ERRORS-IN-VARIABLES IDENTIFICATION OF NOISY MOVING AVERAGE MODELS

Abdelhakim Youcef [* "], *Roberto Diversi* [†] *and Eric Grivel* [*]

["] Mckay Brothers LLC, 50 Rue de Paradis, 75010 Paris, France
E-mail: abdelhakim.youcef@mckay-brothers.com
[†] DEI - University of Bologna, Viale del Risorgimento 2, 40136 Bologna, Italy
E-mail: roberto.diversi@unibo.it
[*] UMR CNRS 5218 IMS, Université de Bordeaux, Bordeaux INP, 33400 Talence, France
E-mail: eric.grivel@ims-bordeaux.fr

ABSTRACT

In this paper, we propose to address the moving average (MA) parameters estimation issue based only on noisy observations and without any knowledge on the variance of the additive stationary white Gaussian measurement noise. For this purpose, the MA process is approximated by a high-order AR process and its parameters are estimated by using an errors-in-variables (EIV) approach, which also makes it possible to derive the variances of both the driving process and the additive white noise. The method is based on the Frisch scheme. One of the main difficulties in this case is to evaluate the minimal AR-process order that must be considered to have a "good" approximation of the MA process. To this end, we propose a way based on K-means method. Simulation results of the proposed method are presented and compared to existing MA-parameter estimation approaches.

***Index Terms*—** Moving average model, autoregressive model, errors-in-variables (EIV), K-means classification.

1. INTRODUCTION

In various applications of signal processing, *a priori* modeling is of interest. Among the models that are often used in speech and audio processing, radar as well as biomedical applications, many authors have focused their attentions on the estimation of the autoregressive (AR) and moving average (MA) parameters for the last decades. Thus, various ways exist to deduce the AR parameters from the observations by using for instance the Yule-Walker (YW) equations, or online approaches based on the LMS, RLS, etc. Concerning the MA parameters, we can organize the estimation methods in the five main following families:

1. The maximum likelihood (ML) estimation of the MA parameters leads to a highly non-linear problem. To reduce the computational cost, Stoica *et al.* [1, 2] have proposed covariance fitting approaches. When using the so-called "basic method", the first step consists in searching a semi-definite positive matrix, whose sum of each diagonal, either main or secondary, corresponds to the covariance function estimated from the data, for successive lags. This leads to a non-convex minimization problem, that can be seen as a semi-definite programming (SDP). It can be solved by using libraries that are available on websites or in some toolboxes such as in [3]. Then, the MA parameters can be obtained by using a spectral factorization. It should be noted that to perform the second step the estimated covariances must form a "valid" MA covariance sequence, i.e. a sequence guaranteeing the positivity of the corresponding power spectral density (PSD), see [4] or [5]. Variants of the basic method have also been proposed for instance in [6] in which the purpose is to estimate the covariance matrix rather than the covariance function itself.

2. In Durbin's method, the MA process is first approximated by a high-order AR process [7]. The AR parameters are estimated by using a least-squares (LS) method. Then, the corresponding MA parameters can be deduced from the estimated AR parameters.

3. An alternative method [8] is to compute the inverse Fourier transform of the inverse of the MA PSD in order to obtain the so-called inverse covariance sequence. Then, the MA parameters can be estimated by means of the Yule-Walker (YW) equations.

4. The fourth family includes the "vocariance" ESPRIT method and "vocariance" recursion method which are based on the cepstrum. For more details, the reader can refer to [9, 10].

5. Finally, some methods proposed in the literature are based on higher-order statistics, such as in [11, 12].

However, in various cases, the observations may be disturbed by an additive noise. This model parameter estimation issue has been widely studied when dealing with an AR process (resp. a time-varying AR process) disturbed by an additive white noise. See for instance [13–17] (resp. [18]). It has been also investigated more recently in [19] when the AR process is disturbed by a MA noise. However, few authors have addressed this problem for MA parameters when the MA process is disturbed by an additive stationary white Gaussian noise (AWGN) [20, 21]. For this reason, this paper deals with the identification of noisy MA processes. More particularly, we propose to approximate the MA process by a high-order AR process. Then, we derive an errors-in-variables (EIV) approach that makes it possible to estimate both the AR model parameters and the variances of the driving process as well as the additive white measurement noise. The main idea of this approach is that the noisy-observation correlation matrix, compensated by a specific diagonal matrix, must be positive semi-definite and its kernel is defined by the AR parameters. According to the Frisch scheme [22], given a trial order for the AR model, there exists a set of diagonal compensation matrices satisfying the above mentioned property that defines a convex curve. Moreover, all the curves associated with orders equal or greater than the true one exhibits a common point in the asymptotic case. The AR parameter estimation can thus be based on this property. The variant we propose includes a way to select the order of the AR process that must be defined to have a "representative" approximation of the MA process.

