Study of Undeterministic Methods for Data Separation in Physics

Jiřina, Marcel
2003

Institute of Computer Science
Academy of Sciences of the Czech Republic

Study of undeterministic methods for data separation in physics.

Marcel Jiřina and František Hakl

Technical report No. 903

December 2003

Pod Vodárenskou věží 2, 182 07 Prague 8, phone: (+4202) 6605 1111, fax: (+4202) 86 585 789,
e-mail:ics@cs.cas.cz

Institute of Computer Science
Academy of Sciences of the Czech Republic

Study of undeterministic methods for data separation in physics.

Marcel Jiřina and František Hakl [1]

Technical report No. 903

December 2003

Abstract:

This report summarizes work done in ICS AS CR in 2003. Of course some partial results and work from previous years were also included. The report consists of two parts, one of them deals with tracking problem, the other with Higgs boson search.

We developed a new method for particle track idebtification. The method is based on the so-called circular regression for approximation of parameters of particle track, especially p_T and impact factor. It is simply a linear regression of variables which are, in fact, quadrats of geometric variables as radius or center distance from the origin.

In Higgs boson search the separation ability of neural network and of new distribution density etimation method were tested first. To get better results it was suggested to try find well recostructed events and badly reconstructed ones. It was found that when using the best reconstructed events of both classes (Higgs/noHiggs) as the learning set, the results of Higgs/noHiggs separation was enhanced.

Keywords:
data classification, neural networks, H–boson search

[1] marcel@cs.cas.cz, hakl@cs.cas.cz

Contents

Chapter 1

Circular regression for approximation of parameters of particle tracks

1.1 Introduction

From physics point of view the B-physics studies include the search for and measurement of so called CP violation through decays $B_d^0 \to J/\psi k_S^0, B_d^0 \to \pi^+\pi^-$ and $B_S^0 \to J/\psi\phi$. The other decay $J/\psi \to e^+e^-$ produces electrons, which can be identified using transition radiation tracker (TRT) [5], [6]. There are several other b-physics events are identified through identification of special electrons in TRT.

Standard algorithms as X K A L M A N [6] or the one used in D E L P H I detector [7], [8], as well as in Z E U S (D E S Y) [9] and A L I C E detector [10], use a transform of circular TRT to polar coordinates. It changes the circular tracks going through the origin to straight lines. The algorithm searches these straight lines systematically in all possible directions and places or build the track starting from the initial hit search for next hit which can be a candidate and at the same time with filtering procedure the outliers are rejected.

In this report a new algorithm is proposed. After a set of hits which may, but need not form a track, is found (so we get the track set), the track approximation is computed and outlier hits are deleted from the set to get track candidate. These track candidates are again tested with respect to other criteria and so we get tracks accepted. The best fit for part of track with hits in TRT as well as in silicon, is the weighted fit with weights inversely proportional to the pitch size of the corresponding layer and we use "circular" linear regression technique.

Circular regression uses the fact that the equation of a circle in a plane (x, y) can be rewritten as a linear equation with respect to squared variables, i.e. with respect to X, Y, where $X = x^2$ and $Y = y^2$.

1.2 Problem formulation

Given TRT barrel geometry (i.e. radii of straw layers, angular position of the first straw in each layer) and list of hits for each event. The straws are ordinarily numbered from 1 to approx. 50.000 starting from inner layer successively one by one to the outer layer. In the list of hits there is ordinal number of hitted straw and intensity of hit–either 1 (low) or 2 (high). The list may contain a drift time, but we do not use it. For testing using simulated data there is also KINE number of the track to which particular hit belongs.

For each event there should be also data from silicon detector – similarly as in TRT: geometry parameters and list of hits with necessary parameters. **The task** is to reconstruct tracks. It means to find for each track a set of hits in TRT which form it and state global parameters of the track (track hypothesis)

- p_T (or radius R, see (1.1)),
- the angle of the track in the origin, if it is a single track coming from origin (primary vertex),

- or the distance ρ of the centre from origin if $\rho > R$.

If Si detector data are given, we have to test this track hypothesis with respect to data, approve it and give more exact parameters of the track, especially p_T, and initial angle, and, finally, list of hits in Si detector and in TRT which form such a track.

In this report we deal with problem of finding tracks in full scan TRT. The tracks looked for are circular tracks either going from the origin (primary vertex) or circular tracks having radius smaller than is the distance of tracks's centre from the origin.

The tracks of the second kind may originate from lot of different processes, and among them from Bremstrahlung phenomenon. This case is not discussed here. Our searching of tracks is based on finding a part of track in TRT, then forming a hypothesis for this track and finally in approving this hypothesis using hits in silicon detector.

Using data from TRT detector only the algorithm forms hypothesis for the track giving possible approximate value of p_{T1}. Also corresponding angles of hits in individual silicon layers and starting angle are computed.

In the second part the algorithm approves or rejects the track hypothesis.

The point of view to the problem of finding tracks is more geometrical than physical. Of course, there is simple relation of circle radius R and transversal momentum p_T of electron or positron moving in magnetic field 2 tesla in ATLAS detector [5]

$$p_{T[GeV]} = 0.006 R_{[cm]}. \tag{1.1}$$

In discussion of method and its testing we limit ourselves to barrel only, more exactly, to one half of barrel.

For searching tracks in TRT several methods are known:

- methods based on Kalman filter [6], [7], [8]
- histogramming method [6]
- method used in Alice detector [9], [10]

To identify initial track set, the wave algorithm [11] has been used.

1.3 Method

The method used is based on wave algorithm [11] in its part of searching track candidate. When the track candidate (the set of hits possibly forming a track) is found, it is tested by new approach. This approach is essentially different from approach in the report cited, where assumption that the track goes through the origin (primary vertex) has been used.

1.3.1 Outline of the procedure

1. Find a set of hits in TRT, which form (possibly) a circular track. This set of hits we call simply a **track set**. The track candidate must obey some simple rules as to have enough hits or do not have "breaks" – missing hits in more than – say – in 10 successive layers. For details see [5], [11].
2. Test the track set for circularity using circular regression without prior assumption that track goes through the origin. The computation using TRT hits only will give several values
   - track radius R_2
   - track radius error $\triangle R_2$
   - distance ρ_2 of centre of the circle from the origin
3. When evaluating these values one can get

- $\rho_2 < R2$ (with respect to track radius error, i.e. more exactly $\rho_2 < R_2 + \triangle R_2$). Then the origin lies inside the circle and the track candidate cannot form a track we are searching for.
- $\rho_2 \geq R_2$. Then if
  - $\triangle R_2 <$ Straw Radius, the error is small enough to accept track candidate as a **track**.
  - $\triangle R_2 \geq$Straw Radius. Then the hit from track set which has the largest error is omitted from the track set. Then the parameters $R_2, \triangle R_2, \rho_2$ are computed again and newly evaluated according to this procedure.

Thus we get track candidate.

4. Track hypothesis is formed, i.e. we estimate positions of hits in seven layers of silicon tracker SCT. In a broad land along hypothetical track hits in SCT are found. Then track including then hits is again tested using circular regression and with weights corresponding to the reciprocal value of pitch size of layers in SCT as well as much larger pitch of straws in TRT.

