Tatra Mt. Math. Publ. **26** (2003), 291–306

TATRA MOUNTAINS
Mathematical Publications

ELASTIC ARM METHODS OF DATA ANALYSIS AS A ROBUST APPROACH

GENNADII A. OSOSKOV

ABSTRACT. One of common problems of experimental physics is the detection of curves, such as charged particle tracks or Cherenkov rings, within a complex pattern of measurements and background produced by various detectors. Methods of elastic tracking (ET) or deformable templates (DT) were developed to satisfy a need in approaches more powerful than conventional tracking methods. It is shown here that both ET and DT methods can be considered as particular cases of a generalized robust approach. A robust method is proposed as a dynamically reweighted iterative least square procedure with a special weight function. Maximum likelihood method is used to optimize this weight function for the given level of contamination. Special study of the joint robust estimate of location and scale parameters provides an effective algorithm to avoid spurious local minima of least square functional. Several examples of effective RM applications are given.

1. Introduction

Advances in theoretical physics in the last decades gave rise to revolution in experimental high energy physics (HEP). Colliders operating in TeV-energy range are in use or going to be used. That means an active increase in the energies of particle beams and in the load of detectors. The facilities and the detection technique drastically changed due to recent achievements in the electronics, computings, programming and the Internet expansion. Two examples of experimental data registered in two of contemporary detectors are depicted in Fig. 1,2 to demonstrate their complex character and peculiarities.

One can see several thousand trajectories (tracks) of charged particles (ions of Pb-Pb collision) produced by the central tracker of the CMS detector [1] in Fig. 1 and Cherenkov radiation photons registered by the RICH detector [2] as numerous rings on a heavy background in Fig. 2.

Problems to be solved are:

(i) curve recognizing within a complex pattern of points;

2000 Mathematics Subject Classification: 65C60, 62F35.
Keywords: robust estimates, elastic arm, deformable template methods.

(ii) curve fitting for track (ring) parameter estimation;

(iii) elaborating fast algorithms to carry out both above items.

These problems could be considered as classical problems of pattern recognition and computer vision, however, taking into account experimental specifics, they must require special approaches.

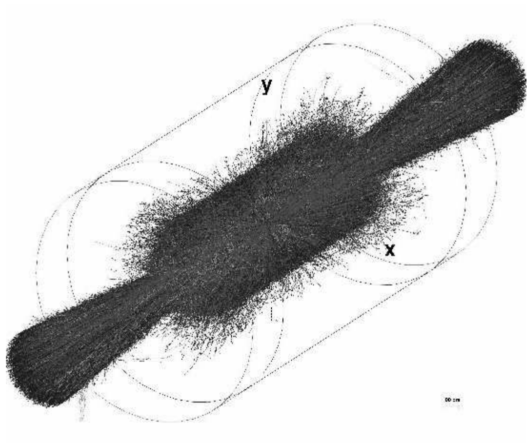

FIGURE 1. Image of a simulated event registered by the CNS detector.

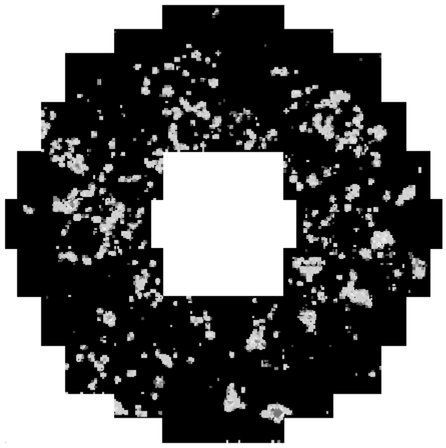

FIGURE 2. Image of a real Au-Pb event registered by the CERES RICH detector.

The following particular features of the HEP data should be emphasized:

- recognized patterns are discrete and have a complex texture, data arrive with the extremely high rate;
- the number of background events, which are similar to "good" events, is larger than the number of the latter events by several orders of magnitude;
- noise counts are numerous and correlated;
- high multiplicity of objects (tracks, Cherenkov radiation rings, electromagnetic showers) is recognized in each event.

The basic requirements to data processing in current experiments are maximum speed of computing due to the mass-production way of data handling in combination with the highest attainable accuracy and high efficiency of methods of estimating physical parameters interesting for experimentalists. The implementation of these requirements under the above conditions inevitably faced the restrictions of traditionally applied classical combinatorial methods, cluster analysis, least squares fitting or even such special track following methods as road finders [3] or artificial neural networks [4], which do not ensure one of these requirements or all of them. Therefore, vital needs for developing new mathematical and algorithmic techniques arose and required the application of new approaches. Methods of elastic tracking or elastic arm (deformable templates) were reported [5, 9] as more powerful than road finders or neural networks. The idea of such "elastic" methods was based on recognition of a known parametric form of curves and consists in a simultaneous accomplishment of both stages: a curve recognition and fitting.

It is shown in this paper that "elastic" methods can be considered as particular cases of a generalization of the robust approach.

