Appl. Math. and Comp. Sci., 1997, Vol.7, No.2, 377–399

PARAMETER IDENTIFICATION IN ROBOT CONTROL[†]

CARSTEN SCHMIDT*, MARTIN PRÜFER*

This paper discusses practical problems of robot dynamics identification. The problems as well as possible solutions are illustrated by records from an experimental direct drive arm and a geared industrial robot. Questions regarding measurement noise, biased estimates, poor identifiability conditions, and the concepts of on-line and off-line identification are also discussed. The main contributions of the paper are a computationally cheap procedure for on-line identifiability supervision and a scheme which permits to reduce the bias of the least-squares estimator.

1. Introduction

1.1. Identification in Robot Control

A large majority of more sophisticated robot control algorithms require knowledge of a mathematical model of the robot's dynamics. Since these algorithms usually assume a particular model structure, it is not necessary to identify the structure of the model but it is sufficient to adapt some model parameters $\underline{x}$ to the actual robot.

The models usually are used to compute the course of torques (resp. forces) $\underline{\tau}(t)$ necessary to achieve a desired movement $\underline{q}_d(t)$. (See notation in Fig. 1.) The demand of a close match between the computed torques and the actually necessary torques provides a criterion for parameter adjustment. Figure 2 shows an application in robot control.

1.2. Peculiarities of Robot Dynamics Identification

The most common application of parameter identification in control engineering is to use it with linearized models of the behaviour of some nonlinear system around a slowly time-varying operation point. This enables one to adjust exactly a linear controller, which would not be possible for the global (nonlinear) model. Furthermore, there is a wealth of theory on identification of linear models available. Since robot dynamics are nonlinear, it is reasonable to use the linearized model depending on the comparatively slowly changing joint positions and velocities in robot control as well.

[†] Parts of this paper appeared in Proc. 3-rd Int. Symp. *Methods and Models in Automation and Robotics* (Prüfer and Schmidt, 1996). The present paper provides a comprehensive view.

* Institute for Robotics and Process Control, Technical University of Braunschweig, Hamburger Straße 267, 38114 Braunschweig, Germany, e-mail: M.Pruefer@tu-bs.de.

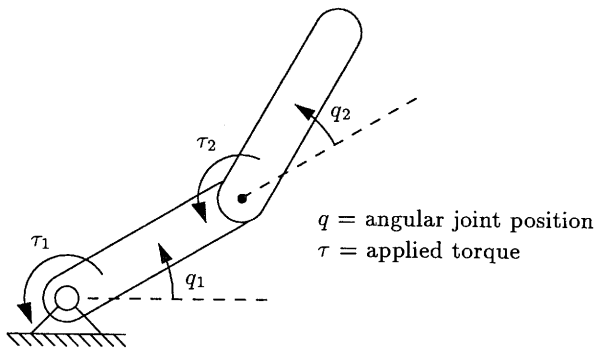

Fig. 1. Example notation for a robot with two rotary joints.

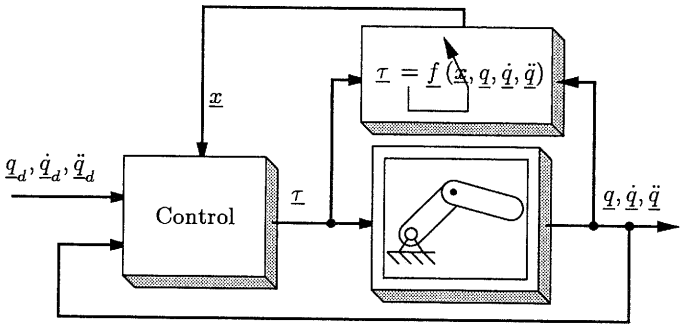

Fig. 2. Application of robot dynamics identification within robot control; $\underline{\tau} = \underline{f}(\underline{x}, \underline{q}, \dot{\underline{q}}, \ddot{\underline{q}})$ represents the mathematical model of the robot dynamics, $\underline{x}$ is a vector of model parameters which are identified and transferred to the controller.

This approach, however, turns out to be difficult since parameter changes are as fast as the setpoint changes which permit to observe the dynamical behaviour to be identified. Therefore it is impossible to gather enough linear independent data to calculate parameter estimates before the parameters change again. This problem cannot be solved by simply increasing the rate of data acquisition because the additional data will contain redundant information only. The usual approach of adding small high-frequency disturbances will not work because the dynamical behaviour of the system is quite different at frequencies above the frequency range of interest: the actuators would not permit to generate the necessary high frequency torque changes.

Therefore, the usual approach in robot control is to use the full nonlinear model of the robot dynamics. The unknown parameters which appear in these equations are inertial and friction properties of the mechanical structure which may be considered to be constant (as compared with the parameters of the linearized model)

(Kozłowski and Prüfer, 1992). This is a great advantage because of two consequences:

1. Since we know in advance—from physical considerations—that parameters do not change, it is possible to predict their values based on previously measured data.
2. The assumed constancy of the parameters provides a means to distinguish them from other effects, in particular setpoint changes, measurement noise and modelling errors. The familiar demand of 'exciting trajectories' only makes sense on this assumption of constant parameters. Otherwise, if parameters were known to change permanently and rapidly, one would try to keep constant everything else, such that the parameter changes could be observed without interference of other effects. The essential requirement is not that of exciting trajectories but that of maximum difference of the frequency range of the parameters as compared with everything else.

But this advantage has to be paid for by the fact that the model is no longer linear. Fortunately, this does not mean that parameters have to be found by some searching algorithm. The model of robot dynamics actually is linear in the parameters which means that the set of optimal parameters can be found in one step by solving a system of linear equations instead of searching a multidimensional parameter space. Still, the fact that there appear nonlinear functions of the measured data in the model equations causes problems which will also be discussed in this paper.

From a theoretical point of view, three properties of robot dynamics are of special importance to the identification:

1. (Nearly) constant parameters.
2. The model is linear in the parameters.
3. The data values in general are nonlinear functions of the measured values.

2. Basic Considerations

2.1. On-line versus Off-line Identification

Because of the first of the above mentioned properties—the constancy of parameters—it is self suggesting to perform the identification off-line and to adjust the robot control before operation accordingly. This approach is frequently taken and has the merit that one can avoid many of the problems of identifiability of parameters, since a series of experiments can be carefully designed, such that single properties are clearly observable in each experiment, until all interesting parameters can be accurately inferred from the accumulated information (Seeger and Leonhard, 1989).

Although this seems quite reasonable and in spite of the apparent simplicity of this approach we abandon it and choose to tackle the problems that are implied with on-line identification. This decision is justified by

- the observation that the dynamic parameters actually are not constant,
- the fact that the problems of on-line identification can be solved, and

- the fact that an on-line identification scheme finally presents a more general approach which can be easily specialized also to the task of off-line identification; thus, we do not necessarily give up the advantages of the off-line identification approach.

2.1.1. Simplicity of Adjustment

It is obvious that some mass parameters of a robot may change, e.g. when tools are changed. The repetition of the above-mentioned series of experiments after each configuration change, including data acquisition, analysis of the data and finally transferring the results to the control system, seems to be highly impractical in real applications. Furthermore, this procedure has to be repeated regularly to verify the values of the parameters.

Parameter changes may also occur when objects are grabbed, moved and released. Finally, friction may change with time, due to aging lubricants.

2.1.2. Temperature Dependence of Friction

Unexpected drastical changes of friction can be observed with geared industrial robots (e.g. Manutec r2,r15) when the temperature of the joints is changing (Prüfer and Wahl, 1994). Figure 3 shows the velocity-dependent friction as a function of the actuator temperature. It should be pointed out that the temperature range shown in the diagram was the result of changes in the speed of robot operation.