To illustrate the performance of our proposed method, a comparative study is carried out with the standard existing methods, recalled in the above state of the art. We also analyze the limits of our proposed method versus the signal-to-noise ratio (SNR), the number of samples available and the positions of the zeros of the MA process[1].

The remainder of this paper is organized as follows: In Section 2 we give the system model. Section 3 describes the proposed identification procedure. In Section 4, simulation results are presented where a comparative study is done with the existing methods. Some conclusions and perspectives are given in Section 5.

2. SYSTEM MODEL

Let $x(n)$ be a qth-order MA process defined as follows:

$$x(n) = \sum_{i=0}^{q} b_i u(n-i) \tag{1}$$

[1]The MA process can be seen as a filtering of a white noise. The corresponding transfer function is only defined by its zeros. They are called the zeros of the MA process.

where $u(n)$ is a zero-mean white Gaussian driving process with variance σ_u^2, $\{b_i\}_{i=0,\dots,q}$ are the MA parameters[2] while $b_0 = 1$ to guarantee the identifiability of the generic driving noise variance. Then, the MA process is assumed to be disturbed by a zero-mean AWGN with variance σ_w^2, denoted by $w(n)$, which is uncorrelated with the driving process $u(n)$, to form the noisy observations $y(n)$, whose nth sample is defined as follows:

$$y(n) = x(n) + w(n) \tag{2}$$

In the following, we assume that N noisy observation samples $\{y(n)\}_{n=1}^{N}$ are available. The nth MA vector of length $r+1$, denoted by $\mathbf{x}_n$, is defined as follows:

$$\mathbf{x}_n = [x(n),\ x(n-1), \dots, x(n-r)]^T \tag{3}$$

Let $\mathbf{y}_n$ be the nth noisy-observation vector of length $r+1$, which is similarly defined as $\mathbf{x}_n$, and let $\mathbf{R}_y^{r+1}$ be the corresponding autocorrelation matrix.

$$\mathbf{R}_y^{r+1} = E[\mathbf{y}_n \mathbf{y}_n^H] = \mathbf{R}_x^{r+1} + \sigma_w^2 \mathbf{I}^{r+1} \tag{4}$$

where $\mathbf{R}_x^{r+1} = E[\mathbf{x}_n \mathbf{x}_n^H]$ is the MA correlation matrix and $\mathbf{I}^{r+1}$ is the identity matrix of size $(r+1) \times (r+1)$.

Here, the objective is to estimate the MA parameters only on the basis of the noisy observations $\{y(n)\}_{n=1}^{N}$ and without any *a priori* information on the AWGN. For this purpose, the MA process is approximated by a rth-order AR process, where r is high enough and is defined as follows:

$$x(n) \approx -\sum_{i=1}^{r} a_i x(n-i) + u(n) \tag{5}$$

where $\{a_i\}_{i=1,\dots,r}$ are the AR model parameters. It should be noted that one way to deduce the set $\{a_i\}_{i=1,\dots,r}$ is to use the YW equations where the correlation function of the MA process, denoted as $\mathbf{R}_{xx}(\tau)$ is defined as mentioned in footnote 2. However, one of the key issues is the selection of the order r. It is known that when the zeros are close to the unit circle in the z-plane, the order must be very high. In the next section, we propose a way to select r and to deduce AR parameters by means of EIV approach.