2. Elastic arm methods

Already in 1991, Gyulassy and Harlander (G&H) [5], who were worried about these problems, compared three tracking methods: classical road finder method (RFM), neural network method (NNM) that was new at that time, and elastic tracking method (ETM) they proposed. To accomplish such a comparison they suggested to use a load-density criterion—the track density ρ_{track} determined as the mean ratio of distances between points measured along a track and points belonging to different tracks or noise points. The analysis conclusively demonstrated that the road finder and neural network methods provided quite satisfactory results for $\rho_{\text{track}} \leq 1$, but RFM ceased to reliably work for $1 \leq \rho_{\text{track}} \leq 4$, where only NNM could compete with the elastic tracking method. For $\rho_{\text{track}} \geq 5$, NNM rapidly ceased to be efficient and only the ETM could succeed in the tracking problem for high load and contamination up to $\rho_{\text{track}} \sim 10$ expected at the new great CERN collider LHC.

Proposing the elastic tracking method, G&H [5] based on the concept of the elastic template, i.e., the track equation depending on varying parameters so that being bent with their variations, a curve described by this equation (this curve is also called deformable template, elastic arm, etc.) passed as close as possible to the points measured in a track. This approach can physically be described as interaction between the positively charged template and negatively charged spatial points measured in the track. The better the elastic template fits points, the lower the energy of their interaction. Let a charge of the track template be distributed with the density $\rho_T(r)$, and the set of measured points have charge density $\rho(r')$. Calculating the energy E of interaction between these two charges, we obtain

$$E = -\int dr' dr\, \rho_T(r) V(r - r') \rho(r') \longrightarrow \min\,, \tag{1}$$

where V is the potential depending on the distance from measured points to the track. Gyulassy and Harlander took the Lorentz potential

$$V(x,t) = \frac{w^2(t)}{x^2 + w^2(t)} \tag{2}$$

with the time-dependent width

$$w(t) = b + (a - b)e^{-t/T}\,, \tag{3}$$

where T is the time constant, a is the maximum distance at which points are still accredited to this template, and $b \approx \sigma_{\text{res}}$ is the accuracy of the spatial resolution of a detector. It is evident that $b \ll a$. Taking measurement discreteness into account, we obtain the track interaction energy as the sum:

$$E(\pi, t) = -\frac{1}{N}\sum_{i}^{N} \frac{w^2(t)}{\left(\vec{x}_i - \vec{r}(\pi, \vec{x})\right)^2 + w^2(t)}\,, \tag{4}$$

instead of the integral (1). Here, N is the number of points in the track, $\vec{x}_i$ and $\vec{r}$ are the ith point measured in the space and its distance from the template, respectively, and π is the set of the track parameters. When a detector is placed in a homogeneous magnetic field, the track is described by a helix with the curvature κ, helix angle λ, and phase Φ at the point $\vec{x}_i$, i.e., $\pi = \{\kappa, \lambda, \Phi, \vec{x}_o\}$. The functional in Eq. (4) depends on the points of only one track. Although all tracks could be fitted simultaneously, G&H did not recommend such a synchronous fit.

The width $w(t)$ in (3) is time dependent to avoid $E(\pi, t)$ getting caught in local spurious minima. It is carried out by the *simulated annealing* procedure [6] applied in the following way: on the first iteration of this procedure for the highest temperature the potential is taken so wide that it would safely hold all the measured points, and we have the only one minimum. Its evaluation is easy and gives us a rough approximation of parameters. Then on the every next iteration the potential is narrowed allowing us to keep the search in a vicinity of the global minimum in order to evaluate parameters more and more accurately. Corresponding evolution of the track interaction energy is shown in Fig. 3.

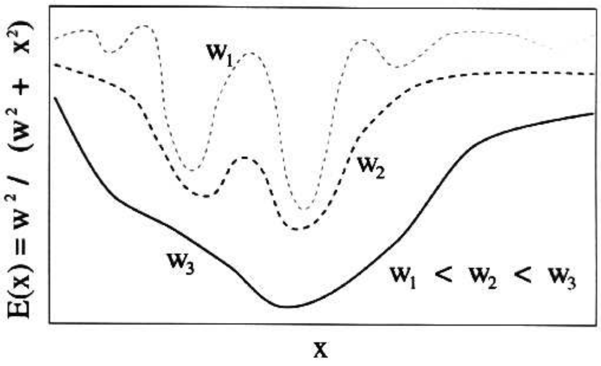

FIGURE 3. Track interaction energy for different potential widths

The ETM in such a formulation was successfully applied in [7] to recognize tracks in data from the time projection chamber of the STAR set-up, where the extremal occupancy conditions were up to 5000 tracks per event ($\rho_{\rm track} \geq 7$). Energy functional (4) was minimized with respect to the parameters $\pi = \{\kappa, \lambda, \Phi, \vec{x}_o\}$ by the steepest descent method

$$\Delta\pi_i(t) = -\eta \bigtriangledown_\pi E(\pi, t) + \alpha\Delta\pi_i(t-1)\,, \tag{5}$$

where η is a small rate parameter and the inertial term is used for suppressing fluctuations when approaching the minimum.

