15th Triennial World Congress, Barcelona, Spain

# A NEW MINIMAL LOCAL PARAMETRIZATION FOR MULTIVARIABLE LINEAR SYSTEMS [1]

**Tomas McKelvey** [*]

[*] *Dept. of Signals and Systems, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden, mckelvey@s2.chalmers.se*

Abstract: A new minimal parametrization of multivariable linear system is proposed. The parametrization is defined as a perturbation around the realization of a nominal transfer function. A particular parameter basis is selected which, for the impulse response identification problem, leads to a Hessian matrix of the criterion function which is equal to the identity matrix.

Keywords: System identification, Parametrization, Linear multivariable systems, Identification algorithms, State-space realization

## 1. INTRODUCTION

This paper introduces a new parametrization of multivariable linear dynamical systems. The parametrization is based on a local affine map around a nominal system. This type of local parametrizations was introduced in the papers (Wolodkin *et al.*, 1997), (McKelvey and Helmersson, 1997) and (McKelvey and Helmersson, 1999). Some system theoretical aspects of such parametrizations has been shown in (Deistler and Ribartits, 2001). The current contribution is a development of these ideas and provides a more computationally efficient way of deriving such parametrizations as well as some additional and alternative insights.

### 1.1 *Preliminaries*

Consider a state-space representation of a linear system

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) \end{aligned} \tag{1}$$

where the vector signals $u(t) \in \mathbb{R}^m$, $y(t) \in \mathbb{R}^p$ and $x(t) \in \mathbb{R}^n$ are the inputs, outputs and states respectively. The dimension of the state vector is equal to the McMillan degree of the linear system if the realization is minimal, i.e., is both controllable and observable (Kailath, 1980). In the sequel we will normally assume that the realization is minimal so the state dimension and the McMillan degree coincides. In (1) we have left out a direct connection (the $D$ matrix) between $u(t)$ and $y(t)$. Inclusion of such a connection is straightforward and does not significantly change the results of the paper.

The matrix triple $(A,B,C)$ in (1) defines a relation between the input signals and output signals via the state. The input output relation is obtained by solving the difference equation (1) which yields the convolution equation

$$y(t) = \sum_{k=1}^{\infty} CA^{k-1}Bu(t-k) = g(k)u(t-k) \tag{2}$$

where

$$g(k) = CA^{(k-1)}B \in \mathbb{R}^{p\times m}, k = 1,2,\ldots \tag{3}$$

is the matrix valued impulse response of the system. From the impulse response, the transfer function

$$G(z) = \sum_{k=1}^{\infty} CA^{k-1}Bz^{-k} = C(zI-A)^{-1}B \tag{4}$$

is defined for all $z \notin \lambda(A)$, where $\lambda(A)$ denotes the spectrum (set of eigenvalues) of $A$. The triple $(A,B,C)$ is known as a state-space *realization* of a linear system. When we refer to a specific linear system we actually refer to its input output relation, i.e., the impulse response or equivalently the transfer function. Hence, one linear system can have many realizations both minimal and non-minimal ones.

[1] Supported by the Swedish Research Council (VR).

### 1.2 *Parametrizations*

A *parametrization* of a state-space model is a mapping from a parameter space to the space of state-space matrices, and we can informally write $(A(\xi), B(\xi), C(\xi))$ to illustrate that the point $\xi$ in the parameters space defines a state-space realization which in turn defines a transfer function.

It is well known (Kailath, 1980) that the input-output relation is unchanged by a linear non-singular basis change of the state variables, i.e., if $T \in \mathbb{R}^{n\times n}$ is non-singular matrix representing the change of basis, then the realizations $(A,B,D)$ and $(TAT^{-1}, TB, CT^{-1})$ represents exactly the same impulse response and hence the same linear system. The change of basis is called a *similarity transformation*. Consequently, if a linear systems of a given order is parametrized using all elements of the state-space matrices the parametrization is not identifiable. Loss of identifiability by such an over-parametrization is not a severe restriction if the identification techniques used is prepared for such a situation (McKelvey, 1994; Pintelon *et al.*, 1996). However the numerical complexity of estimating more parameters than necessary leads to an increased computation time. Therefore, a minimal parametrization is desirable.