1.3.2 Procedure details

Track tolerance, worst case and statistical hit error

From TRT geometry[5] follows that average distance of axes of neighbour straws in given layer is 0.67 cm (0.66–0.68 cm). Straw diameter is 0.4 cm, but evaluating drift values one can find, that radius is a little bit smaller – straw radius is 0.1949 cm (maximal value of drift time). If there is a hit then a particle went around the centre of hitted straw in tolerance less than 0.1949 cm from one or other side.

¿From it follows simply, that if given a circle, the hits can be accepted if a distance of centre of corresponding straw from the circle is less than 0.1949 cm. If not, the hit is rejected, i.e. excluded from track candidate. If the track candidate is formed of only good hits (i.e. hits belonging really to the track), then worst case error is just 0.1949 cm – we consider the track as practically perpendicular to the circle of straws in a layer:

$$\epsilon_{\text{worst case}} = R_{\text{straw}},$$

where R_{straw} is effective straw radius, i.e.R_{straw} =0.1949 cm.

The position of straw with respect to real track is random. If only good hits are considered, the individual deviations of straw centers from the track may vary from $-R_{\text{straw}}$ to $+R_{\text{straw}}$ with uniform distribution. Let us consider L_1 norm for deviation. Then mean error in this norm is

$$\epsilon_{L_1} = \frac{1}{2} R_{\text{straw}}.$$

One can consider L_2 norm, and in this case

$$\epsilon_{L_2} = \frac{1}{3} R_{\text{straw}}.$$

In fact, we found that the tracks in simulated events are a little broader than it corresponds to this case. In fact, electron of $p_T > 0.5GeV$ never hits two neighbour straws in one layer. On the other hand it hits straws which have distance of its centre from ideal track larger than R_{straw}. It corresponds to R_{straw} larger than 0.1949 cm. Then we must use in these consideration $R_{\text{straw}} = 0.3$ cm, a little hit larger than 0.1949 cm.

Circular regression

Let n points be given in a plane in polar coordinates $(\rho_1, \phi_1), (\rho_2, \phi_2), \ldots (\rho_n, \phi_n)$. These points we wish to approximate by a circle so that the error measured as suitable measure of distance of points from the approximating circle was minimal. The circle k looked for is given by its radius R and centre $S = (\rho, \phi)$, then $k = (R, \rho, \phi)$.

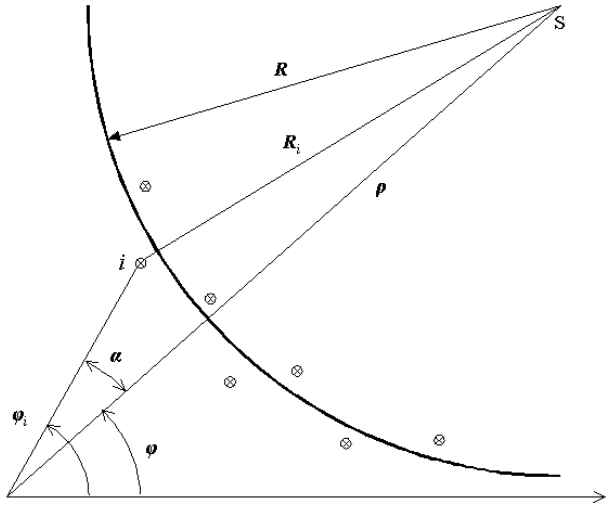

Fig. 1

It holds (see Fig. 1) $\alpha = \phi_i - \phi$ and according to cosinus theorem in triangle OS_i there is

$$R_i^2 = \rho_i^2 + \rho^2 - 2\rho_i\rho\cos\alpha. \tag{1.2}$$

We define an error by formula

$$\epsilon = \sum_{i=1}^{n}(R_i^2 - R^2)^2 \tag{1.3}$$

and let $A_i = R_i^2 - R^2$. (The error is then measured by sum of distances squared and the distance we measure not by difference but by difference of squares; nonetheles $|A_i|$ is a metrics).

As there is

$$|R_i - R| \ll R,$$

one can write $R_i = R + \delta_i$, where δ_i is small number with respect to $R, |\delta_i| \ll R$. Then

$$R_i^2 - R^2 = 2R\delta_i + \delta_i^2$$

and with sufficient exactness

$$R_i^2 - R^2 = 2R\delta_i. \tag{1.4}$$

Because R is a constant then, under the assumption that δ_i is a small number, it is a linear relation with respect to variable R_i^2. In (1.2) we have then standard sum of squares of differences of the second power of track radius because with respect to (1.3) there is

$$\epsilon = 4R^2\sum_{i=1}^{n}\delta_i. \tag{1.5}$$

Function $\epsilon = \epsilon(R, \rho, \phi)$ is continuous and differentiable. To find its local minimum, we find its partial derivatives with respect to R, ρ, ϕ and formulas found we set equal to zero. We will get successively (we write $\sum$ instead of $\sum_{i=1}^{n}$):

$$\epsilon = \sum[\rho_i^2 + \rho^2 - R^2 - 2\rho_i\rho\cos(\phi_i - \phi)]^2.$$

$$\begin{aligned}\frac{\partial\epsilon}{\partial R} &= \sum 2A_i(-2R)\\ \frac{\partial\epsilon}{\partial\rho} &= \sum 2A_i(2\rho - 2\rho_i\cos(\phi_i-\phi))\\ \frac{\partial\epsilon}{\partial\phi} &= \sum 2A_i(-2\rho_i\rho\sin(\phi_i-\phi))\end{aligned}$$

these derivatives are equal to zero in local minimum. Then

$$\sum A_i = 0 \tag{1.6}$$
$$\sum A_i(\rho - \rho_i\cos(\phi_i-\phi)) = 0 \tag{1.7}$$
$$\sum A_i(\rho - \rho_i\sin(\phi_i-\phi)) = 0 \tag{1.8}$$

We rewrite (1.6) using (1.5) in form:

$$\sum A_i\rho_i\cos(\phi_i-\phi) = 0 \tag{1.9}$$

Using formulas for sinus and cosinus of difference of angles we get equations (1.6) a (1.7) in form

$$\begin{aligned}\cos\phi\sum A_i\rho_i(\sin\phi_i+\cos\phi_i) + \sin\phi\sum A_i\rho_i(\sin\phi_i-\cos\phi_i) &= 0 &(1.10)\\ \cos\phi\sum A_i\rho_i(\sin\phi_i-\cos\phi_i) + \sin\phi\sum A_i\rho_i(\sin\phi_i+\cos\phi_i) &= 0 &(1.11)\end{aligned}$$

¿From sum and difference of equations (1.9) a (1.10) one gets after a little algebra

$$\begin{aligned}\sum A_i\rho_i(\sin\phi_i-\cos\phi_i) &= 0 &(1.12)\\ \sum A_i\rho_i(\sin\phi_i+\cos\phi_i) &= 0 &(1.13)\end{aligned}$$

Let us denote

$$\begin{aligned}a_i &= \rho_i(\sin\phi_i-\cos\phi_i)\\ b_i &= \rho_i(\sin\phi_i+\cos\phi_i).\end{aligned}$$