Choosing an initial approximation is a very important step of the elastic tracking method. In [7], this choice was made by a variant of the Hough local transform (see for example, [8]). Although the choice of three spatial points allows the determination of the helix parameters, the corresponding accuracy of determining the curvature κ was not enough to obtain the reliably continuing template under trying conditions in the track density. An analysis revealed that the initial triple should be supplemented with three more points in the adjacent coordinate planes to ensure such an extendibility.

One more similar approach called the method of *elastic arms* (*or deformable templates*) was proposed by Ohlsson and Peterson (O&P) from Lund University [9]. Instead of functional (4) in the G&H approach, they suggested to globalize the problem by using the following energy function including the effect of all M tracks of an event:

$$E(S_{ia}, \pi) = \sum_i^N \sum_a^M S_{ia} D_{ia}(\pi, \vec{x}) + \lambda \sum_i^N \left(\sum_a^M S_{ia} \right)^2, \tag{6}$$

where π are the helix parameters, $D_{ia}(\pi, \vec{x})$ is the square distance from the point to the track. To facilitate the problem of associating tracks with corresponding points O&P invented a special logic decision unit (binary neuron) S_{ia}, which is equal to 1 or 0 when ith point belongs or not to the ath track, respectively. Functional (6) is minimized under the condition that each point can belong to one track or no one. In the latter case, the functional under minimization is penalized by the value λ determining the critical distance $\sqrt{\lambda}$: the points, whose distances from a track are less or larger than $\sqrt{\lambda}$, are energetically profitable to be accredited to the track, otherwise they are considered as noise points ($S_{ia} = 0, \forall i$).

Further calculations on seeking the global minimum are performed according to the well-known Hopfield neural network scheme: network is thermalized and the mean-field theory is applied (see [10, 11]). The *simulated annealing* scheme is applied again: for a sequence of decreasing temperatures carrying, the steepest descent method provides the iterative algorithm [9] of seeking the minimum of (6)

$$\Delta\pi_a^{(k)} = -\eta_a^{(k)} \sum_i V_{ia} \frac{\partial D_{ia}}{\partial \pi_a^{(k)}}\,. \tag{7}$$

Here, the potential V_{ia} called Potts factor has the form

$$V_{ia} = \frac{e^{-\beta\, D_{ia}}}{e^{-\beta\,\lambda} + \sum\limits_{b=1}^{M} e^{-\beta\, D_{ib}}}\,, \tag{8}$$

where $\beta = 1/T$. O&P recommended to use the local variant of the Hough transform for initializing procedure.

The method of deformable templates was successfully applied in [12, 13] to process data parametrized by circles: Cherenkov radiation rings and tracks in the homogeneous magnetic field.

3. Robust estimates for heavy contaminated samples

Let us take for the sake of simplicity a linear model

$$y_i = y_0 + t_y x_i + \varepsilon_i\,, \qquad i = \overline{1,n}\,, \tag{9}$$

where x_i, y_i is the ith measured point, ε_i is a measurement error supposed to be normally distributed $\varepsilon_i \in \mathcal{N}(0,\sigma)$, $\bar{p}^T = (y_0, t_y) = (p_1, p_2)$ is the vector of unknown parameters.

The solution of the problem of parameter estimating can be obtained by the least square method (LSM), i.e., by minimizing the sum of squared residuals ε_i normalized by their standard deviations σ

$$S(\bar{p}) = \sum_i \varepsilon_i^2 \implies \min\,. \tag{10}$$

To find a minimum of $S(\bar{p})$ one should equate to zero its derivatives by parameters $\bar{p}$ that gives a set of LSM normal equations:

$$\frac{\partial S}{\partial p_j} = \sum_i \varepsilon_i \frac{\partial \varepsilon_i}{\partial p_j} = 0\,, \qquad j = 1,2\,. \tag{11}$$

These LSM estimations possess all the optimal properties inhered to maximum likelihood solution only under the assumption of residual normality. However, such the crucial assumption is in real life violated very often due to contaminating of the measurements by points of noise or background. Then due to quadratic view of the functional S in (10) the contribution to it of any point-outlier could seriously distort parameter estimations what causes a dramatic loss of accuracy. To calculate correct values of parameters one should use only the measurements from a close vicinity of the fitted function. All others should have much less impact or be completely negligible. This idea can be implemented by attributing *special weights* to each of measurements. Values of these weights must decrease with the growth of residuals ε_i, i.e., of the distance from the fitted curve. This approach called *robust* was proposed by P. Huber [14]. He suggested M-estimates by going over from the sum of squares in Eq. (10) to the sum of certain *contribution functions* $\rho(\varepsilon)$, which also depend on the deviation e of a point from the line but increase slower than a quadratic parabola.

As a result, the functional to be minimized has the form

$$L(p,\sigma) = \sum_{i=1}^{n} \rho\Big(\frac{x_i - p}{\sigma}\Big) \tag{12}$$

(since our problem is indifferent to the view of parametrization, we reduce it to a one-parameter model).

The usual requirements on the M-estimate (12) are: the function $\rho(\varepsilon)$ must be even, C^2-smooth, not decreasing for $\varepsilon > 0, \rho(\varepsilon) = 0, \rho(\varepsilon) \sim \varepsilon^2/2$ as $e \to 0$, and the estimate must be shift- and scale-invariant, as well as the estimate of σ, if its value is unknown.