Minimal parametrizations of linear systems has always played an important role in system identification and was and early topic in the development of linear system theory. The single input single output case turns out the be rather easy while the multivariable case where both $m > 1$ and $p > 1$ is much more involved. Especially for such systems there exists no single continuous minimal parametrization which covers the whole space of linear systems (i.e. transfer functions) with a fixed McMillan degree and input output configuration (Luenberger, 1967; Guidorzi, 1975) and (Hannan and Deistler, 1988). Hence it is necessary to use several (possibly overlapping) parametrizations when dealing with MIMO systems.

In this paper the focus will be on parametrization where the state-space matrices are an affine function of the parameters. Consider the following class of affine parametrizations

$$\begin{bmatrix} \mathrm{vec} A(\xi) \\ \mathrm{vec} B(\xi) \\ \mathrm{vec} C(\xi) \end{bmatrix} = p_0 + P\xi \tag{5}$$

Here vec denotes the vectorization operator which converts a matrix to a vector by stacking all columns. The vector $p_0$ is a fixed vector of size $n(n+m+p)$ and $P$ is a matrix of size $n(n+m+p)\times d$ where $d$ is the dimension of the free parameter vector $\xi$. The parametrization is uniquely defined by the pair $(p_0, P)$ and we interpret $p_0$ as the realization which the (local) parametrization is based around.

If the parametrization is of minimal dimension then $d = d_{\min} = n(m+p)$. Parametrizations which are over-parametrized have $d > d_{\min}$ parameters. Examples of affine parametrizations are the (trivial) full parametrization with $P$ as the identity matrix (McKelvey, 1994), the classical observable or controllable parametrizations (Kailath, 1980; Ljung, 1999) (which are minimal ones) and the tridiagonal parametrization (McKelvey and Helmersson, 1996). The recent local parametrizations (McKelvey and Helmersson, 1997) and (Wolodkin *et al.*, 1997) also belong to this class.

The local parametrization described in (McKelvey and Helmersson, 1997) here called *OrthPar* uses the affine subspace in the full parameter space which is orthogonal to the tangent space of the corresponding equivalence class of a nominal realization. In short the range space of the matrix

$$Q \triangleq \frac{\partial}{\partial(\mathrm{vec}\, T)} \begin{bmatrix} \mathrm{vec}\,(TAT^{-1}) \\ \mathrm{vec}\,(TB) \\ \mathrm{vec}\,(CT^{-1}) \end{bmatrix}\Bigg|_{T=I_n} = \begin{bmatrix} A^T\otimes I_n - I_n \otimes A \\ B^T \otimes I_n \\ -I_n \otimes C \end{bmatrix} \tag{6}$$

spans the tangent plane of the equivalent class defined by by the nominal realization $(A,B,C)$. OrthPar uses

$$p_0 = \begin{bmatrix} \mathrm{vec} A \\ \mathrm{vec} B \\ \mathrm{vec} C \end{bmatrix} \quad \text{and} \quad P = Q^{\perp}$$

as a parametrization where $Q^{\perp}$ is an orthonormal matrix which spans the orthogonal complement to the range space of $Q$. For more details see (McKelvey and Helmersson, 1997; McKelvey and Helmersson, 1999).

## 2. IMPULSE RESPONSE BASED PARAMETRIZATION - ORTHIMP

In this section we develop an alternative way of calculating the matrix $P$ in (5). Define the impulse response vector based on the $2n$ vectorized impulse response matrices, i.e.

$$\mathbf{g} \triangleq [\mathrm{vec}(g(1))^T, \ldots \mathrm{vec}(g(2n))^T]^T \in \mathbb{R}^{2nmp} \tag{7}$$

From realization theory it is clear that a transfer function of McMillan degree $n$ is completely specified by the $2n$ first samples of the impulse response (Ho and Kalman, 1966). Hence, $\mathbf{g}$ is an alternative representation of the linear system of order $n$.

Consider a full parametrization of a state-space model

$$\begin{bmatrix} \mathrm{vec} A(\theta) \\ \mathrm{vec} B(\theta) \\ \mathrm{vec} C(\theta) \end{bmatrix} = \theta \in \mathbb{R}^{n(n+m+p)} \tag{8}$$

Pick an arbitrary transfer function of McMillan degree $n$ and let the parametrized matrix triple $(A(\theta), B(\theta), C(\theta))$ for $\theta = \theta_0$ be a minimal realization of it. Let $\mathbf{g}(\theta)$ be the corresponding impulse response vector according to the definition in (7).

