International Journal of Control
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tcon20

On system identification from noise-obscured input and output measurements

A. E. ROGERS [a] & K. STEIGLITZ [b]

[a] Department of Statistics and Computer Science, University of Delaware

[b] Department of Electrical Engineering, Princeton University

Available online: 27 Mar 2007

INT. J. CONTROL, 1970, VOL. 12, NO. 4, 625–635

On system identification from noise-obscured input and output measurements†

A. E. ROGERS
Department of Statistics and Computer Science, University of Delaware

K. STEIGLITZ
Department of Electrical Engineering, Princeton University

[Received 30 August 1969]

This paper deals with maximum-likelihood system identification when both the input and the output signals are corrupted by Gaussian observation noise. A derivation of exact maximum-likelihood estimation for this problem is included, but the difficulty of implementing it numerically precludes its practical evaluation at this time. A new approximate method is introduced, called the 'output reference' method, in which the input noise is referred to the output, and an iterative gradient search method used. This technique requires no *a priori* knowledge of the noise covariance matrix. The method of Koopmans–Levin, which does require knowledge of the noise covariance matrix, is then reviewed in detail, and experimental results are presented for the white noise case which indicate that the output reference method is more accurate.

1. Introduction

We consider in this paper maximum-likelihood system identification when both the input and output signals are corrupted by Gaussian observation noise. We assume that the unknown system can be modelled by a linear time-invariant discrete-time transfer function with both zeros and poles and that the available records are uninterrupted sequences of evenly spaced simultaneous samples of input and output.

Koopmans (1937) described an eigenvalue procedure which produces maximum-likelihood estimates provided the data can be separated into independent blocks, each block containing $P/2$ successive samples of the input and corresponding output signals, where there are $P-1$ unknown parameters. Levin (1964) suggested using the procedure on uninterrupted data, organizing it into adjacent or even overlapping blocks. Ignoring the requirement for independence between the blocks of data produces estimates which are not truly maximum likelihood, but which approximate the property in some sense. Smith and Hilton (1967) have reported that the use of overlapping data blocks in the eigenvalue procedure is preferable to the use of adjacent blocks for the estimates from a finite record deviate less from known values in Monte Carlo computer experiments. Their results are corroborated here. It is important to note that this method assumes that the noise covariance matrix is known in advance.

Åström and Bohlin (1965) and McBride and the present authors (Rogers and Steiglitz 1967, Steiglitz and McBride 1965) have dealt with the case when noise is present only in the output measurements, and have presented iterative gradient search methods for achieving maximum-likelihood estimation in this case.

† Communicated by the Authors. This work was supported by U.S. Army Research Office–Durham under Contract DAHC 04–69–C–0012.

We put forth here the idea of referring the input noise to the output and using the iterative gradient search methods for the transformed problem. The unknown noise colouring filter is in this case related to the unknown system transfer function, and this constraint must be omitted from the computation. The resulting procedure, called the 'output reference' method, therefore produces estimates which are only approximately maximum likelihood, as does the Koopmans–Levin method. It has the advantage over the Koopmans–Levin method of requiring no advance knowledge of the noise covariance matrix.

The formulation of a true maximum-likelihood estimate for this problem, without any approximation, is also discussed in this paper. It is shown that although theoretically appealing, it leads to a numerical problem which is not now practical to solve on a digital computer.

Finally, the results of computational experiments are given, comparing the output reference method with the Koopmans–Levin method. It is found that the output reference method is more accurate in the white noise case investigated. Thus, the output reference method seems preferable on all counts, since it appears more accurate, and does not require knowledge of the noise covariance matrix.

2. The output reference method

Consider the situation shown in fig. 1. To relate the input–output sequences $x(t), y(t)$, and the corresponding corrupted observations $p(t), q(t)$, define the following linear model in z-transform notation:

$$\left.\begin{aligned} Y(z) &= \frac{A(z)}{B(z)} X(z) \\ P(z) &= X(z) + \lambda U(z), \\ Q(z) &= Y(z) + \mu V(z), \end{aligned}\right\} \qquad (1)$$

where

$$A(z) = a_0 + a_1 z^{-1} + \ldots + a_K z^{-K},$$

$$B(z) = 1 + b_1 z^{-1} + \ldots + b_L z^{-L}$$

and $U(z), V(z)$ are z-transforms of mutually independent, Gaussian noise sequences with unknown spectra.