We consider the following likelihood equation for the functional (12)

$$\frac{\partial L(p)}{\partial p} = \sum_{i=1}^{n} \frac{\partial \rho(\varepsilon_i)}{\partial \varepsilon_i} \frac{\partial \varepsilon_i}{\partial p} = 0\,. \tag{13}$$

By introducing

$$w(\varepsilon) = \frac{1}{\varepsilon} \frac{\partial \rho(\varepsilon)}{\partial \varepsilon}\,, \tag{14}$$

it can be modified to the form

$$\frac{\partial L(p)}{\partial p} = \sum_{i=1}^{n} w(\varepsilon_i) \frac{\partial \varepsilon_i}{\partial p} \varepsilon_i = 0\,, \tag{15}$$

which is similar to the normal LSM equations (11), but with replacement of the numerical weight coefficients to *weight functions* $w(\varepsilon)$.

The solution of (13) is now the M-estimate

$$\hat{p} = \frac{\sum_i w_i x_i}{\sum_i w_i}\,. \tag{16}$$

We describe the contaminated distribution of measurement errors ε_i following J. Tukey's *gross-error model* [15]

$$f_\epsilon(\varepsilon) = (1-\epsilon)\cdot \phi(\varepsilon) + \epsilon \cdot h(\varepsilon)\,, \tag{17}$$

where ϵ is a parameter of contamination, $\phi(\varepsilon)$ is the Gauss distribution

$$\phi(\varepsilon) = \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(-\frac{\varepsilon^2}{2\sigma^2}\Big)\,, \tag{18}$$

and $h(\varepsilon)$ is some long-tailed noise distribution density.

The problem of the choice of the function $\rho(\varepsilon)$ (or $w(\varepsilon)$) is widely discussed in the literature on robust statistics. Unbounded convex functions $\rho(\varepsilon)$ provide the uniqueness of the estimate (12), its consistency, asymptotic normality in some models and a certain minimax efficiency [14, 16]. But these estimates are practically unsuitable for heavy contaminated data models with $\epsilon > 1/2$ and asymmetric, not unimodal contamination density function $h(\varepsilon)$.

The M-estimators with bounded functions $\rho(\varepsilon)$ [18, 19] are very robust in these cases, but there are many difficulties in their use.

The first one is that the weight functions are to be recalculated at each iteration of the resulting iteration procedure referred to as Fletcher–Grant–Hebden (FGH) procedure [17]. Each iteration of the FGH-procedure involves weighted LS analysis but with weight functions. If *a priori* hints on the choice of input weight values are absent, FGH-procedure can be started by means of usual LSM with unit weights: $w_i^{(0)} \equiv 1$.

The second difficulty is that there is almost no theoretical foundation for the use of such weight functions. In particular, they all are obtained by their authors heuristically. In any case there are certain objections against their application (see [14]). We shall demonstrate that the maximum likelihood estimation in the framework of our model straightforwardly leads to a bounded function $\rho(\varepsilon)$ in (12). Evaluating the corresponding likelihood equation

$$L = \prod_{i=1}^{N} f(\varepsilon_i) \implies \max$$

we obtain the following system of non-linear equations

$$\frac{\partial \ln(L)}{\partial p} = \sum_i \frac{1}{f(\varepsilon_i)} \frac{\partial f}{\partial \varepsilon_i} \frac{\partial \varepsilon_i}{\partial p} = 0 \,. \tag{19}$$

Taking $\phi(\varepsilon)$ from (18) and $h(\varepsilon)$ to be uniform $h = h_0$ for distinctness one obtains an equation quite similar to (11):

$$\frac{\partial \ln(L)}{\partial p} = \sum_i w(\varepsilon_i)\, \varepsilon_i \frac{\partial \varepsilon_i}{\partial p} = 0 \,,$$

with some optimal weights $w(\varepsilon)$ depending on relations of ϕ and h distributions

$$w_{\text{opt}}(\varepsilon) = \frac{1 + c}{1 + c \cdot \exp\left(\frac{\varepsilon^2}{2}\right)} \,, \tag{20}$$

with the constant

$$c = \frac{h_0 \epsilon \sigma \sqrt{2\pi}}{1 - \epsilon} \,.$$

The weight function (20) corresponds to the bounded function

$$\rho_{\mathrm{opt}}(\varepsilon) = (1+c)\ln\frac{1+c}{c+\exp(\varepsilon^2/2)}\,. \tag{21}$$

The only parameter c is the ratio of the mean number of noise observations within an interval of the length $\sigma\sqrt{2\pi}$ to the mean number of useful observations in the sample. Thus it is determined by the contamination of data not in the whole range of the sample but within its essential part where all useful observations are practically concentrated. The value of c is often approximately known in experimental models.

The upper bound of (21) $(1+c)\ln(l+l/c)$ increases without limit as $c\to 0$ (with the noise diminishing). Hence the boundedness of this function is significant only for $c>0.1$, which corresponds to heavy contamination.

