it was stable. It was therefore appropriate to use the "Lyapunov balanced" parameterization of Section IV to improve the initial model. Extremely good results were obtained using McMillan degrees as low as six.

An initial indication of a suitable state dimension for the linearized model was obtained by examining the Hankel singular values of a high-order model obtained by Kung's method [24]. This high-order model is in fact an FIR model, whose impulse response matches the initial segment of the nonlinear model's impulse response exactly and is zero subsequently. For this example we supplied the algorithm with an impulse response over 250 time steps, the interval between steps representing 10 s. Since there are three inputs and two outputs, the FIR model has $\min(2, 3) \times 250 = 500$ states. The first 15 Hankel singular values (of 500) are shown in Fig. 1. Assessing these "by eye," they indicate that about eight states should be sufficient, and even six may be enough. Low-order models of these state dimensions were obtained by balanced truncation. An alternative initial model was obtained by using a slightly different algorithm, due to Al-Saggaf [2], in which steady-state properties are preserved when model reduction by truncation is performed.

The parameter estimates were first improved by optimizing the criterion

$$V(\theta) = \sum_{t=0}^{N} ||D_o Y(t) D_i - \hat{Y}(t, \theta)||_F^2 \qquad (66)$$

where $Y(t)$ and $\hat{Y}(t, \theta)$ are the step response matrices, at time t, of the nonlinear model and of the linearized model, respectively, D_o and D_i are diagonal scaling matrices, and $||.||_F$ denotes the Frobenius norm.

TABLE III
THE MODEL ACCURACY BEFORE AND AFTER OPTIMIZATION

	Kung's		Al-Saggaf's	
Objective function $V(\theta)$	6^{th} order	8^{th} order	6^{th} order	8^{th} order
Before optimization	116.57	32.68	42.73	16.55
After optimization	27.54	9.94	23.84	13.68

Table III shows the values of $V(\theta)$ before and after optimization for sixth- and eighth-order models obtained using both Kung's and Al-Saggaf's algorithms. The fact that each optimization depends on the initial estimate, even though the behavior of both initial models appears very similar, emphasizes the susceptibility of parameter estimation to local optima. Fig. 2 shows the step responses of the original nonlinear model (the "real data"), the initial eight-state model obtained from Kung's algorithm, and the final eight-state model obtained after parameter estimation. It can be seen that parameter optimization gives a significant improvement on the initial model.

Replicating the step response may not be particularly meaningful for control, of course, since much attention is paid to modeling low-frequency behavior, and it is precisely this behavior that will be changed by feedback action. From Parseval's theorem one expects that reproducing the impulse response should give good reproduction of the frequency response across all frequencies. Since we are starting with a nonlinear model, and one for which computing even one time trajectory takes considerable computing resources, there is no "true frequency response." We therefore consider the frequency response of the 500-state FIR model, which matches the initial segment of the impulse response exactly, as the "true frequency response." Fig. 3 shows the frequency responses of eighth-order models obtained by optimizing the impulse and step response, respectively, and compares these with the "true frequency response." It also shows the frequency response of the eight-state "initial" model obtained using Kung's algorithm. All the models can be seen to have similar frequency responses. For a more careful comparison, Fig. 4 shows $\overline{\sigma}[f(\omega) - \hat{f}(\omega)]$ as a function of ω, where $\overline{\sigma}$ denotes the greatest singular value, f denotes the "true" frequency response, and $\hat{f}$ the response of a model, for three different models. As expected, the optimized impulse response gives a better approximation than the optimized step response, except at low frequencies. The "initial" (that is, unoptimized) eight-state Kung model is superior to the optimized one up to $\omega = 0.02$ rad/s, approximately, and is as good as the optimized step-response model at very low frequencies.

C. Industrial Dryer

The second example concerns an industrial drying process with three inputs and three outputs. Raw material enters the dryer continuously, and the moisture in it is removed by passing hot gas over it. The efficiency of the process depends on both the temperature and the rate of flow of the hot gas. The process inputs are: the fuel flow rate, the hot gas exhaust