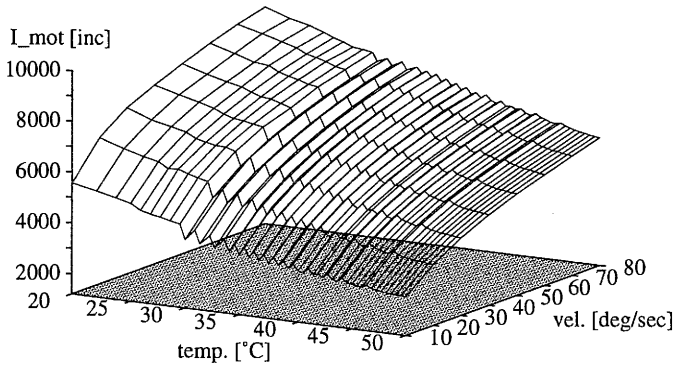

Fig. 3. Motor current v. temperature and velocity.

In order to deal with these changes by off-line identification, it is necessary to model these temperature dependencies and to identify them along with the other parameters of the robot dynamics (Prüfer and Wahl, 1994). This means that temperature sensors have to be placed at each joint at positions with a high correlation between temperature and friction changes.

On-line identification of the friction parameters, on the other hand, could be accomplished easily because the temperature changes are very slow. This would avoid the placement of temperature sensors and the mathematical model would be simpler.

2.2. Parameter Estimator

It has already been pointed out that the linearity of the model in the parameters has the consequence that the parameters can be found by solving a system of linear equations rather than searching the parameter space. This is the reason why on-line identification is feasible at all: it permits to process the measured data at the high rates necessary on account of the relatively rapid movements of robots (Fig. 4).

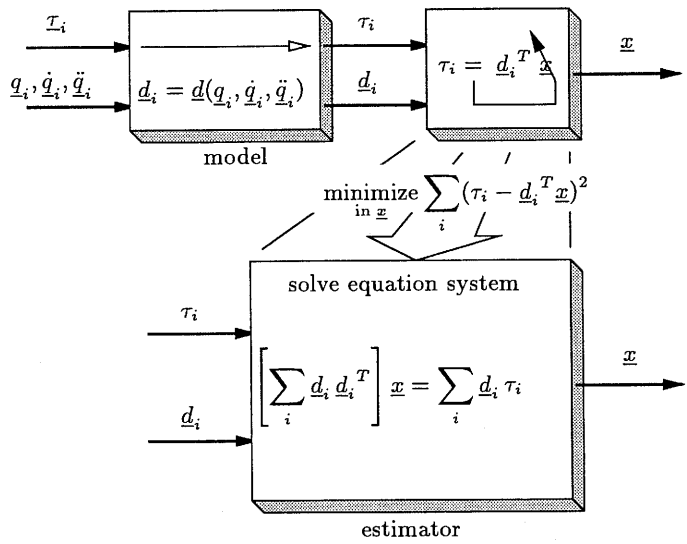

Fig. 4. On-line identification is realized as a pipeline of the model and the LS-estimator.

In order to represent this linearity, the model equation for one joint shall be written in the form

$$\tau_i = \underline{d}(\underline{q}_i, \underline{\dot{q}}_i, \underline{\ddot{q}}_i)^T \underline{x} \tag{1}$$

where i is the index of sample instant, $\underline{q}$, $\underline{\dot{q}}$, $\underline{\ddot{q}}$ denote the vector of measured joint positions and its derivatives, respectively, τ is the steered torque that is output to the driver, $\underline{d}(\underline{q}_i, \underline{\dot{q}}_i, \underline{\ddot{q}}_i)$ stands for the vector of model functions of the joint applied to the measured values from sample instant t_i, $\underline{d}_i$ is the abbreviation for $\underline{d}(\underline{q}_i, \underline{\dot{q}}_i, \underline{\ddot{q}}_i)$, and $\underline{x}$ denotes the vector of model parameters to be estimated.

The equation error e_i, resulting from mismatched parameters and from measurement noise or modeling errors, is

$$e_i = \tau_i - \underline{d}_i^T \underline{x} \tag{2}$$

The criterion for parameter adjustment is the minimization of the sum of squared equation errors:

$$\sum_i e_i^2 = \sum_i (\tau_i - \underline{d}_i^T \underline{x})^2 \stackrel{!}{=} \min_{\text{in } \underline{x}} \tag{3}$$

It should be noted that the equation errors of *all* equations (one per joint) have to be summed up. The distinction between the different model equations is not necessary for the following discussion. Hence, for the benefit of simpler expressions, we write them only in index i, assuming that i is used to distinguish between sample instants as well as between model equations.

The minimum of this sum can be found by setting its derivatives with respect to the unknown parameters to zero, which yields the system of equations

$$\left[\sum_i \underline{d}_i \underline{d}_i^T\right] \underline{x} = \sum_i \underline{d}_i \tau_i \tag{4}$$

3. Practical Problems

Since commercial robots usually are equipped with position sensors only, it is common to restrict the identification input to position values (and to the torques that are output to the actuators). In this way, the general problem of disturbed measurement becomes a special problem with robot dynamics since the necessary acceleration signal has to be derived numerically from the position values (Fig. 5).

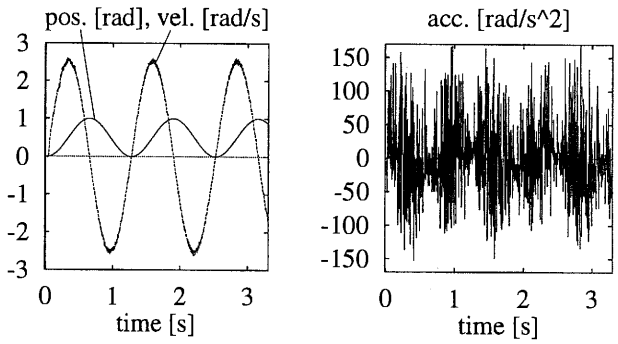

Fig. 5. Acceleration signal (right) derived numerically from the measured joint positions (left).

3.1. Measurement Noise

The consequence of the disturbed acceleration signal might be surprising if one expects to see disturbed estimates and slower convergence during data acquisition. Figure 6 shows[1] recursive least squares estimates (UD-Covariance filter (Bierman, 1977)) in comparison with a reference method, which—in this case—gives optimal estimates.

[1] Experimental details: Each diagram shows six repetitions of the same identification experiment to visualize the reproducibility. Identification starts at time zero with no information. The sample interval is one millisecond. If not specified, a two-link direct drive arm is used. Joint 1 performs the movement from Fig. 5 (0.8 Hz). Joint 2 performs a similar movement at 1.1 Hz.

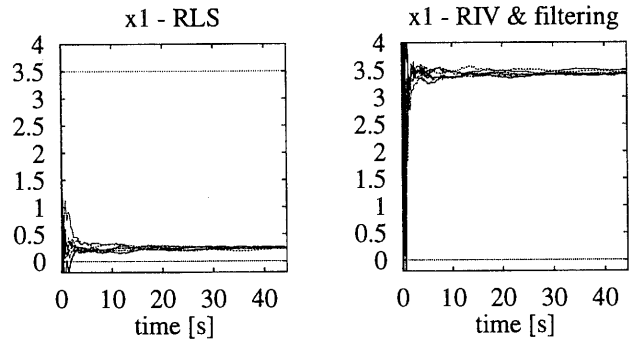

Fig. 6. Convergence of least-squares parameter estimates (RLS, left) in comparison with an optimal reference method (right).

