STABLE ALGORITHM FOR RECONSTRUCTING THE SHAPE OF A PULSED LIDAR RETURN SIGNAL

Yu.F. Dem'yanov

Received July 29, 1993

A method for reconstructing the shape of a pulsed lidar return signal in the case the law of error distribution differs from the normal distribution law is proposed. A stable algorithm of the reconstruction is presented and its investigation using Monte Carlo method is discussed. A theorem is presented that allows, for a particular case of estimating the shift parameter, assessment of optimal properties of the algorithm to be made.

Information about the shape of reflected lidar pulse is important in the case of remote determining the different parameters of the atmosphere by means of a pulsed lidar. The information about distortion of the regular component of the sounding pulse makes it possible to obtain the concentration and particle spectrum of atmospheric aerosol, volume coefficients of cloud scattering, and some other parameters.[1] However, known methods for processing of the reflected signals based on linear filtration principles can be efficient under conditions of explicit or implicit assumptions about the form (normal view, in particular) of the laws of error distribution. Practically, because of the influence of noise of different origin, for example, monitor noise, errors of the channels of digital conversion and information transfer, and others, these laws differ from the normal distribution law.

Deficit of information as well as the character of the statistical criteria for check of distribution laws for adequacy and certain probability of the presence of rough errors (outliers) does not allow elucidation of the formulation of the law of its distribution. Thus, use of known methods can result in considerable distorsion of the results of the interpretation of the signal shape measurement.[2–4] There is a problem to develop such methods of the measurement processing that would be insensitive to the formulation of the distribution law and, in particular, poorly responsive to outliers.

Let us assume that portions of observation are limited in duration, and the regular signal compound is limited in the first and the second derivatives. In this case, the regular signal compound can be considered as polynomial, because, according to Wieirstrass theorem, any continuous on the closed segment function can be approximated by polynomial within any given accuracy. The linear regression on polynomial can be chosen as the stochastic mathematical model of the problem solution.

Let polynomial describing the regular signal compound be as follows

$$F(t) = C_0 + C_1 t + \ldots + C_{p-1} t^{p-1},$$

where $p-1$ is the degree of polynomial (given); $C_{p-1} \neq 0$, $0 \le t \le T$; T is the signal duration. It is necessary to determine unknown coefficients C_j, $j = 1, 2, \ldots, p-1$ from the results of b_i measurements in the moments of time t_i, $i = 1, 2, \ldots, n$. It is seen from Ref. 5 that formulated problem reduces to three mathematical models in the form of the systems of linear algebraic equations, through the solution of which we determine required coefficients.

1. Stochastic model. In this case the regression coefficients are determined as a result of solution of the stochastic system of linear algebraic equations (SLAE) of the form

$$A\,x = b, \tag{1}$$

where $A = (a_{ij})$ is the $n \times p$ matrix, $x = (x_j)$ is the $1 \times p$ column, and $b = (b_i)$ is the $1 \times n$ column. Elements a_{ij} and b_i of the system are the random variables with the unknown law of distribution. Further, we assume that $n \ge p$ and

$$\text{rang}\, A = p \tag{2}$$

is almost correct.

2. Determinate model. This is determinate SLAE of the form

$$\overline{A}\,\overline{x} = \overline{b}, \tag{3}$$

where elements $(\overline{a}_{ij})$ and $(\overline{b}_{ij})$ are known precisely. This most theoretically developed model is not considered here because there are no measurements without errors.

3. Intermediate model. This is stochastic SLAE of the form

$$\overline{A}\,x = b, \tag{4}$$

where $\overline{A} = (\overline{a}_{ij})$, $\overline{a}_{ij} = \overline{t}_i^{\,j-1}$, $\overline{t}_i$ are the moments of measurements known precisely, $b_i = \overline{b}_i + \Delta b_i$, with the unknown law of signal measurement error Δb_i distribution.

Let $\overline{b}_{i+1} - \overline{b}_i = \overline{\theta}$ $(i = 1, \ldots, n-1)$. If the number of equations is large then it is possible to obtain the system (5) using sequential mutual subtraction.

$$\overline{D}\,y = e, \tag{5}$$

where

$$\overline{D} = (d_{2l-1,j}) = (\overline{a}_{2l,j} - \overline{a}_{2l-1,j});$$

$$l = 1, \ldots, [n/2];\ j = 1, \ldots, p-1;\ y_1 = x_2, \ldots, y_{p-1} = x_p;$$

$$e = (e_{2l-1}) = (\overline{q} + \Delta e_{2l-1});\ \Delta e_{2l-1} = \Delta b_{2l} - \Delta b_{2l-1}.$$

0235-6880/94/09 677-03 $02.00

Unknown x_1 can be found after unknown x_2, ..., x_p from the additional equation.