3. PROPOSED APPROACH

3.1. EIV for true rth-order AR process identification

For a true rth-order AR process, (5) becomes:

$$\begin{cases} x(n) = -\sum_{i=1}^{r} a_i x(n-i) + u(n) \\ a_i = 0 \text{ (for } i > r) \end{cases} \tag{6}$$

[2]It is well known that the correlation of $x(n)$ satisfies :

$$\mathbf{R}_{xx}(\tau) = \begin{cases} \sigma_u^2 \sum_{i=\tau}^{q} b_i b_{i-\tau} & (\text{if } |\tau| \leq q) \\ 0 \text{ (otherwise)} \end{cases}$$

Let $\mathbf{\Theta}_{r+1}$ be the AR parameters vector,

$$\mathbf{\Theta}_{r+1} = [1, a_1, \ldots, a_r]^T \tag{7}$$

By defining the following vector,

$$\overline{\mathbf{x}}_n = [x(n) - u(n),\ x(n-1), \ldots, x(n-r)]^T \tag{8}$$

(5) can be written as follows:

$$\overline{\mathbf{R}}_{x,u}^{r+1} \mathbf{\Theta}_{r+1} = \mathbf{0}_{(r+1)\times 1} \tag{9}$$

where $\overline{\mathbf{R}}_{x,u}^{r+1} = E[\overline{\mathbf{x}}_n \overline{\mathbf{x}}_n^H]$ and $\mathbf{\Theta}_{r+1}$ is hence the kernel of the matrix $\overline{\mathbf{R}}_{x,u}^{r+1}$. However, in practice, only noisy observations are available (see (2)). Since $E[x(n)u(n)] = \sigma_u^2$, it follows that,

$$\overline{\mathbf{R}}_{x,u}^{r+1} = \mathbf{R}_x^{r+1} - diag[\sigma_u^2, \mathbf{0}_{1\times r}] \tag{10}$$

Therefore, by using (4) and (10), we can rewrite (9) as follows:

$$(\mathbf{R}_y^{r+1} - diag[\sigma_w^2 + \sigma_u^2, \sigma_w^2 \mathbf{1}_{1\times r}])\mathbf{\Theta}_{r+1} = \mathbf{0}_{(r+1)\times 1}\ . \tag{11}$$

Given (11), the AR model parameters as well as the variances of both the driving process and the AWGN can be derived by using the EIV approach [14, 23]. To this end, given a trial order ρ, let us consider the set of couples (α, β) with $\beta \geq \alpha$ such that:

$$\mathbf{R}_y^{\rho+1} - diag[\beta, \alpha \mathbf{1}_{1\times \rho}] \geq 0 \tag{12}$$

i.e. the matrix (12) must be positive semidefinite and singular. This set defines the points of a convex curve belonging to the first quadrant of the $(\alpha\beta)$-plane[3]. Every point $P = (\alpha, \beta)$ of this curve can be associated with a coefficient vector $\Theta_{\rho+1}$ that can be obtained from the kernel of the matrix (12).

Because of (11), the point $P^a = (\sigma_w^2, \sigma_w^2 + \sigma_u^2)$, whose elements are the true driving noise and measurement noise variances, belongs to all the curves related to orders $\rho \geq r$ and the following relations hold:

$$(\mathbf{R}_y^{\rho+1} - diag[\sigma_w^2 + \sigma_u^2, \sigma_w^2 \mathbf{1}_{1\times \rho}])\mathbf{\Theta}_{\rho+1}^a = \mathbf{0}_{(\rho+1)\times 1} \tag{13}$$

where

$$\mathbf{\Theta}_{\rho+1}^a = [1, a_1, \ldots, a_r, \underbrace{0, .., 0}_{\rho - r}]^T \tag{14}$$

The above mentioned property makes it possible to determine, in the asymptotic case, the point P^a and then, the true AR coefficients by means of (11).

In practice, the covariance matrices of the noisy observations $\mathbf{R}_y^{\rho+1}$, $(\rho = 1, 2, ...)$ must be replaced by their sample estimates. The resulting curves for $\rho \geq r$ no longer share any common point. To identify the AR parameters, it is thus necessary to introduce suitable selection criteria, such as the shifted relation criterion or the Yule-Walker equations based criterion described in [14].