The stable and reproducible convergence may be misleading. Only comparison with other methods reveals that the estimates are biased up to 100 %. In fact, the least-squares method takes the measurement noise to some extent as an information signal.

This can be seen when representing the model function values $\underline{d}$ as the sum of 'true' values $\underline{d}_0$ and the measurement noise $\underline{n}_d$, and similarly the steered torques τ as the sum of the effective values τ_0 and the difference n_τ that results from imperfect actuators:

$$\underline{d} = \underline{d}_0 + \underline{n}_d, \qquad \tau = \tau_0 + n_\tau \tag{5}$$

With this decomposition, the sum of squared errors (3) becomes

$$\sum_i e_i^2 = \sum_i \Big(\tau_{0i} + n_{\tau i} - (\underline{d}_{0i} + \underline{n}_{d\,i})^T(\underline{x}_0 + \Delta\underline{x})\Big)^2 \tag{6}$$

where the set of parameter estimates $\underline{x}$ is composed of the 'true' set $\underline{x}_0$, for which the model $\tau_0 = \underline{d}_0^T\underline{x}_0$ is valid, and of the estimation error $\Delta\underline{x}$.

Application of the relationship $\tau_0 = \underline{d}_0^T\underline{x}_0$ gives

$$\sum_i e_i^2 = \sum_i \Big(\underbrace{n_{\tau i} - \underline{n}_{d\,i}^T\underline{x}}_{\text{error noise}} - \underbrace{\underline{d}_{0i}^T\Delta\underline{x}}_{\substack{\text{parameter}\\ \text{mismatch}}} \Big)^2 \tag{7}$$

If the noise signals n_τ and $\underline{n}_d$ are uncorrelated, have zero mean values and are statistically independent from the information signals $\underline{d}_0$ and τ_0, then the expectation of the sum of squared errors is

$$E\left\{\sum_i e_i^2\right\} = \sum_i E\{n_\tau^2\} + E\Big\{(\underline{n}_d^T\underline{x})^2\Big\} + E\Big\{(\underline{d}_0^T\Delta x)^2\Big\} \tag{8}$$

$E\{(\underline{d}_0^T \Delta x)^2\}$ becomes minimal if the estimation error $\Delta \underline{x}$ is zero. This explains, why minimization of this sum has been chosen as an estimation criterion. The first term, $E\{n_\tau^2\}$, increases the minimal value of the sum, but it does not shift the position of the minimum within the parameter space—unlike the second term $E\{(\underline{n}^T \underline{x})^2\}$ which generally becomes minimal if the parameter values are zero. Hence the parameter estimates are shifted closer to zero (Fig. 6).[2]

The most common way to avoid bias (which is to impose a relationship between the components of $\underline{n}$ such that the term $E\{(\underline{n}^T \underline{x})^2\}$ becomes zero when $\underline{x} = \underline{x}_0$) is possible only if the components of $\underline{n}$ are time-shifted values of the same signal. This is not the case with robot dynamics, however, and hence this approach is not applicable.

Since in robot dynamics identification the problem is mainly caused by the acceleration signals, it would be self-suggesting to rearrange the model equation such that the acceleration signals appear in the position of the torques

$$\sum_i \left(\ddot{q}_i - f(\underline{\tau}_i, \underline{q}_i, \underline{\dot{q}}_i, \underline{x})\right)^2 \stackrel{!}{=} \min_{\text{in } \underline{x}} \tag{9}$$

In this way, the noise of $\ddot{q}$ appears in place of n_τ in (8) and would not contribute to the displacement of the minimum within the parameter space. Yet, f in eqn. (9) would not be linear in the parameters any more because it results from inverting the inertia matrix of the robot dynamics equation. As pointed out in Section 2.2, this would require employing a minimum-searching algorithm and, thus, computation in real time would not be possible any more.

Other measures against bias can be read from the expectation of the linear equation system yielding the parameter estimates:

$$E\left\{\sum_i \underline{d}_i \underline{d}_i^T\right\} \underline{x} = E\left\{\sum_i \underline{d}_i \tau_i\right\} \tag{10}$$

The expectation of the information matrix on the left-hand side decomposes into

$$E\left\{\sum_i \underline{d}_i \underline{d}_i^T\right\} = \sum_i \left[E\{\underline{d}_0 \underline{d}_0^T\} + E\{\underline{d}_0 \underline{n}_d^T + \underline{n}_d \underline{d}_0^T\} + E\{\underline{n}_d \underline{n}_d^T\}\right] \tag{11}$$

while the right-hand side of the equation system is

$$E\left\{\sum_i \underline{d}_i \tau_i\right\} = \sum_i \left[E\{\underline{d}_0 \tau_0\} + E\{\underline{n}_d \tau_0 + \underline{d}_0 n_\tau + \underline{n}_d n_\tau\}\right] \tag{12}$$

If τ_0, n_τ, $\underline{d}_0$ and $\underline{n}_d$ are statistically independent of each other and if the noise signals n_τ and $\underline{n}_d$ have zero mean values, then the covariance terms vanish and

[2] Of course, the point in the parameter space where $E\{\sum_i e_i^2\}$ becomes minimal does not permit to conclude anything about the statistical properties of the parameter estimates, in particular about the expectation $E\{\underline{x}\}$. But it can be shown that $E\{\underline{x}\}$ approaches this point asymptotically with data acquisition because the parameter estimates finally become constants.

the variance of the vector of measurement noise $\underline{n}_d$ appears to account for biased parameter estimates:

$$\left[\sum_i E\{\underline{d}_0\,\underline{d}_0^T\} + E\{\underline{n}_d\,\underline{n}_d^T\}\right]\underline{x} = \sum_i E\{\underline{d}_0\tau_0\} \tag{13}$$

Again the disturbance n_τ does not contribute to any bias because it only appears in the covariance terms. This suggests that something similar should be accomplished with the noise signal $\underline{n}_d$. If e.g. one of the $\underline{d}$-vectors in the expression for the information matrix $\sum_i \underline{d}_i\,\underline{d}_i^T$ could be replaced by a vector of signals $\underline{d}^*$ which also contains the information signals $\underline{d}_0$ but a different vector of noise signals, $\underline{n}_d^*$ instead of $\underline{n}_d$, which is not correlated with $\underline{n}_d$, then there would be no variance terms left. This is the basic idea of the instrumental-variable method.

3.1.1. Instrumental-Variable Method

The instrumental-variable method (Söderström and Stoica, 1983) uses a modified version of eqn. (4):

$$\left[\sum_i \underline{d}_i^*\,\underline{d}_i^T\right]\underline{x} = \sum_i \underline{d}_i^*\,\tau_i \tag{14}$$

where the instrumental variable $\underline{d}^*$ has to satisfy

$$\begin{aligned} \underline{d}^* &= \underline{d}_0 + \underline{n}_d^* \\ E\{\underline{n}_d^*\} &= \underline{0} \\ E\{\underline{n}_d^*\,\underline{n}_d^T\} &= \text{zero matrix} \end{aligned} \tag{15}$$

The problem of this method is to find such instrumental variables. The most common approach is to simulate the system based on already identified parameters and to use the signals from the simulation as instrumental variables. There are usually convergence problems with this approach if the initial parameter values are not in the vicinity of the 'true' values $\underline{x}_0$. A simple alternative would be to use the control loop input to derive the instrumental variables. In robot control applications one would apply the functions $\underline{d}(\underline{q},\dot{\underline{q}},\ddot{\underline{q}})$ to the desired values of $\underline{q}$, $\dot{\underline{q}}$ and $\ddot{\underline{q}}$ (instead of their measured values) to obtain the instrumental variables $\underline{d}^*$.

