IEEE Transactions on Automatic Control, Vol. 46, No. 10, 2001, pp. 1584-1589.

Performance-optimized Applied Identification of Separable Distributed-Parameter Processes

Dimitry Gorinevsky* and Michael Heaven†

Abstract

This paper studies practical algorithms for parametric identification of cross-directional processes from input/output data. Instead of working directly with the original two-dimensional array of the high-resolution profile scans, the proposed algorithms use separation properties of the problem. It is demonstrated that by estimating and identifying in turn cross-directional and time responses of the process, it is possible to obtain unbiased least-square error estimates of the model parameters. At each step, a single data sequence is used for identification which ensures high computational performance of the proposed algorithm. A theoretical proof of algorithm convergence is presented. The discussed algorithms are implemented in an industrial identification tool and the paper includes a real-life example using paper machine data.

*Honeywell Systems and Control Laboratory, 1 Resuts Way, Cupertino, CA 95014; dimitry.gorinevsky@honeywell.com; gorinevsky@ieee.org

†was with Honeywell ASCa Inc., Measurex Devron Operation, North Vancouver, BC, Canada; currently with Norsat International Inc., Burnaby, BC, Canada

1 Introduction

The paper considers an application-oriented model of a distributed-parameter process where distributed measurements and control inputs are discretized both in space and time. The process model assumes separation between the spatial response shape and time-response of the process. That is the process transfer function is a product of a scalar transfer function describing process dynamical response and a constant matrix describing spatial response of the process. Both the transfer function and spatial response matrix need to be identified from the collected data. Transfer functions that are a product of two conceptually different multipliers are encountered in classical problems of simultaneous identification of process and disturbance dynamics. Such problems can be solved by using the Generalized Least Squares (GLS) algorithm, e.g., see [12]. The GLS algorithm iteratively identifies one of the two unknown transfer functions in turn, while assuming that another one is known. The same general idea is used in this paper.

While the considered model and the algorithm are general for a class of problems, the paper is geared towards an application to the industrial process of manufacturing a web of paper or plastic. In cross-directional (CD) control of a paper or plastic film machine, an array of actuators controlling web properties are placed across the moving web. Such web properties as weight and thickness are usually required to be uniform across the web. In paper manufacturing, parameters such as moisture, smoothness, and gloss can also be controlled. Downstream from the CD actuators, the profile of the controlled web property is determined by a high-resolution measurement system and fed back to the CD control system on a periodic basis.

In modern industrial CD control systems, control of a single profile, such as a paper weight profile, might include more than 300 actuators and 1000 databoxes or samples across the profile. A few hundred profiles can be collected in a single identification experiment. Processing and identification of such two-dimensional arrays of data requires highly optimized computational algorithms.

Several prior papers on identification of CD processes have been published. Many of them, e.g., [8, 3, 9, 14] consider application of modern time-series analysis methods to CD process identification. Such methods, however, are not very practical because of an enormous volume of computations associated with processing time series for each of the profile databoxes and a significant amount of process noise typically present in the data. Furthermore, a direct application of time-series analysis might result in highly complicated models, which are difficult to use in control design and tuning.

In this paper, we describe a practical approach to distributed process identification based on identification of a parametric model. The proposed approach describes the identification results in an intuitive way (through the process spatial and time responses), provides for optimal least square identification, and is very efficient computationally. This is achieved by exploiting such a feature of the problem as a separation between spatial and time-response models. The approach is based on a separable model of the distributed process, which is introduced in Section 2.

The contributions of this paper are as follows. The paper demonstrates that the general problem of the distributed-parameter process identification from two-dimensional data can be separated into two one-dimensional least square problems of the spatial response and time-response identification. It proposes an iterative algorithm for solving the overall identification problem with a two-dimensional data set as a series of one-dimensional problems. The algorithm is demonstrated to be convergent and give unbiased estimate of the process parameters.

The paper presents an example of the algorithm application to paper machine model identification. The design and practical applications of an industrial identification tool based on the discussed algorithms are described in more detail in [6].

2 Identification Problem

Measurements of a distributed parameter process, such as CD process, are characterized by a profile (vector) of the process data sampled at regular spatial intervals. The profiles are obtained from a measurement device and are available once every control sample. The profile at discrete time t is described by a vector $p \in \Re^m$. The manipulated variables of the process are setpoints of the actuators influencing the process. These actuators are usually placed at regular spatial intervals across the web. It is assumed that the actuators move once per control sample interval. The control input to the process is defined by the actuator profile $u(t) \in \Re^n$.