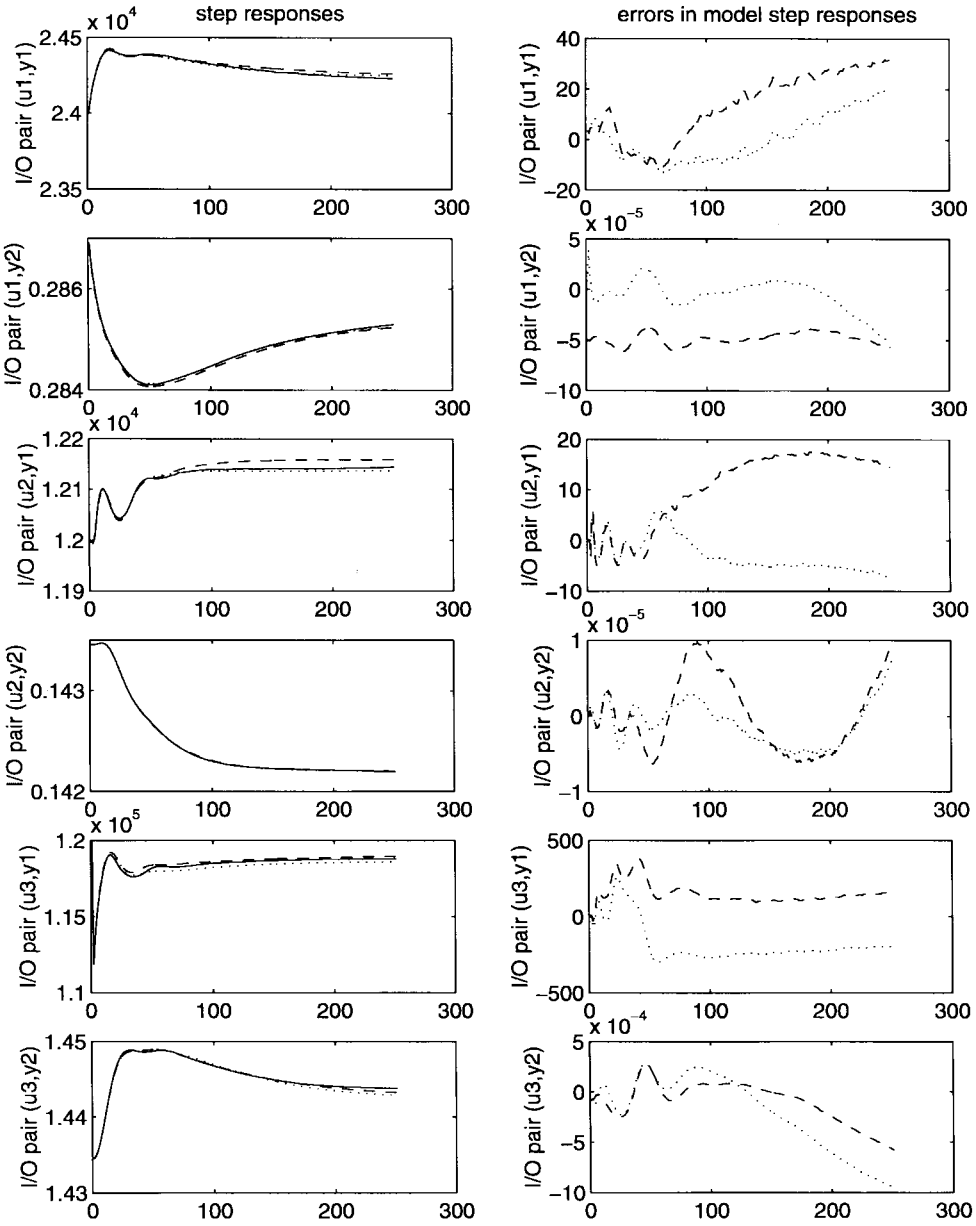

Fig. 2. The left-hand column shows the step responses of the simulation model (solid), the eight-state Kung's estimate obtained from scaled data (dotted), and the optimized model (dashed) for each input–output pair. The errors in step responses of both models are plotted on the right-hand column. Unit time interval = 10 s.

fan speed, and the rate of flow of raw material. The three outputs are: dry bulb temperature, wet bulb temperature, and the moisture content of the raw material when it leaves the dryer.

The data for this example consists of measurements taken on a real process. The first two inputs could be manipulated, and pseudo-random binary signals were applied to these. The third input was an external disturbance which could not be manipulated but could be measured, and its actual value was recorded. After a certain amount of preprocessing (detrending, prefiltering and decimation), the data consisted of 867 samples for each input and output, with sampling period 10 s. The first 600 samples were used to obtain a model, and the rest of the data was used for model validation.

In this case it was appropriate to obtain a model in the "innovations" form (1), (2), with the noise model (A_d, K_d, C_d, I) being minimum-phase. We used the N4SID subspace algorithm [48] to obtain an initial model. We then used the balanced parameterization of minimum-phase systems (the parameters in B_d and D_d being unrestricted) to improve the model by prediction-error optimization. The optimization criterion was the unweighted one-step-ahead prediction error

$$V(\theta) = \sum_{i=0}^{N} \|y(t) - \hat{y}(t)\|^2 \tag{67}$$

and the results were judged according to the ability of the models to reproduce the validation output. A possible numerical

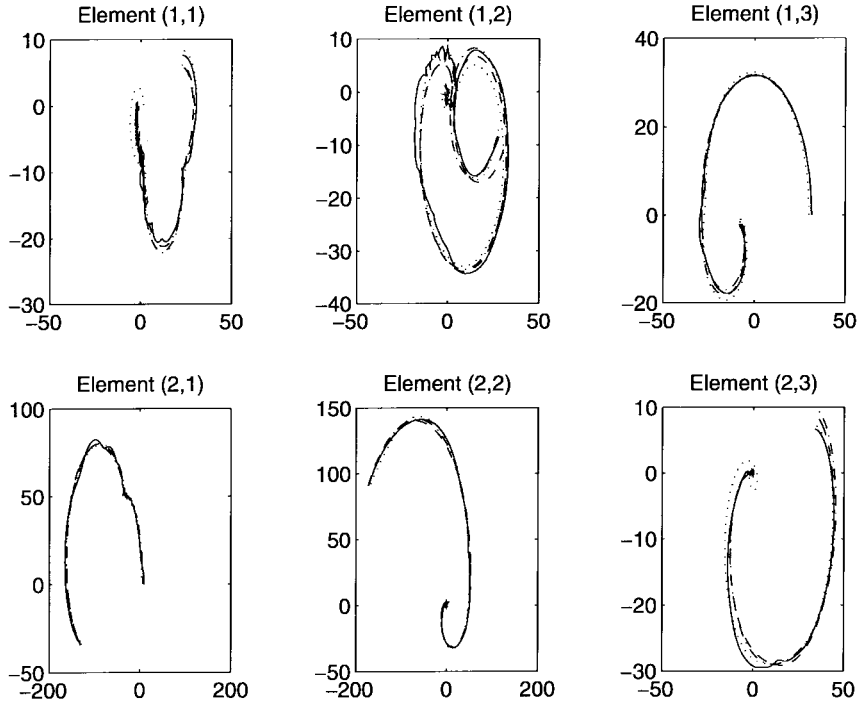

Fig. 3. Comparison of frequency responses of alternative eight-state models. Solid line: FIR model; dashed line: eight-state, optimized impulse response; chained line: eight-state initial model; dotted line: eight-state, optimized step response.