In this way, the variance term in (11), $E\{\underline{n}_d\,\underline{n}_d^T\}$, can be avoided:

$$E\left\{\sum_i \underline{d}_i^*\,\underline{d}_i^T\right\} = \sum_i \left[E\{\underline{d}_0\,\underline{d}_0^T\} + E\{\underline{d}_0\,\underline{n}_d^T + \underline{n}_d^*\,\underline{d}_0^T\} + E\{\underline{n}_d^*\,\underline{n}_d^T\}\right] \tag{16}$$

Unfortunately, this does not mean that bias can be entirely eliminated. The model functions $\underline{d}(\underline{q},\dot{\underline{q}},\ddot{\underline{q}})$ are non-linear functions of the measured $\underline{q}$, $\dot{\underline{q}}$, and $\ddot{\underline{q}}$, which means that the measurement noise of $\underline{q}$, $\dot{\underline{q}}$, and $\ddot{\underline{q}}$ and the information signals have passed the generally non-linear model functions $\underline{d}(\underline{q},\dot{\underline{q}},\ddot{\underline{q}})$ together such that the error noise $\underline{n}_d$ is not completely uncorrelated with the information signal $\underline{d}_0$. Hence the

covariance term $E\{\underline{d}_0\,\underline{n}_d^T + \underline{n}_d^*\,\underline{d}_0^T\}$ does not vanish. This theoretical consideration is the main reason why the problem of bias has not been addressed previously in robot dynamics identification. Nevertheless, there will be an improvement because the bias is dominantly caused by the variance terms. In order to get an idea of the magnitude of the difference, we calculated estimates of $E\{(\underline{d}-\underline{d}^*)(\underline{d}-\underline{d}^*)^T\} = E\{(\underline{n}_d-\underline{n}_d^*)(\underline{n}_d-\underline{n}_d^*)^T\}$ and compared them with estimates of $E\{(\underline{d}-\underline{d}^*)\underline{d}^{*T}\} = E\{(\underline{n}_d-\underline{n}_d^*)(\underline{d}_0+\underline{n}_d^*)^T\}$. The dominant elements of the first expression which contains $E\{\underline{n}_d\,\underline{n}_d^T\}$ appear to be of about two orders of magnitude higher than the elements of the second expression, which contains $E\{\underline{n}_d\,\underline{d}_0^T\}$ instead (with a factor of 65 to 300).

Figure 7 shows that the parameter estimates produced by the instrumental variable method behave in a way which one would expect: the disturbances of the acceleration signal (Fig. 5) are not taken as part of the information signal and hence cause the expected disturbances of the parameter estimates. On the other hand, it can be seen that the bias from Fig. 6 disappeared.

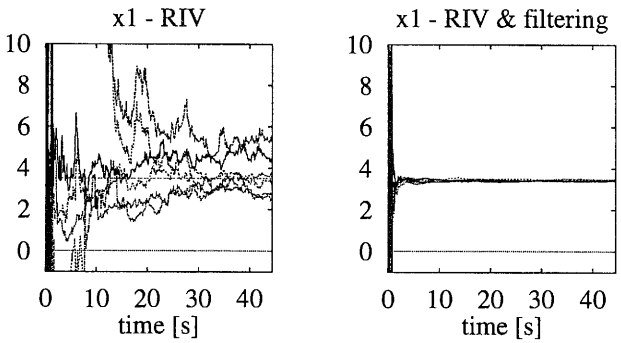

Fig. 7. The Recursive Instrumental-Variable (RIV) method (left) in comparison with an optimal reference identification (right).

There is still something to be done regarding the disturbances of the parameter estimates. This leads to filtering.

3.1.2. Data Filtering

It is a relative simple task to reject the noise of the acceleration signals by low-pass filtering because it results from the numerical derivation and, hence, is by its very nature a high-frequency noise (depending on the sample rate, Fig. 8 and 9).

It should be pointed out, however, that simply filtering the acceleration signal represents a change in the model equation. It yields an equation which is, in a mathematical sense, not equivalent to the original model equation. Although, at a first glance, this seems to be a minor change, it can be shown that it affects the parameter estimates in a clearly recognizable way. The correct method would be to apply the low-pass filter to the complete model (Fig. 10 and 11). Figure 12 shows the model functions with filtering of the input data and with filtering of the model equations. Figure 13 shows one of the resulting parameters with both methods.

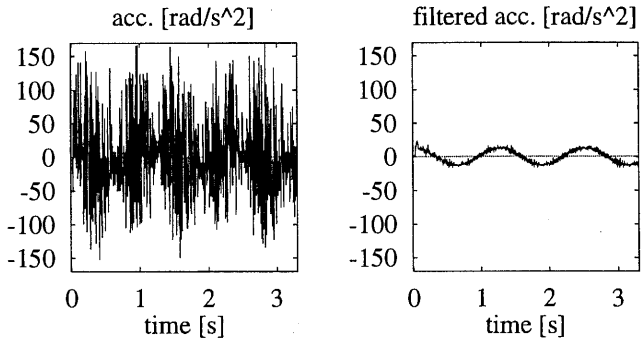

Fig. 8. Unfiltered and filtered acceleration signal.

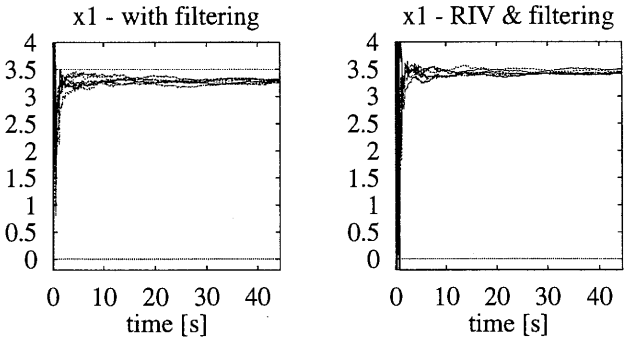

Fig. 9. Estimation with filtered acceleration signal (left) in comparison with a reference identification (right).

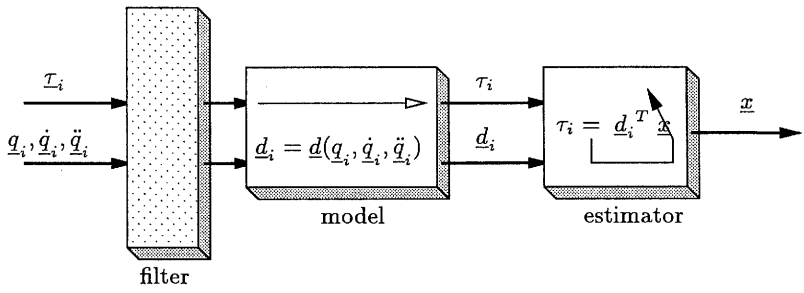

Fig. 10. Input filtering.

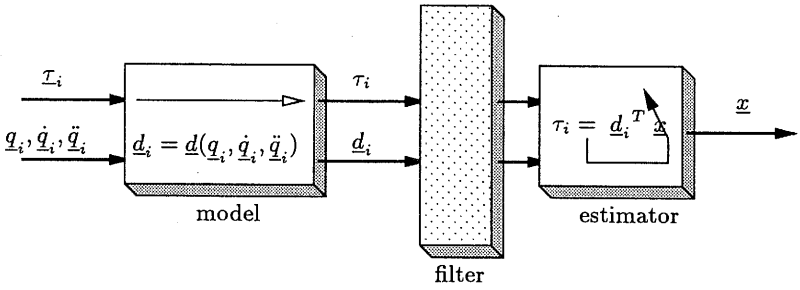

Fig. 11. Model filtering.

