Fig. 3. Independent jitter. Comparison of theoretical average periodogram and simulation result of one sample.

Fig. 4. Accumulated jitter. Comparison of theoretical average periodogram and simulation result of one sample.

VII. CONCLUSIONS

The influence of sampling instabilities on spectral estimation by FFT was considered. The analytic results were confirmed with simulation. Two types of random sampling were treated. For the accumulated jitter in sampling instants, the distortion level is relatively high and the resolution capability is considerably degraded. Use of the FFT with this kind of instability is limited to a small amount of jitter, low input frequencies, and short sequences. In the independent jitter case, the distortion level is relatively low and the resolution capability is considerably conserved, even for a relatively large amount of jitter and high input frequencies.

REFERENCES


The Effect of Identifier Structure on Parameter Convergence

A. A. (Louis) Beex and Victor E. DeBrunner

Abstract—We consider the effects of a chosen system structure on the identification of its parameters. In particular, we compare the convergence rates for the parameters of three state-space structures: direct form II, parallel, and dual generalized Hessenberg representation structures. We show that the chosen structure does indeed influence identification algorithms, and that this influence is measured by examining the information contained in the structural parameters. We offer a conjecture about the relative convergence of the parameters, and provide evidence in its support. An important result is that identification of the usually identified direct form II parameters (the standard ARMA parameters) does not necessarily yield the fastest parameter convergence for the system being identified.

I. INTRODUCTION

Work on reduced sensitivity digital filter structures [1], [6], [7] leads to the question: How does parameter sensitivity affect identification convergence? The more sensitive a structure is to a parameter change, the more that parameter affects the system input/output description [8], [9]. We explore the connection of parameter sensitivity to parameter estimate convergence rates for a Gauss-Newton identification procedure [16]. System parameters are identified from a (partial) impulse response realization for 3 canonical single-input single-output (SISO) state-space structures (direct form II (DII), parallel (PPLL II) form, and dual generalized Hessenberg representation (DGHR) [14]). Parameter sensitivity is closely related to output error. High sensitivity models have parameters which provide much information about the system. However, maximizing the available information is not enough to ensure quick parameter convergence, parameter interdependence needs to be considered as well. This interdependence is explored via Martingale theory [4], [5]. Parameter information and interdependence are combined into one measure $r$, a self-styled "relative time constant" of system convergence. Numerical analysis arguments support the use of $r$. Illustrative examples are provided.

II. BACKGROUND

Define the parameter sensitivity as [6], [7], [19], [20]

$$S = \sum_{i=1}^{Np} \frac{1}{2\pi\gamma} \int \frac{\partial H(z)\partial H(z^{-1})}{\partial r_i} \frac{dz}{\gamma}$$

The contour is along the unit circle of the $z$ plane in a counterclockwise direction.

A. A. Beex is with the Bradley Department of Electrical Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0111.
V. E. DeBrunner was with the Bradley Department of Electrical Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0111. He is now with the School of Electrical Engineering and Computer Science, University of Oklahoma, Norman, OK 73019-0631.
IEEE Log Number 9202799.
where $H(z)$ is the system transfer function and $NP$ the number of system parameters $\gamma$. If the $\gamma$ are state-space parameters, then the controllability gramian $K$ and observability gramian $W$ are non-linearly related to the sensitivity [6], [7], [25]-[27]

$$S \leq \text{tr}(K) + \text{tr}(W) + \text{tr}(KW).$$

(2)

Grammians indicate system controllability and observability. It is well known that a system must be both controllable and observable for that system to be identifiable.

In optimal input design methods, the input is designed so as to most efficiently excite (and therefore make most determinable) the system parameters [13], [16], [22]. Optimal input design can entail the use of the system parameter correlation matrix $R$ and its inverse $F = R^{-1}$, the Fisher information matrix. In the presence of additive Gaussian noise, the $ij$th element of $F$ is [28]

$$f_{ij} = \frac{1}{2\pi \sigma^2} \int \frac{\partial H(z)}{\partial \gamma_i} \frac{\partial H(z^{-1})}{\partial \gamma_j} \frac{dz}{z} \quad i, j = 1, \ldots, NP.$$  

(3)

Comparison of (3) and (1) shows that

$$S = \text{tr}(F).$$

(4)

Optimal input design seeks to maximize $F$ by some appropriate measure of the size of $F$ (e.g., tr($F$)).

