Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Direction change detection from noisy data and application to mud logging data processing

Qinghua Zhang, Nicolas Fréchin and Nicolas Guezé

***Abstract*— Motivated by mud logging data processing in petrol exploitation, the detection of changes in data line direction is studied in this paper. Because of noise corruption to all measured variables, the classical regression model is not suitable to this problem. After the formulation of an appropriate model, the problem of noise covariance matrix estimation is first considered, then a particular generalized likelihood ratio test is derived for direction change detection. Examples of its application to mud logging data processing are presented for illustration.**

***Index Terms*— direction change detection, generalized likelihood ratio test, covariance estimation, singular value decomposition.**

I. INTRODUCTION

Given a set of vector-valued data samples resulting from measurements on the coordinates of points along a straight line, the main purpose of this paper is to propose a method for the detection of changes in the direction of the straight line. Due to the presence of noises in the coordinates measurements, the non trivial detection problem is addressed in a statistical framework.

This study has been directly motivated by the processing of mud logging data recorded during oil drilling: the data line direction changes are related to geological layer changes encountered during the drilling procedure. See Section IV-B for more details about mud-logging data processing. Nevertheless, the method presented in this paper is general enough to be applicable to other problems of direction change detection, *e.g.*, in navigation direction monitoring.

Let $x = [x_1, \ldots, x_n]^T$ be the coordinates of a point in $\mathbb{R}^n$. A straight line along the direction of a vector $\theta \in \mathbb{R}^n$ is characterized by the equation

$$x = t\,\theta \tag{1}$$

with an auxiliary free variable in $t \in \mathbb{R}$. When t goes through $\mathbb{R}$, the point $x \in \mathbb{R}^n$ moves along the straight line in the direction of θ. Notice that such a line necessarily goes through the origin of the coordinate system.

Now assume that some vector-valued data $y_k \in \mathbb{R}^n$, corresponding to noise corrupted measurements of the coordinates of a point belonging to a straight line, satisfy the equation

$$y_k = \beta_k \theta + e_k \tag{2}$$

with some *unknown* value $\beta_k \in \mathbb{R}$ and $e_k \in \mathbb{R}^n$ representing measurement noises. Different data samples are indexed by $k = 1, \ldots, N$. In order to remove the scale ambiguity between β_k and θ, it is assumed that

$$\theta^T \theta = 1 \tag{3}$$

REMARK. This straight line model is different from the classical linear regression model, due to the *unknown* nature of the value β_k at the right hand side of equation (2) (which would be the regressor in a regression model). Though there exists a (noise perturbed) linear equation between the components of y_k, these components are all perturbed by noises. In contrast, in a regression model, the regressors (inputs) are assumed to be noise-free, only the output is subject to noise. □

The only available data are y_k for $k = 1, \ldots, N$. The noise vector e_k is assumed to be Gaussian, independent between different data samples indexed by k. The covariance matrix of e_k is unknown and independent of k.

For the detection of changes in the direction characterized by θ, knowledge about the covariance matrix of e_k is needed. Therefore, the first problem considered in this paper is the estimation of the noise covariance matrix which is closely related to the estimation of the line direction vector θ. At this point, the reader may ask if the detection of changes in θ can be simply performed by comparing estimated values of θ. Such an approach would require the assessment of the uncertainty in the estimation of θ. Unfortunately, this assessment is a very difficult problem in the framework formulated above. Only approximations under some particular assumption (for instance, small noise assumption) can be reasonably computed. In terms of decision errors, the statistical detection method presented in this paper performs better.

The main detection tool used in this paper is the generalized likelihood ratio (GLR) test [1], [2]. For a typical application of the GLR test, the unknown noise covariance matrix should be considered as a nuisance parameter. However, its implementation would require the numerical solution of a constrained non convex optimization problem. For this reason, in this paper, the covariance matrix is first estimated and then considered as known in the implementation of the GLR test. This method is not optimal, but has two advantages. First, the resulting test has a simple computation form and is numerically efficient. Second, the

Qinghua Zhang is with INRIA-IRISA, Campus de Beaulieu, 35042 Rennes Cedex, France. zhang@irisa.fr

Nicolas Fréchin is with Geoservices, Z.I. du Coudray, 93150 Le Blanc-Mesnil, France. Nicolas.Frechin@ geoservices.com

Nicolas Guezé was with Geoservices during the period of the reported work.