Then the equations (1.5), (1.11) and (1.12) can be rewritten in matrix form:

$$\begin{pmatrix} n & \sum\rho_i\cos\phi_i & \sum\rho_i\sin\phi_i\\ \sum a_i & \sum a_i\rho_i\cos\phi_i & \sum a_i\rho_i\sin\phi_i\\ \sum b_i & \sum b_i\rho_i\cos\phi_i & \sum b_i\rho_i\sin\phi_i\end{pmatrix}\cdot\begin{pmatrix}\rho^2-R^2\\ 2\rho\cos\phi\\ 2\rho\sin\phi\end{pmatrix}\begin{pmatrix}-\sum\rho_i^2\\ -\sum\rho_i^2a_i^2\\ -\sum\rho_i^2b_i^2\end{pmatrix} \tag{1.14}$$

Solving these equations we should get

$$\begin{aligned}\rho^2-R^2 &= A=\beta_1\\ 2\rho\cos\phi &= B=\beta_2\\ 2\rho\sin\phi &= C=\beta_3\end{aligned} \tag{1.15}$$

and from it

$$\rho\frac{1}{2}\sqrt{B^2+C^2},\qquad R=\sqrt{\rho^2-A},\qquad \cos\phi B/(2\rho).$$

The system (1.13) can be considered as sum of n equations in form

$$\begin{pmatrix} 1 & \rho_i \cos\phi_i & \rho_i \sin\phi_i \\ a_i & a_i\rho_i \cos\phi_i & a_i\rho_i \sin\phi_i \\ b_i & b_i\rho_i \cos\phi_i & b_i\rho_i \sin\phi_i \end{pmatrix} \cdot \begin{pmatrix} A \\ B \\ C \end{pmatrix} \begin{pmatrix} -\rho_i^2 \\ -a_i\rho_i^2 \\ -b_i\rho_i^2 \end{pmatrix} \tag{1.16}$$

All these equations are singular because the second and the third row are multiples of the first row. For equation (1.12) it does not hold generally but it will be bad conditioned equation. For this reason we consider only the first equation of (1.13). Such equations one can write n and it is possible to write them in standard form of system of linear regression equations with respect to vector $\beta = (A, B, C)^t$:

$$X \cdot \beta = y.$$

Solving it we will get (under assumption of regularity of matrix X^tX):

$$\beta = (X^tX)^{-1}X^ty.$$

Error estimations

For estimation of error of regression coefficients the theorem holds [12], [13].

Lemma: Under assumption of linear independent columns of matrix X it holds:

1. regression coefficients $\beta = (\beta_1, \beta_2, \dots \beta_n)$ are normally distributed
2. regression coefficients have variance matrix

$$\sigma^2 \cdot (X^tX)^{-1} \tag{1.17}$$

3. unbiased estimation of parameter σ^2 is

$$s^2 = \frac{S_e}{r-m}, \tag{1.18}$$

where

$$S_e = (y - y')^t(y - y') = y^ty - \beta^tX^ty, \tag{1.19}$$

and where $y' = X\beta$.

Lemma[12], [13]
Let v_{ij} be element of matrix

$$(X^tX)^{-1}, \tag{1.20}$$

$i, j = 1, 2, \dots m$. Then random variable

$$T_i = (b_i - \beta_i)\sqrt{s^2v_{ii}}$$

has distribution t_{r-m}.

Lemma [12]: The half width of confidence interval on level $1 - \alpha$ for β_i is

$$\frac{k_i}{2} \leq t_{r-m}(\alpha) \cdot \sqrt{s^2v_{ii}}. \tag{1.21}$$

We set (1.17) and (1.19) to (1.20) and thus we have just all m values $k_i/2$.

In our case $m = 3$ (R, ρ, ϕ).

For practical error estimation we use formula (1.20) and assumption that β_i is normally distributed and its σ correponds to half width of confidence interval on level $0.683 = 1 - 0.317; \alpha = 0.317$, then

$$\sigma_i \leq t_{r-m}(0.317)\sqrt{s^2v_{ii}}$$

here for $r - m \geq 7$ value of $t_{r-m}(0.317)$ is between 1 and 1.0767. In the following we use value 1, then

$$\sigma_i = \sqrt{s^2 v_{ii}},$$

where we use for v_{ii} eq. (1.19) and for s^2 successively (1.18) a (1.17).

We will get σ_i for variables A,B,C and we recompute them to $\sigma_\rho, \sigma_R, \sigma_\phi$ using assumption that σ is small enough with respect to mean value and that the curvature of the function near mean value is small so that distribution will keep its normality after transformation, and only the mean value and σ will be changed.

Then after a little algebra we get sigma squared for center distance ρ

$$\begin{aligned}
\frac{\partial \rho}{\partial B} &= \frac{B}{4\rho} \\
\frac{\partial \rho}{\partial C^2} &= \frac{C}{4\rho} \\
\sigma_\rho^2 &= (\frac{B\sigma_B}{4\rho})^2 + (\frac{C\sigma_C}{4\rho})^2
\end{aligned}$$

$\frac{B\sigma_b}{4\rho}$ and $\frac{C\sigma_c}{4\rho}$ are marginal distributions to corresponding directions.
Similarly for sigma squared of circle radius R

$$\begin{aligned}
\frac{\partial R}{\partial \rho} &= \frac{\rho}{R} \\
\frac{\partial R}{\partial A} &= -\frac{1}{2R} \\
\sigma_R^2 &= (\frac{\rho}{R}\sigma_\rho^2)^2 + (\frac{\sigma_A}{2R})^2
\end{aligned}$$

$$\phi = arccos\frac{B}{2\rho}$$

and sigma squared for angle ϕ in the origin

$$\begin{aligned}
\frac{\partial \phi}{\partial B} &= \frac{-1}{\sqrt{1-(\frac{B}{2\rho})^2}} \cdot \frac{1}{2\rho} \\
\frac{\partial \phi}{\partial \rho} &= \frac{B}{2\rho^2\sqrt{1-(\frac{B}{2\rho})^2}} \\
\sigma_\phi^2 & (\frac{-1}{2\rho\sqrt{1-(\frac{B}{2\rho})^2}} \cdot \sigma_{\rho_B})^2 + (\frac{B}{2\rho^2\sqrt{1-(\frac{B}{2\rho})^2}} \cdot \sigma_\rho)^2 \\
\sigma_\phi^2 &= \frac{1}{4\rho^2(1-(\frac{B}{2\rho})^2)} \cdot (\sigma_B^2 + \frac{B^2}{\rho^2}\sigma_\rho^2) \\
\sigma_\phi^2 &= \frac{1}{4\rho^2 - B^2}(\sigma_B^2 + \frac{B^2}{\rho^2}\sigma_\rho^2).
\end{aligned}$$

At the same time, the following condition must hold:

$$\rho^2 > B^2.$$

We are interested in circle radius which corresponds to tangential momentum p_T of a particle according to formula

$$p_T 0.006R,$$

and tangent angle ψ of the track in the point nearest to the origin. In Fig. 2 is seen that it holds

$$\psi = \phi + \pi/2$$

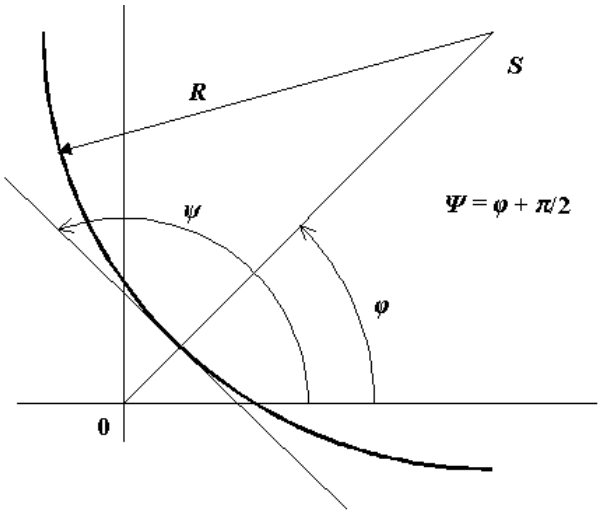

Fig. 2

Circular regression - simplified case

When the approximation circle goes through the origin.

