Superresolution algorithms in discrete detectors and its testing for the silicon drift detectors data

E.A.Kolganova [a], E.L.Kosarev [b], G.A.Ososkov [a]

[a] *Joint Institute for Nuclear Research, Dubna, 141980 Moscow Region, Russia*

[b] *P.L.Kapitza Institute for Physical Problems, Russian Academy of Sciences, ul.Kosygina, 2, Moscow 117334, Russia*

Abstract

Two approaches: parametric and nonparametric ones are proposed and investigated for the problem of particle coordinate recovering from measurements detected as histograms. Algorithms and their testing for both approaches are sketched. They can resolve coordinates of particles with resolving power better than the bin-size of a detector granulation unit and the instrumental function characteristic scale in terms of the Cramer-Rao limit for the parametric algorithms and the Shannon's limit for the nonparametric algorithms. Results obtained on simulated and real data are demonstrated.

Keywords: Superresolution; Histograms; Silicon Drift Detectors; Accuracy; Parametrization; Algorithm

1 Introduction

The problem of the most exact determination of the observed signal position from noisy experimental data distorted by a measuring device is considered on one of the most typical example of the measurements of an elementary particles that passes a discrete detector producing an electron shower. Then during the process of registering the signal is to be analog-to-digit converted and stored as a histogram. **The basic problem is:** to reconstruct the original signal position and other its parameters from the registered histogram. However, such hindering factors, as the signal distortions due to the digitizing process, the presence of background noise and signal overlapping due to the high occupancy of the majority of modern experimental systems, demand to combine two statistic approaches: parametric and non-parametric ones, in order to find an efficient solution of the problem.

2 Non-parametric approach and Shannon's superresolution limit

The non-parametric approach, also named *the unfolding problem*, is applied when the problem parametrization is unknown. Mathematically the problem is reduced to the solution of the following integral equation of the first kind

$$F_i = \int\limits_{\Delta_i} dx \int K(x-y)f(y)dy = \int\limits_{x_i-\Delta/2}^{x_i+\Delta/2} dx \int K(x-y)f(y)dy, \tag{1}$$

where the unknown function is $f(y) > 0$ and so-called instrumental function $K(x)$ is, in general, determined by a specific of the physical problem. Our contribution to the problem is the explicit including to the kernel $K(r)$ the effect of the information loss due to the histograming process. To take into account the finite size of our histogram bin we substitute the kernel $K(x)$ in (1) for the convolution $K_1(s) = \int h(s-x)K(x)dx$ with the characteristic function of the interval $[-\Delta/2, \Delta/2]$, i.e. $h(x) = 1$ for $|x| \leq \Delta/2$ and 0 otherwise. Although for continuous kernels this problem is ill-posed (see for example [1]), however as it was there proved, when one looks for a solution on a compact set the problem can be solved for a sufficiently wide class of kernels including the convolution $K_1(s)$. Due to always presenting noise there is the resolution limit for close signals, which is non-improved of principle. That follows from the well-known Shannon's theory on the maximum speed of the data transmission via the channel having a noise. Such the limiting superresolution factor is (see [2])

$$SR = \frac{1}{3}\log_2(1 + E_s/E_n). \qquad (2)$$

Here E_s is the signal energy: $E_s = \int\limits_{-\infty}^{\infty} F^2(x)dx$, and E_n is the noise energy: $E_n = n\sigma^2$, where n is the number of experimental data points, and σ^2 is the variance of input noise. If the signal-to-noise ratio is expressed in decibels $dB = 10\log(E_s/E_n)$, the approximate expression for the superresolution limit is $SR \simeq dB/10$.

In the paper [3] the computer program package RECOVERY was described, which can reach the Shannon superresolution limit. The programs from the RECOVERY package are based on the maximum likelihood principle and they look for maximum of the likelihood function in the many-dimensional set of solutions, which is always a compact set. The RECOVERY code is accessible from the CPC Program Library. The results of non-parametric approach described in this paper was obtained by use the modified DCONV program from the RECOVERY package. Further we refer to non-parametric method as to NP.

3 Parametric approach and Cramer-Rao accuracy limit

We consider here the approach formalism as it was given in [4] in its applications to the CERES Silicon Drift detector (SiDD). As it was shown in [5] the bell-shape form of the signal with reasonable assumption can be treated as a two-dimensional Gaussian with the maximum amplitude A. The digitized signal can be considered as two-dimensional histogram $\{a_{ij}\}$ formed by extracting cluster of adjacent cells with amplitudes exceeding a given threshold. However due to the factorial view of 2D Gaussian we can reduce this 2D problem to several one-dimensional ones $\{a_i\}$ for $i = 1, 2, ..., n_k$, which we suppose having the unit bin size, i.e. $\Delta r = r_{i+1} - r_i = 1$.