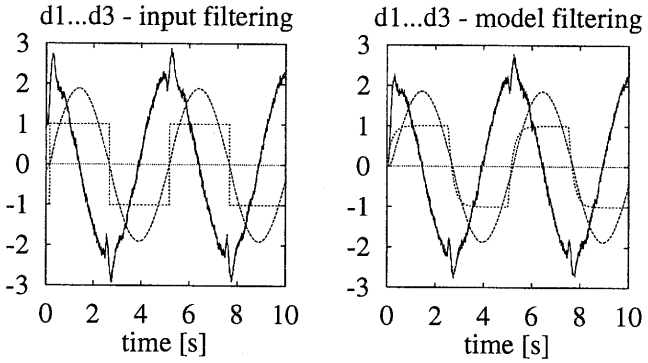

Fig. 12. Model function values of a simple one-link model: acceleration (inertia), velocity (dynamic friction) and sign of velocity (static friction). For this experiment a geared industrial robot (r15) has been used.

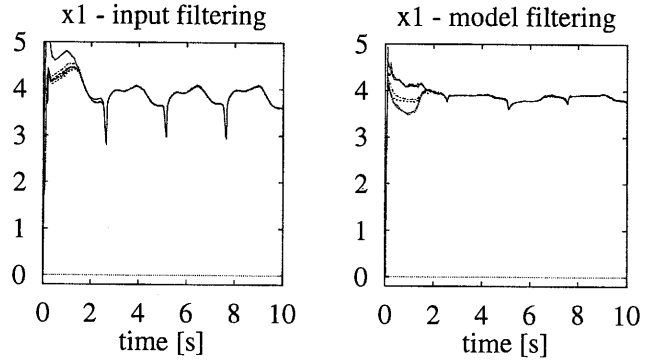

Fig. 13. Parameter estimates (inertia) of a simple one-link model with filtering of measured data (left) and with filtering of model function values (right).

3.1.3. Integral Model

Another approach to avoid the problems caused by noisy acceleration signals has been proposed in the literature on robot dynamics identification (Gautier and Khalil, 1989). It has been suggested to use a model of the mechanical energy of the robot's dynamical state (Fig. 14). The parameters are identical with those of (1):

$$\int_0^{t_i} \underline{\tau}^T \underline{\dot{q}} \, \mathrm{d}t = \underline{d}(\underline{q}_i, \underline{\dot{q}}_i)^T \underline{x} \tag{17}$$

Like the model from eqn. (1), this model is linear in the parameters. But the vector of model functions $\underline{d}$ does not depend on the disturbed acceleration signal. As a result, the parameter estimates (Fig. 15) are much less biased than the parameter estimates from the 'differential' model (Fig. 6).

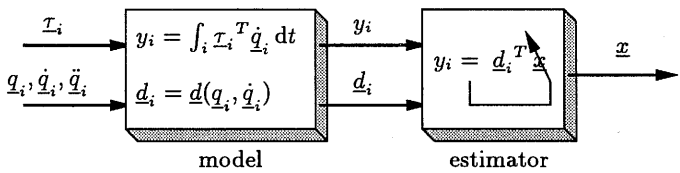

Fig. 14. Identification based on the integral model.

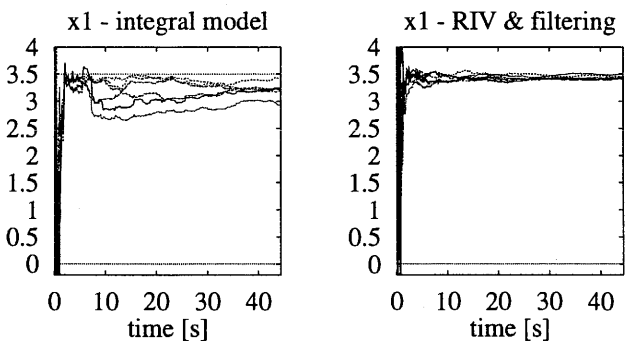

Fig. 15. Estimation based on the integral model (left), which does not depend on the acceleration signal, in comparison with a reference identification (right).

Yet, the integration by which the integral model can be derived from the differential model equations enhances low-frequency signals and attenuates high-frequency signals. By this integration the acceleration signal is eliminated. However, this shift of the frequency range of the model function signals to the lower frequencies reduces the frequency distance to the parameters. As already pointed out (Section 1.2), the parameters can be identified and distinguished from disturbances only because of their

constancy or slowly changing behaviour as compared with the supposed rapid changes of measurement or modeling errors. Figure 16 shows extremely slow parameter drift, which is caused by very low-frequency disturbances, to which the integral model is especially sensitive (Prüfer *et al.*, 1994).

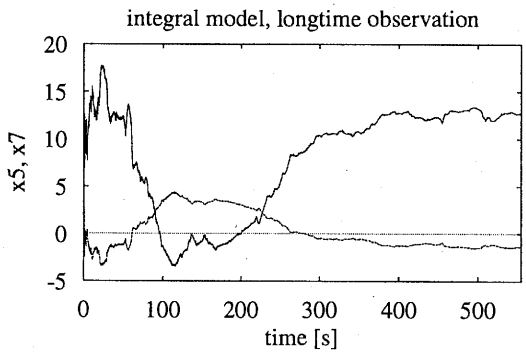

Fig. 16. Observation of two frictional parameters during 10 minutes of data acquistion. Identification with the integral model is affected by low-frequency disturbances. The algorithm has difficulties in distinguishing two frictional parameters.

3.2. Identifiability

So far, we have only discussed problems that are common to both on-line and off-line identification. It has already been mentioned that the advantage of off-line identification is that one can easily design series of experiments that allow for observation of individual dynamical parameters. The complete model can then be calculated from the observed parameters.

With on-line identification, in contrast, all parameters of the complete model have to be identifiable at once. A lot of work has been carried out to find trajectories that meet this requirement ('exciting trajectories' in (Armstrong, 1988; Gautier and Khalil, 1992; Kozłowski, 1996)).

This is a problem because we do not intend to use on-line identification during specially designed test movements only but also during normal operation (see Section 2.1). Thus, it cannot be assumed that identifiability conditions will always be optimal. It is, on the contrary, much more likely that good identifiability conditions for all parameters will be an exception during normal operation.

However, the lack of identifiability during robot operation is not the core of the problem: the reason why a parameter cannot be identified is that changes in the parameter do not affect the computed torque (which means that there is no criterion for the parameter's adjustment). This, of course, means that the parameter is not required to compute the correct torques and hence its identification can be abandoned at the moment. A problem arises when the parameter becomes identifiable again and thus regains influence on the computed torque. It will take some time to

readjust the parameter, and, depending on how far it has been drifting off, it will cause disturbances or even instability of the control system.

Our requirement therefore would be that poor identifiability conditions should be recognized automatically and parameters preserved, e.g. by turning the identification off.

An even harder requirement could be that not only unidentifiable parameters be preserved but that the available information be exploited as far as possible. This can be achieved e.g. by restricting the identification to an identifiable parameter subset rather than turning it completely off.

If it is possible to meet this requirement, then we are also able to take advantage of the off-line identification approach: we could simply execute a series of test movements designed for off-line identification. The on-line identification process should be able to detect automatically which parameters are identifiable and adjust them one after another, as we step from one test movement to the next. Having initialized estimates of all parameters in this way, they are used during robot operation and updated whenever a deviation from the actual values becomes evident. (In case an actual deviation is not evident, it will not harm the quality of control.)

In the following it is shown that it is possible to meet these requirements (and maybe even more demanding ones). First, we have to turn to the reason behind poor identifiability conditions.