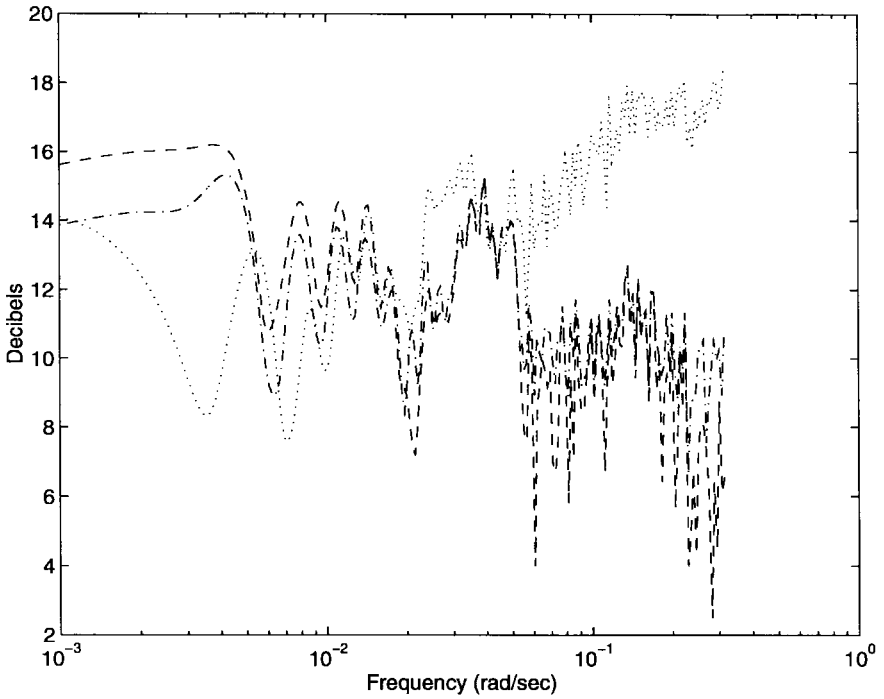

Fig. 4. Maximum singular value of difference from FIR model. Key as for Fig. 3.

indicator for this is

$$\epsilon = 100\left[\frac{1}{3}\sum_{i=1}^{3}\sqrt{\frac{\sum_{t=1}^{N_v}(y_i(t)-\hat{y}_i(t))^2}{\sum_{t=1}^{N_v}(y_i(t))^2}}\right]\% \qquad (68)$$

where $y_i(t)$ and $\hat{y}_i(t)$ are the measured ith output and the one-step ahead predicted ith output, respectively, and N_v is the number of samples in the validation data set. In Table IV, the values of ϵ for initial models and optimized models of order 2–10 are given. It can be seen from the table that the best model is the optimized model of order six. In Fig. 5, the simulated output of the sixth order subspace model and