A Taylor expansion of $\mathbf{g}(\theta)$ around the minimal realization defined by $\theta_0$ is given by

$$\mathbf{g}(\theta_0 + \delta\theta) = \mathbf{g}(\theta_0 + \delta\theta) + \mathbf{g}'(\theta_0)\delta\theta + o(|\delta|) \tag{9}$$

where $\mathbf{g}'(\theta_0) = \frac{\partial}{\partial\theta}\mathbf{g}(\theta)\Big|_{\theta=\theta_0} \in \mathbb{R}^{2nmp\times n(n+m+p)}$ is the Jacobian matrix of the impulse response function $\mathbf{g}(\theta)$ at $\theta = \theta_0$.

*Lemma 1.* Let $\mathbf{g}(\theta_0)$ represent a minimal realization of McMillan degree $n$. Then the range space of $\mathbf{g}'(\theta_0)$ has dimension $n(m+p)$

**Proof:** Since the equivalence class of realizations with identical impulse responses is of dimension $n^2$ (represented by the similarity transform $T$) the Jacobian matrix $\mathbf{g}'(\theta_0)$ must have a null-space of at least dimension $n^2$. Hence the dimension of the range space is bounded from above by $n(n+m+p)-n^2 = n(m+p)$.

Now pick a minimal parametrization such that the point $\xi$ in the parameter space, which represents the transfer function, is in the interior of the parameter space. This is always possible for example using the partially overlapping parametrizations described in (Ljung, 1999, Appendix 4A). Let the triple $(A_m, B_m, C_m)$ represent the resulting realization defined by the parameter vector $\xi$. Explicitly we can write

$$\begin{bmatrix} \operatorname{vec} A_m \\ \operatorname{vec} B_m \\ \operatorname{vec} C_m \end{bmatrix} = p_{m0} + P_m\xi \tag{10}$$

where $\xi \in \mathbb{R}^{n(m+p)}$ is the parameter vector of the minimal parametrization and $P_m$ is matrix with linearly independent columns which distributes the parameters to the right places in the matrices. The vector $p_{m0}$ represent the constant elements of the state-space matrices implied by the particular parametrization chosen.

Since both $(A,B,C)$ and $(A_m,B_m,C_m)$ have the same transfer function there exists a non-singular similarity transformation $T$ such that

$$(T^{-1}A_mT, T^{-1}B_m, C_mT) = (A,B,C)$$

which in a vectorized form is given by (remember the formula $\operatorname{vec}(XYZ) = Z^T \otimes X \operatorname{vec} Y$)

$$\begin{bmatrix} \operatorname{vec} A \\ \operatorname{vec} B \\ \operatorname{vec} C \end{bmatrix} = \bar{T}(T) \begin{bmatrix} \operatorname{vec} A_m \\ \operatorname{vec} B_m \\ \operatorname{vec} C_m \end{bmatrix} \tag{11}$$

$$\bar{T}(T) = \begin{bmatrix} T^T \otimes T^{-1} & 0 & 0 \\ 0 & I_m \otimes T^{-1} & 0 \\ 0 & 0 & T^T \otimes I_p \end{bmatrix}$$

where $\otimes$ is the Kronecker product and $I_m$ denotes the identity matrix of dimension $m \times m$. It is easy to establish that if $T$ is non-singular so is also $\bar{T}(T)$. We can now explicitly parametrize the neighborhood of the original realization with the aid of the minimal parametrization.

Let

$$\begin{bmatrix} \operatorname{vec} A(\xi) \\ \operatorname{vec} B(\xi) \\ \operatorname{vec} C(\xi) \end{bmatrix} = \theta(\xi) = \bar{T}(T)(p_{m0} + P_m\xi)$$

which implies that we can represent the impulse response vector $\mathbf{g}$ using the minimal parametrization: $\mathbf{g}(\xi+\delta\xi) \triangleq$

$$\mathbf{g}(\theta(\xi+\delta\xi)) = \mathbf{g}(\theta_0) + \mathbf{g}'(\theta_0)\bar{T}(T)P_m\delta\xi + o(|\delta\xi|)$$

Since the parametrization is minimal any perturbation in the parameter space yields a new transfer function and hence a new impulse response vector, i.e. $\mathbf{g}(\xi+\delta\xi) = \mathbf{g}(\xi)$ if and only if $\delta\xi = 0$. This shows that the range space of $\mathbf{g}'(\theta_0)$ is at least of the same dimension as $\delta\xi$ which concludes the proof. $\square$

Lemma 1 establishes that the rank of the Jacobian matrix of the impulse response vector $\mathbf{g}$ equals the number of parameters in a minimal parametrization. This property we now exploit when deriving a local minimal parametrization.