0-7803-9568-9/05/$20.00 ©2005 IEEE

designed statistic has a simple known theoretic distribution which provides a useful guideline for threshold tuning following a desired false detection rate.

Let us mention that model (2) is similar to the one considered in the total least squares (TLS) problem [3], [4] which assumes a linear relationship with both inputs and output corrupted by noises. Usually in the TLS problem the output variable is distinguished from input variables, whereas no such distinction is made in model (2). The basic TLS problem considers a single scalar equation (corresponding to a hyper-plane in $\mathbb{R}^n$), whereas model (2) is a vectorial equation (corresponding to a straight line in $\mathbb{R}^n$). Another related topic is system identification with errors-in-variable models which mainly considers the problem of model estimation [5].

The problem of noise covariance estimation is considered in Section 2. The detection of direction change based on the generalized likelihood ratio (GLR) test is presented in Section 3. Numerical results obtained with simulated data and with real data (mud logging data) will be presented in Section 4. Finally, some concluding remarks are drawn in Section 5.

II. Noise covariance matrix estimation

Let us pack the data samples $y_k \in \mathbb{R}^n$ (column vectors) for $k = 1, \dots, N$ into a matrix $Y \in \mathbb{R}^{N\times n}$. Similarly, pack the noises e_k into a matrix $E \in \mathbb{R}^{N\times n}$, and the scalar values β_k into a column vector $\beta \in \mathbb{R}^N$:

$$Y \triangleq \begin{bmatrix} y_1^T \\ y_2^T \\ \vdots \\ y_N^T \end{bmatrix} \quad E \triangleq \begin{bmatrix} e_1^T \\ e_2^T \\ \vdots \\ e_N^T \end{bmatrix} \quad \beta \triangleq \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix}$$

With these notations the assumed model is rewritten as

$$Y = \beta\theta^T + E \tag{4}$$

The noises are assumed independent between the rows of E. Each row of E, that is e_k, has an *unknown* covariance matrix Σ independent of k. The log-likelihood function for the data matrix Y then writes

$$\log p(\theta,\beta,\Sigma,Y) = -\frac{N}{2}\log[(2\pi)^n \det\Sigma] - \frac{1}{2}\text{trace}[(Y-\beta\theta^T)\Sigma^{-1}(Y-\beta\theta^T)^T] \tag{5}$$

One may try to estimate Σ by maximizing the log-likelihood with respect to Σ, θ and β. Unfortunately, this estimation is impossible due to the singular property stated in the following proposition.

Proposition 1: For any fixed value of θ and some value of β depending on θ and Σ, the log-likelihood function $\log p(\theta,\beta,\Sigma,Y)$ tends to $+\infty$ if any of the eigenvalues of Σ, associated to an eigenvector non orthogonal to θ, tends to zero, while the other eigenvalues remain bounded between two positive constants. □

A proof of this proposition can be found in the appendix of this paper. It indicates that, under the above formulated assumptions, the data Y is not sufficient for the estimation of Σ. In order to solve this illposed problem, one possibility is to introduce constraints on the parameters to be estimated. Let us choose the constraint

$$\theta^T\Sigma^{-1}\theta = \alpha \tag{6}$$

with some positive constant α. The choice for the value of α will be discussed later at the end of this section. Notice that an alternative approach is to add an penalty term into the log-likelihood function. In particular, if the penalty term is proportional to trace(Σ^{-1}), then the result is similar to the following one.