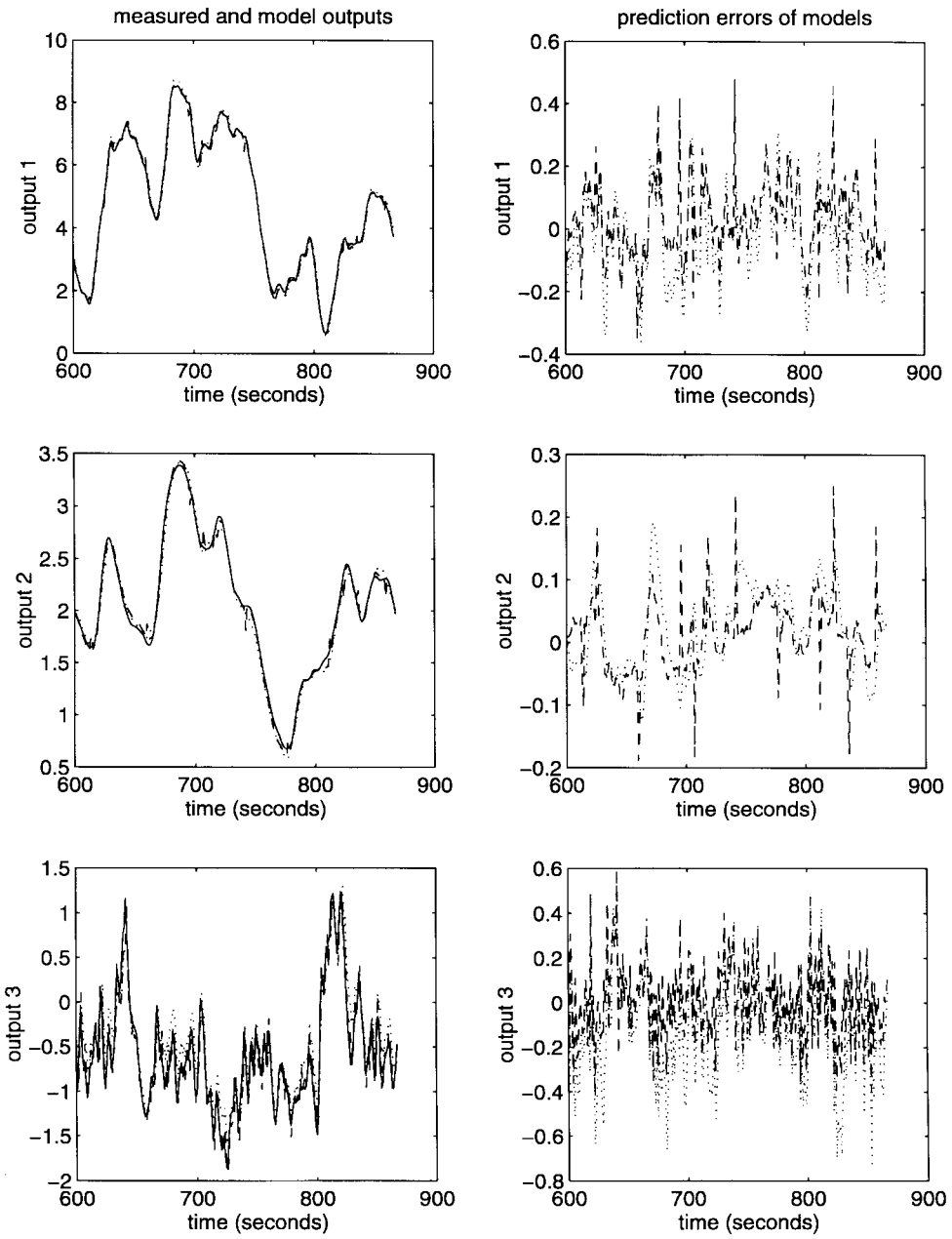

Fig. 5. The left-hand column shows the predicted outputs on the validation data set, and the right-hand column shows the prediction errors of the models: measured output (solid), six-state subspace model (dotted), six-state optimized model (dashed).

that of the sixth order optimized model are plotted against the validation data. It can be seen that both the subspace model and the optimized model are able to predict the first two outputs very well, but the optimized model predicts the third output significantly better than the subspace model.

TABLE IV
MODEL ERROR OF SUBSPACE MODEL AND OPTIMIZED MODEL

Model order	Subspace models	Optimized models
2	26.3406	15.6249
3	14.6125	9.2929
4	13.9288	9.6979
5	12.5058	9.1138
6	12.3669	* 8.7947
7	11.0762	9.0579
8	11.1219	9.6294
9	13.3647	9.1706
10	9.3996	9.0817

D. Sea Wave Spectrum

Hydrodynamicists derive spectra of ocean waves on the basis of "first-principles" reasoning. A widely used example is the Pierson–Moskwitz spectrum, which has the form

$$\Phi(\omega) = \frac{4\pi^3 h^2}{\omega^5 z^4} \exp\left(\frac{-16\pi^3}{\omega^4 z^4}\right) \tag{69}$$

where h is the wave height, z is the zero-crossing period, and ω is the frequency [43]. In applications such as the control of ships, particularly for station-keeping (e.g., when recovering sunken objects or surveying), it is usually necessary

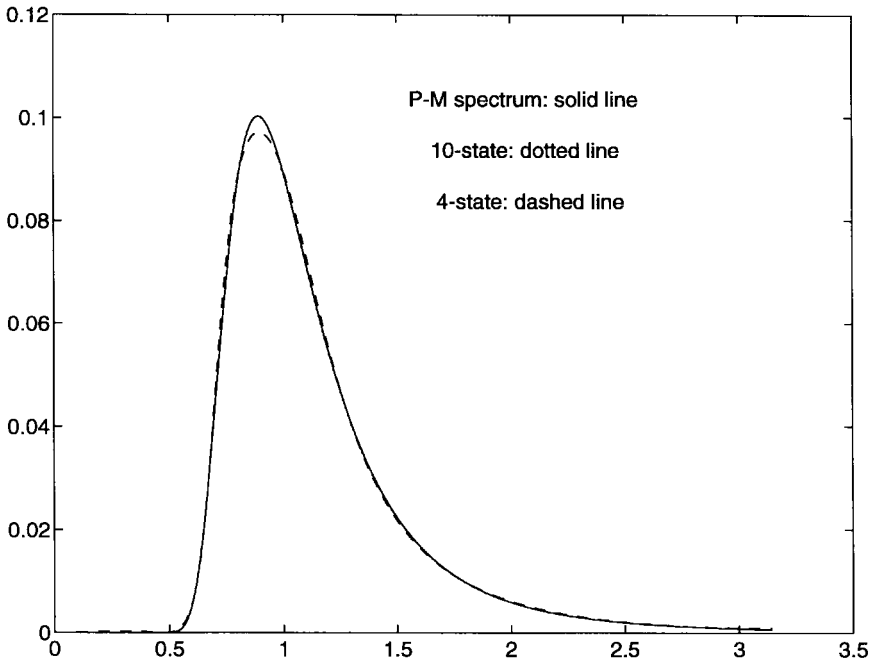

Fig. 6. Spectra of initial four- and ten-state models.

to have finite-dimensional models of disturbances such as waves. It is therefore of great importance to find finite-dimensional approximations of irrational spectra such as this one. It is also important to do so with techniques which can be generalized to multivariate spectra, since directional wave spectrum information is frequently available.

The approach we use is to obtain the autocovariance function $\phi(\tau)$ corresponding to the given spectrum $\Phi(\omega)$ (by inverse discrete Fourier transformation) and then to use the approximate stochastic realization algorithm of [12] to find an initial positive-real spectral summand, namely a finite-dimensional system whose frequency response $G(e^{i\omega T})$ satisfies $\Phi(\omega) \approx G(e^{i\omega T}) + G(e^{-i\omega T}) \geq 0$.