If the approximation circle goes through the origin, then the distance ρ of the centre from the origin is equal to radius of the circle R. Equation (1.1) changes its form to

$$R_i^2 = \rho_i^2 + R^2 - 2\rho_i R\cos\alpha. \tag{1.22}$$

Then there is

$$\epsilon = \sum_{i=1}^{n}(\rho_i^2 - 2\rho_i R\cos\alpha)^2$$

and

$$A_i = \rho_i(\rho_i - 2R\cos\alpha).$$

Instead of three unknown variables we have two only, R and ϕ. It holds

$$\begin{aligned}\frac{\partial\epsilon}{\partial R} &= \sum 2A_i(2\rho - 2\rho_i\cos(\phi_i - \phi))\\ \frac{\partial\epsilon}{\partial\phi} &= \sum 2A_i(-2\rho_i R\sin(\phi_i - \phi))\end{aligned}$$

and we get a system

$$\begin{pmatrix} \sum a_i & \sum a_i\rho_i\cos\phi_i & \sum a_i\rho_i\sin\phi_i \\ \sum b_i & \sum b_i\rho_i\cos\phi_i & \sum b_i\rho_i\sin\phi_i \end{pmatrix}\cdot\begin{pmatrix} 2\rho\cos\phi \\ 2\rho\sin\phi \end{pmatrix}\begin{pmatrix} -\sum\rho_i^2a_i^2 \\ -\sum\rho_i^2b_i^2 \end{pmatrix}$$

¿From it we should get values of

$$\begin{aligned}2\rho\cos\phi &= B\\ 2\rho\sin\phi &= C\end{aligned} \tag{1.23}$$

and finally

$$\rho = \frac{1}{2}\sqrt{B^2 + C^2}, \qquad \cos\phi = B/(2\rho)$$

and due to assumption mentioned there is $R = \rho$. For different angle ψ of the track and momentum p_T it holds

$$\begin{aligned} \psi &= \phi + \frac{\pi}{2} \\ p_T &= 0.006R. \end{aligned}$$

With respect to badly conditioned system matrix we use again linear regression.

A little pessimism about error

Let us consider three hits only: the first (1), medium hit (2) and the last hit (3), and let straw diameter be taken into account. Then we can construct three circles as shown in Fig. 3.

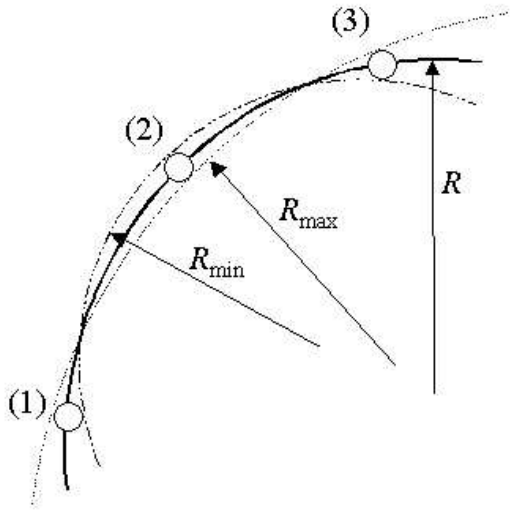

Fig. 3

In Fig.4 is easily seen that

$$\begin{aligned} \sin\gamma_s = \tfrac{d}{R} \qquad & \sin\alpha \tfrac{R_{\text{straw}}}{d/2} \\ \sin\gamma_m \tfrac{d}{R_{\text{min}}} \qquad & \gamma_m = \gamma_s + \alpha \end{aligned}$$

Form it, after a little algebra and using approximately $\cos\alpha = 1, \cos\gamma_s = 1$ we get

$$\frac{\sin\gamma_s}{\sin\gamma_m} = \frac{1}{1 + 2\frac{R - R_{\text{straw}}}{d^2}}$$

On the other hand simply

$$\frac{\sin\gamma_s}{\sin\gamma_m} = \frac{R\text{min}}{R}$$

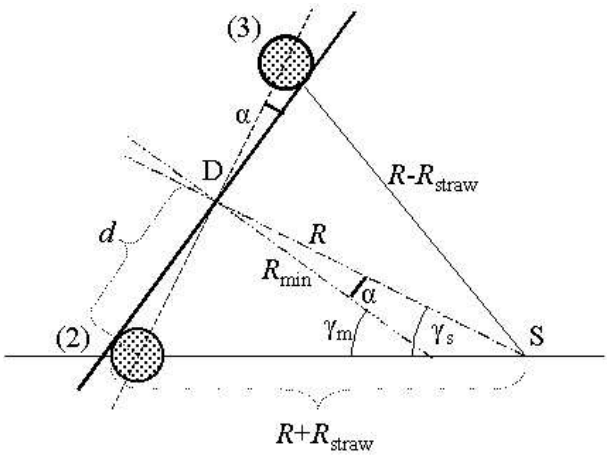

Fig. 4

and combining last two equations there is

$$R_{\min} = \frac{R}{1 + 2\frac{R - R_{\text{straw}}}{d^2}}$$

Similarly we get

$$R_{\max} = \frac{R}{1 - \frac{R - R_{\text{straw}}}{d^2}}$$

In the end

$$R_{\max} - R_{\min} = 2\epsilon_R = 2 \cdot R \frac{4\frac{R - R_{\text{straw}}}{d^2}}{1 - 4(\frac{R - R_{\text{straw}}}{d^2})^2}.$$

Example:

$$d \sim \frac{1}{2}(106 - 56)cm = 25cm, R \sim 100cm, R_{\text{straw}} \sim 0,2cm.$$

Then

$$\epsilon_R = 12.8cm.$$

The worst case error of track radius is then rather large, nearly 13% and thus it has no sense to use the worst case analysis for any other estimation. Similarly one cannot use regression model error which is approximately of the same size.

Therefore, we will use the regression approximation of circle centre coordinations as "good" (fixed centre assumption), and we will compute average distance of all track candidate hits as track radius. Then individual deviations of hits from this average radius can be computed and finally we can get mean quadratic error of radius. This error is two orders of magnitude less than error of regression approximation or than worst case error. Moreover, knowing individual deviations of hits, one can exclude hit having largest deviation in process of "clearing" the track candidate, see the next text.