Proposition 2: Given a data matrix $Y \in \mathbb{R}^{N\times n}$ with full column rank and distinct largest singular value, the arguments $\hat\theta, \hat\beta, \hat\Sigma$ maximizing the log likelihood function (5) under the constraints (3) and (6) is uniquely determined as follows. Let the singular value decomposition (SVD) of Y be

$$Y = \sum_{i=1}^{n} s_i\, u_i\, v_i^T$$

with s_1 the largest singular value of Y. Then

$$\hat\theta = v_1 \tag{7}$$

$$\hat\beta = s_1\, u_1 \tag{8}$$

$$\hat\Sigma = \frac{1}{\alpha}v_1 v_1^T + \frac{1}{N}\sum_{i=2}^{n} s_i^2\, v_i\, v_i^T \tag{9}$$

Proof (sketch). To solve the constrained maximization problem, let us first formulate the Lagrangian

$$L = \log p(\theta,\Sigma,\beta,Y) + \frac{1}{2}\lambda(\theta^T\theta - 1) + \frac{1}{2}\mu(\alpha - \theta^T\Sigma^{-1}\theta)$$

where λ and μ are Lagrangian multipliers. The first order optimality condition can then be derived:

$$\begin{aligned}
\frac{\partial L}{\partial\theta} &= \beta^T Y\Sigma^{-1} - \beta^T\beta\theta^T\Sigma^{-1} + \lambda\theta^T - \mu\hat\Sigma\theta^T\Sigma^{-1} = 0\\
\frac{\partial L}{\partial\beta} &= \theta^T\Sigma^{-1}Y^T - \theta^T\Sigma^{-1}\theta\beta^T = 0\\
\frac{\partial L}{\partial\Sigma} &= -\frac{N}{2}\Sigma^{-1} + \frac{1}{2}\Sigma^{-1}(Y-\beta\theta^T)^T(Y-\beta\theta^T)\Sigma^{-1}\\
&\quad + \frac{1}{2}\mu\Sigma^{-1}\theta\theta^T\Sigma^{-1} = 0\\
\theta^T\theta &= 0\\
\theta^T\Sigma^{-1}\theta &= \alpha
\end{aligned}$$

The solutions to these 5 equations can be computed by the SVD of Y in the following way. Let the SVD of Y be

$$Y = \sum_{i=1}^{n} s_i\, u_i\, v_i^T$$

For the moment assume that the singular values s_i are arranged in an arbitrary order. Then it is straightforward to verify that the solution

$$\begin{aligned}
\hat{\theta} &= v_1 \\
\hat{\beta} &= s_1 u_1 \\
\hat{\Sigma} &= \frac{1}{\alpha} v_1 v_1^T + \frac{1}{N} \sum_{i=2}^{n} s_i^2 v_i v_i^T \\
\lambda &= N \\
\mu &= \frac{N}{\alpha}
\end{aligned}$$

satisfy the above 5 conditions. Some computations then yield the value of the log-likelihood function

$$\log p(\hat{\theta}, \hat{\beta}, \hat{\Sigma}, Y) = -\frac{N}{2} \log\left((2\pi)^n \frac{s_2^2 \cdots s_n^2}{N^{n-1}\alpha} \right) - \frac{N}{2}$$

In order to maximize this value, $s_2, \ldots, s_n$ should be the $n-1$ smallest singular values of Y, then s_1 is the largest singular value of Y, as stated in Proposition 2. □

The above result has been presented mainly for the purpose of noise covariance estimation, though it also provides an estimation for the other parameters involved in the model. Notice that, among the three estimates, only $\hat{\Sigma}$ depends on the choice of α.

Now let us discuss on the choice for the value of α. In the estimation $\hat{\Sigma}$ given by equation (9), the computed quantities s_i, v_i for $i = 1, \ldots, n$ are independent of α. Proposition 1 states that, when constraint (6) is *not* considered, the maximization of the likelihood function leads to solutions with singular covariance matrix. The introduction of the constraint thus avoids singular covariance matrix. It is thus clear that the term in equation (9) involving α plays the role of regularization in the covariance estimation. This term should be relatively small compared to the other terms. A large value of α will make the regularization term small, but leads to an almost singular (ill-conditioned) $\hat{\Sigma}$. A quantitative criterion, the matrix condition number (the ratio between the largest and the smallest singular values), should be considered at the place of singularity property. It is therefore recommended to choose the value of α corresponding to the best matrix condition number of $\hat{\Sigma}$ and to the smallest regularization term, as given by

$$\alpha = \frac{N}{s_n^2}$$

with s_n being the smallest singular value of Y.