For an arbitrary transfer function of McMillan degree $n$ pick a minimal realization $(A,B,C)$. As before let

$$\begin{bmatrix} \operatorname{vec} A \\ \operatorname{vec} B \\ \operatorname{vec} C \end{bmatrix} = \theta_0 \tag{12}$$

Consider the affine parametrization given by (5) with $p_0 = \theta_0$ which means we define a local parametrization in the neighborhood of the realization $(A,B,C)$. A minimal parametrization has $n(m+p)$ parameters so in this case the dimension of $P$ is $n(n+m+p) \times n(m+p)$. Any choice $P$ such that the product $\mathbf{g}'(\theta_0)P$ has rank $n(m+p)$ will correspond to a valid minimal parametrization.

A suitable matrix $P$ defining the directions of the affine parametrization can be determined in the following way. Calculate the QR factorization of the Jacobian matrix

$$\mathbf{g}'(\theta_0) = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \end{bmatrix} \tag{13}$$

where $R_{11}$ is of dimension $n(m+p) \times n(m+p)$ since $\mathbf{g}'(\theta_0)$ has rank $n(m+p)$ according to Lemma 1. A good candidate for $P$ is given by the right generalized inverse to the matrix $\begin{bmatrix} R_{11} & R_{12} \end{bmatrix}$. A second QR-factorization of the form

$$\begin{bmatrix} R_{11} & R_{12} \end{bmatrix}^T = \tilde{Q}\tilde{R} \tag{14}$$

gives us the final expression for $P$

$$P = \tilde{Q}\tilde{R}^{-T} \tag{15}$$

where $(\cdot)^{-T}$ denotes matrix inverse and transpose. Since $\tilde{R}$ is triangular the matrix inverse is easily determined.

An important property of the derived parametrization is the following result.

*Lemma 2.* Consider a nominal state-space realization $(A,B,C)$ which is minimal and define a local parametrization according to (5), (12) and where $P$ is given by (13), (14) and (15). Let $\mathbf{g}(\xi)$ denote the $2nmp$ long vectorized impulse response vector function parametrized by the local parametrization. Then

$$\mathbf{g}'(\xi)^T\mathbf{g}'(\xi)\big|_{\xi=0} = I$$

**Proof:** The chain rule gives

$$\mathbf{g}'(\xi)\big|_{\xi=0} = \mathbf{g}'(\theta_0)P = Q_1 \left[ R_{11} \;\; R_{12} \right] P = Q_1 \tilde{R}^T \tilde{Q}^T P = Q_1 \tilde{R}^T \tilde{Q}^T \tilde{Q} \tilde{R}^{-T} = Q_1$$

since $\tilde{Q}$ has orthonormal columns. The proof is concluded by noting that $Q_1$ also has orthonormal columns. □

The local parametrization as defined in Lemma 2 will be named *OrthImp parametrization*

*Impulse criterion*

This section discuss the use of the new parametrization when identifying a linear system from a measured impulse response. The simple setup enables us to clearly exemplify some properties of the parametrization. In Section 2.1 we point out some generalizations for other identification settings.

Consider the non-linear least-squares identification criterion

$$V(\xi) = \sum_{k=1}^{N} \|g_m(k) - g(k,\xi)\|_F^2 \tag{16}$$

where

$$g(k,\xi) = C(\xi)A(\xi)^{k-1}B(\xi), \quad k = 1,2,\ldots \tag{17}$$

is the impulse response of the parametrized model, $g_m(k)$ is the measured impulse response of an unknown system and $\|\cdot\|_F$ is the Frobenius matrix norm $\|X\|_F^2 = \mathrm{tr}(X^T X)$.