3.2. EIV for MA process identification

In this work, as the MA process is approximated by a high-order AR process, there is in theory no common point between the convex curves. Nevertheless, when the AR process order ρ is above a certain threshold ρ_{min} (which is high enough), the AR parameters $\{a_i\}_{i=\rho_{min},\ldots,\rho}$ tend to be close to zero and the corresponding spectrum and correlation properties do not change much. Therefore, the resulting convex curves tend to be close one another, especially in a certain region of the $(\alpha\beta)$-plane. In this case, the main issue when using the EIV approach consists in determining the minimal order of the high–order AR process that is required to have a "good" approximation of the MA process. To ensure this latter, we suggest using a method based on the k-means classification algorithm [24]. For this purpose, the EIV approach is first applied for different values of ρ varying from a lower bound (e.g. $\rho = 4$) to an upper bound (e.g. $\rho = 30$), both chosen by the user. Then, the resulting convex curves are classified into two classes, as shown in the example of Fig. 1. To give some details about the classification, it operates as follows: as depicted in Fig. 1, a specific angle ψ is defined to generate a straight line in the first quadrant of the $(\alpha\beta)$-plane. Its intersections with the convex curves provide a set of points that are then separated into two clusters. The first one, denoted C^1, corresponds to the AR orders which are below a certain minimal required order, denoted by r_0. The second one, denoted C^2, corresponds to the AR orders which are greater than or equal to r_0. Once the minimal order is estimated, the EIV-approach, initially proposed for the AR process, is applied on the second class C^2. In this work, the point P^a is obtained by minimizing the sum of distances between successive curves belonging to the class C^2. In Fig. 2, we give the true MA spectrum as well as the corresponding AR spectrum based on (5) using YW method and the estimated AR spectrum using the EIV approach. The three curves are very close one another.

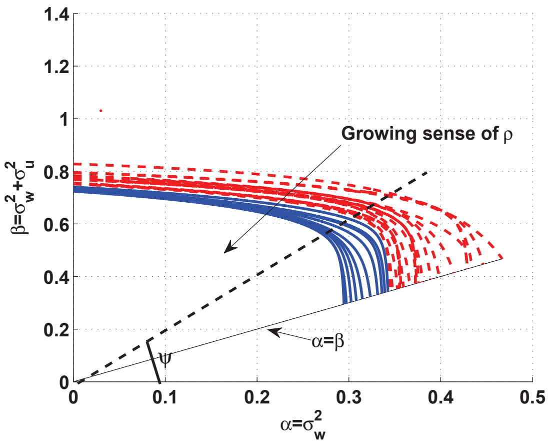

Fig. 1. Classification of the EIV curves of rth-order AR process using k-means algorithm, $r = 4, .., 30$, dotted curves: class 1 (C^1), solid curves: class 2 (C^2), $N = 256$ SNR= 15 dB

[3] α stands for the abscissa while β stands for the ordinate.

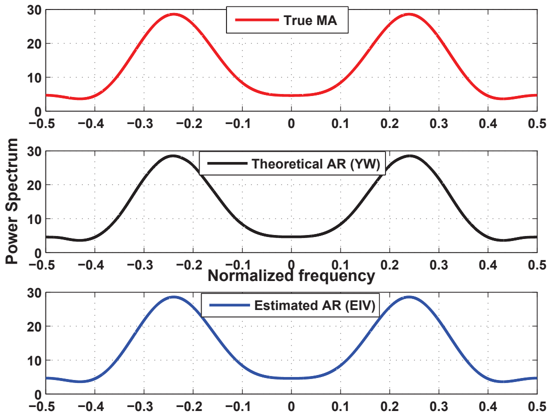

Fig. 2. MA process spectrum approximated by AR process of order $r = 20$, $z_1 = 0.6e^{jpi/8}$, $z_2 = 0.7e^{j5pi/6}$ SNR= 15 dB, 100 RUNS

Table 1. Itakura-Saito spectral distance when the zeros are far from the unit circle, $z_1 = 0.3e^{j\pi/8}$, $z_1 = 0.4e^{j5\pi/6}$ $N = 4096$, $\rho = 17$, $\sigma_u^2 = 1$, RUNS= 30

Methods \ SNR	15 dB	12 dB	8 dB	5 dB
Proposed method	**0.4519**	**0.41**	**0.5747**	**2.8611**
ICM	0.59	0.604	0.770	3.071
DM	0.549	0.550	0.710	3.014

Table 2. Itakura-Saito spectral distance when the zeros are far from the unit circle, $z_1 = 0.3e^{j\pi/8}$, $z_1 = 0.4e^{j5\pi/6}$ $N = 2048$, $\rho = 15$, $\sigma_u^2 = 1$, RUNS= 100

Methods \ SNR	15 dB	12 dB	8 dB	5 dB
Proposed method	**0.3773**	**0.2648**	**0.3919**	**1.179**
DM	0.514	0.565	0.593	1.521
ICM	0.504	0.557	0.594	1.527

4. SIMULATION RESULTS

4.1. Simulation protocol

The proposed method is compared to the following MA estimation methods: Durbin's method (DM) and the inverse covariance matrix (ICM).