III. Direction change detection

Now we are given two data matrices $\tilde{Y}_1$ and $\tilde{Y}_2$, each being modeled in the same way as Y in the previous sections. The problem is then to decide if the two straight lines behind the two data matrices are along the same direction or not. It is assumed that there is a natural way to delimit the two sets of data forming the matrices $\tilde{Y}_1$ and $\tilde{Y}_2$, which is indeed the case of mud logging data processing reported in Section IV-B.

Typically in parametric change detection, a model is first estimated with one data set and then a statistical test is applied to check the validity of the model for the other data set. This approach treats the two data sets in an asymmetric manner. Since there is no natural reason to choose one of the two data sets for model estimation, the method proposed below treats the two data sets *symmetrically*.

It is assumed that $\tilde{Y}_1$ and $\tilde{Y}_2$ have the same, but unknown, noise covariance matrix Σ. Let us first estimate this covariance matrix and use it to normalize the data $\tilde{Y}_1$ and $\tilde{Y}_2$. The covariance matrix is not considered as a nuisance parameter in the design of the statistical test presented below. As already explained in the introduction, this choice has been made in order to achieve a numerically efficient algorithm and also to easily evaluate the probability of false detection associated to the designed statistical test.

Assume that $\hat{\Sigma}_1$ is the noise covariance estimated from $\tilde{Y}_1$, with the method presented in the previous section. Similarly $\hat{\Sigma}_2$ is estimated from $\tilde{Y}_2$. Let N_1 and N_2 be respectively the numbers of rows in $\tilde{Y}_1$ and $\tilde{Y}_2$. Then the overall noise covariance estimation is given by the weighted average

$$\hat{\Sigma} = \frac{1}{N_1 + N_2} (N_1 \hat{\Sigma}_1 + N_2 \hat{\Sigma}_2)$$

This estimation has been chosen for simplicity. The optimal estimation jointly based on $\tilde{Y}_1, \tilde{Y}_2$ would requires the solution of a non convex optimization problem.

Now the data matrices are normalized as follows

$$Y_1 = \tilde{Y}_1 \hat{\Sigma}^{-\frac{1}{2}} \tag{10}$$

$$Y_2 = \tilde{Y}_2 \hat{\Sigma}^{-\frac{1}{2}} \tag{11}$$

By neglecting the estimation error of $\hat{\Sigma}$, it is assumed in the following that the normalized data Y_1, Y_2 have a unitary noise covariance matrix.

From now on, assume that

$$\begin{aligned}
Y_1 &= \beta_1 \theta_1^T + E_1 \\
Y_2 &= \beta_2 \theta_2^T + E_2
\end{aligned}$$

where the noise matrices E_1, E_2 have independent rows, each row being Gaussian with a unitary covariance matrix.

The problem of direction change detection is then formulated as the decision between the two hypotheses

$$\begin{aligned}
H_0 &: \theta_1 = \theta_2 \\
H_1 &: \theta_1 \neq \theta_2
\end{aligned}$$

The generalized likelihood ratio (GLR) test [1], [2] is chosen for this purpose for its nice asymptotic property under hypothesis H_0. Define the notations

$$Y_{12} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \qquad \beta_{12} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

The statistic for the GLR test then writes

$$g(Y_1, Y_2) = 2\log \frac{\max_{\theta_1^T\theta_1=1,\ \beta_1} p(\theta_1, \beta_1, Y_1) \cdot \max_{\theta_2^T\theta_2=1,\ \beta_2} p(\theta_2, \beta_2, Y_2)}{\max_{\theta^T\theta=1,\ \beta_{12}} p(\theta, \beta_{12}, Y_{12})}$$

where the probability density function $p(\cdot)$ is similar to the one in (5), except that now the noise covariances are identity matrices and are omitted from the notations. The computation of $g(Y_1, Y_2)$ is summarized in the following proposition.