If $N = 2n$ we can use the vectorized representation of the impulse response

$$V(\xi) = \|\mathbf{g}_m - \mathbf{g}(\xi)\|^2 \tag{18}$$

where $\mathbf{g}_m$ represents the vectorized measured impulse response.

*Theorem 1.* Let $(A,B,C)$ be a minimal realization of a transfer function and consider the OrthImp parametrization as defined in Lemma 2. Let $N = 2n$ and assume $g_m(k)$ is the impulse response from the realization $(A,B,C)$. Then

$$V''(\xi)\big|_{\xi=0} = I \tag{19}$$

where $V''(\xi)$ denotes the Hessian of the criterion $V(\xi)$.

**Proof:** The assumption that $g_m(k)$ was generated by $(A,B,C)$ implies that $V(0) = 0$ which in turn leads to $V''(\xi)|_{\xi=0} = [\mathbf{g}'(\xi)^T \mathbf{g}'(\xi)]\big|_{\xi=0} = I$ where the last equality follows from Lemma 2. □

The result shows that if $(A,B,C)$ is chosen sufficiently close to the true system then the Hessian will be almost diagonal. In a stochastic setting this also tells us that the Fisher information matrix and the covariance matrix of the parameter estimate will both also be proportional to the identity matrix. Such a covariance matrix implies that all parameters are independent of each other and equally affected by the noise.

2.1 *Generalizations*

The methodology of defining the parametrization based on the impulse response can be generalized to arbitrary quantities which uniquely represents the system. In this way a whole class of local parametrizations can be derived. For example another local parametrization can be defined suitable for the prediction error method (PEM) (Ljung, 1999). Instead of using the impulse responses the parametrization is based on the sequence of prediction errors which appears in the least-squares criterion. A more detailed discussion of this is beyond the scope of this paper and is part of ongoing work.

2.2 *Reduced numerical complexity*

In a previous paper a similar local parametrization, OrthPar, was suggested (McKelvey and Helmersson, 1997). However, the calculation of the basis $P$ for the OrthPar parametrization requires the calculation of a null-space of a matrix of size $n^2 \times n(n+m+p)$. The new parametrization introduced in this paper requires a calculation of the range space of a matrix only of size $(2nmp) \times n(n+m+p)$. This requires significantly less number of operations if the number of inputs and outputs ($m$ and $p$) are reasonably small and the system order $n$ is large. Figure 1 illustrates the difference in numerical complexity in computing the basis $P$ for the OrthPar and OrthImp parametrizations for a system with 2 inputs and 2 outputs and varying state size $n$.

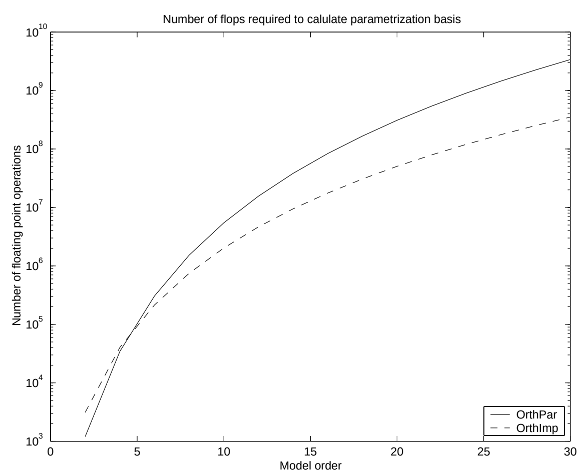



Fig. 1. The graph illustrates the number of floating points operations required for calculating the parametrization basis $P$ for the OrthPar and the new OrthImp parametrizations.

## 3. EXAMPLE

A numerical example of a multivariable identification problem is considered which consists of estimating a linear discrete time dynamical system from a measured impulse response. The system is determined by minimizing the sum of squares of the difference between noisy impulse response data and the impulse response of the parametrized system. Hence, the criterion function $V(\xi)$ in (16) is minimized with respect to the parameter vector $\xi$ and where $g_m(k)$ is the noisy impulse response of the true system and $g(k,\xi)$ is the impulse response of the parametrized system. Note that (16) is (generally) a non-quadratic function of the parameters and a nonlinear optimization algorithm must be employed to perform a local search for the optimum.