It is convenient to assume that all noise spectra result from filtering white noise through linear, time-invariant filters. For simplicity these colouring filters are taken to be stable and all-pole where, of course, the pole positions are unknown and have to be determined from the observations along with the system parameters. One could equally well allow the filters to have both poles and zeros, but here we assume:

$$\left.\begin{aligned} U(z) &= \frac{1}{F(z)} E_1(z), \\ V(z) &= \frac{1}{G(z)} E_2(z), \end{aligned}\right\} \qquad (2)$$

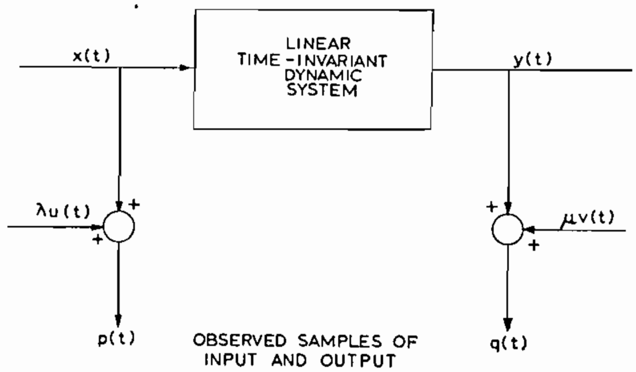
where

$$F(z) = 1+f_1 z^{-1}+\ldots+f_m z^{-m},$$

$$G(z) = 1+g_1 z^{-1}+\ldots+g_n z^{-n}$$

and $E_1(z)$, $E_2(z)$ are z-transforms of mutually independent, Gaussian, white, noise sequences.

Fig. 1



Block diagram illustrating the assumed situation in which system parameter estimation is performed.

Our primary requirement is to estimate the coefficients a_j, b_j and the values of λ, μ.

From (1) and (2) we write:

$$Q(z) = \frac{A(z)}{B(z)}\left[P(z)-\frac{\lambda}{F(z)}E_1(z)\right]+\frac{\mu}{G(z)}E_2(z)$$

$$= \frac{A(z)}{B(z)}P(z)+\sigma\frac{C(z)}{D(z)}E(z), \tag{3}$$

where

$$\sigma\frac{C(z)}{D(z)}E(z) = \frac{\mu}{G(z)}E_2(z)-\frac{\lambda A(z)}{B(z)\,F(z)}E_1(z). \tag{4}$$

Since both sequences $e_1(t)$ and $e_2(t)$ are Gaussian, and all transfer functions are linear and time-invariant, the signal $e(t)$ in eqn. (4) is also Gaussian. Furthermore, by proper choice of $C(z)/D(z)$, $e(t)$ can be made white. Thus it is possible to view $p(t)$ as a precisely measurable input sequence and $q(t)$ as an output sequence which includes additive, coloured, Gaussian noise, the result of filtering white noise by an unknown filter C/D.

Let:

$$\left.\begin{aligned} C(z) &= 1+c_1 z^{-1}+\ldots+c_M z^{-M}, \\ D(z) &= 1+d_1 z^{-1}+\ldots+d_N z^{-N}. \end{aligned}\right\} \tag{5}$$

The task of estimating from input–output records $p(t), q(t)$, $t = 1, 2, \ldots, T$, the values of the parameters $a_j, b_j, c_j, d_j, \sigma$ has been treated elsewhere (Åström and Bohlin 1965, Rogers 1968, Rogers and Steiglitz 1967), where a practical gradient search technique, the damped Gauss–Newton algorithm, is described.