Martingale theory (the study of stochastic differential equations) is one approach used to determine the convergence of parameter estimates. Recursive least squares and stochastic gradient identification schemes consistently converge exponentially and asymptotically if the eigenvalue spread of the approximation to the parameter correlation matrix $R$ is not too great [4], [5]. As $R$ is positive definite symmetric, the eigenvalue spread of $F$ is equal to that of $R$. The actual rate of exponential convergence is shown to depend inversely on the spread, i.e., the closer the minimum and maximum eigenvalues of $R$ are to each other, the faster the convergence [5]. The ratio of maximum-to-minimum eigenvalue of $F$ is strongly tied to the idea of persistent excitation [15], [18], [24].

III. THE GENERAL OUTPUT ERROR METHOD

A deterministic gradient algorithm suitable for identifying the system parameters of a stable, but otherwise arbitrary, state-space structure from a (noisy) impulse response measurement was presented [2], [13], [8], [9], [16]. Define the impulse response error at time $n$ as $e_n = y_n - h_n(\Theta)$. $y_n$ is the noisy impulse response measurement and $h_n(\Theta)$ is the impulse response corresponding to parameterization $\Theta$. The identification criterion is

$$J_N = \frac{1}{2} \sum_{n=0}^{N} e_n^2.$$  

(5)

When the measurements are corrupted by additive Gaussian noise, the estimate of the Hessian is an approximation to $F$. Convergence of the algorithm to the "true" parameters is not guaranteed, as $J_N$ may have local minima [10], [23]. Note that the same noisy impulse response record is used repeatedly in this method, so that convergence is to that set of system parameters which minimizes $J_N$ [15].

IV. ERROR GRADIENT-SENSITIVITY RELATIONSHIP

It was shown in [8] that a high error gradient of $J_N$ in the $\Theta_i$ direction implies high sensitivity $s_i$ and vice versa. Figs. 1 and 2 indicate how the error for each parameter (keeping all other parameters at their true value) versus the normalized parameter deviation

Fig. 1. Error criterion for DGHR parameters: $J_{05}$ slightly correlated to $S$.

![Fig. 1](image)

Fig. 2. Error criterion for DGHR parameters: $J_{05}$ highly correlated to $S$.

![Fig. 2](image)

of the LP2 DGHG system (described in the next section) is related to the parameter sensitivity. In Fig. 1 the error plotted is $J_{05}$, while in Fig. 2 the error shown is $J_{05}$. Notice that the slopes in Fig. 2 correspond much better with the sensitivities than do those in Fig. 1. The parameter-error/sensitivity behavior in Fig. 2 is representative for all the examples given in Section VI.

V. A CONJECTURE FOR ANALYZING CONVERGENCE AND SENSITIVITY

Viewing parameter convergence as a quadratic minimization [21] also leads to a numerical reason for exploring the Fisher information matrix. The error criterion $J_N$ is nearly quadratic in the parameters with the ellipsoid defined by the Fisher information matrix [8], i.e.,

$$\Delta J_N \approx \frac{1}{2} \Delta \Theta^T F \Delta \Theta.$$  

(6)

Figs. 1 and 2 also suggest that $J_N$ is nearly quadratic in the vicinity of the optimal solution.
We claim that the local convergence properties of our identification algorithms depend to a large extent on $F$, a matrix dependent on the identified structure. Furthermore, this convergence depends upon two characteristics of the matrix: 1) $\text{tr}(F)$—the sensitivity; and 2) $\kappa_2(F)$—the condition number (which is the ratio of its maximum-to-minimum eigenvalues).

When the matrix is well conditioned, the isolines of $J_N$ are nearly circular. When the condition number is large, the isolines are elongated hyperellipsoids. When the trace is high, the "valleys" of $J_N$ are steep, while if the trace is low, the "valleys" are broad. If the trace is large (high sensitivity) and the structure is also ill-conditioned (high $\kappa_2$), slow convergence of the quadratic minimization in (8) results [11]. We combine these two measures in the algorithm relative time constant

$$\tau = \frac{\kappa_2(F)}{\text{tr}(F)} \tag{7}$$

When $\tau$ is small, the convergence should be quite fast, while for $\tau$ large the convergence is slow. Our measure $\tau$ combines the ideas of high sensitivity and low eigenvalue spread. Formally, we conjecture as follows.

**Conjecture:** Let $S_1$ and $S_2$ be two different parameterizations of the same system of known order. Assume the optimization criterion $J_N$ possesses no local minima. Apply a gradient based algorithm to minimize $J_N$. Then, if $\tau_1 \leq \tau_2$, for the same measurement data and equivalent local parameter initializations and covariances, the parameters of $S_1$ will have a higher asymptotic convergence rate toward the optimal values than those of $S_2$.