Consider the following control-oriented model of the distributed process relating the manipulated variables and the obtained profiles:

$$p(t) = \gamma G g(z^{-1}) u(t), \qquad (1)$$

where $G \in \Re^{m,n}$ is a spatial interaction matrix, and $g(z^{-1})$ is a scalar function of the unit delay operator z^{-1} defining the process pulse response. The scalar parameter γ defines the overall gain of the process and the role of this parameter will be explained further on. The model (1) assumes that the process time-response does not change in space.

Models of the form (1) are commonly used in the industry. They have been studied in many papers on cross-directional process control and identification, e.g., see [2, 8, 3, 7, 14]. The identification approach we are going to pursue for the model (1) is based on the fact that the time response model $g(z^{-1})$ and the spatial interaction model G enter the overall transfer function (1) as separate multipliers.

Let us assume that the process time response can be described by a parametric model of the form

$$g(z^{-1}) = g(z^{-1}; \theta_T), \tag{2}$$

where θ_T is a vector of the model parameters.

To describe the spatial interaction model, we assume that the process response to each actuator has the same shape, so that the entries of G have the form

$$G_k^j = b(d_s k - c_j, \theta_S), \tag{3}$$

where b is a continuous function of the spatial coordinate, d_s is a spatial distance between the measurement samples (databox width), c_j is a coordinate of the response center for the actuator j, and θ_S is a parameter vector describing the model of the spatial response shape. We assume that the spatial coordinate is counted from the beginning of the first databox.

In addition to the response shape, the spatial model is described by a mapping (alignment) model. The alignment model describes how the coordinate c_j of the spatial response center depends on the spatial position of the respective actuator. This model is

$$c = f(\theta_A); \qquad c = [c_1 \ \ldots c_n], \tag{4}$$

where the vector c collects the response center coordinates and θ_A is a vector of the alignment parameters. In addition to the overall offset and shrinkage of the web between the actuators and the measurement site, θ_A may include parameters describing the paper web shrinkage due to the drying process or the expansion of a plastic film in the extrusion process. For industrial CD control systems, the CD mapping model (4) is the most important part of the CD process model for such systems.

In accordance with (3) and (4), the overall matrix G is defined by the parameter vector

$$\theta_C = [\theta_S^T \ \ \theta_A^T]^T \tag{5}$$

The above described model can be identified from the data collected in a specially arranged identification experiment. The industry practice is to perform such an experiment in the form of a "bump test". In a bump test, the control loop is open and several actuators are stepped (bumped) one or more times.

We will further assume that in the identification experiment, the actuator setpoints change such that

$$u(t) = Ua(t), \qquad U \in \Re^n, \;\; a(t) \in \Re \tag{6}$$

where U is the bump profile and $a(t)$ is the amplitude of the bump. This is in accordance with industrial practice. Typically, $a(t)$ would take one of the values $\{-1, 0, 1\}$. The measured profile data collected in the identification experiment are put together to form a data matrix

$$P = [p(1) \;\ldots\; p(N)] \tag{7}$$

Based on the above discussion, the identification problem can be formulated as follows.

Let the process input (6), where $t = 1, \ldots, N$, and the process output data (7) be given. Assuming the process model (1)–(4), find the vector of the model parameters

$$\theta = [\gamma \;\; \theta_T^T \;\; \theta_C^T]^T, \tag{8}$$

where θ_C is given by (5).

The identification algorithm should generate unbiased estimates of the parameter vector. This means that, in case where the data (6), (7) is indeed obtained for the process described by (1)–(4) for the parameter vector

$$\theta^* = [\gamma^* \;\; \theta_T^{*\,T} \;\; \theta_C^{*\,T}]^T, \tag{9}$$

then the obtained estimate of θ should coincide with θ^*.

It is practically important that the identification algorithm is simple and has high computational performance. For the data set sizes typical in the distributed-parameter control problems, algorithms might take an unreasonable execution time, unless they are carefully designed. An algorithm for distributed-parameter process identification satisfying the above requirements is proposed in the next section.