The performance are measured in terms of spectral error distance between the original spectrum $S(f)$ and its estimation $\hat{S}(f)$. The Itakura-Saito (IS) spectral distance is considered and defined as follows:

$$D_{IS}(S, \hat{S}) = \frac{1}{N} \sum_{m=1}^{N} \left[\frac{S(f_m)}{\hat{S}(f_m)} - log \frac{S(f_m)}{\hat{S}(f_m)} - 1 \right] \quad (15)$$

According to preliminary tests, we noticed that the estimation of the driving process variance with DM and ICM methods fails on account of a numerical problem (negative values of $\hat{\sigma}_u^2$). For this reason, and unlike our proposed method, the variance of the driving process is assumed to be known in DM and ICM. The MA process is defined by its zeros whose modulus are not too close to 1. The performance are measured as a function of SNR and number of available noisy observations N. The AR order was selected by using our proposed method.

4.2. Results and discussion

Tab. 1 and Tab. 2 give the IS spectral distance measurements where we compare our proposed approach with DM and ICM methods.

Results in Tab. 1 and Tab. 2 are given respectively for a number of noisy observations N of 4096 and 2048.
In Tab. 1, the proposed approach outperforms the other methods for SNR values of about 8 to 15 dB. When the SNR is relatively low (SNR=5 dB), the performance degrade but still remain advantageous for our proposed method. The same performance analysis can be concluded from Tab. 2 when the number of noisy observations decreases. It should be noted that our proposed approach jointly estimates the AR parameters with both driving noise and AWGN variances without any *a priori* information. This is not the case of DM and ICM methods where only the AR parameters are estimated. In addition, the driving process have to be preliminary estimated with another approach (for the results above, this value is provided by our proposed approach). It can be concluded that the proposed identification procedure performs better than DM and ICM methods; this is all the more significant as both the SNR and the available number of observations increase.

We also note that the use of the clustering k-means method to classify the EIV curves and select the minimal required AR order is necessary for the "good" behavior of the proposed method.

5. CONCLUSION AND PERSPECTIVES

In this paper, we address the estimation of MA models disturbed by an AWGN. For this purpose, the MA process is approximated by a high-order AR process. The greater the value of AR order r is, the better the approximation of the MA process is. However, when the AR order is very high, the computational cost of the whole identification procedure increases. To this end, an approach based on the K-means clustering method combined with an EIV approach is used to determine the minimal order of the AR process. Given the proposed method, we can get an estimation of the MA spectrum. Under some assumptions regarding the zeros of the MA process (i.e. in the unit circle in the z-plane), we could then deduce the MA parameters by using a method such as the spectral factorization.

6. REFERENCES

[1] P. Stoica, T. McKelvey, and J. Mari, "MA Estimation in Polynomial Time", *IEEE Trans. on Signal Processing*, vol. 48, no. 7, pp. 1999-2012, 2000.

[2] B. Dumitrescu, I. Tabus, and P. Stoica, "On The Parameterization of Positive Real Sequences and MA Parameter Estimation", *IEEE Trans. on Signal Processing*, vol. 49, no. 11, pp. 2630-2639, 2001.

[3] "http://users.isy.liu.se/johanl/yalmip".

[4] R. Moses and D. Liu, "Optimal Nonnegative Definite Approximations of Estimated Moving Average Covariance Sequences", *IEEE Trans. on Signal Processing*, vol. 39, no. 9, pp. 2007-2015, 1991.

[5] A. Stoica, R. Moses, and P. Stoica, "Enforcing Positiveness on Estimated Spectral Densities", *Electronics Letters*, vol. 29, no. 23, pp. 2009-2011, 1993.