How local the initial guess must be to the true system is an open question at this point. Practical considerations are illustrated in the example section.

Prior to exploring example identifications, we first discuss some convergence results with direct bearing on the above conjecture. Let $C^{(2)}$ be the class of continuous functions with continuous derivatives of orders 1 and 2. Since the identification procedures use a Gauss direction, the convergence is superlinear if $J_N$ is not in $C^{(2)}$, and quadratic otherwise [12]. The convergence rate is bounded above by the condition number of the Hessian approximation (i.e., by $\kappa_2(F)$ as $N \to \infty$) [17]. In this case, the higher the condition number, the higher the error bound, implying that slower convergence may occur. Another upper bound on the convergence rate of the estimated parameters can be found by examining the rate at which the Hessian approximation converges to the true Hessian [12]. In this case, the faster the Hessian approximation converges to the true Hessian, the lower the upper bound on the parameter estimates, implying that faster convergence may occur. Examination of the identification algorithms [16] shows that for quick Hessian convergence (and $N$ large), $S = \text{tr}(F)$ must be large. Since these results are based on upper bounds for the errors, we cannot make a quantitative statement about the conjecture; however, these two ideas strongly suggest that $\tau$ is indicative of convergence rates.

**VI. Examples**

Since the nature of the relative time constant $\tau$ is based on heuristic arguments, many (and carefully chosen) simulations must be
TABLE I
LP1 FISHER INFORMATION MATRIX DATA

<table>
<thead>
<tr>
<th>$S$</th>
<th>$k_2$</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DII</td>
<td>$6.12 \times 10^2$</td>
<td>$3.65 \times 10^2$</td>
</tr>
<tr>
<td>PRLL</td>
<td>$1.14 \times 10^2$</td>
<td>$4.66 \times 10^2$</td>
</tr>
<tr>
<td>DGHR</td>
<td>$9.70 \times 10^2$</td>
<td>$2.39 \times 10^2$</td>
</tr>
</tbody>
</table>

TABLE II
BP1 FISHER INFORMATION MATRIX DATA

<table>
<thead>
<tr>
<th>$S$</th>
<th>$k_2$</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DII</td>
<td>415.49</td>
<td>1164.62</td>
</tr>
<tr>
<td>PRLL</td>
<td>441.24</td>
<td>2687.16</td>
</tr>
<tr>
<td>DGHR</td>
<td>981.41</td>
<td>1922.94</td>
</tr>
</tbody>
</table>

examined. Due to the brevity of this correspondence item, only 2 representative systems are presented. The first system, LP1, has poles well within the unit circle. The number of impulse response samples $N$ was chosen to be 6 since $k_2$ is strongly correlated to $S$, even at this low number of samples. The convergence shown in Fig. 3 indicates that the time constant accurately predicts the relative convergence rate (see Table I). Note that while the DII structure is more ill-conditioned than the PRLL structure, the sensitivity overcomes the conditioning to the point that the DII structure actually converges faster than the PRLL structure. Also, sensitivity by itself is not enough to predict the convergence, as the DII structure has a higher sensitivity than the DGHR structure, and yet the DGHR structure is so well-conditioned that it overcomes the low sensitivity to converge faster than the DII structure. The initial parameter estimate $\theta_0$ is close to the true parameters $\theta_{true}$.

System BP1 illustrates how the convergence results can be influenced by the initial parameter guesses. In Fig. 4, one can see that the convergence paths do not follow $r$ (from Table II). The DGHR structure has the lowest $r$, and yet the DII structure actually outperforms it. However, by restarting the identification procedure of all three structures at the system given by the DII estimates at iteration 400, we see (Fig. 5) that the DGHR structure outperforms the DII structure (note that the DGHR structure of this estimated system has a lower $r$ than the DII estimated system, indicating that convergence rates can be increased by monitoring $r$ of multiple structures, and switching to that structure with the lowest $r$). The relative convergence is now in agreement with the conjecture.

VII. CONCLUSIONS

We have shown that model structure makes a difference in the convergence rates of the identified parameters. The structural influence is examined via the Fisher information matrix. The sensitivity of the system to its parameters indicates the degree to which the individual parameters of the system are identifiable. The parameter interconnection also affects parameter estimate convergence. It is measured by the condition number of the Fisher information matrix. Based on observation, a measure $r$ was introduced and a convergence conjecture was offered.

The results with 3 different identifier structures indicated that each structure yields the best performance for a different system. None is consistently the best. We also noted that if we start the identification of all three structures with an arbitrary initial parameter choice, a change to the structure whose $r(\theta_0)$ is smaller yields the fastest convergence.

REFERENCES