3 Iterative identification algorithm

We assume that initially, at $t = 0$, the process is at a steady state. By combining (1) and (6), we can present the modeled process output $\hat{p}$ in the identification experiment in the form.

$$\begin{aligned} \hat{p}(t) &= \gamma p_u(\theta_C) h(t;\theta_T) && (10) \\ p_u(\theta_C) &= G(\theta_C) U \in \Re^m, && (11) \\ h(t;\theta_T) &= g(z^{-1}, \theta_T) a(t), \quad (t = 1, \ldots, N), && (12) \end{aligned}$$

where the profile p_u defines a spatial response to the actuator excitation profile U in (6), $h(t)$ defines a time-response of the process to the excitation sequence $a(t)$, z^{-1} should be considered as a unit delay operator, and N is the number of the collected profiles.

The identification problem to be solved can be formulated as a least square fit of the model (10) against the data (7). The parameter vector θ (8) can be found by minimizing the following loss index

$$\{\gamma, \theta_T, \theta_C\} = \arg\min \|P - \gamma p_u(\theta_C) h(\theta_T)\|_F^2, \tag{13}$$

where $h(\theta_T) = [h(1,\theta_T) \ \ldots \ h(T,\theta_T)]^T$, $\|\cdot\|_F^2$ denotes the squared Frobenius norm of the matrix (sum of all squared matrix entries).

By using standard equalities relating the Frobenius norm and the matrix trace [1], the loss index (13) can be presented in the form

$$\|P - \gamma p_u h^T\|_F^2 = \mathrm{Tr}\left((P - \gamma p_u h^T)^T (P - \gamma p_u h^T)\right) = \|P\|_F^2 + \gamma^2 \|h\|^2 \|p_u\|^2 - 2\gamma p_u^T P h \tag{14}$$

where $\|\cdot\|$ denotes the Euclidean norm of a vector.

Without a loss of generality, it is assumed that the parameters θ_C and θ_T are chosen such that always

$$\|h(\theta_T)\|^2 \equiv \|p_u(\theta_C)\|^2 \equiv 1 \tag{15}$$

This is possible for the model (1), because the scaling of the overall gain is done by the parameter γ.

For any given parameter vectors θ_C and θ_T, we can find γ by minimizing (14). Taking (15) into account, the estimate of γ minimizing (14) is given by

$$\gamma = p_u^T P h = h^T P^T p_u \tag{16}$$

Assume that the parameter vector θ_C is known (G is known). By computing γ in (14) according to (16), the conditional estimate of the time-response model parameters θ_T in (13) can be found from

$$\theta_T = \arg\min(\|h(\theta_T) - \hat{h}\|^2 + \dots) \tag{17}$$

$$\hat{h} = P^T p_u / \gamma \tag{18}$$

where $\dots$ denotes terms independent of θ_T and $\hat{h} \in \Re^N$ is an estimate of the process time response $h(t)$. The vector $\hat{h}$ (18) is obtained by cross-correlating the collected spatial profiles $p(t)$ with the predicted spatial response shape p_u. The problem (17) is a standard single-input, single-output identification problem that can be readily solved by one of the standard parametric identification methods. The reader is referred to [10] for discussion of the standard parametric identification approaches and to [6] for an application example similar to problem (17).

Let us now assume that the time-response parameter vector θ_T is given ($h = h(\theta_T)$ is known). By using (16), the optimal conditional estimate of the spatial model parameter vector θ_C in (13) can be found by solving the following problem

$$\theta_C = \arg\min(\|p_u(\theta_C) - \hat{p}_u\|^2 + \dots) \tag{19}$$

$$\hat{p}_u = Ph/\gamma \tag{20}$$

where $\dots$ denotes terms independent of θ_C, γ is given by (16), and $\hat{p}_u$ is an estimate of the process spatial response p_u. The spatial response estimate $\hat{p}_u$ is obtained by correlating the time history for each databox with the single predicted time-response $h(t)$. The problem (19) deals with identification of the spatial shape parameters θ_C for a single spatial response profile and is much simpler than (13). Such problem is discussed in [4, 5].

We propose the following iterative algorithm for solving the identification problem (13).