This initial model is then improved by optimizing the parameters in the positive-real balanced parameterization. Finally, an approximate spectral factor is found from the improved spectral summand, namely a stable, minimum-phase system whose frequency response $F(e^{i\omega T})$ satisfies

$$F(e^{i\omega T})F(e^{-i\omega T}) = G(e^{i\omega T}) + G(e^{-i\omega T}) \approx \Phi(\omega). \quad (70)$$

This is the model usually needed in applications such as Kalman filtering and optimal control. An alternative approach would be to find an initial spectral factor from the initial spectral summand and then optimize; in this case the balanced parameterization of minimum-phase systems, rather than positive-real systems, would be needed. It is well known how to obtain the spectral factor from the spectral summand [12], [13], but numerical difficulties can be encountered at this stage.

Fig. 6 shows the exact spectrum and spectra of ten-state and four-state *initial* models for the case $h = 1$ m, $z = 5$ s. The ten-state spectrum is indistinguishable by eye from the exact one. (Of course no rational spectrum can match the exact one at low frequencies, and the comparison would not look so favorable on logarithmic axes.) The parameters of the four-state model were optimized, the loss function being the peak error between the exact spectrum and the model spectrum. Fig. 7 shows the spectra of the initial four-state model, the optimized four-state model, and the exact spectrum. It can be seen, by comparing Figs. 6 and 7, that the four-state optimized model gives almost as good an approximation to the spectrum as the initial ten-state model. The optimized (discrete-time) spectral factor in this case has transfer function as in (70a), shown at the bottom of the page, where it is assumed that the sampling period is $T = 1$ s.

It should be remarked that the approximation we have obtained is appropriate for tasks such as linear-quadratic (LQ) control, because the increase in cost due to mismodeling the spectrum is bounded by a linear function of $\sup_\omega ||\Phi(\omega) - \hat{\Phi}(\omega)||$. It is quite inappropriate for some of the traditional concerns of hydrodynamicists, such as predicting the "100-year wave," because the very low-frequency information in the spectrum which is crucial for such concerns is modeled extremely badly (as it must be, with any rational spectrum).

Although our example concerns a scalar spectrum, the algorithm extends to multivariate spectra with no modifications of the ideas and only minor modifications of the computational steps.

$$F(z) = \frac{0.001(5.0260z^4 - 9.7454z^3 + 4.1354z^2 + 3.4071z - 2.4411)}{z^4 - 2.1837z^3 + 2.3391z^2 - 1.2533z + 0.3435} \quad (70a)$$

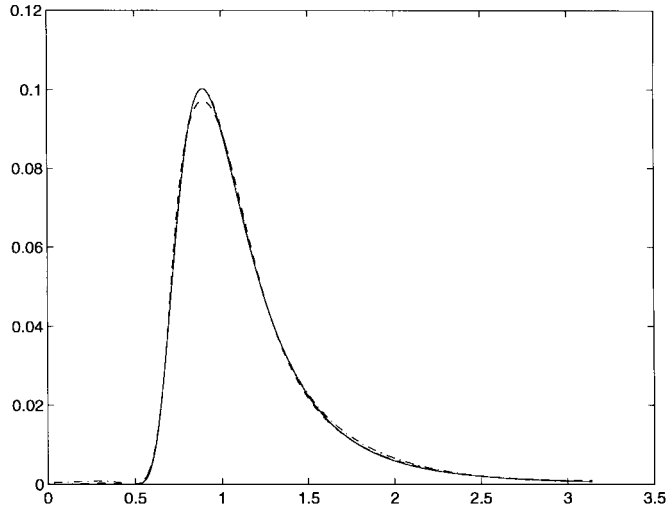

Fig. 7. Spectra of four-state initial and optimized models (solid: original; broken: initial model; chained: optimized).

VIII. Conclusions

Balanced parameterizations seem to offer an effective and convenient solution to the identifiability problem which occurs when identifying multivariable linear systems. The relatively recent development of subspace methods has already increased the attractiveness of using state-space models for multivariable identification, and we believe that the use of balanced parameterizations will reinforce this trend.

The ability to focus on specific subclasses of linear systems is a very nice feature of balanced parameterizations. We have shown how this can be applied to particular examples. There is scope to exploit this ability further in applications, such as identification of slowly time-varying systems, adaptive signal processing, and control schemes.

We have not discussed the statistical consequences of using balanced parameterizations. Asymptotically (as the amount of data grows), the particular parameterization used does not affect the statistics of the estimated transfer function. But in many applications small-sample behavior is more important than asymptotic behavior, and this can be expected to be related to the computational efficiency of the parameter optimization, since both are affected by the numerical conditioning of the optimization problem. These questions remain to be investigated for balanced parameterizations. We have found examples for which balanced parameterizations give more accurate results than the more traditional observable forms, but also examples for which the opposite is true.

Another aspect which has not been thoroughly investigated yet is the question of structure estimation. Choosing a suitable McMillan degree seems to be relatively straightforward with realization and subspace methods. These methods also seem to avoid difficulties with the estimation of other structural parameters, such as the sign parameters for one-output systems in the balanced forms. We emphasize that this is an attribute of the methods for finding initial models, not of balanced parameterizations themselves. Indeed, without such effective methods for finding initial models, the structure estimation problem with balanced parameterizations may well be worse than with other parameterizations, particularly for multivariable systems, since the number of possible structures is very large for balanced forms. It should therefore be emphasized that it is the combination of balanced parameterizations with realization and subspace methods that gives effective algorithms for identification.