Considerations about regression error

Using fixed centre assumption, the individual error of i-th hit is

$$\epsilon_i = R_i^2 - R^2,$$

where

$$R_i = \rho_i^2 + \rho^2 - 2\rho_i\rho\cos(\phi_i - \phi) \tag{1.24}$$

(see(1.2)). Note that ϵ_i has dimensionality cm^2. The mean quadratic error of squared track radius is given by

$$\epsilon_0^2 = \frac{1}{n}\sum_{i=1}^{n}(R_i^2 - R^2).$$

Let

$$R_i = R + \delta_i \tag{1.25}$$

where δ_i is individual deviation of i-th hit from mean track circle. Then

$$\epsilon_i = R_i^2 - R^2 = 2R\delta_i + \delta_i^2 \approx 2R\delta_i \tag{1.26}$$

and then

$$\delta_i = \frac{\epsilon_i}{2R}. \tag{1.27}$$

The mean value of $\epsilon_i, i = 1, 2, \ldots, n$ is the mean quadratic error ϵ_0 of R^2. Due to (1.27), the mean value of $\delta_i, i = 1, 2, \ldots, n$ is

$$\delta_0 = \frac{\epsilon_0}{2R}$$

and δ_0 is mean quadratic error of track radius R (with uncertainty given by approximation used in (1.26)).

For practical computation we use relation

$$\delta_o = \frac{1}{2R}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(R_i^2 - R^2)^2}.$$

Transversal momentum error The mean quadratic error δ_0 of track radius R gives, at the same time, the mean quadratic error of p_t, due to formula $p_T[GeV] = 0.006R_{[cm]}$, then $\delta_{p_{T[GeV]}} = 0.006\delta_{0[cm]}$.

Cleaning the track

In cleaning process we exclude from track candidate successively hits with largest individual error until for all remaining hits absolute individual errors are less than straw radius:

$$\delta_i = R_{\text{straw}}$$

$i = 1, 2, \ldots, N_0$ of remaining hits of track candidate. The individual errors δ_i follow from (1.25) using (1.24).

One bad hit case: Let in the track candidate be n "good" hits and one bad hit. The good hits in average correspond to radius R. The bad hit to the radius, say, $R + \delta$. The radius R' computed from all $n + 1$ hits will be

$$R' = \frac{1}{n+1}(n \cdot R + R + \delta) = R + \frac{\delta}{n+1}.$$

¿From it follows that for large n we need not count essential influence of one hit to track radius R. Let 2δ be approximately span of straws in a layer and $\delta \doteq 2R_{\text{straw}}$, see Fig 5.

Then

$$\frac{\delta}{n+1} = \frac{2R_{\text{straw}}}{n+1}.$$

Minimal number of hits in track must be $N_{\min}$. The largest δ will be

$$\delta = \frac{R_{\text{straw}}}{N_{\min} + 1}$$

and for $N_{\min} = 10 \quad \delta = 0.1 \cdot R_{\text{straw}} \doteq 0.02cm$. It is seen, that bad hit near the track can cause only very slight error in track radius or in p_T.

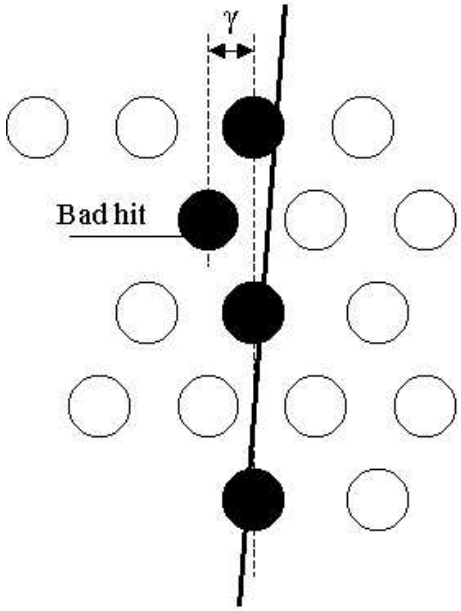

Fig. 5

Cleaning procedure:

1. Given track candidate.
2. Compute track parameters R, ρ, ϕ.
3. Compute individual deviations δ_i for all hits in the track candidate.
4. If there is a hit with largest δ_i and its $\delta_i > R_{\text{straw}}$, then exclude this hit from the track candidate. Go to 1.
5. If track candidate follows other conditions for track, accept it as resulting track; stop.

Evaluation procedure:

The next step is testing the results of computation. The most essential parameter is mean square deviation δ_0 of track radius computed directly according to formula

$$\delta_o = \sqrt{\frac{1}{n}\delta_i^2},$$

where δ_i are computed according to (1.24) and (1.27).

Testing procedure:

- if $\delta_0 > R_{\text{straw}}$ then
  - if $|R_2| > \rho + \delta_0$ then it is a bad circle, $T = -1$, end
  - else it is perhaps good circle but with rather large error. Exclude the worst hit and go back to the computation of δ_0 without that hit; $T = 0$.
- else it is sufficiently exact circle
  - if $|R_2| > \rho + \delta_0$ then it is a bad circle, $T = -2$, end

- else it is a good and exact circle
  - if $|R_2| < 70$ then circle is too small, $T = -4$, end
  - if $||R_2| - \rho| \leqq R_{\text{straw}}$ then the circle goes through the origin, $T = 1$, end
  - else it is circle with positive impact factor.

1.4 Testing the Method

Data at Hand

Data used for evaluation was generally data after reconstruction with old geometry. There were total 92 events with total 2194 tracks. The analysis was performed for all tracks, for hadrons and for electrons. Table 1 gives number of tracks under different conditions.

Type of data	All data saved after reconstruction without • Fake tracks • Multiplicity	All data saved after reconstruction without • Fake tracks • Multiplicity • too large error	All data saved after reconstruction without • Fake tracks • Multiplicity • too large error • nothing was identified in silicon layers
All tracks	2194	1762	1362
Electrons (+positrons)	207	155	107
Hadrons	1650	1335	1137

Tab. 1. Number of entries (tracks).

Fake tracks are reconstructed tracks, which cannot be assigned to any true track as they are formed by hits from several different true tracks. Multiplicities are the cases of two or more tracks assigned to one true track. Too large error means that error criterion was not fulfilled during reconstruction process (this is not a comparison with true track).

Efficiency

To evaluate efficiency there is necessary to know corresponding true tracks. There is the fact that not all tracks in the file correspond to elementary criteria to be reconstructed, especially they have mostly too small number of hits or more hits than the number of layers in TRT. The set of TRT hits to be considered as a track set must

a. have sufficient number of hits (HitsCountMin = 10)

b. cannot have more hits than there is TRT layers (NumOfLayers73; max. one hit in a layer)

c. must not have gaps, i.e. the number of successive layers with no hits must not exceed a given number (MaxEmpty = 10). A derived condition of this condition is that the track must start sufficiently early and not end too early, i.e. it have the first hit not further than in layer No. MaxEmpty = 10, and the last hit in layer at least No. NumOfLayers - MaxEmpty = 63.

The such set of hits is the "track set" not yet the track found. It may be a rather random set of hits not forming a circular track. To become the "track found" it must fulfill best-fit criterion more. The best fit is the weighted fit with weights inversely proportional to the pitch size of the corresponding layer and we use "circular" linear regression technique.

Reconstruction Efficiency

Reconstruction Efficiency is the percentage of tracks found in all tracks looked for. To estimate number of true tracks fulfilling conditions a., b., c. above, we use all track candidates without multiplicity as the true tracks. Multiplicities are found using KINE number that prevails in hits of the track set or track found. In fact we use the program to find all tracks fulfilling conditions a., b., c.

and we suppress track multiplicities. Reconstructed tracks are only those which comply error criteria more. From it reconstruction efficiency less than one follows. The average value found is 80.31 as ratio of 1762 and 2194 tracks.