3.2.1. Poor Identifiability Conditions

It has been mentioned that poor identifiability conditions arise when any parameter looses influence on the computed torque because in this case there is no criterion for the parameter's adjustment. The 'computed torque' is a linear combination of the model functions $d_1(\underline{q},\underline{\dot{q}},\underline{\ddot{q}})\,x_1 + d_2(\underline{q},\underline{\dot{q}},\underline{\ddot{q}})\,x_2 + \cdots$ that is compared with the observed torque τ in order to adjust the parameters (Fig. 17, compare with eqn. (1)).

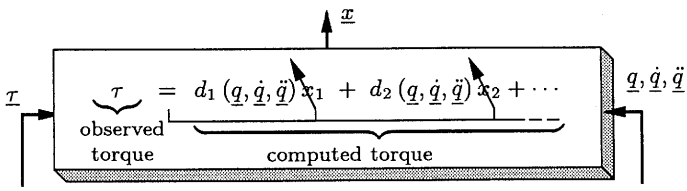

Fig. 17. Principle of parameter identification.

In general, poor identifiability conditions occur when there are linear combinations of parameters which do not contribute to the computed torque (see e.g. Figs. 18 and 19). Any multiple of such linear combination $\underline{x}_{\text{null}}$ can be added to or removed from a set of parameter estimates $\underline{x}$ without changing the computed torque:

$$d_1(\underline{q}_i,\underline{\dot{q}}_i,\underline{\ddot{q}}_i)\,x_{1\,\text{null}} + d_2(\underline{q}_i,\underline{\dot{q}}_i,\underline{\ddot{q}}_i)\,x_{2\,\text{null}} + \cdots = 0, \quad \text{for each } i \qquad (18)$$

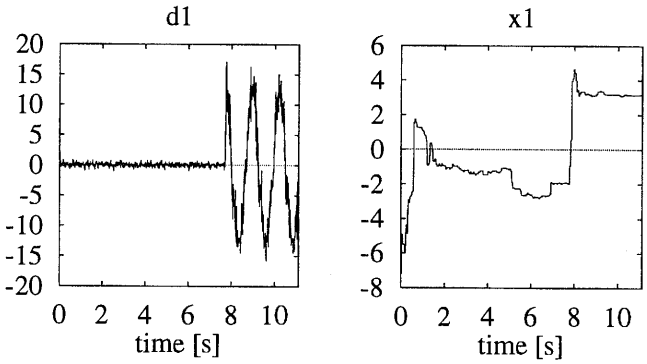

Fig. 18. Poor identifiability of one parameter x_1 because the corresponding model function d_1 is permanently zero during first eight seconds. When the model function starts changing values, the parameter estimate stops drifting and settles at its actual value.

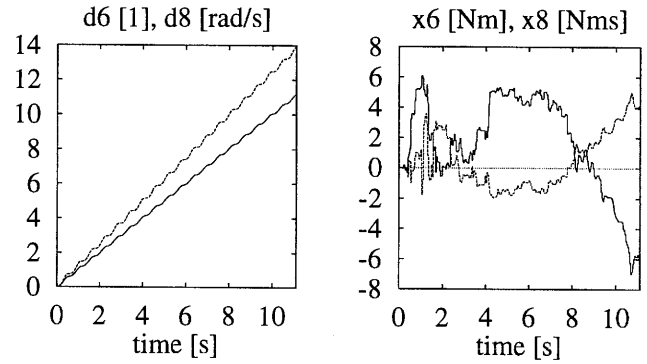

Fig. 19. Poor identifiability of a linear combination of two parameters x_6 and x_8: a change in one parameter estimate can be compensated for by a change in the other, because the model functions d_6 and d_8 are not linear independent of each other. (Such model functions can be observed with the integral model (17).)

Thus, in order to judge the identifiability conditions, we have to discover each non-trivial $\underline{x}_{\text{null}}$ that satisfies (18).

The same ambiguities turn up when the system of eqns. (4) is solved to find the parameter estimates $\underline{x}$: each $\underline{x}_{\text{null}}$ that satisfies

$$\left[\sum_i \underline{d}_i \underline{d}_i^T\right] \underline{x}_{\text{null}} = \underline{0} \tag{19}$$

can be added to or removed from $\underline{x}$ without rendering (4) invalid. Thus (4) yields no criterion for adjustment of $\underline{x}$ in direction of $\underline{x}_{\text{null}}$. Furthermore, each $\underline{x}_{\text{null}}$ that satisfies (19) also satisfies (18) and vice versa.[3] Hence, it is not necessary to analyse the system of equations of (18), which may have thousands of equations and is growing with each new sample, but it is sufficient to analyse (19), which has only as many equations as there are parameters.

3.2.2. Recognition of Poor Identifiability

In order to determine the null-space defined in (19), we can evaluate eigenvectors and eigenvalues of the information matrix (as suggested in (Posten and Munack, 1990)), which is a standard task in numerical analysis. If any eigenvalue λ is zero, then the corresponding eigenvector $\underline{v}$ lies within the null-space of $\sum_i \underline{d}_i \underline{d}_i^T$. (Compare the definition $[\sum_i \underline{d}_i \underline{d}_i^T]\,\underline{v} = \lambda\,\underline{v}$ with (19).) Since the information matrix is symmetric, we know that all eigenvalues and eigenvectors exist. Thus the eigenvectors form a complete orthogonal basis of the parameter space, and all $\underline{x}_{\text{null}}$ which satisfy (19) are discovered.

However, since the eigenvectors form an orthogonal basis, we can use them as coordinate vectors. After transformation into this coordinate system, the infomation matrix becomes a diagonal matrix which has the eigenvalues as diagonal elements. If we transform eqn. (4) into this coordinate system, we get a set of decoupled equations:

$$\lambda_1 x_1^* = y_1^*, \qquad \lambda_2 x_2^* = y_2^*, \qquad \lambda_3 x_3^* = y_3^*, \qquad \ldots \tag{20}$$

where λ is one of the eigenvalues of the information matrix $\sum_i \underline{d}_i \underline{d}_i^T$ of (4); x^* is one element of the transformed parameter vector $\underline{V}^T \underline{x}$, $\underline{V}^T$ being the transformation matrix (see below); y^* stands for one element of the transformed right-hand side of 4), $\underline{V}^T [\sum_i \underline{d}_i \tau_i]$; $\underline{V}$ denotes the matrix composed of the normalized eigenvectors as column vectors. Multiplication by $\underline{V}^T$ transformes into the coordinate system which renders the information matrix diagonal, and multiplication by $\underline{V}$ transforms back into the original parameter space. This representation of (4) clearly suggests how to proceed when parameter estimates have to be calculated in the presence of poor identifiability conditions:

We calculate each x^* from y^*/λ if the corresponding eigenvalue is well above the noise level, otherwise we leave x^* at a value from some previous successfull identification. The resulting vector $\underline{x}^*$ is then transformed back into the original parameter space.