A polynomial expansion of the optimal weights (20) up to the fourth order leads to the approximation

$$w_T(\varepsilon) = \begin{cases} \left(1-\left(\frac{\varepsilon}{c_T\sigma}\right)^2\right)^2, & \text{if } |\varepsilon| < c_T\sigma\,,\\ 0\,, & \text{otherwise}\,. \end{cases} \tag{22}$$

We obtain, in fact, the famous Tukey's bi-weights [15], which are easier to calculate than optimal ones.

The third problem caused by the use of bounded functions $\rho(\varepsilon)$ is connected with the non-uniqueness of the estimate (12). The functional (12) often has several minima. It is difficult to find them all and to choose one of them for parameter estimating. The number of minima and their location depends on the value of p, i.e., the problems of M-estimating p and σ are closely related. Ramsay (19) also notes that separate procedures for their estimating are unwise. Our approach is based upon the joint M-estimates of p and σ.

It is a difficult problem to estimate the parameter σ in our model. The common robust estimate $\sigma = \mathrm{const} * \mathrm{med}\,|\varepsilon_i|$ is unavailable for $\epsilon > 1/2$ as well as the other estimates based on the order statistics. Moreover, the likelihood estimate of σ proposed in [19, 20]

$$\hat\sigma^2 = \frac{\sum\limits_i w_i\varepsilon_i^2}{\sum\limits_i w_i}\,, \tag{23}$$

with w_i different from (20) does not exist in some often cases.

To clarify that let us consider some geometric properties of $L(p,\sigma)$ in (12) as the function of two parameters. The set of local conditional minima of $L(p,\sigma)$ for all fixed $\sigma>0$ forms a finite collection of smooth curves in the semi-plane $\{(p,\sigma)\colon\, \sigma>0\}$ Denote them $\gamma_1,\gamma_2,\ldots,\gamma_m$. There is sufficiently large $\sigma_1>0$

such that the semi-plane $\{(p,\sigma)\colon \sigma > \sigma_1\}$ contains only one of these curves which is unbounded and has the asymptote $p_0 = (x_1 + x_2 + \cdots + x_n)/n$ as $\sigma \to \infty$ (we denote this curve by γ_1).

In Fig. 4 there are two examples of the surface $z = L(p,\sigma)$ for the samples each containing 20 points grouped into three and two clusters, correspondingly. On these surfaces one can see "ravines" (sometimes quite curved), the number of which increases as $\sigma \to 0$. Fig. 5 represents the curves $\gamma_1, \gamma_2, \ldots, \gamma_m$. which are the "bottoms" of these ravines (the corresponding samples are marked with asterisks on the axis 0p).

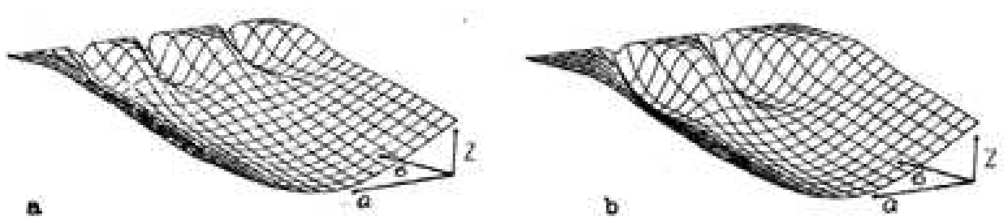

FIGURE 4. Two examples of the surface $z = L(p,\sigma)$.

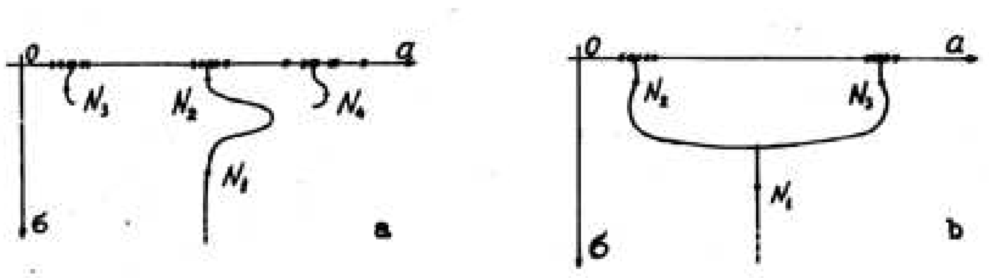

FIGURE 5. Curves $\gamma_1, \gamma_2, \ldots, \gamma_m$, which are the "bottoms" of ravines from the previous figure.