Algorithm 1 *Assume that initial estimates of the parameters θ_T and θ_C are given. At a step k of the algorithm use the estimates $\theta_T^{(k)}$ and $\theta_C^{(k)}$ to compute the predicted responses h and p_u according to (11), (12). Use $h(\theta_T^{(k)})$ and expression (16) for optimal γ to compute $\hat{p}_u$ in (20). Using the obtained vector $\hat{p}_u$, solve the identification problem (19) to obtain the estimate $\theta_C^{(k+1)}$. In the same way, use $p_u(\theta_C^{(k)})$ and (16) to solve the identification problem (17), (18) and obtain the estimate $\theta_T^{(k+1)}$. Repeat the iterations as described above till the convergence is achieved.*

In practice, initial estimates for identification parameters can be obtained from physical system consideration or by using a coarse global search procedure. The experience in applying the proposed algorithm to real-life paper machine data shows that it converges very fast: after 2-3 iterations. One of the reasons for the fast convergence is that the identification of θ_T assuming inaccurately known θ_C and vice versa can give fairly accurate result. A formal discussion and references on such "nuisance parameter" problem can be found in [13].

Algorithm 1 is very efficient computationally, as it reduces the identification problem (13), which is essentially a two-dimensional problem, to a series of one-dimensional identification problems. The intermediate and final identification results for the proposed sequential iterative algorithm are intuitively clear, since they can be presented to a user as a fit of the model spatial response $p_u(\theta_C)$ against the estimated process spatial response $\hat{p}_u$ and the model time-response $h(\theta_T)$ against the estimated process time-response $\hat{h}_u$.

Theoretical justification of the algorithm convergence is the subject of the following section.

4 Iterative algorithm convergence

Let us describe the iterative update of Algorithm 1 in a more formal mathematical way. As a part of this algorithm, θ_C is determined by solving (17), where $h = h(\theta_T)$, $p_u = p_u(\theta_C)$, and γ is defined by (16). Given θ_T, the optimal estimate vector θ_C can be determined by solving the equation

$$f_C(\theta_T, \theta_C) \equiv \frac{\partial \|p_u - \hat{p}_u\|^2}{\partial \theta_C} = 0 \tag{21}$$

Similarly, the optimal estimate for θ_T is determined by solving (19), where $h = h(\theta_T)$, $p_u = p_u(\theta_C)$, and γ is defined by (16). The loss function minimum condition in this case has the form

$$f_T(\theta_T, \theta_C) \equiv \frac{\partial \|h - \hat{h}\|^2}{\partial \theta_T} = 0 \tag{22}$$

By using (21) and (22), an iteration of Algorithm 1 can be presented in the form

$$\begin{array}{rcl} \theta_C^{(k+1)} & = & \phi_C(\theta_T^{(k)}) \\ \theta_T^{(k+1)} & = & \phi_T(\theta_C^{(k)}) \end{array}, \tag{23}$$

where $\phi_C(\theta_T)$ is the implicit function obtained by solving (21) with respect to θ_T and $\phi_T(\theta_C)$ is the implicit function obtained by solving (22) with respect to θ_C.

The convergence properties of Algorithm 1 are given by the following Theorem proved in the Appendix.

Theorem 1 *Assume that the identification data (7) is generated by the process of the form (1)–(5) with the input (6)*

$$P = \gamma_* p_u(\theta_C^*) h^T(\theta_T^*) + \Xi, \tag{24}$$

where the matrix $\Xi \in \Re^{m,T}$ *contains the measurement noise and* $\theta_* = [\gamma_* \ \theta_C^{*T} \ \theta_T^{*T}]^T$ *is a vector of the true process parameters in (24). Assume further that the Jacobian matrices*

$$G_C = \frac{\partial p_u(\theta_C)}{\partial \theta_C}, \tag{25}$$

$$G_T = \frac{\partial h(\theta_T)}{\partial \theta_T} \tag{26}$$

are of the full column rank for the parameter vector θ *(8) being in an open neighborhood of the parameter vector* θ_**. This assumption means that the parameters* θ_C *and* θ_T *are nominally identifiable from the process data produced with the excitation input used. Assume that functions* $p_u(\theta_C)$ *and* $h(\theta_T)$ *in (14), (21), (22) are twice continuously differentiable.*

By characterizing the noise Ξ *through the following parameters*

$$\xi_1 = \|\Xi^T p_u(\theta_C^*)\|, \quad \xi_2 = \|\Xi h(\theta_T^*)\|, \quad \xi_3 = \|\Xi^T G_C(\theta_C^*)\|, \quad \xi_4 = \|\Xi G_T(\theta_T^*)\|, \tag{27}$$

the following statements can be established

1. *A positive constant* ξ_0 *can be found such that for* $\max(\xi_1, \xi_2, \xi_3, \xi_4) < \xi_0$ *the Algorithm 1 iterations converge in an open neighborhood of the point* $\theta = \theta^*$ *to a single solution - an estimate of the parameter vector* θ^**.*

2. *This estimate is not biased, i.e., for* $\xi_0 \to 0$*, the estimate converges to* θ^**.*

Note that the noise parameters (27) will be vanishing small for a broad class of the bounded random noise models provided the dimensions of the identification data array are sufficiently large.