In order to estimate the center and the amplitude of a signal we have to fit this histogram with the function $\psi(r; A, r_0) = A\exp(-(r - r_0)^2/D^2)$, where D is supposed to be known. The iterative solution of the corresponding least square functional $\mathcal{L}(A, r_0) = \sum\limits_i \left(a_i - \int_{r_i}^{r_{i+1}} \psi(r; A, r_0)dr\right)^2$ is elaborated in [4] on the basis of approximating $\mathcal{L}(A, r_0)$ by an elliptic paraboloid. As initial values A^0, r_0^0 of unknown parameters for this procedure in a single signal case we use maximum amplitude of input data and its position or the histogram center of gravity. Further we refer to this method as to Paraboloidal Approximation Method (PAM).

Analysing the PAM accuracy we use the standard maximum likelihood method (MLM) for estimating the signal amplitude A and location parameter r_0 and apply the famous Cramer-Rao inequality [6] for low limit of these parameters estimation accuracy. Under conditions of the negligible correlation between x_0 and A and for

$\Delta \ll D$ and the gaussian shape of signals $\psi(x) = \exp(-(x/D)^2)$ we obtain

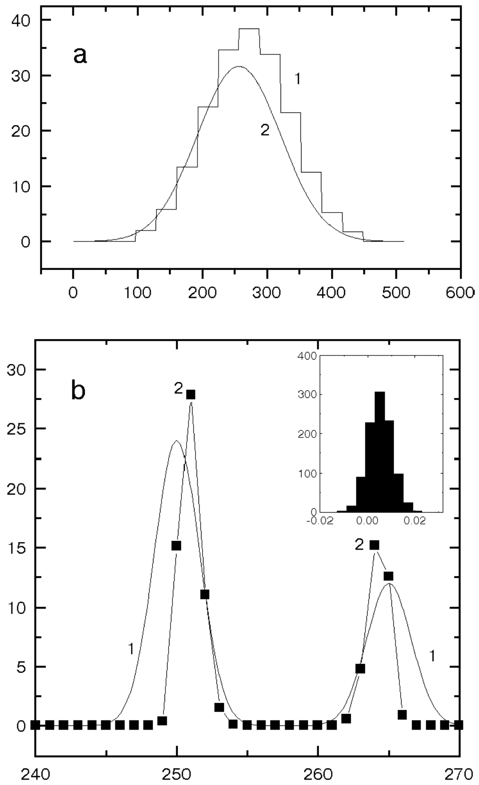
$$D(x_0) \geq \frac{\sigma^2}{A^2} \cdot \frac{1}{\sum_{i=1}^{n} \left[\frac{\partial \psi_i}{\partial x_0} \right]^2} \approx \frac{\sigma^2}{A^2} \cdot \frac{D\Delta}{\sqrt{\frac{\pi}{2}}} \tag{3}$$

and its approximate estimation

$$\delta x_0 = \sqrt{D(x_0)} \approx \frac{\sigma}{A} \cdot \frac{\sqrt{D\Delta}}{(\pi/2)^{1/4}} \approx 0.89 \frac{\sigma}{A} \sqrt{D\Delta}\,. \tag{4}$$

If $\sigma/A = 5\%$, $D = 4$, $\Delta = 1$, one has $\delta x_0 \approx 0.1$
The shape of a signal produced by two superimposing peaks can be described like

$$\psi(x; A_1, x_1^{(0)}, A_2, x_2^{(0)}) = A_1 \exp((x - x_1^{(0)})^2/D^2) + A_2 \exp((x - x_2^{(0)})^2/D^2) \tag{5}$$



The minimized functional depends now on four parameters. Its minimization was done in [4] by direct generalization of the PAM procedure. The important innovation of the given paper is the idea to use, as initial values of parameters, results obtained by the non-parametric algorithm. Such combined NP-PAM method improves considerably the accuracy and reliability of the whole procedure.

Fig.1 The numerical test of two peaks reconstruction from the 16 bins histogram, size of 1 bin equal 32; a) input data: histogram - 1, instrumental function - 2, SNR=60 dB; b) output data: solid line 1 is presented the original peaks, black squares 2 are presented the reconstructed peaks by DCONV program [3]. The insert shows distribution of deviations between simulated and estimated peak positions obtained by the parametric PAM algorithm with starting values provided by non-parametric algorithm.