The example is based on simulated data and illustrates that the choice of parametrization can have a large impact on the convergence properties of the local search. For each data set four different parametrizations are used when minimizing the criterion function (16). The new OrthImp parametrization is compared with both the classical observable canonical parametrization (Ljung, 1999, Appendix 4b), the tridiagonal parametrization (McKelvey and Helmersson, 1996) and the more recently proposed OrthPar parametrization (McKelvey and Helmersson, 1997). For the observable canonical form we select structural indices such that the true transfer function is in the interior of the set of realizable transfer functions.

### 3.1 *Simulation setup*

A range of different systems of increasing McMillan degree is identified. The impulse response of the true system of order $2M$, $g_0(k)$, is generated from a system with complex conjugate poles defined by

$$p_k = \frac{(0.1 \pm i\sqrt{1-0.1^2})}{10}\left(1+\frac{4(k-1)}{M-1}\right)$$

for $k = 1,\ldots,M$ The zero configuration is generated by using random values of the elements in the matrices $C$ and $B$. For each system order $2M$, $M = 1,\ldots,4$, a sampled impulse response of length 200 is generated with a sampling interval of 0.1. The identification data $g_m(k)$ is obtained by adding Gaussian white noise with variance $10^{-4}$ to the generated impulse response. The Gaussian distribution of the noise implies that the model which minimizes the quadratic criterion (16) is the maximum-likelihood estimate (Ljung, 1999).

### 3.2 *Identification method and results*

The non-linear least-squares criterion (16) is minimized by iterative parametric optimization using the method of Levenberg-Marquardt (Dennis and Schnabel, 1983). The MATLAB implementation `lsqnonlin` of the MATLAB Optimization Toolbox ver 2.1 (R12) is used. The minimization is terminated if the criterion function decreases less than $10^{-6}$ between two iterations or if more than 1000 criterion evaluations has occurred.

The optimization algorithm is initiated with a parameter value derived from a perturbed system. The perturbation is constructed by starting from a balanced realization of the true system and perturbing all matrix elements with zero mean Gaussian random noise with variance $4\times10^{-4}$. The perturbed balanced form is converted to the form associated with the particular parametrization which then defines the initial point in the parameter space from where the optimization is started.

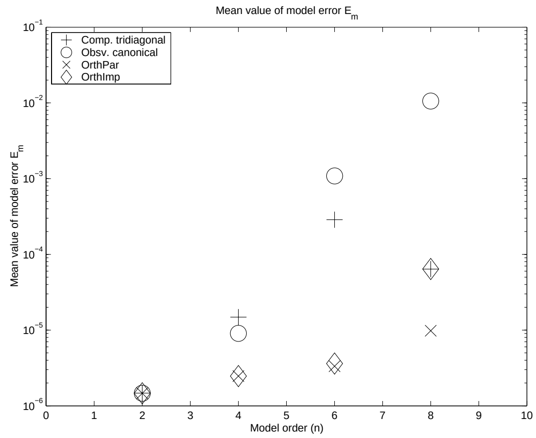
For each model order, 100 different random initial parameter points are generated and the criterion function (16) is minimized using the four different parametrizations. The quality of the estimated model is determined by calculating the model error as

$$E_m(\hat{g}) = \frac{1}{Npm}\sum_{k=1}^{N} \|g_0(k)-\hat{g}(k)\|_F^2 \qquad (20)$$

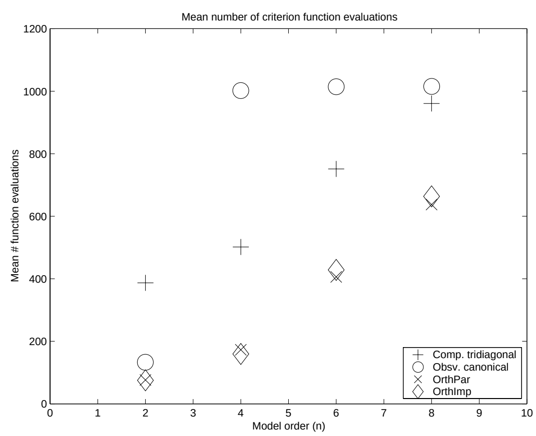
where $g_0$ is the true noise free impulse response and $\hat{g}$ is the impulse response of the identified model. Figure 2 reports the results of the simulations averaged over the 100 different initializations. For model order 2 the performance of all parametrizations are comparable . For model order 4 OrthImp and OrthPar have best performance while the Observable canonical and the Compact tridiagonal ones are comparable when comparing model error. However the Observable parametrization requires significantly more function evaluations. It might be expected that the tridiagonal parametrization which is non-minimal should require more calculations but instead it is the observable canonical form which requires most effort. For model orders 6 and 8 the observable canonical parametrization fails to converge for all 100 optimizations. A possible reason is that the optimization problem for that parametrization has become so ill-conditioned so the numerical optimization method fails completely to converge within the 1000 evaluation limit. The other three parametrization performs much better. The new OrthImp and the OrthPar parametrization have almost identical results while the tridiagonal parametrization requires more evaluations. The additional advantage of the new OrthImp parametrization is that the calculation of the parametrization basis $P$ requires less floating point operations than the OrthPar parametrization.