Estimation of the correctness of SLAE solution by means of algorithm proposed below was carried out with the help of Monte Carlo method, and realized during model estimation expression for errors of matrix definition (right–hand part) of the systems (1) and (or) (4) is of the form

$$\Delta t(\Delta b) = \gamma_{a(b)} + \delta_{a(b)}\, \xi_{a(b)}, \qquad (6)$$

where $\gamma_i \in \Phi(0, \sigma_\Phi)$, $\delta_i \in B(\varepsilon)$, $\xi_i \in H(0, \sigma_\text{H})$ are normal, Bernully, and symmetrical unimodal distributions, respectively;

$$E(\Delta t_i) = E(\Delta b_i) = 0;\ \ E(\Delta t_i \Delta b_j) = E(\Delta t_i)\, E(\Delta b_j);$$
$$E(\Delta t_i \Delta t_j) = E(\Delta t_i)\, E(\Delta t_j);\ \ E(\Delta b_i \Delta b_j) = E(\Delta b_i)\, E(\Delta b_j),$$
$$i \neq j,\ \ \varepsilon \in [0;\ 0.5],$$

where E is mathematical expectation (mean value).

Let us consider the next (referred to as separating) algorithm of solution of problem (4).

A1. Form the right–hand part of so called "base" system, or $\mathbf{b}_\text{M}$ vector. This is the vector the components of which are order statistics $b(r)$, $b(r+l)$, ..., $b(s)$ of $\mathbf{b}$ vector. Quantities r and s are determined from the equation (6)

$$\pi\left(r, n-r+1, n, \frac{1}{2}\right) = 2\, I_{1/2}(\, r, n-r+1) - 1 = 2^{-n} \sum_{i=r}^{n-r} C_n^i,$$

where $\pi\left(r, s, n, \frac{1}{2}\right)$ is the confidence coefficient; $s = n - r + 1$; n is the size of $\mathbf{b}$ sampling; $I_{1/2}$ is incomplete β–function.

Quantities r and s virtually are nonparametric confidence boundaries with $\alpha_0 = 1 - \pi(\cdot)$ regression level for median of independent identically distributed random quantities corresponding to sampling values b_i, $i = 1, \ldots, n$.

Further we form the matrix of the base system

$$\overline{A}_\text{M} = \{\overline{a}_{ij}\},$$

where i is the row number of initial matrix $\overline{A}$, corresponding to k element of $\mathbf{b}_M$, $j = 1, \ldots, p$; $k = r, r+1, \ldots, s$.

Let us assume that we know precisely the degree of approximating polynomial p and number of equations n.

A2. Estimate the components vector of solution $\hat{\mathbf{x}}_M$ by least–square method (LSM).

A3. Calculate the estimation of rms deviation of model distribution law σ_Φ

$$\hat{S}_\Phi = \frac{1}{s-r-p} (\mathbf{b}_\text{M} - \overline{A}_\text{M} \hat{\mathbf{x}}_\text{M})^\text{T} (\mathbf{b}_\text{M} - \overline{A}_\text{M} \hat{\mathbf{x}}_\text{M}),$$

as well as $100(1-\alpha)$–per cent confidence interval for predicted value of response b_*, according to the expression[7]

$$\hat{\Delta} b_{*2} = t_{s-r-p} \hat{S}_\Phi (\upsilon_* + 1)^{1/2},$$

where t_{s-r-p} is the top $100(\alpha/2)$–per cent point of distribution of t_{s-r-p}, that is

$$P\left\{\frac{\overline{b}_* - b_*}{\hat{S}_\Phi (\upsilon_* + 1)^{1/2}} > t^{\alpha/2}_{s-r-p}\right\} = \alpha/2;$$

$$\upsilon_* = \overline{a}^\text{T}_* (\overline{A}_\text{M}^\text{T} \overline{A}_\text{M})^{-1} \overline{a}_*;\quad \overline{a}_* = (1, \overline{t}_*, \overline{t}^2_*, \ldots, \overline{t}^p_*),$$

asterisk denotes the raw number of initial matrix $\overline{A}$ corresponding to the order $b_{(r-i)}$ or $b_{(s+i)}$ statistics (left–hand and right–hand parts of the interval are taken alternatively; $i = 1, 2, \ldots, r-1$ for the left–hand part and $i = 1, 2, \ldots, n-s$ for the right–hand part). Further, i value changes by one at the fourth step of the algorithm. If σ^2_Φ value is *a priori* known, then we calculate

$$\hat{\Delta} b_{*1} = \psi(1-\alpha/2)\, \sigma_\Phi\, (\upsilon_* + 1)^{1/2},$$

where $\psi(1-\alpha/2) = N^{-1}(1-\alpha/2;\ 0.1)$.

A4. Check the element $b_{(r-i)}$ or $b_{(s+i)}$ for falling within confidence interval found at the third step of the algorithm. In the case of positive result, $b_{(r-i)}(b_{(s+i)})$ element is included to $\mathbf{b}_\text{M}$ vector composition, $\overline{A}_\text{M}$ matrix is added with a^T_* row, r and s values are changed for $r-i$ and $s+i$ respectively; and change to the second step of the algorithm is performed. In the case of negative result $b_{(r-i)}(b_{(s+i)})$ element is rejected, i increases by one, and return to the third step is performed.