5 Application to industrial paper machine process application

The iterative identification algorithm discussed in the two preceding sections is based on solving the problems of spatial and time response identification. This section provides illustrative examples of such one-dimensional identification models. The algorithms described above and models described below have been implemented in an industrial control application product and deployed on paper machines in many paper mills around the world.

As a model (2) of the process dynamics, a first-order process model with a dead time is used. Such models are overwhelmingly accepted in the industry for paper machine processes. The parameters of the model include T_{rise} - the process rise time and T_{del} - the process dead time. For such model, the pulse response in (12) has the form $\left(1 - e^{-(T_s(t-k) - T_{\text{del}})/T_{\text{rise}}}\right)$, where T_s is the sample time. The parameter vector $\theta_T = [T_{\text{rise}} \;\; T_{\text{del}}]^T$ can be identified by solving the problem of the form (17).

Consider now the model of the spatial response of the form (3). As discussed in [4, 5], the experience shows that for a vast majority of paper machine processes the spatial response shapes can be well described by a simple parametric expression of the form $b(x) = \gamma_S[r(x + \delta w) + r(x - \delta w)], \;\; r(x) = \exp\frac{-ax^2}{w^2} \cos\frac{\pi x}{w}$,, where γ_S, w, a, and δ are scalar parameters. For this spatial response shape model, the parameters in (3) have the form $\theta_S = [w \;\; a \;\; \delta]^T$. The response amplitude scaling parameter γ_S is absent from θ_S. It is chosen such that the normalization condition (15) is satisfied. In accordance with the Algorithm 1, the components of the parameter vector θ_S, as a part of the parameter vector θ_C (5), should be identified by solving the problem (19).

For illustrative purposes, consider further a linear mapping (4) between the spatial actuator position and its spatial response center position. Such mapping describes a uniform shrinkage (or expansion) of the web between the actuators and a profile measuring device. A discussion of a parametric model and practical identification results for nonlinear (nonuniform) shrinkage of paper web can be found in [4, 5]. Denote by a_k a spatial coordinate of the k-th actuator in the actuator array. For uniform web shrinkage, the response center c_k for this actuator can be computed as $c_k = \alpha_0 + a_k\alpha_1$. This leads to a model of the form (4), where $\theta_A = [\alpha_0 \;\; \alpha_1]^T$. The optimal estimate of θ_A from the data is very insensitive to change of the response shape parameters θ_S. In practice, a 50% error in the estimated response width w usually results in a negligible error of the mapping identification. This means an accurate estimate of the spatial mapping parameter θ_A can be in effect separated from the estimate of the spatial response shape parameter factor θ_S. A more detailed analysis and explanation of this fact is given in [4].

Figure 1 shows identification results obtained for a weight CD process on a US newsprint machine. The estimated process responses $\hat{h}$ (18) (dashed line in the upper plot) and $\hat{p}_u$ (20) (dashed line in the middle plot) are plotted together with the predicted model responses p_u (11) and h (12) (smooth lines) for the identified parameters. The lower plot in Figure 1 shows the actuator excitation profile U in (6).