Overall reconstruction efficiency gives 80.31 %, 74.88 %, and 80.90 % for all tracks, electron tracks and hadron tracks, respectively. Dependence on p_T and eta is shown in Fig. 6.

Tracks Multiplication

Tracks multiplication is defined as the ratio of the number all tracks found over the number of all tracks found without multiplicities. Overall Tracks multiplication gives 2.60 %, 4.37 %, and 1.71 % rise above one for all tracks, electron tracks and hadron tracks, respectively.

Hits Finding Efficiency

Hits finding efficiency is defined as ratio of true hits found on the track by our algorithm with respect to true number of hits forming the track (i.e. true track). The tracks considered are the accepted tracks as defined in Introduction. Hits finding efficiency for all tracks, electrons and hadrons is shown in Figs. 7-9.

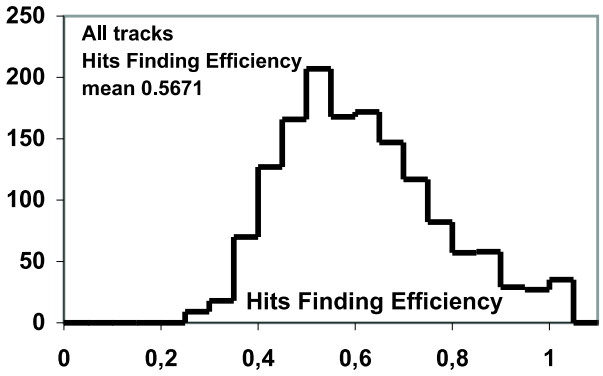

Fig. 7 Hits finding efficiency, all tracks.

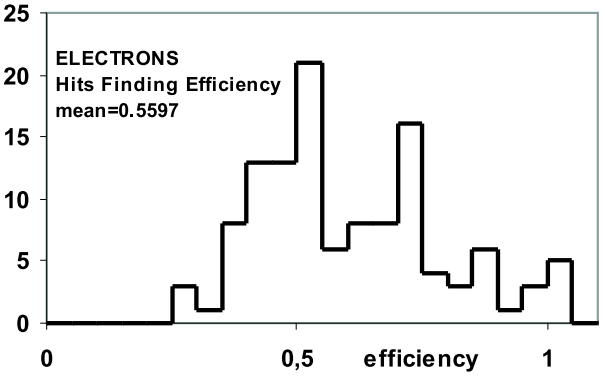

Fig. 8 Hits finding efficiency, electrons.

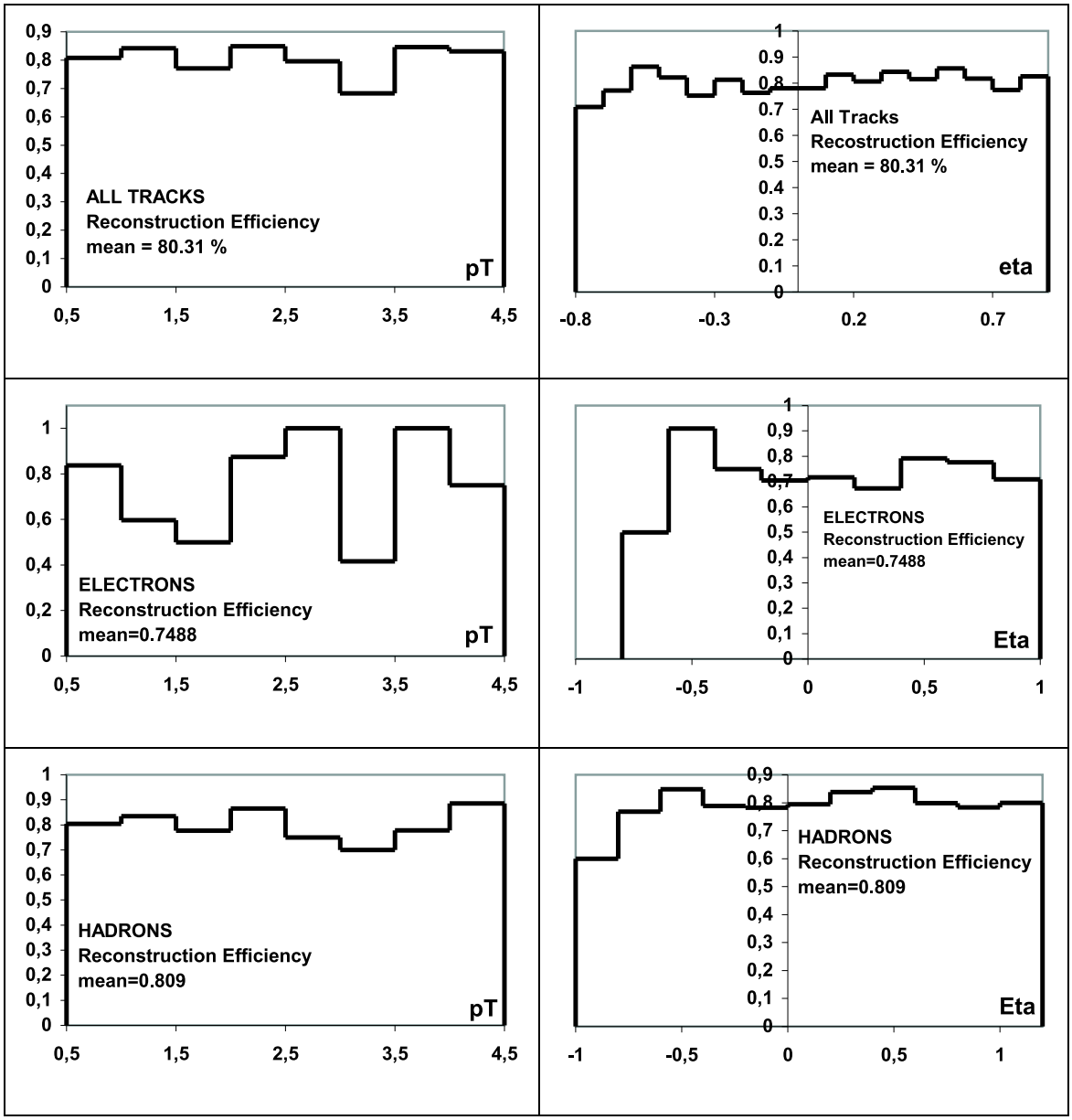

Fig. 6 Reconstruction Efficiency as function of pT and eta, all tracks.

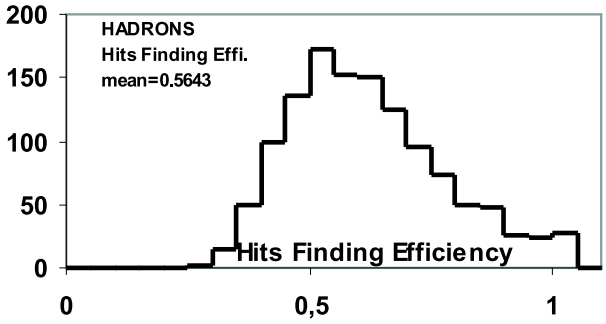

Fig. 9 Hits finding efficiency, hadrons.