In order to decide whether an eigenvalue λ is 'well above the noise level', we have to compare it with some threshold. This threshold can be determined by exploiting

[3] Rewrite (18) as matrix equation $\underline{D}\,\underline{x}_{\text{null}} = \underline{0}$, where row i of matrix $\underline{D}$ is made up of $\underline{d}_i^T$. Likewise, (19) can be rewritten as $\underline{D}^T \underline{D}\,\underline{x}_{\text{null}} = \underline{0}$. From (19) we have $\underline{D}^T \underline{D}\,\underline{x}_{\text{null}} = \underline{0} \Rightarrow \underline{x}_{\text{null}}^T \underline{D}^T \underline{D}\,\underline{x}_{\text{null}} \equiv [\underline{D}\,\underline{x}_{\text{null}}]^T [\underline{D}\,\underline{x}_{\text{null}}] = 0 \Rightarrow \underline{D}\,\underline{x}_{\text{null}} = \underline{0}$, which is another notation for (18). On the other hand, starting from (18) we have $\underline{D}\,\underline{x}_{\text{null}} = \underline{0} \Rightarrow \underline{D}^T [\underline{D}\,\underline{x}_{\text{null}}] = \underline{0}$, which is another notation for (19). Thus (18) and (19) are equivalent.

the fact that the information matrix provides an estimate of the covariance matrix of the estimation error. More precisely,

$$\mathrm{cov}\{\underline{x}\} \equiv E\left\{\left[\underline{x} - E\{\underline{x}\}\right]\left[\underline{x} - E\{\underline{x}\}\right]^T\right\} = \sigma_e^2\, E\left\{\left[\textstyle\sum_i \underline{d}_i \underline{d}_i^T\right]^{-1}\right\} \quad (21)$$

where σ_e^2 denotes the variance of the equation error, which can be estimated by

$$\sigma_e^2 = \frac{1}{N}\sum_{i=1}^{N} e_i^2 \quad (22)$$

N being the number of samples.[4] Inverting $\sum_i \underline{d}_i \underline{d}_i^T$ after transformation is especially easy: we get σ_e^2/λ as an estimate of the variance $\sigma_{x^*}^2$ of one component of the transformed estimation error $[\underline{x}^* - E\{\underline{x}^*\}]$, and all covariances are zero. We can specify a threshold $\sigma_{x^*,\max}$ for the standard deviation of each element of the transformed parameter vector, say 10% of the nominal values from some initial identification (e.g. $\underline{\sigma}_{x^*,\max} = 0.1\underline{x}^*_{\mathrm{init}}$). If a λ is greater than its threshold

$$\lambda > \frac{\sigma_e^2}{\sigma_{x^*,\max}^2} \quad (23)$$

then the corresponding x^* will be calculated from y^*/λ, otherwise x^* will be left at some previous value.

For the purpose of any reasonable on-line application, we have to provide for the capability to follow parameter changes. This means that we do not start the identification together with robot operation and let it accumulate ever more and more information. We would rather specify a kind of observation window. Afterwards this puts a more precise meaning to 'the value x^* from some previous successful identification'. With on-line identification we observe those values of λ which result from the data within the window. They will be rising and dropping according to the data entering and leaving the window. The last successfully identified value of x^* is the value it had at the last time, when the corresponding λ was above the threshold.

[4] The proof is taken from (Isermann, 1992). If we subtract $\left[\sum_i \underline{d}_i \underline{d}_i^T\right] E\{\underline{x}\}$ from both sides of (4), we get an equation for the current estimation error $\left[\sum_i \underline{d}_i \underline{d}_i^T\right][\underline{x} - E\{\underline{x}\}] = \sum_i \underline{d}_i[\tau_i - \underline{d}_i^T E\{\underline{x}\}]$. Rewrite this as the matrix equation $\underline{D}^T\underline{D}\,[\underline{x} - E\{\underline{x}\}] = \underline{D}^T\underline{e}$, where row i of matrix $\underline{D}$ is made up of $\underline{d}_i^T$, and element i of vector $\underline{e}$ contains $[\tau_i - \underline{d}_i^T E\{\underline{x}\}]$, which is that part of the equation error which results from measurement and modelling noise but not from parameter mismatch. With this equation for $[\underline{x} - E\{\underline{x}\}]$, the covariance matrix can be written in form $\mathrm{cov}\{\underline{x}\} = E\{[\underline{D}^T\underline{D}]^{-1}\underline{D}^T\underline{e}\,\underline{e}^T\,\underline{D}\,[\underline{D}^T\underline{D}]^{-1}\}$. If we assume that $\underline{e}$ and $\underline{D}$ are uncorrelated and that $\underline{e}$ contains only a time-invariant white-noise signal such that $\underline{e}\,\underline{e}^T$ is a diagonal matrix with σ_e^2 as diagonal elements, then we can simplify this to $\mathrm{cov}\{\underline{x}\} = \sigma_e^2 E\{[\underline{D}^T\underline{D}]^{-1}\}$, which is another notation for (21). This expression is not strictly true since $\underline{e}$ and $\underline{D}$ contain the same noise signal $\underline{n}_d$ and hence are not entirely uncorrelated (see eqn. (5)). The smaller the noise signal $\underline{n}_d$ in comparison with the information signal $\underline{d}_0$, the better the estimate for $\mathrm{cov}\{\underline{x}\}$. The estimates are also biased because we do not have $E\{\underline{x}\}$ for calculating an estimate of σ_e^2 but must use the current estimate of $\underline{x}$ instead.

3.2.3. Realization of Identifiability Supervision

Unfortunately, the computational cost of calculating eigenvalues and eigenvectors is much higher than that of inverting a matrix. With n being the number of parameters, we have about $O(n^3)/3$ floating-point operations for the solution of (4) by Cholesky factorization (Golub and Van Loan, 1989). Furthermore, we are using recursive algorithms which require only $3\,O(n^2)$ floating-point operations for updating the inverse of $\sum_i \underline{d}_i \underline{d}_i^T$ with each new sample and for calculating a new estimate of $\underline{x}$ (UD-factorization in (Bierman, 1977)). A divide-and-conquer algorithm for calculating eigenvalues and eigenvectors, in contrast, requires $9\,O(n^3)$ floating-point operations (Golub and Van Loan, 1989).

If such reduction of data throughput is not acceptable,[5] we can opt for a less efficient scheme, which excludes parameters from identification more often than actually would be necessary, thus sacrificing some of the information present in the data.

The means available to judge identifiability conditions without recurring to computation of eigenvalues and eigenvectors are as follows:

- The sum of diagonal elements of the information matrix $\sum_i \underline{d}_i \underline{d}_i^T$ (the trace of the matrix) is invariant with respect to orthogonal transformations (coordinate transformations) and hence it is equal to the sum of eigenvalues.
- The determinant of the information matrix is equal to the product of all eigenvalues. Most factorization methods for recursive estimation update triangular factors of the information matrix, which allows for calculation of the determinant by multiplying the diagonal elements of those triangular matrices.
- We can calculate an estimate of the condition number, which is the ratio of the largest to the smallest eigenvalue. The most common way to do this is to calculate the infinity norm condition number instead of the 2-norm condition. Any condition number requires to evaluate a matrix norm of the matrix and of its inverse. If we use an identification algorithm in covariance form, like the Kalman algorithm, which updates the inverse of the information matrix (or factors of the inverse of the information matrix), then we can avoid explicitly inverting the information matrix, which would be computationally too expensive.
- We can test for certain special cases of poor identifiability conditions, in particular the case of a single model function being permanently zero (or close to zero) and the case of linear dependencies among pairs of model functions.

A scheme that we have adopted for experiments comprises the following steps:

1. A test for special cases of poor identifiablility is performed by comparing the norm of each column of the information matrix with some threshold and the angle between pairs of column vectors with another threshold. The information from this test is then used to exclude single parameters from identification. (This feature

[5] The greater the density of samples that are entering the identification process, the more information can be accumulated in short periods, which permits the reduction of the temporal size of the observation window accordingly. The lower limit of the size of the observation window depends on the speed of the movement. But even if this limit is reached, higher sample rates still can reduce the effects of measurement noise.

can easily be added to existing recursive estimation algorithms by manipulating the information matrix.) If the norm of any column is too small, then the corresponding parameter is excluded. If the angle between any two columns of the information matrix is too small, then the parameter corresponding to the column with the largest off-diagonal elements is excluded.