## 4. CONCLUSIONS

A new local minimal parametrization is presented which is based on parametrization of the impulse response with a basis such that the Jacobian of the finite impulse response of length $2n$ is an orthonormal matrix. Consequently it can be argued that the parametrization is locally optimal. Furthermore, if the parametrization is used in a least-squares impulse response criterion then, under certain conditions, the Hessian of the criterion is equal to the identity matrix. The parametrization is closely related to the recent OrthPar parametrization and in an example shows similar good convergence performance for high order systems. Furthermore the new parametrization can be calculated with less floating point operations than OrthPar which is important when dealing with models of high order.




(a) Mean value of model error $E_m$




(b) Mean number of evaluations

Fig. 2. Simulation results based on averaging over 100 optimizations from randomly perturbed initial models. Sub-figure (a) shows the mean value of the resulting model error $E_m$. Sub-figure (b) presents the average number of evaluations of the function $V(\xi)$ the Levenberg-Marquardt optimization routine used.

## 5. REFERENCES


Deistler, M. and T. Ribartits (2001). Parametrizations of linear systems by data driven local coordinates. In: *Proc. 40th IEEE Conference on Decision and Control, Orlando, Florida*.

Dennis, J. E. and R. B. Schnabel (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall. Englewood Cliffs, New Jersey.

Guidorzi, R. (1975). Canonical structures in the identification of multivariable systems. *Automatica* **11**, 361–374.

Hannan, E. J. and M. Deistler (1988). *The Statistical Theory of Linear Systems*. Wiley.

Ho, B. L. and R. E. Kalman (1966). Effective construction of linear state-variable models from input/output functions. *Regelungstechnik* **14**(12), 545–548.

Kailath, T. (1980). *Linear Systems*. Prentice-Hall. Englewood Cliffs, New Jersey.

Ljung, L. (1999). *System Identification: Theory for the User*. second ed.. Prentice-Hall. Englewood Cliffs, New Jersey.

Luenberger, D. G. (1967). Canonical forms for linear multivariable systems. *IEEE Trans. on Automatic Control* **AC-12**, 290.

McKelvey, T. (1994). Fully parametrized state-space models in system identification. In: *Proc. of the 10th IFAC Symposium on System Identification*. Vol. 2. Copenhagen, Denmark. pp. 373–378.

McKelvey, T. and A. Helmersson (1996). State-space parametrizations of multivariable linear systems using tridiagonal matrix forms. In: *Proc. 35th IEEE Conference on Decision and Control.* Kobe, Japan. pp. 3654–3659.

McKelvey, T. and A. Helmersson (1997). System identification using an over-parametrized model class - improving the optimization algorithm. In: *Proc. 36th IEEE Conference on Decision and Control*. San Diego, California, USA. pp. 2984–2989.

McKelvey, T. and A. Helmersson (1999). A dynamic minimal parametrization of multivariable linear systems and its applications to optimization and system identification. In: *Preprints of the 14th World Congress of IFAC, Beijing, P. R. China* (H.F. Chen and B. Wahlberg, Eds.). Vol. H. Elsevier Science. pp. 7–12.

Pintelon, R., J. Schoukens, T. McKelvey and Y. Rolain (1996). Minimum variance bounds for overparameterized models. *IEEE Trans. on Automatic Control* **41**(5), 719–720.

Wolodkin, G., S. Rangan and K. Poolla (1997). An LFT approach to parameter estimation. In: *Preprints to 11th IFAC Symposium on System Identification*. Vol. 1. Kitakyushu, Japan. pp. 87–92.