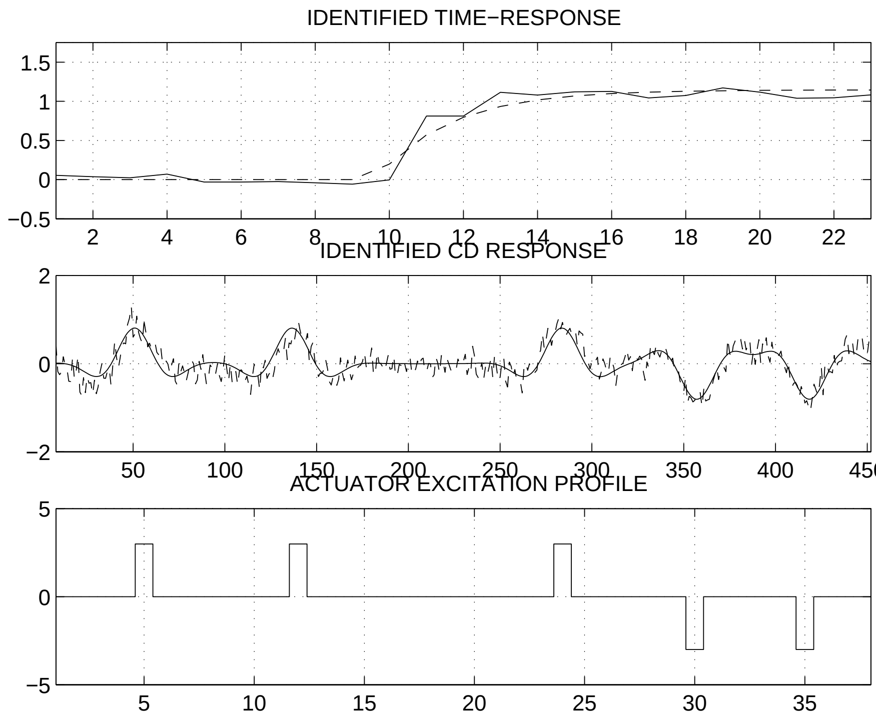

Figure 1: Identification of paper weight response for a stock extrusion gap (slice lip) actuator on a US newsprint machine. Upper plot: identified and estimated time response depending on the time sample number, in relative units; middle plot: identified and estimated spatial weight response depending on the measurement databox number, in g/m^2; lower plot: actuator excitation profile depending on the actuator number, in percent of the actuator range.

The excitation amplitude $a(t)$ in (6) was an unit step applied at sample 8. The proposed algorithm identifies CD process model in a fast, accurate, and reliable way.

6 Conclusions

This paper proposed high-performance algorithms for identification of cross-directional processes from input/output data. The main approach is reminiscent of the GLS identification method and is based on iterative alternating identification of the parametric models for time-response and spatial response of the process. This allows to achieve high performance, when processing two-dimensional identification data. It was theoretically demonstrated that the proposed algorithm is asymptotically convergent and provides unbiased estimates of the distributed process parameters.

The proposed general approach was used for specific parametric models in an industrial identification tool. The identification results for diverse types of the paper machine CD process data were demonstrated. The proposed algorithms work very well for these data, which confirms their practical value. In the application the algorithms provide very reliable identification of the process model from short sequences of very noisy identification experiment data. This, together with the high computational performance, is very desirable in practice and ensures successful application of the algorithms in industrial conditions.

Appendix: Proof of Theorem 1

The proof of Theorem 1 will require the following Lemma.

Lemma 1 *Under the conditions of Theorem 1 the following estimates hold*

$$\|\hat{h} - h\| \leq C_1^h \|h_* - h\| + C_2^h \|p_* - p\| + C_3^h \xi_0 \tag{28}$$

$$\|\hat{p} - p_u\| \leq C_1^p \|h_* - h\| + C_2^p \|p_* - p\| + C_3^p \xi_0 \tag{29}$$

$$\|\gamma - \gamma_*\| \leq C_1^\gamma \|h_* - h\| + C_2^\gamma \|p_* - p\| + C_3^\gamma \xi_0 \tag{30}$$

where C_j^h, C_j^p, and C_j^γ (j = 1,2,3) are constants independent of the parameter vector θ and the noise realization Ξ.

Note that estimates similar to (28)–(30) can be obtained for $\|\hat{h} - h_*\|$, and $\|\hat{p}_u - p_*\|$. This is because $\|\hat{h} - h_*\| \leq \|\hat{h} - h\| + \|h - h_*\|$

Proof of Lemma 1. The estimates (28)–(30) can be obtained from (15), (16), (18), (20), using (24) and the Theorem 1 conditions of θ being limited to a neighborhood of θ^* and ξ_1, ξ_2, ξ_3, ξ_4, in (27) being sufficiently small. ■

Proof of Theorem 1. The iterations of Algorithm 1 converge to a single solution (stationary point) provided that in a certain domain containing this point the mapping is contracting. Let us show that this holds in a vicinity of the exact solution, $\theta = \theta_*$, provided that the measurement noise Ξ is small enough. We assume that in the considered problem all functions are infinitely continuously differentiable. It is therefore sufficient to show that for the linearization of the mapping (23) at the exact solution $\theta = \theta_*$, the singular values of the transformation matrix for such linearization are strictly less than unity.