To be precisely maximum likelihood, the output-reference method needs to include the constraint relating $C(z)/D(z)$ to $A(z)/B(z)$, eqn. (4). For practical purposes the constraint is omitted and the method thereby becomes only approximately maximum likelihood.

The gradient search method employs the log-likelihood function:

$$L(\theta) = \ln [P(q(1), \ldots, q(T) \mid p(1), \ldots, p(T), \boldsymbol{\theta}, \sigma)]$$
$$= -\frac{T}{2} \ln 2\pi - T \ln \sigma - \frac{1}{2\sigma^2} \sum_{t=1}^{T} \tilde{e}^2(t), \tag{6}$$

where

$$\boldsymbol{\theta} = [a_0, \ldots, a_K, b_1, \ldots, b_L, c_1, \ldots, c_M, d_1, \ldots, d_N]$$

and

$$\tilde{E}(z) = \frac{D(z)}{C(z)} \left[Q(z) - \frac{A(z)}{B(z)} P(z) \right].$$

The function $L(\boldsymbol{\theta})$ is maximized by a choice $\boldsymbol{\theta}$ and σ.

3. The Koopmans–Levin method

We choose to derive the eigenvalue method since the details are not easily found elsewhere (they were omitted from Levin (1964) and Smith and Hilton (1967)), and they are very relevant to the present discussion of maximum-likelihood estimation.

The sequences $x(t), y(t), p(t), q(t), u(t), v(t)$ shown in fig. 1 are arranged into data blocks and the following P-vectors defined:

$$\left.\begin{aligned} \boldsymbol{\mu}(t) &= [y(t), y(t-1), \ldots, y(t-L), x(t), \ldots, x(t-K)]', \\ \boldsymbol{\lambda}(t) &= [q(t), q(t-1), \ldots, q(t-L), p(t), \ldots, p(t-K)]', \\ \boldsymbol{\eta}(t) &= [v(t), v(t-1), \ldots, v(t-L), u(t), \ldots, u(t-K)]', \\ \boldsymbol{\theta} &= [1, b_1, \ldots, b_L, -a_0, \ldots, -a_K]', \end{aligned}\right\} \tag{7}$$

where $P = K + L + 2$.

The structure of the system is then given by:

$$\boldsymbol{\mu}'(t)\, \boldsymbol{\theta} = 0, \tag{8}$$

equivalent to $Y(z) = A(z)\, X(z)/B(z)$.

The corrupting noises represented in $\boldsymbol{\eta}(t)$ are assumed to be Gaussian with zero mean and known covariance matrix:

$$E[\boldsymbol{\eta}(t)\, \boldsymbol{\eta}'(t)] = Z. \tag{9}$$

Thus for each data block there is a conditional probability density function:

$$P(\boldsymbol{\lambda} \mid \boldsymbol{\theta}) = [2\pi]^{-P/2} |Z|^{-1/2} \exp\{-\tfrac{1}{2} \| \boldsymbol{\lambda}(t) - \boldsymbol{\mu}(t) \|_{Z^{-1}}^2\}. \tag{10}$$

If we choose N independent blocks of data at times t_j then we can define a log-likelihood function:

$$L(\boldsymbol{\theta}) = -\frac{PN}{2} \ln 2\pi - \frac{N}{2} \ln |Z| - \frac{1}{2} \sum_{j=1}^{N} \| \boldsymbol{\lambda}(t_j) - \boldsymbol{\mu}(t_j) \|_{Z^{-1}}^2 \tag{11}$$

and for maximum-likelihood estimation, a value $\hat{\boldsymbol{\theta}}$ is chosen to maximize L subject to the structural constraints:

$$\boldsymbol{\mu}'(t_j)\, \boldsymbol{\theta} = 0, \quad j = 1, 2, \ldots, N. \tag{12}$$

Clearly, due to its Gaussian form, $L(\boldsymbol{\theta})$ is a maximum when

$$J = \frac{1}{2}\sum_{j=1}^{N}[\|\boldsymbol{\lambda}(t_j)-\boldsymbol{\mu}(t_j)\|_{Z^{-1}}^2 + 2\Psi_j \boldsymbol{\mu}'(t_j)\boldsymbol{\theta}] \tag{13}$$

is a minimum, Ψ_j being Lagrange multipliers.