Quality

1/pTrec / 1/pTtrue

This is computed for tracks having at least one hit in silicon tracker. Diferences of 1/pT reconstructed and true are shown in Figs. 10-13. It can be found that for all tracks but without outliers, i.e. for

$$abs\,(1/pTrec - 1/pTtrue) < 0.25\ GeV^{-1},$$

there is shift of histogram peek to the right, i.e. systematic error of pT is 0.00410 GeV (not 1/pT). Mean squared error 0.627 GeV^{-1} for hadrons and 1.347 GeV^{-1} for electrons can be reduced by using more strict condition on minimal number of hits in silicon detector. At the same time, tail will be reduced.

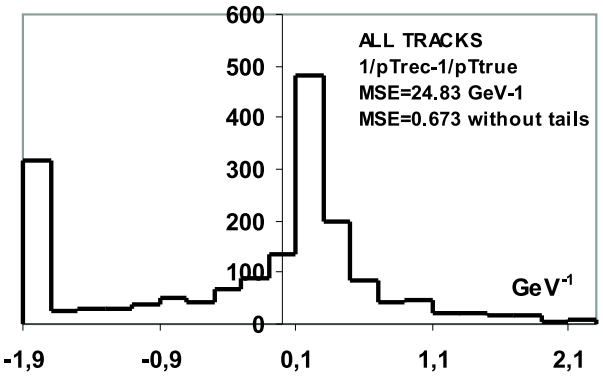

Fig. 10 Difference between 1/pT reconstructed and 1/pT true, all tracks.

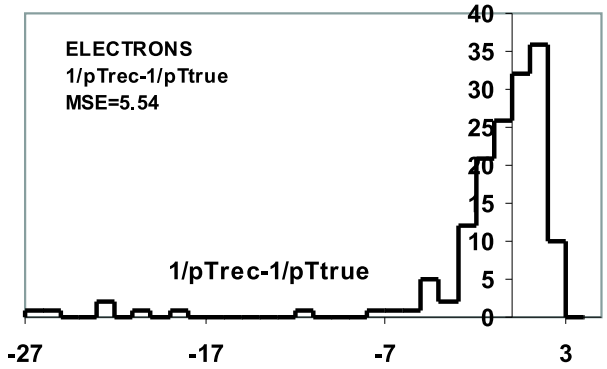

Fig. 11 Difference between 1/pT reconstructed and 1/pT true, electrons.

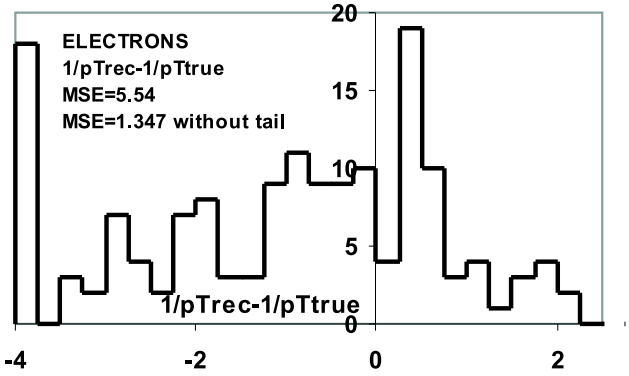

Fig. 12 Difference between 1/pT reconstructed and 1/pT true, detail, electrons.

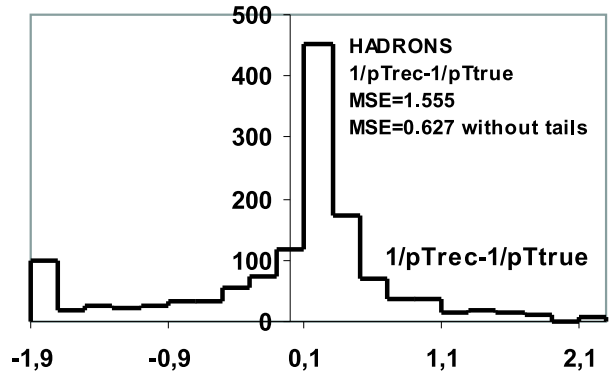

Fig. 13 Difference between 1/pT reconstructed and 1/pT true, hadrons.

Impact Parameter Reconstructed
In this case the true impact parameter is computed the same way in which it can be established for reconstructed tracks, i.e. as the negative difference between the radius of the track and the distance of the circle center from the origin. It is seen that there is a systematic error 0.001 cm. This is best visible in Fig. 16. The impact parameter reconstructed is 10 μm larger than true one. On the other hand, with this systematic error the error distribution without tails is very close to the normal distribution N(0.001 cm, 0.0297 cm) with shift 30 times smaller than dispersion σ = MSE = 0.0297 cm.

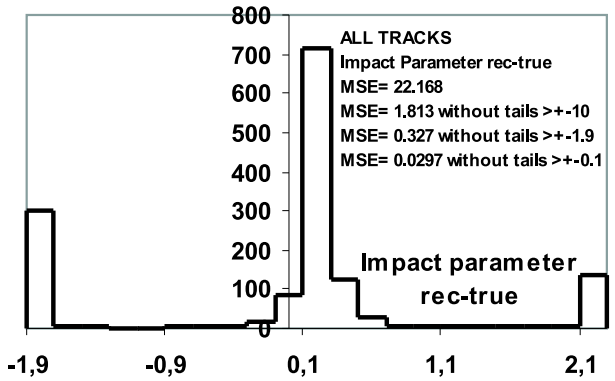

Fig. 14 The difference between Impact parameter Reconstructed and the Impact parameter True in cm, all tracks.

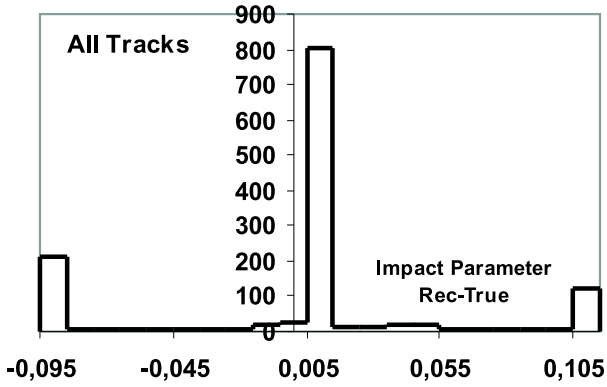

Fig. 15 The difference between Impact parameter Reconstructed and the Impact parameter True in cm, detail, all tracks.

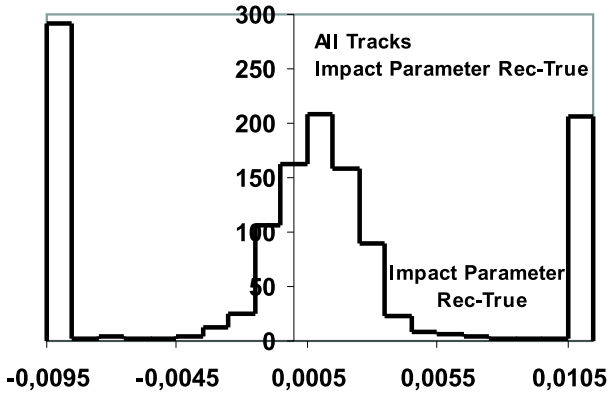

Fig. 16 The difference between Impact parameter Reconstructed and the Impact parameter True, all tracks - more detail.