Appendix
Derivatives of MIMO State-Space Matrices

The formulas for the derivatives of the state-space matrices defined in Section IV are as follows. We assume that the row vectors $\tilde{B}_j$ are parameterized as $[\tilde{b}_{j1}, \cdots, \tilde{b}_{jm}]$ and the unit vectors U_j are parameterized according to (11).

1) Derivatives of B_c:
   a) $\partial B_c/\partial x = 0$ if $x = \sigma_i$ or $x = \phi_{i,j}$.
   b) Let $\partial B_c/\partial \tilde{b}_{ij} = [\bar{b}_{kl}]$, then

$$\bar{b}_{kl} = \begin{cases} 1 & \text{if } k = i \text{ and } l = j \\ 0 & \text{otherwise.} \end{cases} \tag{71}$$

2) Derivatives of C_c:
   The matrix C_c depends on $\tilde{b}_{ij}$ and $\phi_{i,j}$ only.
   a) $\partial C_c/\partial x = 0$ if $x = \sigma_i$.
   b) Let $\partial C_c/\partial \tilde{b}_{ij} = [\tilde{c}_{kl}]$, then

$$\tilde{c}_{kl} = \begin{cases} c_{ij}\dfrac{\tilde{b}_{ij}}{||\tilde{B}_i||^2}, & \text{if } k = i \\ 0, & \text{otherwise.} \end{cases} \tag{72}$$

   In other words

$$\frac{\partial C_c}{\partial \tilde{b}_{ij}} = \left[0_{p\times(i-1)} \quad C_i \frac{\tilde{b}_{ij}}{||\tilde{B}_i||^2} \quad 0_{p\times(n-i)} \right]. \tag{73}$$

   c) Before we give an expression for the derivative $\partial C/\partial \phi_{i,j}$, we define the following function:

$$\begin{aligned} g_q: \quad & \mathbf{R} \times \mathbf{R}^q \to \mathbf{R}^{q+1} \\ & (K, \beta) \mapsto w \end{aligned}$$

   such that

$$\begin{cases} w_1 = K\cos(\beta_1)\cos(\beta_2)\cos(\beta_3)\cdots \\ \qquad \cos(\beta_{q-1})\cos(\beta_q) \\ w_2 = K\cos(\beta_1)\cos(\beta_2)\cos(\beta_3)\cdots \\ \qquad \cos(\beta_{q-1})\sin(\beta_q) \\ w_3 = K\cos(\beta_1)\cos(\beta_2)\cos(\beta_3)\cdots \\ \qquad \sin(\beta_{q-1}) \\ \vdots \quad \vdots \\ w_q = K\cos(\beta_1)\sin(\beta_2) \\ w_{q+1} = K\sin(\beta_1). \end{cases} \tag{74}$$

   Also let

$$K_j = \begin{cases} 1, & \text{if } j = 1 \\ \cos(\phi_{i,p-1})\cos(\phi_{i,p-2})\cdots\cos(\phi_{i,j}), & \text{if } j > 1 \end{cases}$$

$$\gamma_j = \left[\phi_{i,j-1} + \frac{\pi}{2} \quad \phi_{i,j-2} \quad \cdots \quad \phi_{i,1} \right].$$

The derivative $\partial C_c/\partial \phi_{i,j}$ is

$$\frac{\partial C_c}{\partial \phi_{i,j}} = ||\tilde{B}_i|| \begin{bmatrix} & g_{j-1}(K_j, \gamma_j) & \\ 0_{p\times(i-1)} & 0_{(p-j)\times 1} & 0_{p\times(n-i)} \end{bmatrix}. \tag{75}$$

3) Derivatives of A_c:

a) Let $\partial A_c/\partial \sigma_i = [\tilde{a}_{kl}]$, then

$$\tilde{a}_{kl} = \begin{cases} -\dfrac{a_{kk}}{\sigma_k}, & \text{if } k = l = i \\ \dfrac{-2\sigma_k\sigma_l}{(\sigma_k^2-\sigma_l^2)^2}\,\tilde{B}_k\tilde{B}_l^T & \\ \quad +\dfrac{(\sigma_k^2+\sigma_l^2)}{(\sigma_k^2-\sigma_l^2)^2}\,C_k^TC_l, & \text{for } k = i \text{ and } l \neq i \\ \dfrac{(\sigma_k^2+\sigma_l^2)}{(\sigma_k^2-\sigma_l^2)^2}\,\tilde{B}_k\tilde{B}_l^T & \\ \quad -\dfrac{2\sigma_k\sigma_l}{(\sigma_k^2-\sigma_l^2)^2}\,C_k^TC_l, & \text{for } l = i \text{ and } k \neq i \\ 0, & \text{otherwise.} \end{cases} \tag{76}$$

b) Let $\partial A_c/\partial \tilde{b}_{ij} = [\overline{a}_{kl}]$, then

$$\overline{a}_{kl} = \begin{cases} -\dfrac{\tilde{b}_{kl}}{\sigma_k}, & \text{for } k = l = i \\ \dfrac{1}{\sigma_k^2-\sigma_l^2} & \\ \quad \cdot\left(\sigma_l\tilde{b}_{lj} - \dfrac{1}{||\tilde{B}_k||^2}\sigma_k\tilde{b}_{kj}C_k^TC_l\right), & \text{for } k = i \text{ and } l \neq i \\ \dfrac{1}{\sigma_k^2-\sigma_l^2} & \\ \quad \cdot\left(\sigma_l\tilde{b}_{kj} - \dfrac{1}{||\tilde{B}_l||^2}\sigma_k\tilde{b}_{lj}C_k^TC_l\right), & \text{for } l = i \text{ and } k \neq i \\ 0, & \text{otherwise.} \end{cases} \tag{77}$$

c) Let $\partial A_c/\partial \phi_{i,j} = [\hat{a}_{kl}]$, then

$$\hat{a}_{kl} = \begin{cases} -\dfrac{\sigma_k}{\sigma_k^2-\sigma_l^2}\left(\dfrac{\partial C_k}{\partial \phi_{k,j}}\right)^T C_l, & \text{for } k = i \text{ and } l \neq i \\ -\dfrac{\sigma_k}{\sigma_k^2-\sigma_l^2}\,C_k^T\left(\dfrac{\partial C_l}{\partial \phi_{l,j}}\right), & \text{for } l = i \text{ and } k \neq i \\ 0, & \text{otherwise.} \end{cases} \tag{78}$$