[6] P. Stoica, D. Lin, L. Jian, and T. Georgiou, "A New Method for Moving-Average Parameter Estimation", in *Proc. of the IEEE Signals, Systems and Computers (ASILOMAR)*, pp 1817-1820 2010.

[7] J. Durbin, "Efficient Estimation of Parameters in Moving-Average Models", *Biometrika*, vol. 46, no. 3/4, pp. 306-316, 1959.

[8] R. Bhansali, "A Simulation Study of Autoregressive and Window Estimators of The Inverse Correlation Function", *Journal of the Royal Statistical Society - Applied Statistics*, vol. 32, no. 2, pp. 141-149, 1983.

[9] C. Byrnes, P. Enqvist, and A. Lindquist, "Cepstral Coefficients, Covariance Lags, and Pole-Zero Models for Finite Data Strings", *IEEE Trans. on Signal Processing*, vol. 49, no. 4, pp. 677-693, 2001.

[10] A. Kaderli and A. Kayhan, "Spectral Estimation of ARMA Processes Using ARMA Cepstrum Recursion", *IEEE Signal Processing Letter*, vol. 7, no. 9, pp. 259-261, 2000.

[11] G. B. Giannakis and J. M. Mendel, "Identification of Nonminimum Phase Systems Using Higher Order Statistics", *IEEE Trans. Acoustics, Speech, and Signal Process.*, vol. 37, no. 3, pp. 360-377, 1989.

[12] J. M. Mendel, "Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications", *IEEE Trans. Acoustics, Speech, and Signal Process.*, vol. 79, no. 3, pp. 278-305, 1991.

[13] K. K. Paliwal, "A Noise-Compensated Long Correlation Matching Method for AR Spectral Estimation of Noisy Signals", *Signal Processing*, vol. 15, pp. 437-440, 1988.

[14] W. Bobillet, R. Diversi, E. Grivel, R. Guidorzi, M. Najim, and U. Soverini, "Speech Enhancement Combining Optimal Smoothing and Errors-In-Variables Identification of Noisy AR Processes", *IEEE Trans. on Signal Processing*, vol. 55, no. 12, pp. 5564-5578, 2007.

[15] C. E. Davila, "A Subspace Approach to Estimation of Autoregressive Parameters from Noisy Measurements", *IEEE Trans. Signal Process.*, vol. 46, no. 2, pp. 531-534, 1998.

[16] J. Petitjean, E. Grivel, W. Bobillet, and P. Roussilhe, "Multichannel AR Parameter Estimation From Noisy Observations as an Errors-In-Variables Issue", *Signal, Image and Video Processing (SIVIP)*, vol. 4, no. 9, pp. 209-220, 2010.

[17] W. X. Zheng, "A Least-Squares Based Method for Autoregressive Signals in The Presence of Noise", *IEEE Trans. Circuits and Systems II: Express Briefs.*, vol. 46, no. 1, pp. 81-85, 1999.

[18] H. Ijima and E. Grivel, "Deterministic Regression Methods for Unbiased Estimation of Time-Varying Autoregressive Parameters from Noisy Observations", *Signal Processing*, vol. 92, no. 4, pp.857-871, 2012.

[19] R. Diversi, H. Ijima, and E. Grivel, "Prediction Error Method to Estimate the AR Parameters when the Process is Disturbed by a Colored Noise", in *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013, pp. 6143–6147.

[20] J. K. Tugnait, "New Results on FIR System Identification Using Higher Order Statistics", *IEEE Trans. Signal Processing.*, vol. 39, no. 10, pp. 2216-2221, 1991.

[21] Y. J. Na, K. S. Kim, I. Song, and T. Kim, "Identification of Nonminimum Phase FIR Systems Using the Third and Fourth-Order Cumulants", *IEEE Trans. Signal Processing.*, vol. 43, no. 8, pp. 2018-2022, 1995.

[22] R. Guidorzi, R. Diversi, and U. Soverini, "The Frisch Scheme in Algebraic and Dynamic Identification Problems.", *Kybernetika*, vol. 44, no. 5, pp. 585-616, 2008.

[23] R. Diversi, U. Soverini, and R. Guidorzi, "A New Estimation Approach for AR Models in Presence of Noise", *Proceedings of the 16th IFAC World Congress*, pp. 160–165, July 2005.

[24] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*, Wiley-interscience, 2001.