Proposition 3: Let $s_{(1)}, s_{(2)}$ and $s_{(12)}$ be respectively the *largest* singular values of Y_1, Y_2 and Y_{12}. Then

$$g(Y_1, Y_2) = s_{(1)}^2 + s_{(2)}^2 - s_{(12)}^2 \tag{12}$$

PROOF (sketch). Following the form of $p(\cdot)$ as shown in equation (5) (with identity Σ), for any fixed values of normalized vectors $\theta_1, \theta_2, \theta$ (satisfying $\theta_1^T\theta_1 = \theta_2^T\theta_2 = \theta^T\theta = 1$), maximize each $p(\cdot)$ with respect to β_1, β_2 or β_{12}. The results lead to

$$g(Y_1, Y_2) = \max_{\theta_1^T\theta_1=1} \theta_1^T Y_1^T Y_1 \theta_1 + \max_{\theta_2^T\theta_2=1} \theta_2^T Y_2^T Y_2 \theta_2 - \max_{\theta^T\theta=1} \theta^T Y_{12}^T Y_{12} \theta$$

The remaining computation requires the solution of three similar constrained maximization. For example, consider the first term. It is well known that the solution is given by θ_1 equal to the right singular vector $v_{(1)}$ of Y_1 associated to the largest singular value $s_{(1)}$, or equivalently, $v_{(1)}$ is the eigenvector of $Y_1^T Y_1$ associated to the largest eigenvalue (well known result in principal component analysis (PCA), see [6]). Then

$$\max_{\theta_1^T\theta_1=1} \theta_1^T Y_1^T Y_1 \theta_1 = v_{(1)}^T Y_1^T Y_1 v_{(1)} = s_{(1)}^2$$

The other two maximization problems are similarly solved, leading to $g(Y_1, Y_2) = s_{(1)}^2 + s_{(2)}^2 - s_{(12)}^2$. □

The computation of $g(Y_1, Y_2)$ is then based on the numerically well developed SVD algorithm [7]. Another advantage is that, asymptotically for large N_1 and N_2, the statistic $g(Y_1, Y_2)$ under H_0 follows the central χ^2 distribution, with the number of degrees of freedom equal to the difference of degrees of freedom in the parameters between H_0 and H_1, namely $n-1$. This knowledge provides a useful guideline for threshold tuning, following a desired false detection probability.

Let λ_t be the chosen threshold for the GLR test, then the absence of direction change (H_0) is decided if $g(Y_1, Y_2) < \lambda_t$. Otherwise the presence of direction change (H_1) is decided. The probability of false detection (H_1 is decided whereas H_0 is true) is given by

$$P_{FD} = 1 - \int_0^{\lambda_t} \chi_{n-1}^2(x)dx \tag{13}$$

where $\chi_{n-1}^2(x)$ is the χ^2 density function with $n-1$ degrees of freedom.

IV. NUMERICAL RESULTS

The GLR test presented in this paper is first applied to some simulated data, then to the mud logging data.

A. Simulation examples

The purpose of the simulations is to verify the statistical behavior of the particular GLR test presented in this paper. The simulations are made with the dimension $n = 7$. For each data matrix $\tilde{Y}_1$ or $\tilde{Y}_2$, the number of rows $N_1 = N_2 = 100$. The noise covariance matrix Σ is a randomly drawn positive diagonal matrix.

Simulations under H_0 are first made with $\theta_1 = \theta_2$. For each pair of generated data matrices $\tilde{Y}_1$ and $\tilde{Y}_2$, the noise covariance matrix is first estimated and the data matrices are normalized according to (10) and (11). The statistic $g(Y_1, Y_2)$ is then computed following (12). The simulation is repeated 10000 times with different random realizations of the data matrices. The distribution of the statistic over these 10000 experiments is illustrated by the histogram in figure 1. The theoretic χ^2 probability density function with 6 degrees of freedom is also plotted in figure 1. As expected, the form of the histogram is quite similar to the density function curve.