Now J depends on $\boldsymbol{\theta}$ indirectly through $\boldsymbol{\mu}(t_j)$ and thus to find $\hat{\boldsymbol{\theta}}$, we choose first to minimize J with respect to $\boldsymbol{\mu}(t_j)$. This is analogous to estimating the input–output sequences, a step we are forced to take in finding an exact maximum-likelihood procedure with uninterrupted data. Provided the observation blocks of data are independent, J can be viewed as the sum of N independent quadratic forms, each having an additional scalar constraint. The minimum of J is given by the sum of the separate minimum values, readily found by applying the necessary conditions:

$$\operatorname*{grad}_{\boldsymbol{\mu}(t_j)} J = -Z^{-1}[\boldsymbol{\lambda}(t_j)-\boldsymbol{\mu}(t_j)] + \Psi_j\boldsymbol{\theta} = 0, \quad j = 1, 2, \ldots, N. \tag{14}$$

Thus

$$\Psi_j Z\boldsymbol{\theta} = \boldsymbol{\lambda}(t_j) - \boldsymbol{\mu}(t_j). \tag{15}$$

Premultiplying by $\boldsymbol{\theta}'$ and using the structural constraint, eqn. (12) gives:

$$\Psi_j = \frac{\boldsymbol{\theta}'\boldsymbol{\lambda}(t_j)}{\boldsymbol{\theta}'Z\boldsymbol{\theta}}. \tag{16}$$

Alternatively from eqn. (14), we have:

$$\boldsymbol{\mu}(t_j) = \boldsymbol{\lambda}(t_j) - \Psi_j Z\boldsymbol{\theta} \tag{17}$$

and thus substituting in eqn. (13):

$$\begin{aligned}\min_{\boldsymbol{\mu}(t_j)} J &= \frac{1}{2}\sum_{j=1}^{N}\Psi_j\boldsymbol{\lambda}'(t_j)\boldsymbol{\theta} \\ &= \frac{1}{2}\frac{\boldsymbol{\theta}'D\boldsymbol{\theta}}{\boldsymbol{\theta}'Z\boldsymbol{\theta}},\end{aligned} \tag{18}$$

where

$$D = \sum_{j=1}^{N}\boldsymbol{\lambda}(t_j)\boldsymbol{\lambda}'(t_j). \tag{19}$$

To minimize J over $\boldsymbol{\theta}$, it is now therefore necessary to choose $\boldsymbol{\theta}$ so that $\min_{\boldsymbol{\mu}(t_j)} J$ is a minimum. Due to the form of $\min_{\boldsymbol{\mu}(t_j)} J$ this is easily accomplished by making $\boldsymbol{\theta}$ the eigenvector, normalized so that its first component is unity, corresponding to the smallest eigenvalue of:

$$(D-\gamma Z)\boldsymbol{\theta} = 0. \tag{20}$$

The analysis hinges on the observation blocks $\boldsymbol{\lambda}(t_j)$ being independent. Were they to be interrelated by constraints arising from the system transfer function then the minimization of J would require the inclusion of these constraints as well as those of eqn. (12). If the observation blocks are adjacent or overlapping, the estimation that results is not truly maximum likelihood, a point noted by McBride and Levin (1965), but not mentioned by Smith and Hilton (1967). With limited data one chooses to ignore the constraints and, as suggested by Levin (1964), solve for $\boldsymbol{\theta}$ through the eigenvalue method making the matrix D:

$$D = \sum_{t=1}^{T}\boldsymbol{\lambda}(t)\boldsymbol{\lambda}'(t). \tag{21}$$

4. True maximum-likelihood estimation

When both input and output records are corrupted by additive Gaussian noise, as represented in fig. 1, there is no difficulty in expressing the log-likelihood function:

$$L = \ln [P(q(1), \ldots, q(T), p(1), \ldots, q(T) \,|\, x(1), \ldots, x(T), \boldsymbol{\theta}, \lambda, \mu)]$$

$$= -T \ln 2\pi - T \ln \lambda\mu - \frac{1}{2\lambda^2} \sum_{t=1}^{T} \tilde{e}_1{}^2(t) - \frac{1}{2\mu^2} \sum_{t=1}^{T} \tilde{e}_2{}^2(t), \tag{22}$$

where in z-transform notation

$$\left.\begin{aligned} \tilde{E}_1(z) &= \lambda E_1(z) = F(z)\,[P(z) - X(z)], \\ \tilde{E}_2(z) &= \mu E_2(z) = G(z)\,[Q(z) - \frac{A(z)}{B(z)} X(z)]. \end{aligned}\right\} \tag{23}$$

Clearly L is a function of the unknown input sequence $x(t)$, as well as λ, μ and the coefficients of A, B, F and G. Thus its maximization requires that the sequence $x(t)$ be estimated along with the parameters. As in the Koopmans–Levin procedure, one possibility is to maximize L with respect to the input sequence $x(t)$, before choosing the parameter vector $\boldsymbol{\theta}$.

For simplicity of discussion we assume the noise to be white

$$(F(z) = G(z) = 1),$$

and omit from expressions the explicit dependence on z writing X for $X(z)$ and $\bar{X}$ for $X(z^{-1})$. Then

$$L = -T \ln 2\pi - T \ln \lambda\mu - J_1 - J_2, \tag{24}$$

where, using Parseval's relation,

$$\left.\begin{aligned} J_1 &= \frac{1}{2\lambda^2} \sum_{t=1}^{T} \tilde{e}_1{}^2(t) = \frac{1}{2\lambda^2} \frac{1}{2\pi j} \oint (P - X)(\bar{P} - \bar{X}) \frac{dz}{z}, \\ J_2 &= \frac{1}{2\mu^2} \sum_{t=1}^{T} \tilde{e}_2{}^2(t) = \frac{1}{2\mu^2} \frac{1}{2\pi j} \oint \left(Q - \frac{A}{B} X\right)\left(\bar{Q} - \frac{\bar{A}}{\bar{B}} \bar{X}\right) \frac{dz}{z}, \end{aligned}\right\} \tag{25}$$

L is a maximum when $J = J_1 + J_2$ is a minimum. Differentiating J with respect to $x(\tau)$ and applying the necessary conditions gives for every τ:

$$\frac{\partial J}{\partial x(\tau)} = \frac{-1}{2\pi j} \oint \left[\frac{1}{\lambda^2}(P - X) + \frac{1}{\mu^2}\left(Q - \frac{A}{B} X\right) \frac{\bar{A}}{\bar{B}}\right] z^{-\tau 1}\, dz = 0. \tag{26}$$

Thus

$$\frac{1}{\lambda^2}(P - X) + \frac{1}{\mu^2}\left(Q - \frac{A}{B} X\right) \frac{\bar{A}}{\bar{B}} = 0$$

or

$$X = \frac{P + (\lambda^2/\mu^2)(\bar{A}/\bar{B}) Q}{1 + (\lambda^2/\mu^2)(\bar{A}A/\bar{B}B)}, \tag{27}$$

which is clearly recognizable as the solution of a Wiener-type filtering problem. Substituting this expression in those for J_1 and J_2 gives:

$$\min_{x(t)} J = \frac{1}{2} \frac{1}{2\pi j} \oint \frac{(AP - BQ)(\overline{AP - BQ})}{(\mu^2 B\bar{B} + \lambda^2 A\bar{A})} \frac{dz}{z}, \tag{28}$$

which is analogous to eqn. (18) above in the discussion of the Koopmans–Levin eigenvalue procedure. Notice that when $\lambda = 0$ then $X = P$ and eqn. (28) can be derived directly as the function to be minimized in a maximum-likelihood estimation of parameters in the presence of white output noise (see Åström and Bohlin (1965), Steiglitz and McBride (1965)).