Note that $\partial C_k/\partial \phi_{i,j}$ is the kth column of $\partial C_c/\partial \phi_{i,j}$.

REFERENCES

[1] H. Akaike, "Canonical correlation analysis of time series and the use of an information criterion," in *System Identification Advances and Case Studies*, R. K. Mehra and D. G. Lainiotis, Eds. New York: Academic, 1976.

[2] U. M. Al-Saggaf and G. F. Franklin, "Model reduction via balanced realizations: An extension and frequency weighted techniques," *IEEE Trans. Automat. Contr.*, vol. 33, pp. 687–692, 1988.

[3] B. D. O. Anderson, "The inverse problem of stationary covariance generation," *J. Stat. Phys.*, vol. 21, pp. 133–147, 1969.

[4] R. W. Brockett, "Some geometric questions in the theory of linear systems," *IEEE Trans. Automat. Contr.*, vol. 21, pp. 449–454, 1976.

[5] P. E. Caines, *Linear Stochastic Systems.* New York: Wiley, 1988.

[6] C. T. Chou, "Geometry of linear systems and identification," Ph.D. dissertation, Cambridge Univ., U.K., 1994.

[7] J. M. C. Clark, "The consistent selection of parameterizations in systems identification," in *Proc. Joint Amer. Contr. Conf.*, 1976.

[8] M. H. A. Davis and R. B. Vinter, *Stochastic Modeling and Control.* London: Chapman and Hall, 1985.

[9] M. Deistler, K. Peternell, and W. Scherrer, "Consistency and relative efficiency of subspace methods," *Automatica*, vol. 31, pp. 1865–1875, 1995.

[10] M. J. Denham, "Canonical forms for the identification of multivariable linear systems," *IEEE Trans. Automat. Contr.*, vol. AC-19, 1974.

[11] U. B. Desai and D. Pal, "A transformation approach to stochastic model reduction," *IEEE Trans. Automat. Contr.*, vol. AC-29, 1984.

[12] U. B. Desai, D. Pal, and R. D. Kirkpatrick, "A realization approach to stochastic model reduction," *Int. J. Contr.*, vol. 42, pp. 821–838, 1985.

[13] P. Faurre, "Stochastic realization algorithms," in *System Identification: Recent Advances and Case Studies*, R. K. Mehra and D. G. Lainiotis, Eds. New York: Academic, 1976.

[14] K. Glover and J. C. Willems, "Parameterization of linear dynamical systems: Canonical forms and identifiability," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 640–645, 1974.

[15] K. Glover, "Some geometrical properties of linear systems with implications in identification," in *Proc. IFAC World Congr.*, Boston, MA, 1975.

[16] M. Green, "Balanced stochastic realizations," *Linear Algebra Appl.*, vol. 98, pp. 211–247, 1988.

[17] E. J. Hannan and M. Deistler, *The Statistical Theory of Linear Systems.* New York: Wiley, 1988.

[18] B. Hanzon and R. J. Ober, "Overlapping block-balanced canonical forms and parameterizations: The stable SISO case," *SIAM J. Contr. Optim.*, vol. 35, pp. 228–242, 1997.

[19] ———," Overlapping block-balanced canonical forms for various classes of systems," *Linear Algebra Appl.*, to be published.

[20] B. L. Ho and R. E. Kalman, "Effective construction of linear, state variable models from input–output functions," *Regelungstechnik*, vol. 14, pp. 545–548, 1966.

[21] E. A. Jonckheere and L. M. Silverman, "A new set of invariants for linear systems—Applications to reduced order compensator design," *IEEE Trans. Automat. Contr.*, vol. 28, pp. 953–964, 1983.

[22] P. T. Kabamba, "Balanced forms: Canonicity and parameterization," *IEEE Trans. Automat. Contr.*, vol. AC-30, pp. 1106–1109, 1985.

[23] T. Kailath, *Linear Systems.* Englewood Cliffs, NJ: Prentice-Hall, 1980.

[24] S. Y. Kung, "A new low-order approximation algorithm via singular value decomposition," in *Proc. 12th Asilomar Conf. Circuits, Syst. Computers*, 1978.

[25] W. E. Larimore, "System identification, reduced-order filtering and modeling via canonical variate analysis," in *Proc. Amer. Contr. Conf.*, San Francisco, CA, 1983.

[26] ———, "Canonical variate analysis in identification, filtering, and adaptive control," in *Proc. 29th IEEE Conf. Decision Contr.*, HI, 1990.

[27] L. Ljung, *System Identification: Theory for the User.* Englewood Cliffs, NJ: Prentice-Hall, 1987.

[28] J. M. Maciejowski and D. A. Vines, "Decoupled control of a macroeconomic model using frequency-domain methods," *J. Econ. Dyn. Contr.*, vol. 7, pp. 55–77, 1984.