These curves are always disconnected except in the special case of artificial symmetric samples like in Fig. 5b, which leads to branching γ_1 into two different curves. The detailed study of positions on the curves $\gamma_1, \gamma_2, \ldots, \gamma_m$ of points corresponding to the estimates (16) and (23) was made in [21]. Some annoying cases were found with the absence of such points on certain curves. After considering different attempts to modify the function $L(p,\sigma)$ in such a way that its minima would be joint M-estimates of p, σ we proposed to replace (23) by

$$\hat{\sigma}^2 = \frac{\sum\limits_i w_i \varepsilon_i^2}{\sum\limits_i H_i}, \tag{24}$$

where $H_i = H\big((x_i - a)/\sigma\big)$, $H(\varepsilon) = \rho''(\varepsilon)$. This estimate is robust and satisfies Huber's definition of M-estimate [14] through the solution of the following equation

$$\sum_i \chi\Big(\frac{x_i - p}{\sigma}\Big) = 0\,, \tag{25}$$

with an even function $\chi(\varepsilon) = \varepsilon^2 w(\varepsilon) - H(\varepsilon)$, hence it is shift- and scale-invariant. The estimate (24) is more suitable for our purposes than (23) because it gives at least one guaranteed solution of the equation

$$M(p,\sigma) = \sigma^2 \sum_i H_i - \sum_i w_i (x_i - p)^2 = 0$$

with the necessary condition $\partial M/\partial\sigma > 0$. The obtained joint estimate (16), (24) can be defined by the equation system containing the function $L(p,\sigma)$ only:

$$\begin{cases} \partial L/\partial p & = 0\,; \\ \partial^2 L/\partial p^2 + \sigma^{-1}\partial l/\partial\sigma & = 0\,, \end{cases} \tag{26}$$

with the following conditions: $\partial^2 L/\partial p^2 > 0$; $\partial M/\partial \sigma > 0$.

The existence of at least one estimate (26) on the curve γ_1 leads to the following algorithm. Starting from some point (p_0, σ_0) on the curve γ_1 with sufficiently large σ_0 and $p_0 = (\sum x_i)/n$ move along γ_1 decreasing σ and looking through all solutions of (26). The solution closest to the axis $\sigma = 0$ must be chosen as the M-estimate. In our model $\sigma_0 \gg l/h_0$, i.e., σ is much less than the range of the sample. Therefore one can simplify this algorithm by moving along γ_1 without looking for solutions of (26) but just stopping when one will reach a small threshold $\sigma = \sigma_{\min}$.

The generalization of one-dimensional to the many-dimensional case is not a difficult problem. Let us take, for instance, a linear regression with parameters $\mathcal{P} = (p_1, p_2, \ldots, p_m)$. In the space of these parameters the function

$$L(p_1, p_2, \ldots, p_m\,;\ \sigma) = \sum_{i=1}^{n} \rho\Big(\frac{y_i - p_1 x_{i1} - p_2 x_{i2} - \cdots - p_{m-1} x_{i,m-1} - p_m}{\sigma}\Big) \tag{27}$$

has the conditional minima for fixed $\sigma > 0$ which satisfy the equations $\partial L/\partial p_j = 0$; $j = 1, \ldots, m$ and lie on the finite number of curves. One of these curves (namely γ_1) is unbounded and converges to $\hat{\mathcal{P}}$, which is the least square estimate of $\mathcal{P}$ as $\sigma \to \infty$. By analogy with (26) the system

$$\begin{cases} \partial L/\partial p_j = 0\,; \quad j = 1, \ldots, m\,, \\ \partial^2 L/\partial p_m^2 + \sigma^{-1}\partial l/\partial\sigma = 0 \end{cases} \tag{28}$$

has at least one solution on the unbounded curve γ_1.

Estimate computing procedure is iterative. The solution of the system (28) is obtained similar to (16) and (24) with the weights expressed via parameter estimates from previous iteration

$$w_i^{(k)} = w\left(\frac{y_i - p_1^{(k-1)}x_{i1} - p_2^{(k-1)}x_{i2} - \cdots - p_{m-1}^{(k-1)}x_{i,m-1} - p_m^{(k-1)}}{\hat\sigma^{(k-1)}}\right). \quad (29)$$

Denoting by $W^{(k-1)}$ the weight matrix $n \times n$ with the weights on its diagonal, by Y—the vector of observations $(y_1, \dots, y_n)$ and by $X = \big((X_{ij});\ i = 1, \dots, n;\ j = 1, \dots, m\big)$-the factor matrix, one can obtain the estimate of $\mathcal{P}$ on the kth iteration as

$$\hat{\mathcal{P}}^{(k)} = \big(X^T W^{(k)} X\big)^{-1} X^T W^{(k)} Y$$

(here ^{T} is the transposition symbol). If there are no a priori initial values the estimate $\hat{\mathcal{P}}^{(1)}$ on the first iteration can be obtained as the least square estimate with the unit weight matrix $W^{(0)} = 1$. Concerning $\hat\sigma$ in the denominator of (29) the proposition was made, in correspondence to the algorithm given above, to take as $\hat\sigma^{(1)}$ the sample range, i.e., $\hat\sigma^{(1)} = y_n - y_1$ and decrease it on the every iteration by the rule: $\hat\sigma^{(k)} = \lambda\hat\sigma^{(k-1)}$; $(0 < \lambda < 1)$. For reducing of the iteration number it is desirable to keep λ close to 1, but not much to prevent loosing the curve γ_1. In our calculations we choose $\lambda \sim 0.9$.

Special scaling of residuals (usually named as *studentizing*) was proposed in [22] to overcome the estimate non-stability due to so-called leverage points.