The problem, of course, is not yet solved for it is now necessary to maximize L with respect to the system parameters and λ, μ:

$$\max_{x(\tau)} L = -T \ln 2\pi - T \ln \lambda\mu - \frac{1}{2}\frac{1}{2\pi j}\oint \frac{(AP-BQ)\overline{(AP-BQ)}}{\mu^2 B\bar{B}+\lambda^2 A\bar{A}}\frac{dz}{z}. \tag{29}$$

The procedure is complicated by the need to factor the denominator of the integrand. Assume:

$$\mu^2 B\bar{B}+\lambda^2 A\bar{A} = H\bar{H}, \tag{30}$$

where all zeros of H lie inside the unit circle. Then

$$\frac{1}{2\pi j}\oint \frac{(AP-BQ)\overline{(AP-BQ)}}{H\bar{H}}\frac{dz}{z} = \sum_{t=1}^{T} e^2(t), \tag{31}$$

where the z-transform of $e(t)$ is:

$$E(z) = \frac{1}{H}(AP-BQ), \tag{32}$$

and therefore the signal $e(t)$ can be obtained numerically by suitably filtering the observations $p(t)$ and $q(t)$—provided only that H is available by factoring $\mu^2 B\bar{B}+\lambda^2 A\bar{A}$. In any iterative search for a maximum of L, the factorization is necessary at each step causing considerable difficulty, one that clearly does not arise if either λ or μ is zero.

The search for the maximum of L requires not only the numerical calculation of $e(t)$, but also in any gradient procedure, the gradient of $e(t)$ with respect to the parameter vector and this too is complicated by the form of the integrand in eqn. (31). In the coloured-noise case the equations become even more complicated. For these reasons the method has not been programmed, despite an interest in comparing the results of approximate techniques with those of true maximum likelihood.

5. Experimental comparison between the output reference and the Koopmans–Levin methods

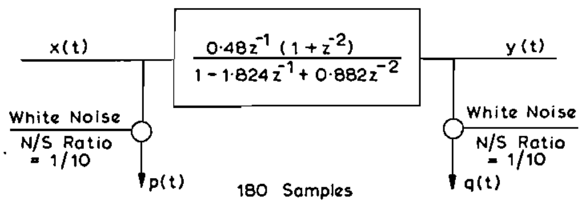
For purposes of comparison, the output reference methods and the Koopmans–Levin were used to estimate system parameter values from data samples generated with the following pulse transfer function:

$$\frac{A(z)}{B(z)} = \frac{0{\cdot}48z^{-1}+0{\cdot}48z^{-2}}{1-1{\cdot}824z^{-1}+0{\cdot}8825z^{-2}}.$$

It is similar to the one used in the work of Smith and Hilton (1967) and represents an under-damped, second-order system. White noise was added to both the input and the output records, with $\lambda^2 = \mu^2 = 0{\cdot}1$, producing the noise-corrupted sequences from which the system parameters were estimated. Ten different noise samples were used, and for each a record of 180 samples of input–output behaviour produced.

Figure 2 shows the results of the Koopmans–Levin method using 60 adjacent data blocks, each containing three consecutive input–output samples to form the necessary six vectors for the method. It was assumed in applying the Koopmans–Levin method that the noise covariance matrix was known to be its true value.

Fig. 2



Koopmans–Levin parameter estimation from finite records, adjacent data blocks.

Koopmans–Levin eigenvalue solution to determine the coefficients of the model:

$$Y(z) = \frac{A(z)}{B(z)} X(z),$$

where

$$A(z) = a_0 + a_1 z^{-1} + a_2 z^{-2},$$

$$B(z) = 1 + b_1 z^{-1} + b_2 z^{-2}.$$

Ten estimates for different white-noise disturbances: 180 samples arranged into 60 adjacent sets