2. The resulting reduced information matrix is then tested for its trace and its condition number. If either of them lies beyond some bounds, identification is suspended completely because this means that there are linear dependencies present in the matrix which were not discovered in the first step.

   The trace of the information matrix, which is proportional to the average value of all eigenvalues, is used to detect the case when no parameter at all can be identified. The condition number, on the other hand, which can only show the ratio between the largest and smallest eigenvalue, is used to make sure that all remaining parameters are of comparable identifiability. Testing the trace of the information matrix can be omitted since the case of no identifiable parameter would be discovered in the first step of the algorithm.

Sripada and Fisher (1987) propose a modified version of suspending identification. They change the size of the observation window in order to keep the trace of the inverse of the information matrix constant. (The inverse of the information matrix, the covariance matrix according to (21), is available with all recursive algorithms in covariance form.) Only when the condition number passes some threshold, identification is completely turned off.

The loss of information implied by this procedure should now be compared with the loss of information that results from reducing the rate of data throughput on account of calculating eigenvalues and eigenvectors. If the temporal size of the observation window is taken to be constant, a decrease in the sample rate will result in a general deterioration of all estimates throughout the experiment. Equation (21) suggests that the variance of the parameter estimates will be reciprocally proportional to the number of samples $\underline{d}_i$ that fall into the window. This means that calculating eigenvalues and eigenvectors will increase the variance of parameter estimates at least by factor $3\,O(n)$ (n denotes the number of parameters).

The implemented procedure will give rise to any increased variance of parameter estimates only in some special cases. This might occur when two column vectors of the information matrix are almost parallel such that one of the two corresponding parameters has to be excluded from identification. It would be optimal to adjust the vector of parameter estimates in the direction of these two column vectors while keeping the position in an orthogonal direction. Yet, only coordinate directions can be choosen for keeping or adjusting the parameter vector, and there will be an angle of maximum $\alpha_{\max} = \arccos(1/\sqrt{n})$ between the direction of actual and optimal adjustment. In the worst case, the variance of parameter estimates is thus increased by factor $1/\cos^2(\alpha_{\max}) = n$.

The main loss of information, however, takes place when there remain unrecognized singularities such that the identification is completely suspended. Thus the

efficiency of this algorithm is more unevenly distributed over time than the efficiency of an algorithm that calculates eigenvalues and eigenvectors.

The question of the temporal size of the observation window can be settled independently from the question of the efficiency: if the variance of parameter estimates is generally too large, then the size of the observation window can be increased. This reduces the variances of parameter estimates at the expense of a slower adjustment of estimates.[6] Note that this does not change the efficiency, which, in the case of on-line identification, has to be a time-related quantity. The suggestion of Sripada and Fisher to keep the trace of the inverse of the information matrix constant means that the average parameter variances are kept constant.

3.2.4. Significance of the Shown Realization

The implemented procedure shows that the goal set at the beginning of this section can be met. Even more efficient ways of exploiting observed data can be expected. For the last three decades the knowledge in numerical analysis has made a significant step. The proof of singular-value decomposition shows that the requirement of a rectangular symmetric matrix is not necessary to apply considerations similar to those of Section 3.2.2. Furthermore, there are building blocks available (Housholder Reflections and Jacobi Rotations) which allow us to design fast updating algorithms for a wide range of purposes (Golub and Van Loan, 1989). It should be possible to construct algorithms which work at data rates comparable with existing recursive algorithms and which are nevertheless able to detect all singularities (singular values), not only those special cases of the procedure shown. Furthermore, it should be possible to design algorithms which allow us to exclude any direction from identification (not only coordinate directions) at only slightly higher computational expense. This is a subject of future work.

There still remains one advantage of the off-line identification approach: in general, the model is changed arbitrarily, when proceeding from one experiment to another, while with the methods treated in this paper, only the dimension of a linear model is changed. It might be considered to increase the efficiency of the identification by employing different sets of model equations and detecting circumstances for which a particular set of equations is more appropriate than others.

4. Summary

The main ideas of this paper can be summarized as follows:

On-line identification provides a means to make model-based control strategies feasible in practice because it automatically performs the necessary supervision and adjustment of the model.

[6] The kind of observation window is actually exponential, owing to the simplicity of introducing an exponential weighting into any recursive estimation algorithm. The manipulation of the size of the window proposed by Sripada and Fisher is in fact a manipulation of the exponent, or rather of the factor by which all previously observed data are reduced relative to the most recent observation.

Yet, the available recursive estimation algorithms are not well-suited to handle ill-conditioned estimation tasks, while they will be certain during common robot operation. The paper shows how to discover and treat poor identifiablility conditions and supplies, for comparison purposes, the notion of theoretically optimal behaviour of an estimation algorithm.

The paper also demonstrates that despite the non-linearity of robot dynamics it is possible to reduce the bias of the least-squares method.

References

Armstrong B. (1988): *Dynamics for robot control: Friction modeling and ensuring excitation during parameter identification.* — Ph.D. Thesis, Dept. of Electrical Engineering, Stanford University.

Bierman G.J. (1977): *Factorization Methods for Discrete Sequential Estimation.* — New York: Academic Press.

Gautier M. and Khalil W. (1989): *Identification of the minimum inertial parameters of robots.* — Proc. Int. Conf. *Robotics and Automation*, Scottsdale, Arizona, pp.1529–1535.

Gautier M. and Khalil W. (1992): *Exciting trajectories for the identification of base inertial parameters of robots.* — Int. J. Robotics Research, Vol.11, No.4, pp.362–375.

Golub G. and Van Loan C. (1989): *Matrix Computations.* — Baltimore, London: John Hopkins University Press.

Isermann R. (1992): *Identifikation dynamischer Systeme, 2-nd Ed.* — Berlin, Heidelberg, New York: Springer-Verlag, Vol.I–II.

Kozłowski K. (1996): *Optimal trajectory design for identification purposes in robotics.* — Proc. 3-rd Int. Symp. *Methods and Models in Automation and Robotics*, Międzyzdroje, Poland, Vol.3, pp.871–878.

Kozłowski K. and Prüfer M. (1992): *Parameter identification of an experimental 2-link direct drive arm.* — Proc. Int. Conf. *Control and Robotics*, Vancouver, Canada, pp.313–316.

Posten C. and Munack A. (1990): *On-line application of parameter estimation accuracy to biotechnical processes.* — Prep. *American Contr. Conf.*, Vol.2, pp.2181–2186.

Prüfer M. and Schmidt C. (1996): *Practical aspects of robot dynamics identification.* — Proc. 3-rd Int. Symp. *Methods and Models in Automation and Robotics*, Międzyzdroje, Poland, Vol.3, pp.971–976.

Prüfer M., Schmidt C. and Wahl F. (1994): *Identification of robot dynamics with differential and integeal models: A comparison.* — Proc. IEEE Int. Conf. *Robotics and Automation*, San Diego, California, pp.340–345.

Prüfer M. and Wahl F. (1994): *Friction analysis and modelling for geared robots.* — Prep., IFAC Symp. *Robot Control*, CApri, Italy, pp.551–556.

Seeger G. and Leonhard W. (1989): *Estimation of rigid body models for a six-axis manipulator with geared electric drives.* — Proc. IEEE Int. Conf. *Robotics and Automation*, Scottsdale, Arizona, pp.1690–1695.

Söderström T. and Stoica P. (1983): *Instrumental Variable Methods for System Identification.* — LNCIS, No.57, Berlin: Springer.

Sripada N.R. and Fisher D.G. (1987): *Improved least squares identification.* — Int. J. Control, Vol.46, No.6, pp.1889–1913.