By using the Implicit Function Theorem for computing the derivatives of the right hand sides in (21) and (22), we obtain

$$B_T = \frac{\partial \phi_T}{\partial \theta_C} = \left(\frac{\partial f_T}{\partial \theta_T}\right)^{-1} \frac{\partial f_T}{\partial \theta_C} \tag{31}$$

$$B_C = \frac{\partial \phi_C}{\partial \theta_T} = \left(\frac{\partial f_C}{\partial \theta_C}\right)^{-1} \frac{\partial f_C}{\partial \theta_T} \tag{32}$$

For θ close to θ_* the system (23) can be represented in the form

$$\theta^{(k+1)} = B\theta^{(k)} + \dots \tag{33}$$

$$\theta^{(k)} = \begin{bmatrix} \theta_C^{(k)} \\ \theta_T^{(k)} \end{bmatrix}, \quad B = \begin{bmatrix} 0 & B_C \\ B_T & 0 \end{bmatrix}, \tag{34}$$

where $\dots$ denotes terms of higher order in the components of $\theta^{(k)} - \theta_*$. One can easily check that the set of squared singular values of the matrix B in (34) (the singular values of the matrix $B^T B$) is the composition of the singular value sets for the matrices $B_T^T B_T$ and $B_C^T B_C$. Therefore, the local convergence condition can be represented in the form

$$\bar{\sigma}(B) = \max\{\bar{\sigma}(B_T), \bar{\sigma}(B_C)\} < 1 \tag{35}$$

Let us now calculate the matrices B_T (31), B_C (31) and show that their maximal singular values are less than unity provided that the measurement noise bound ξ_0 is small enough. This would guarantee that the convergence condition (35) holds. One can observe that the expressions for B_T and B_C are symmetric with respect to θ_T and θ_C. Therefore, we will evaluate singular values for matrix B_T only. Calculations for B_C are similar to these for B_T.

For the calculations we need the following expression

$$\frac{\partial \gamma}{\partial \theta_C} = h^T P^T G_C \tag{36}$$

$$\frac{\partial \gamma}{\partial \theta_T} = p_u^T P G_T \tag{37}$$

By using (25), (26) and (36), (37) we obtain

$$\frac{\partial \hat{p}_u}{\partial \theta_C} = -\gamma^{-2} P h h^T P^T G_C = \hat{h}\hat{h}^T G_C \tag{38}$$

$$\frac{\partial \hat{h}}{\partial \theta_C} = -\gamma^{-2} P^T p_u p_u^T P G_T = \hat{h}\hat{h}^T G_T \tag{39}$$

Thus, in (32), (31) the inverted matrix (Hessian) has the form

$$\frac{\partial f_T}{\partial \theta_T} = \frac{\partial}{\partial \theta_T}\left(G_T^T(I+\hat{h}\hat{h}^T)\right)(h-\hat{h}) \\ +G_T^T(I+2\hat{h}\hat{h}^T+\|\hat{h}\|^2\hat{h}\hat{h}^T)G_T \tag{40}$$

The first summand in (40) is proportional to $h-\hat{h}$. According to Lemma 1, $h-\hat{h}$ can be made as small as needed if ξ_0 and $\theta-\theta_*$ are small. The second summand is a strictly positive definite matrix with the lowest singular value no less than that of the matrix $G_T^T G_T$. Thus, the inverted Hessian matrix in (40) has bounded norm provided that ξ_0 is small enough.

The second multiplicand in (31) can be calculated as follows

$$\frac{\partial f_T}{\partial \theta_C} = \frac{\partial}{\partial \theta_T}\left(G_T^T(I+\hat{h}\hat{h}^T)\right)(h-\hat{h}) \\ +G_T^T(I+\hat{h}\hat{h}^T)(P-\gamma\hat{p}\hat{h}^T)\frac{1}{\gamma}G_C \tag{41}$$

By substituting (24) into the expressions for $(P-\gamma\hat{p}\hat{h}^T)G_C$ and $h-\hat{h}$, where γ, $\hat{p}$, and $\hat{h}$ are given by (16), (18), and (20), and using Lemma 1) one can see that these expressions are bounded by weighted sums of $\|h(\theta_*)-h(\theta)\|$, $\|p_u(\theta_*)-p_u(\theta)\|$, and ξ_0. Therefore, by imposing a small bound ξ_0 on the noise parameters (27) and assuming that θ is in a sufficiently small neighborhood of θ_*, the values $(P-\gamma\hat{p}\hat{h}^T)G_C$ and $h-\hat{h}$ can be made such that $\bar{\sigma}(B_T)<1$ in (35). A similar argument is valid for $\bar{\sigma}(B_C)$. This proves Assertion 1 of Theorem 1.