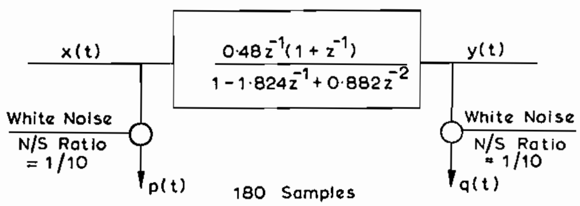
Estimate	a_0	a_1	a_2	$-b_1$	b_2
1	0·688	0·370	1·363	2·442	1·325
2	−0·045	0·727	0·136	1·604	0·859
3	−0·163	−0·130	0·162	1·181	0·332
4	−0·958	−1·370	1·024	1·885	1·160
5	−1·555	0·231	0·423	1·955	1·130
6	0·476	1·159	1·919	2·038	1·114
7	−0·475	−0·863	0·483	2·673	1·999
8	−0·638	1·833	0·945	1·896	0·966
9	−0·181	0·522	0·609	1·313	0·399
10†	0·374	−2·774	−6·817	4·973	4·203
Mean	−0·317	0·275	0·785	1·700	0·929
Variance	0·478	0·956	0·346	0·233	0·247

† Values ignored in calculating mean and variance.

Figure 3 shows the results of the same method using 178 overlapping data blocks, each again containing three consecutive input–output samples. The estimated values of system parameters are clearly closer, on average, to the true values by this use of the data.

Figures 4 and 5 show the results obtained with the output-reference method using three and four parameter noise-colouring filters, respectively. The estimated values of system parameters are more accurate than those obtained above, and the method does not seem very sensitive to the assumed order of the noise-colouring filters. Thus, it appears that the output reference method is preferable over the Koopmans–Levin method, especially when little is known about the covariance of the observation noise.

Fig. 3



Koopmans–Levin parameter estimation from finite records, overlapping data blocks.

Koopmans–Levin eigenvalue solution to determine the coefficients of the model:

$$Y(z) = \frac{A(z)}{B(z)} X(z),$$

where

$$A(z) = a_0 + a_1 z^{-1} + a_2 z^{-2},$$

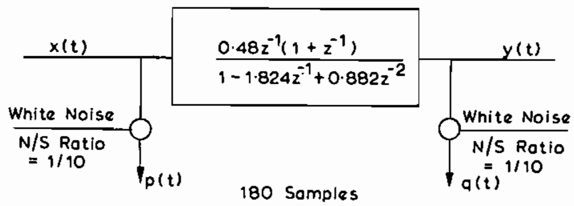
$$B(z) = 1 + b_1 z^{-1} + b_2 z^{-2}.$$

Ten estimates for different white-noise disturbances: 180 samples arranged into 178 overlapping sets

Estimate	a_0	a_1	a_2	$-b_1$	b_2
1	0·560	−0·245	1·048	1·805	0·868
2	−0·864	0·689	0·670	1·834	0·915
3	0·185	−0·274	0·644	1·862	0·935
4	0·087	−0·016	0·986	1·807	0·876
5	−0·341	0·691	0·500	1·812	0·885
6	−0·186	−0·078	1·151	1·807	0·875
7	0·015	0·038	0·979	1·777	0·842
8	−0·753	0·917	0·769	1·777	0·846
9	−0·010	0·498	0·635	1·788	0·851
10†	0·567	0·419	−0·391	1·926	0·992
Mean	−0·145	0·247	0·820	1·807	0·877
Variance	0·203	0·205	0·051	0·0007	0·001

† Values ignored in calculating mean and variance.

Fig. 4



Output reference parameter estimation from finite records, three-parameter noise filter.