4. Applications and concluding remarks

Proposed robust algorithm was quite successfully applied to handle data in many experiments of particle physics [23–29]. Problems considered there relate to recognition and fitting of tracks and vertices, Cherenkov rings and ellipses in quite heavy background.

One of important problems inherent to contemporary experimental physics is the *alignment problem*, i.e., revealing of misaligned components in an experimental set-up among thousands of them, in order to align them mathematically. It can be done in a very sophisticated statistical data-driven analysis, which is impossible to realize without applying robust methods, as it was done in [24, 27, 28].

One of the problems arising in experimental applications should be also emphasized. It is the problem of handling data from highly granulated detectors. A particle or Cherenkov photon passing through a detector of such types deposits its energy between a cluster of adjacent pads (detector sensitive elements) near the pad that it hits as some amplitudes attributed to each pad. Thus one should fit now a curve not to a set of distinctive points, but to a few pad clusters, complemented additionally by amplitudes. This picture sharply contrasts with traditional fitting models based on the random and independent deviations of measured points from the curve to be fitted. Physicists reduce this problem to traditional fitting algorithms by special preprocessing of raw data: by clustering and then calculating points as the center of gravity of each cluster. Robust approach provides fitting in terms of raw data excluding the preprocessing stage and taking into account additional experimental information carried in data on the amplitudes of detected signals.

Let us suppose that the total amount of the released energy, E, in each cluster is a random variable with some p.d.f. $f(E)$, $0 < E < \infty$. The spatial distribution of the deposited energy among pads is bell-shaped, say, a Gaussian, centered at the actual hit point a with a constant variance σ_0^2. The calculation gave [25]

$$w(d;A) = \frac{g\,f(gA) + g^2 A\,f'(gA)}{g\,f(gA) + u(A)}\,, \tag{30}$$

where $g = \sqrt{2\pi\sigma_0^2}\,e^{x^2/2\sigma_0^2}$, $u(A)$ is p.d.f. of a background amplitude A measured in the given pad.

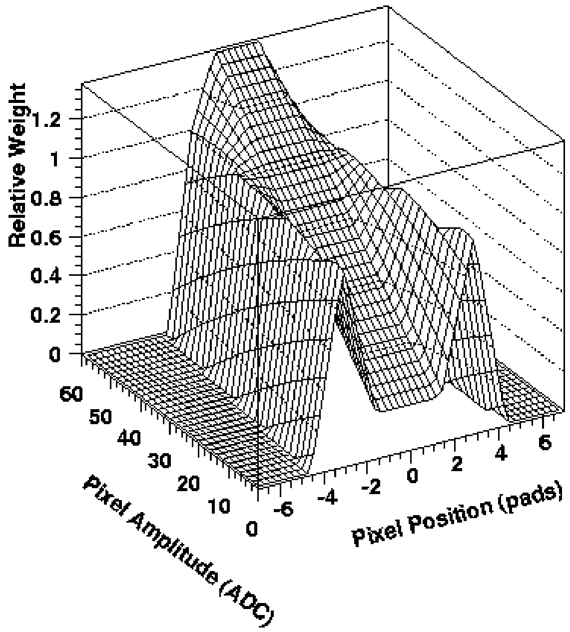

FIGURE 6. A bihorn 2D weight function piecewise linearly approximated.

If the distribution $f(E)$ is specified as an exponential one with the mean E_0, i.e., $f(E) = E_0^{-1} \cdot e^{-E/E_0}$, $E > 0$, then for a wide choice of the background component $u(d)$ (uniform or Gaussian with a wide σ_{noise} or even an exponential one with a large mean value in the cases of δ-electrons) we obtain quite an interesting 2D-weight "surface", see Fig. 6. For weak signals with a small amplitude A the weight as a function of the distance d has a "bi-horn" shape with two pronounced peaks merging together when A is growing. In a track-finding problem, this bimodal weight function is quite natural, it indicates that weaker signals are less likely to appear in the closest vicinity of the track line. Stronger signals are most likely to appear right on the track line, hence the weight function has a narrow single peak. Another problem appears in some experiments when in a certain pad row all the signals are very weak. They will be all rejected by the bimodal weight function, even though within that row the signals may have a nice bell-shaped distribution with a local peak right on the track.

One more remark relates to two possible ways of carrying out the robust approach: either by reweighted least squares with optimal weights (28) or by direct minimization of the functional (12) with a very compact contribution function $\rho(\varepsilon) = \exp(-\varepsilon^2/2\sigma^2)$. The comparison of both approaches made in [23, 29] shows that, although the second way is slightly more efficient in terms of estimate rms, it is much more time consuming due to too strong non-linearity of chosen $\rho(\varepsilon)$.

The most intriguing thing shown by various applications of the robust algorithm is its almost full similarity to elastic methods and *simulated annealing* scheme described above in Section 2. It is especially notable if one would compare levels of the energy function in Fig. 3 for different temperatures with sections of the L-surface in Fig. 4.

Elastic methods, although effective, are based of different physical and heuristical reasonings, while the proposed robust approach is based on clear statistical inferences and can be reduced to any of elastic method as to a particular case.