[29] J. M. Maciejowski, "Balanced realizations in system identification," in *Proc. 7th IFAC Symp. Identification Parameter Estimation*, York, U.K., 1985.

[30] J. M. Maciejowski and R. J. Ober, "Balanced parameterizations and canonical forms for system identification," in *Proc. 8th IFAC Symp. Identification Parameter Estimation*, Beijing, China, 1988.

[31] J. M. Maciejowski and C. T. Chou, "Identification for control using balanced parameterizations," in *Proc. 2nd European Contr. Conf.*, Groningen, The Netherlands, 1993.

[32] J. M. Maciejowski, "Guaranteed stability with subspace algorithms," *Syst. Contr. Lett.*, vol. 26, pp. 153–156, 1995.

[33] B. P. McGinnie, "A balanced view of system identification," Ph.D. dissertation, Cambridge Univ., U.K., 1994.

[34] T. McKelvey, "Identification of state-space models from time and frequency data," Ph.D. dissertation, Linköping Univ., 1995.

[35] M. Moonen, B. De Moor, L. Vandenberghe, and J. Vandewalle, "On and off line identification of linear state-space models," *Int. J. Contr.*, vol. 49, pp. 219–232, 1989.

[36] B. C. Moore, "Principal component analysis in linear systems: Controllability, observability and model reduction," *IEEE Trans. Automat. Contr.*, vol. AC-26, pp. 17–32, 1981.

[37] C. T. Mullis and R. A. Roberts, "Synthesis of minimum roundoff noise fixed point digital filters," *IEEE Trans. Circuits Syst.*, vol. 23, pp. 551–562, 1976.
[38] J. P. Norton, *An Introduction to Identification.* New York: Academic, 1986.
[39] R. J. Ober, "Topology of the set of asymptotically stable minimal systems," *Int. J. Contr.*, vol. 46, pp. 263–280, 1987.
[40] _____, "Connectivity properties of classes of linear systems," *Int. J. Contr.*, vol. 50, pp. 2049–2073, 1989.
[41] R. J. Ober and D. McFarlane, "Balanced canonical forms for minimal systems: A normalized coprime factor approach," *Linear Algebra Appl.*, pp. 23–64, 1989.
[42] R. J. Ober, "Balanced parameterization of classes of linear systems," *SIAM J. Contr. Optim.*, vol. 29, pp. 1251–1287, 1991.
[43] M. H. Patel, *Dynamics of Offshore Structures.* London: Butterworths, 1989.
[44] R. Pintelon, J. Schoukens, T. McKelvey, and Y. Rolain, "Minimum variance bounds for overparametrized models," *IEEE Trans. Automat. Contr.*, vol. 41, pp. 719–720, 1996.
[45] J. Rissanen, "Basis of invariants and canonical forms for linear dynamic systems," *Automatica*, vol. 10, pp. 175–182, 1974.
[46] T. Söderström and P. Stoica, *System Identification.* Englewood Cliffs, NJ: Prentice-Hall, 1989.
[47] L. Thiele, "On the sensitivity of linear state-space systems," *IEEE Trans. Circuits Syst.*, vol. CAS-33, pp. 502–510, 1986.
[48] P. van Overschee and B. DeMoor, "N4SID: Subspace algorithms for the identification of combined deterministic stochastic systems," *Automatica*, vol. 30, pp. 75–93, 1994.
[49] S. Veres, *Structure Selection of Stochastic Dynamic Systems.* New York: Gordon and Breach, 1991.
[50] M. Verhaegen and P. M. Dewilde, "Subspace model identification, Parts 1 and 2," *Int. J. Contr.*, vol. 56, pp. 1187–1241, 1992.
[51] M. Verhaegen, "Identification of the deterministic part of MIMO state space models given in innovations form from input–output data," *Automatica*, vol. 30, pp. 61–74, 1994.
[52] E. Verriest and T. Kailath, "On generalized balanced realizations," *IEEE Trans. Automat. Contr.*, vol. AC-28, pp. 833–844, 1983.
[53] M. Vidyasagar, *Control System Synthesis: A Factorization Approach.* Cambridge, MA: MIT Univ. Press, 1985.
[54] B. Wahlberg, "On model reduction in system identification," in *Proc. Amer. Contr. Conf.*, Seattle, WA, 1986.
[55] H. P. Zeiger and A. J. McEwen, "Approximate linear realization of given dimension via Ho's algorithm," *IEEE Trans. Automat. Contr.*, vol. AC-19, p. 153, 1974.

Chun Tung Chou was born in Hong Kong in 1967. He received the B.A. degree in engineering science from the University of Oxford, U.K., in 1988, and the Ph.D. degree in control engineering from the University of Cambridge, U.K., in 1994.

He is currently a Post-Doctoral Research Fellow at the Electrical Engineering Department of Delft University of Technology, the Netherlands. His current research interests include system identification of linear and nonlinear systems.

Jan M. Maciejowski (M'81–SM'96) received the B.Sc. and Ph.D. degrees in control engineering from the Universities of Sussex and Cambridge, U.K., respectively.

Before entering the academic world, he worked in industry on attitude control systems. He is currently at the Engineering Department, Cambridge University, U.K. His research interests include reconfigurable control, predictive control, system modeling and identification, and computer-aided control engineering. He is the author of *Multivariable Feedback Design*, which was awarded the 1996 IFAC Control Textbook Prize.

Dr. Maciejowski is a Fellow of the Institute of Measurement and Control, a member of the IEE, and a member of the Council of the European Union Control Association (EUCA).