Simulations under H_1 are then made, with an angle of 5 degrees between θ_1 and θ_2 for each pair of generated data matrices $\tilde{Y}_1$ and $\tilde{Y}_2$. The simulation is also repeated 10000 times with different random realizations. The resulting histogram is shown in figure 2. When this histogram is compared with that of figure 1, the difference in the scales of their abscissas (40 and 150) should be noticed. When the threshold $\lambda_t = 12.59$ is chosen for the GLR test, the probability of false detection (H_1 is decided whereas H_0 is true) as computed according to (13) is equal to 0.05. With the same threshold, the probability of mis-detection (H_0 is decided whereas H_1 is true) is 0.0009 according to the simulations.

B. Application to mud logging data processing

During oil drilling, the drilling mud (mixture of water, clay, weighting material and chemicals) is injected through the drill line and then extracted from the drilled hole to lift rock cuttings from the drill bit to the surface. The extracted mud is analyzed for the purposes of

- monitoring the integrity of the well,
- optimizing the drilling procedure, and
- early detection and analysis of hydrocarbons.

In particular, the gas components contained in the drilling mud are analyzed. The recorded data on these gas components corresponding to mud extracted from different depth is called *mud logging data*. These data follow the straight line model (2) inside a relatively homogeneous geological layer. One of the tasks of mud logging data processing is

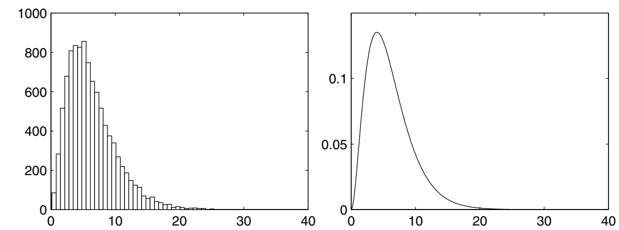

Fig. 1. Histogram of $g(Y_1, Y_2)$ under H_0 (left) and the χ^2 probability density function with 6 degrees of freedom (right).

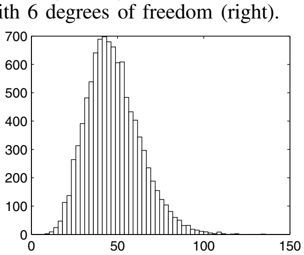

Fig. 2. Histogram of $g(Y_1, Y_2)$ under H_1.

to detect changes of geological layer. Usually this work is done manually by visualizing data points corresponding to a pair of gas components recorded in the mud logging data. Inside a geological layer, these data points form a noise corrupted straight line. Geological layer changes are detected through visual inspection of the data line change. The visual inspection is time consuming and the result is quite subjective. Since the detection is based on a small number of views of the mud logging data corresponding to some chosen pairs of gas components, the information contained in the data is not efficiently exploited.

The variation in the concentration of the gas components contained in the extracted mud allows to simply segment the mud logging data. However, the data is often over-segmented with respect to geological layer changes: two successive segments of the data may correspond to the same geological layer. The statistical change detection method presented in this paper is then applied to such data segments for geological layer change detection. The algorithm has been integrated into a software used by Geoservices for mud logging data processing. Below some examples are given for illustration.

Typically the extracted mud are analyzed for more than 10 gas components. In order to make illustrations with two dimensional plots, the examples shown below are based on only two of the analyzed gas components in the mud logging data. In theory, for a probability of false detection equal to 0.01, the threshold for the GLR test should be $\lambda_t = 6.635$, according to the χ^2 table. Of course, the algorithm implemented in the software of Geoservices uses more gas components in order to fully exploit the information contained in the mud logging data.