REFERENCES

[1] *The Compact Muon Solenoid, Technical Proposal*, CERN/LHCC, CERN, Geneva, 1994, pp. 94–38.

[2] BAUR, R. et al.: *The* CERES RICH *detector system*, Nucl. Instr. and Meth. **A343** (1994), 87–98.

[3] GROTE, H.: *Pattern recognition in high energy physics*, Rep. Prog. Phys. **50** (1987), 473–500.

[4] DENBY, B.: *Neural networks and cellular automata in experimental high energy physics*, Comput. Phys. Commun. **49** (1988), 429–438.

[5] GYULASSY, M.—HARLANDER, M.: *Elastic tracking and neural networks algorithms for complex pattern recognition*, Comp. Phys. Comm. **66** (1991), 31–46.

[6] KIRKPATRICK, S. et al: *Optimization by simulated annealing*, Science **220** (1983), 671–680.

[7] LASIUK, B. et al: *Development of an Elastic Tracking Package*, Proc. of CHEP'98, Chicago, 1998.

[8] ILLINGWORTH, J.—KITTER, J.: *A survey of Hough transform*, Comput. Vision Graphics Image Process. **44** (1988), 87–116.

[9] OHLSSON, M.—PETERSON, C.—YUILLE, A.: *Track finding with deformable templates—the elastic arms approach*, Comput. Phys. Commun. **71** (1992), 77–98.

[10] HOPFIELD, J.: *Neural Networks and Physical Systems with Emergent Collective Computational Abilities*, Proc. Nat. Acad. Sci., USA, Vol. 79, 1982, pp. 2554–2588.

[11] PETERSON, C.: *Parallel distributed approaches to combinatorial optimization problems—benchmark studies on TSP*, Neural Comput. **2** (1990), 261–269.

[12] BAGINYAN, S.—OSOSKOV, G.: *Finding tracks detected by a drift tube system*, Comp. Phys. Comm. **108** (1998), 20–28.

[13] MURESAN, L. et al.: *Deformable templates for circle recognition*, JINR Rapid Comm., Dubna **1** (1997), 81–97.

[14] HUBER, P.: *Robust Statistics*, Wiley, New York, 1981.

[15] TUKEY, J. W.: *Introduction to Today's Data Analysis*, Critical Evaluation of Chemical and Physical Structural Information, Nat. Acad. Sci., Washington, 1974, pp. 3–14.

[16] YOHAI, V.—MARONNA, R.: *Asymptotic behaviour of M-estimators for the linear model*, Ann. Stat. **7** (1979), 258–268.

[17] FLETCHER, R. et al.: *The calculation of linear best L_p approximations*, Comp. J. **72** (1971), 276–279.

[18] ANDREWS, D.: *A robust method for multiple linear regression*, Technometrics **16** (1974), 523–531.

[19] RAMSEY, J.: *A comparative study of several robust estimators of slope, intercept and scale in linear regression*, J. Amer. Stat. Assoc. **72** (1977), 608–615.

[20] AIVASYAN, S. et al.: *Applied Statistics. Study of Relationships*, Finansy i Statistika, Moscow, 1985. (In Russian)

[21] CHERNOV, N.—OSOSKOV, G.: *Joint robust estimates of location and scale parameters*, in: Transac. 10th Prague Conf. on Information Theory, Statistical Decision Functions, Random Processes, Publ. of Czech Acad. of Sci., Prague, 1988, pp. 245–252.

[22] OSOSKOV, G.: *Robust regression for the heavy contaminated sample*, in: Proc. 2-nd Intern. Tampere Confer. in Statistics, (Tampere, Finland), 1987, pp. 615–626.

[23] AGAKICHIEV, G. et al.: *Cherenkov ring fitting techniques for the* CERES RICH *detectors*, Nucl.Instr. and Methods **A371** (1996), 243–247.

[24] AGAKISHIEV, H. et al.: *New robust fitting algorithm for vertex reconstruction in the* CERES *experiment*, Nucl. Instr. and Methods **A394** (1997), 225–231.

[25] CHERNOV, N.—KOLGANOVA, E.—OSOSKOV, G.: *Robust methods for the* RICH *ring recognition and particle identification*, Nuclear Inst. Meth. **A433** (1999), 274–278.

[26] GOLUTVIN, I. et al.: *Robust estimates of track parameters and spatial resolution* CMS *muon chambers*, Comp.Phys.Comm. **126** (2000), 72–76.

[27] AGAKISHIEV, H. et al.: *Alignment of Detectors at* CERES-*NA45*, JINR Commun. E-10-98-277, Dubna, 1998.

[28] BARANNIKOVA, O. et al.: *SVT Internal Alignment Package*, STAR Note 0356, BNL, 1998.

[29] CHERNOV, N.—OSOSKOV, G.—SILIN, I.: *Robust fitting of ellipses to non-complete and contaminated data*, Czech. J. Phys. **50** (2000), 347–354.

Received April 3, 2002

Laboratory of Informational Technologies
Joint Institute for Nuclear Research
R–141980 Dubna, Moscow region
RUSSIA

E-mail: ososkov@jinr.ru