Gradient search to find the maximum likelihood estimate using the model:

$$Q(z) = \frac{A(z)}{B(z)} P(z) + \lambda \frac{C(z)}{D(z)} E(z),$$

where

$$A(z) = a_0 + a_1 z^{-1} + a_2 z^{-2},$$

$$B(z) = 1 + b_1 z^{-1} + b_2 z^{-2},$$

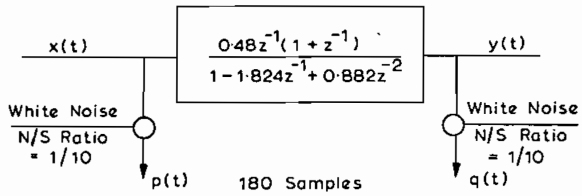
$$C(z) = 1 + c_1 z^{-1},$$

$$D(z) = 1 + d_1 z^{-1} + d_2 z^{-2}.$$

Ten estimates for different white-noise disturbances

Estimate	a_0	a_1	a_2	$-b_1$	b_2	λ^2
1	−0·256	0·742	0·326	1·834	0·897	0·115
2	0·319	−0·146	0·809	1·819	0·880	0·122
3	0·164	0·290	0·627	1·810	0·865	0·134
4	−0·257	1·060	0·047	1·836	0·894	0·107
5	−0·008	0·180	0·785	1·822	0·880	0·132
6	0·056	0·202	0·560	1·824	0·881	0·118
7	−0·149	0·707	0·325	1·816	0·878	0·122
8	−0·093	0·492	0·551	1·815	0·871	0·123
9	−0·317	0·717	0·368	1·821	0·882	0·126
10	0·398	−0·056	0·635	1·826	0·883	0·132
Mean	−0·014	0·420	0·503	1·822	0·881	0·123
Variance	0·061	0·150	0·056	0·0013	0·0003	—

Fig. 5



Output reference parameter estimation from finite records, four-parameter noise filter.

Gradient search to find maximum-likelihood estimate using the model:

$$Q(z) = \frac{A(z)}{B(z)} P(z) + \lambda \frac{C(z)}{D(z)} E(z),$$

where

$$A(z) = a_0 + a_1 z^{-1} + a_2 z^{-2},$$
$$B(z) = 1 + b_1 z^{-1} + b_2 z^{-2},$$
$$C(z) = 1 + c_1 z^{-1} + c_2 z^{-2},$$
$$D(z) = 1 + d_1 z^{-1} + d_2 z^{-2},$$

Ten estimates for different white-noise disturbances

Estimate	a_0	a_1	a_2	$-b_1$	b_2	λ^2
1	−0·253	0·735	0·331	1·834	0·897	0·115
2	0·337	−0·155	0·797	1·820	0·881	0·122
3	0·105	0·391	0·457	1·828	0·886	0·128
4	−0·201	1·019	0·084	1·831	0·888	0·101
5	0·001	0·198	0·604	1·846	0·904	0·123
6	0·136	0·112	0·532	1·830	0·886	0·117
7	−0·142	0·717	0·311	1·817	0·878	0·124
8	−0·082	0·488	0·535	1·817	0·873	0·123
9	−0·320	0·722	0·364	1·821	0·882	0·126
10	0·339	0·201	0·449	1·826	0·883	0·124
Mean	−0·008	0·443	0·446	1·827	0·886	0·120
Variance	0·055	0·129	0·037	0·00008	0·0000	—

References

Åström, K. J., and Bohlin, T., 1965, *proc. IFAC Symp. in Adaptive Control* (Teddington, Middlesex, England).

Koopmans, T., 1937, *Linear Regression Analysis of Economic Time Series* (Haarlem, The Netherlands: De Erven F. Bohn, N.V.).

Levin, M. J., 1964, *I.E.E.E. Trans. autom. Control*, **AC–9**, p. 229.

McBride, L. E., Jr., and Levin, M. J., 1965, *I.E.E.E. Trans. autom. Control*, **AC–10**, p. 214 (correspondence: discussion on Levin (1964).

Rogers, A. E., 1968, Ph.D. Thesis, Department of Electrical Engineering, Princeton University.

Rogers, A. E., and Steiglitz, K., 1967, *I.E.E.E. Trans. autom. Control*, **AC–12**, 594.

Smith, F. W., and Hilton, W. B., 1967, *I.E.E.E. Trans. autom. Control*, **AC–12**, 568.

Steiglitz, K., and McBride, L. E., Jr., 1965, *I.E.E.E. Trans. autom. Control*, **AC–10**, p. 461.