In figure 3-(a) are plotted two data segments, one represented by the crosses, the other by the circles. The two straight lines fitted to these two data segments (following proposition 2) are drawn respectively by a solid line and a

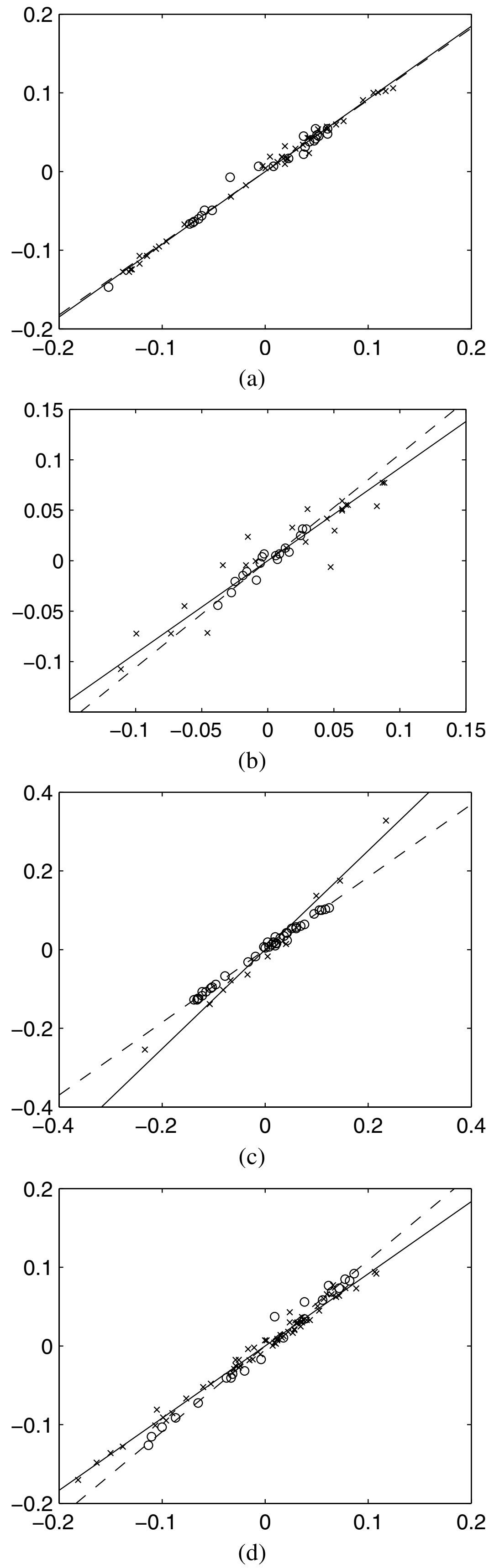

Fig. 3. Examples of mud logging data

dashed line (in this example the two lines are too close to be distinguishable). The statistic value $g(Y_1, Y_2) = 0.2701$ is obtained. It is lower than the threshold $\lambda_t = 6.635$ and it is thus decided that the two segments belong to the same geological layer.

In figure 3-(b) another pair of data segments are illustrated together with their fitted straight lines. The angle between the two lines is larger than in the previous example, but the noise level is also higher. It is decided that the two segments belong to the same geological layer, since the statistic $g(Y_1, Y_2) = 0.5737$ is computed.

For the data of figure 3-(c), the result is $g(Y_1, Y_2) = 98.7239$, clearly over $\lambda_t = 6.635$. A layer change is thus detected.

For the last example shown in figure 3-(d), the value of $g(Y_1, Y_2) = 41.1973$ also allows to detect a layer change. Notice that the angle between the two fitted lines in this example is similar to the angle in figure 3-(b). It is mainly the difference in the noise levels of the two examples that has led to the different decisions. With these examples, it would be difficult to make decisions based on the estimated angles between the fitted lines.

V. Conclusion

A method has been presented in this paper for the detection of direction changes in noise corrupted data line. The main detection tool is the generalized likelihood ratio test. The computation of the test is mainly based on the numerically well developed SVD algorithm. Though this work has been directly motivated by mud logging data processing, it can also be applied to other direction change detection problems.