4 Results and concluding remarks

The method was tested using both Monte-Carlo simulated and real data. To simulate a histogram data set for testing we integrate sequentially the Gaussian sum (5) in each bin for various D. Then a random normally distributed noise with σ_{noise} is added to each bin.
Our calculations show the excellent performance NP-PAM algorithm, in particular, in resolving double peaks from the unimodal histogram. The corresponding example is presented in Fig.1, where two peaks lying apart only for one half of the bin-size were simulated for σ_{noise}=0.1% of the mean amplitude value and signal half-width equal to 2 bin-size (Fig.1-a). In Fig.1-b it is shown how NP-method resolves both peaks and the accuracy of PAM is presented as the distribution of deviations

from modelled and estimated position of one of these peaks. Combine NP-PAM algorithm reconstruct pulse positions with accuracy better than **0.01** of bin size.

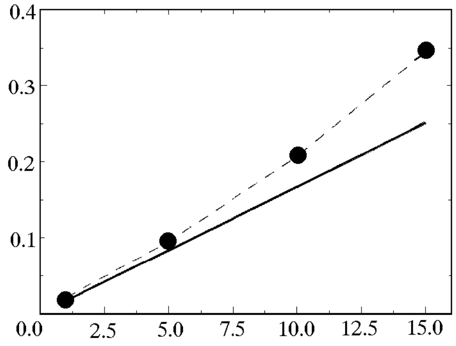


Fig.2 The signal RMS (black circles) and the Cramer-Rao lower bound (4) as the function of the noise standard deviation. On Y axis is δx_0, on X axis is σ/A in [%].

One can see that for σ_{noise} less than **10%** PAM procedure reaches almost its limiting accuracy.

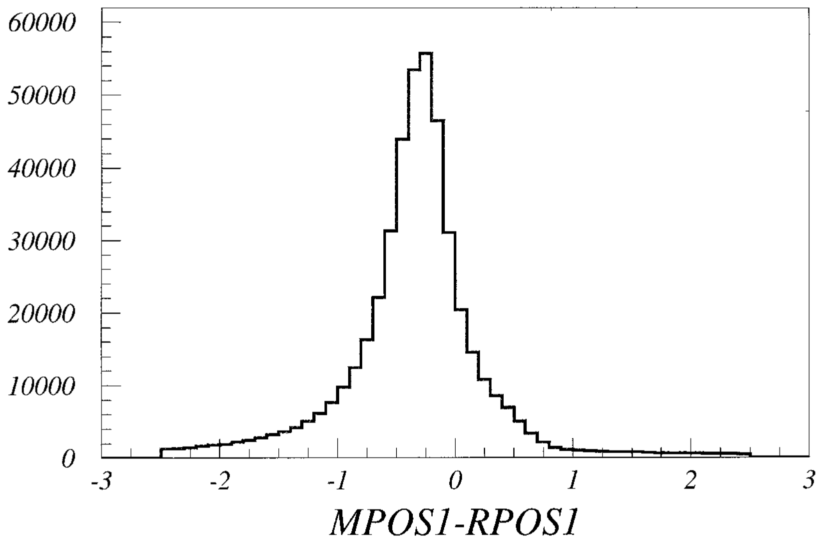


Fig.3 The distribution of differences of two pulse positions obtained by two methods (PAM and a conventional using the MINUIT program) for 160GeV/n Pb-Au data from CERES experiment [5].

Among **5000** events the only clusters were selected with the distance between pulses less or equal than signal shape half-width. For σ_{noise} over **10%** the results are quite satisfactory.

Authors would like to thank Professor J.P.Wurm and Dr H.Agakishiev for the valuable help and providing the experimental data.

References

[1] A.N.Tikhonov, V.Ya.Arsenin. Solution of ill-posed problems, Wiley, 1977

[2] E.L.Kosarev. Shannon's superresolution limit for signal recovery. *Inverse Problems*, Vol.6, No.1, pp.55-76, 1990

[3] V.I.Gelfgat, E.L.Kosarev, E.R.Podolyak. Programs for signal recovery from noisy data using the maximum likelihood principle. Part 1: General description. *Computer Physics Communications*, Vol.74, No.3, pp.335-348, 1993, Part 2: Program implementation. ibid, pp.349-357, 1993

[4] G.Agakichiev, E.Kolganova, G.Ososkov, Yu.Panebratsev. Effective Pulse Resolution Algorithms for Detectors with Gaussian-like Signal Shape, *JINR Comm.*, E10-97-105, Dubna, 1997

[5] W. Chen, H.Kraner, P.Rehak, et al. Large Area Cylindrical Silicon Drift Detector. *IEEE Trans. Nucl. Sci.*, Vol.NS-39, No.4, pp.619-628, 1992

[6] H.Cramér. Mathematical Methods of Statistics. Princeton University Press, Princeton, N.J., 1946, ch.32.