To prove Assertion 2 of Theorem 1, consider the update (23). In order for the stationary point of the update to be θ^*, the following should hold

$$\begin{array}{rcl} \theta_C^* & = & \phi_C(\theta_T^*) \\ \theta_T^* & = & \phi_T(\theta_C^*) \end{array}, \tag{42}$$

In accordance with the definition of $\phi_C(\theta_T)$ and $\phi_T(\theta_C)$ as implicit functions in (21, (22), the condition (42) is equivalent to

$$\begin{array}{rcl} f_C(\theta_T^*,\theta_C^*) & \equiv & G_C^T(p_u(\theta_C^*)-\hat{p_u})=0 \\ f_T(\theta_T^*,\theta_C^*) & \equiv & G_C^T(h(\theta_T^*)-\hat{h})=0 \end{array}, \tag{43}$$

Lemma 1 can be used to estimate the l.h.s. in (43). Recalling that $p_*=p_u(\theta_C^*)$ and $h_*=h_u(\theta_T^*)$ we obtain the bounds for the l.h.s. in (43) proportional to ξ_0. Hence for $\xi_0\to 0$ the stationary (convergence) point of (42) tends to the exact estimates θ_T^*, θ_C^*. ■

References

[1] Barnett, S. *Matrices. Methods and Applications*, Clarendon Press, Oxford, 1990.

[2] Braatz, R.D., Featherstone A.P., and Ogunnaike, B.A. "Identification, estimation, and control of sheet and film processes," *13th World IFAC Congress*, San Francisco, CA, June 1996, Vol. 7 pp. 319–324.

[3] Duncan, S.R. "Estimating the response of actuators in a cross-directional control system," *Control Systems '96*, Halifax, N.S., Canada, May 1996, pp. 19-22.

[4] Gorinevsky, D. and Heaven, E.M., "Automated identification of actuator mapping in cross-directional control of paper machine," *Americal Control Conf.*, Albuquerque, NM, June 1997, pp. 3400–3404.

[5] Gorinevsky, D., Heaven, E.M., et al., "New algorithms for intelligent identification of paper alignment and nonlinear shrinkage," *Pulp and Paper Canada*, vol. 98, no. 7, 1997, pp. 76–81.

[6] Gorinevsky, D., Heaven, E.M., et al., "Advanced on-line automated identification tool for mapped and multivariable cross-direction control systems," *Int. CD Symposium'97, XIV IMEKO World Congress*, Tampere, Finland, June 1997

[7] Heaven, E.M., et al. "Recent advances in cross-machine profile control," *IEEE Control Systems Magazine*, October 1994, pp. 36–46.

[8] Kristinsson, K. and Chen, S-C. "Identification of cross directional behaviour in web production: techniques and experiences," *Int. CD Symposium'97, XIV IMEKO World Congress*, Tampere, Finland, June 1997

[9] Heaven, E.M., et al. "Application of system identification to paper machine model development and simulation," *Pulp and Paper Canada*, April 1996, pp. 49–54

[10] Ljung, L. *System Identification. Theory for the User.* Prentice Hall, Englewood Cliffs, N.J., 1987

[11] McFarlin, D.A. "Control of cross-machine sheet properties on paper machines," *Process Control Symposium*, 1983, pp. 49–54.

[12] Söderström, T. "Convergence properties of the Generalized Least Square identification method," *Automatica*, Vol. 10, 1974, pp. 617–626

[13] Spall, J.C., and Garner, J.P. "Parameter identification for state-space models with nuisance parameters," *IEEE Tr. on Aerospace and Electronic Systems*, Vol. 26, 1990, pp. 992–998.

[14] Wellstead, P.E., Heath W.P., and Kjaer, A.P. "Identification and control for web forming processes," *13th World IFAC Congress*, San Francisco, CA, June 1996, Vol. 7 pp. 325–330.