Appendix: proof of Proposition 1

Let us first maximize $\log p(\theta, \beta, \Sigma, Y)$ with respect to β, for any given values of θ and Σ. Denote by $\hat{\beta}$ the corresponding value of β, it is then straightforward to obtain

$$\begin{aligned}\log p(\theta, \hat{\beta}, \Sigma, Y) = &-\frac{N}{2} \log[(2\pi)^n \det \Sigma] \\ &- \frac{1}{2} \text{trace}(Y \Sigma^{-1} Y^T) \\ &+ \frac{1}{2} \frac{1}{\theta^T \Sigma^{-1} \theta} \theta^T \Sigma^{-1} Y^T Y \Sigma^{-1} \theta \quad (14)\end{aligned}$$

Let the spectral decomposition of Σ be

$$\Sigma = V \Lambda V^T$$

with $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ and $V = [v_1, \ldots, v_n]$. Without loss of generality, let us consider the case $\lambda_1 \to 0$ and assume that the associated eigenvector is not orthogonal to θ, i.e., $v_1^T \theta \neq 0$.

It will be shown that

$$\lim_{\lambda_1 \to 0} p(\theta, \hat{\beta}, \Sigma, Y) = +\infty$$

Remark that there is *no* need to assume $[v_2, \ldots, v_n]^T \theta = 0$ in the following reasoning (though it would have implied useful simplifications).

Let us start by observing that

$$\begin{aligned}\theta^T \Sigma^{-1} \theta &= \theta^T V \Lambda^{-1} V^T \theta \\ &= \sum_{i=1}^{n} \lambda_i^{-1} (\theta^T v_i)^2\end{aligned}$$

and also

$$\begin{aligned}\theta^T \Sigma^{-1} Y^T Y \Sigma^{-1} \theta &= \theta^T V \Lambda^{-1} V^T Y^T Y V \Lambda^{-1} V^T \theta \\ &= \left(\sum_{i=1}^{n} \lambda_i^{-1} \theta^T v_i v_i^T \right) (Y^T Y) \left(\sum_{i=1}^{n} \lambda_i^{-1} v_i v_i^T \theta \right)\end{aligned}$$

Therefore, when $\lambda_1 \to 0$ (remind that $v_1^T \theta \neq 0$ and that the other eigenvalues remain bounded between two positive constants),

$$\frac{1}{\theta^T \Sigma^{-1} \theta} \theta^T \Sigma^{-1} Y^T Y \Sigma^{-1} \theta = \lambda_1^{-1} v_1^T Y^T Y v_1 + O(1)$$

Now

$$\begin{aligned}\text{trace}(Y \Sigma^{-1} Y^T) &= \text{trace}(Y V \Lambda^{-1} V^T Y^T) \\ &= \text{trace}(\Lambda^{-1} V^T Y^T Y V) \\ &= \sum_{i=1}^{n} \lambda_i^{-1} v_i^T Y^T Y v_i\end{aligned}$$

and

$$\log \det \Sigma = \sum_{i=1}^{n} \log \lambda_i$$

Then, when $\lambda_1 \to 0$,

$$\begin{aligned}\log p(\theta, \hat{\beta}, \Sigma, Y) &= -\frac{N}{2} \log \lambda_1 - \frac{1}{2} \lambda_1^{-1} v_1^T Y^T Y v_1 \\ &\quad + \frac{1}{2} \lambda_1^{-1} v_1^T Y^T Y v_1 + O(1) \\ &= -\frac{N}{2} \log \lambda_1 + O(1)\end{aligned}$$

that will indeed tend to $+\infty$.

References

[1] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes – Theory and Applications*, ser. Information and System Sciences Serie. Englewood Cliffs, N.J: Prentice Hall, 1993.

[2] A. W. Van der Vaart, *Asymptotic statistics*. Cambridge: Cambridge University Press, 1998.

[3] S. V. Huffel and J. Vandewalle, *the Total least squares problem – computational aspects and analysis*. Philadelphia, Pennsylvania: S.I.A.M., 1991.

[4] A. Björck, *Numerical methods for Least Squares Problems*. Linkoping, Sweden: Society for Industrial and Applied Mathematics, 1996.

[5] T. Sderstrm, U. Soverini, and K. Mahata, “Perspectives on errors-in-variables estimation for dynamic systems,” *Signal Processing*, vol. 82, no. 8, pp. 1139–1154, 2002.

[6] I. T. Jolliffe, *Principal Component Analysis*. New York: Springer-Verlag, 2002, 2nd edition.

[7] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, London: The Johns Hopkins University Press, 1996, third edition.