At any result we make checking: $i = r-1$ and $i = n-s$. If these conditions hold, then the algorithm accomplishes its work.

The algorithm is considered principally only for model (4), because for system (1) it can be reduced to sequent application of one and the same algorithm firstly to the matrix and then to right–hand part with following analysis of LSM solution of the system which is made up of a set of superpositions.

In Ref. 5 the version of the problem of estimation of shift parameter (model (5)) has been studied extensively, and the following theorem has been proved.

Theorem. Let ratio (5) hold, and conditions imposed on the errors of definition of the right–hand parts Δe_i of the system (5) are the same as the conditions (6) for the errors of definition of the left–hand parts Δb_i of the system (4). In this case solution of the system (5) by means of the separating algorithm when $\varepsilon \neq 0$, $0 \leq \varepsilon \leq 1/2$ is the most B–robust and the most V–robust value, with the following relations: $\gamma^* = \sqrt{\pi/2}$, $\kappa^* = 2$, $\varepsilon = 1/2$. Moreover, given algorithm determines optimal B– and V–robust estimation of solutions.

If $\varepsilon = 0$, then separating algorithm determines the sole empirical assessment with minimum variance of other ones. Here γ^* is the sensitivity to substantial error, κ^* is the sensitivity to the change of variance, and ε is the threshold point.[8]

Unknown form of the error distribution law (6) at the finite size of sample of variously distributed random values make the theoretical study of solution estimation results impossible. Thus, we developed specific algorithm of Monte Carlo method, which realizes numerical approach to the analysis of the investigation results.

Methods of estimation of accuracy of the signal reconstruction and concrete parameters of the simulation algorithm are presented in Ref. 5. Let us note that signal envelope was approximated by the second–degree polynomial (sampling size of 100 observations), parameter of irregularity scale was two orders greater than rms deviations of the noise, and the number of independent realizations, imitating the signal, was chosen to be equal to 100. The results of estimation of the algorithm accuracy are presented in Table I together with accuracy characteristics of the linear algorithm of the least–square method and well–known robust algorithm of Hewber (RH) obtained for

the same model presented for comparison. The numerator of presented data is the estimation of the mean deviation of the signal shape estimation from the true shape, denominator is the rms deviation estimation.

TABLE I. Results of estimation of accuracy of methods under study.

Method	Model											
	Stochastic						Intermediate					
	0	0.1	0.2	0.3	0.4	0.5	0	0.1	0.2	0.3	0.4	0.5
LSM	1.0 / 9.5	0.02 / 4.9	− 0.05 / 4.1	0.3 / 4.4	− 0.06 / 4.3	0.2 / 4.4	0.1 / 1.0	2 / 50	8 / 65	17 / 87	13 / 107	− 3 / 114
RH	–	–	–	–	–	–	0.2 / 1.03	− 0.03 / 1.1	− 0.3 / 1.6	0.3 / 1.8	RNO*	RNO
RA**	0.7 / 9.9	2.0 / 11.2	1.3 / 11.7	− 1.8 / 13.3	0.8 / 12.3	0.3 / 13.8	0.1 / 1.0	− 0.05 / 1.1	0.08 / 1.3	0.1 / 1.3	0.04 / 1.3	0.08 / 1.5

* Result is not obtained.
** Robust algorithm, suggested by the author.

Analysis of the results of the simulation shows that for both (1) and (4) models and "contamination" severity varying from 0 to 0.5 the suggested stable algorithm has the most favorable characteristics. Slight increase in variance (in comparison with optimal one) of the estimation using suggested algorithm at $\varepsilon = 0.5$ can be explained by a finite size of the processed sample.

At the present we modify the algorithm invoking smoothing splines as the basis functions.

REFERENCES

1. V.E. Zuev, *Laser Conquers the Sky* (North–Siberian Publishing House, Novosibirsk, 1972), 192 pp.
2. G.V. Stogov, A.V. Makshanov, and A.A. Musaev, Zarub. Radioelektron., No. 9, 3–46 (1982).
3. P. Hewber, *Robustness in Statistics* [Russian translation] (Mir, Moscow, 1984), 304 p.
4. A.V. Bronnikov and Yu.E. Voskoboinikov, Autometry, No. 1, 21–26 (1990).
5. Yu.F. Dem'janov, *"Study and development of the stable methods of polynomial regression,"* Candidate's Dissertation in Physical and Mathematical Sciences, Priozersk (1992), 88 pp.
6. G. David, *Order Statistics* [Russian translation] (Nauka, Moscow, 1980), 335 pp.
7. J. Seber, *Linear Regressive Analysis* [Russian translation] (Mir, Moscow, 1980), 456 pp.
8. F. Hampel et al., *Robustness in Statistics (Approach on the Base of the Influence Function)* [Russian translation] (Mir, Moscow, 1989), 512 pp.