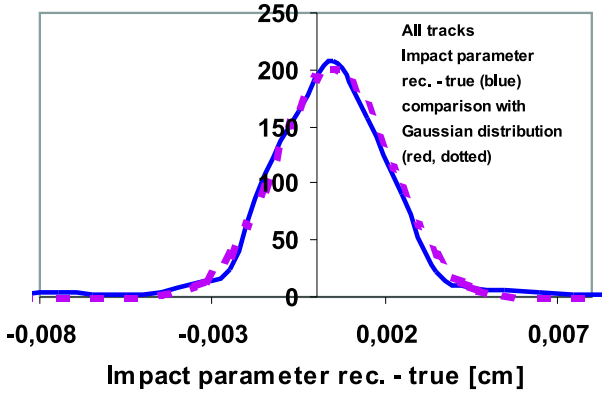

Fig. 17 The difference between Impact parameter Reconstructed and the Impact parameter True, all tracks - more detail - comparison with Gaussian distribution.

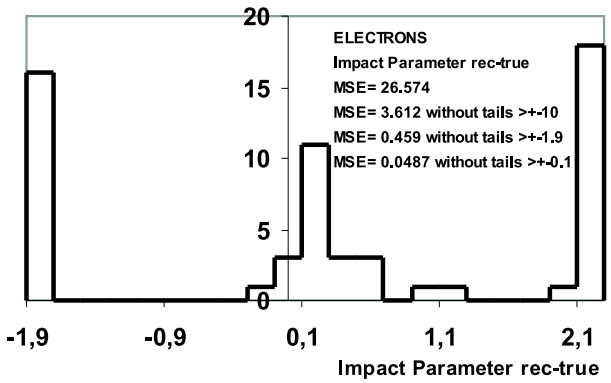

Fig. 18 The difference between Impact parameter Reconstructed and the Impact parameter True, electrons.

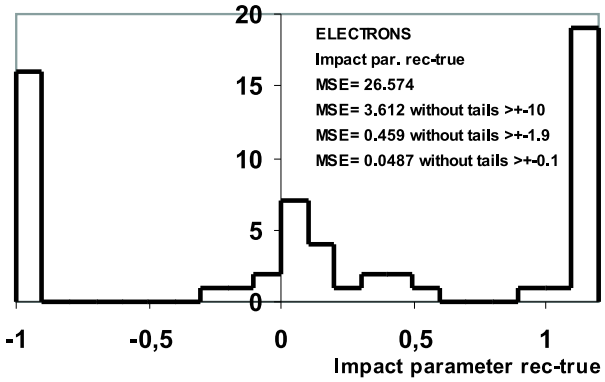

Fig. 19 The difference between Impact parameter Reconstructed and the Impact parameter True, electrons.

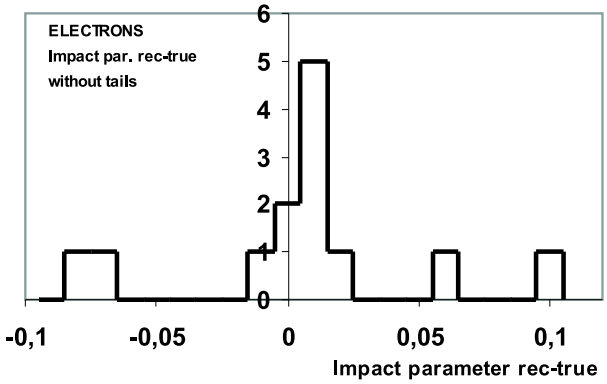

Fig. 20 The difference between Impact parameter Reconstructed and the Impact parameter True - detail, electrons.

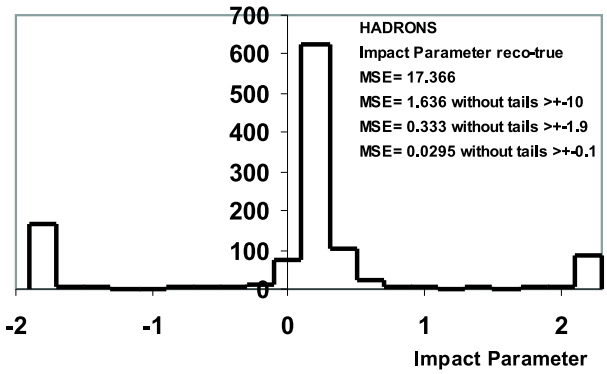

Fig. 21 The difference between Impact parameter Reconstructed and the Impact parameter True, hadrons.

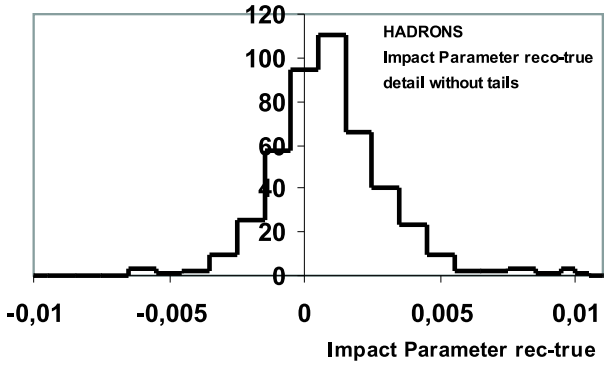

Fig. 22 The difference between Impact parameter Reconstructed and the Impact parameter True - detail, hadrons.

1.5 Conclusion to the track regression

The algorithm developed is based on idea to use instead of distances their squares. Thus most of dependencies related to circular track in elementary particle detectors changes to linear. It was shown that especially linear regression of squared variables is applicable in this case. The whole algorithm of elementary particle track identification has essentially three steps, identification of track using circular regression and rather rough data from TRT detector, forming track hypothesis using geometrical conditions and verification of the hypothesis for each track using the circular regression once more using data from both TRT and silicon detector.

These results are preliminary from the following reasons:

1. data used are related to old geometry, but we mean that it does influence neither behavior of the tracking algorithm nor the algorithm itself,

2. statistics are rather small.

Using this algorithm we found that it is able to reconstruct all "standard" reconstruction parameters as pT and impact factor. It is also able to find that track is "broken" due to energy loss, in the case of electrons due to Bremsstrahlung effect. This case was not discussed here but we have also found that approximation error (not error reconstructed - true) when using set of hits finally selected (also in silicon detector) when supposing a "broken" track can be smaller than approximation error when the same set of hits in silicon detector and in TRT is forced to approximate a single circular track without any break.

Chapter 2

Higgs boson search

2.1 Problem formulation

The partial objective of our work is a search for $H \rightarrow b\bar{b}$ decay. There is a certain probability that Higgs bosons are produced in considered collision. In the case that the mass of the Higgs boson is $m \leq 200GeV/c^2$ a decay $H \rightarrow b\bar{b}$ dominate. The main background involves a gluon instead of the Higgs boson, producing the same final state.

The products of this type of collision are 2 jets from one W decay, 4 b-jets, and one lepton (electron or muon) plus missing energy from an unobserved neutrino. Each visible particle (2 jets, 4 b-jets and lepton) is described by three values P_T – transverse momentum $[GeV/c^2]$, η – pseudorapidity and ϕ – direction angles, corresponding to particle vectors with negligible mass. The neutrino is described by two missing energy values E_x^{miss}, E_y^{miss}. A fundamental variable which can be used in the Higgs boson search is a effective mass M_{bb} of two b-'s which can arise either form Higgs boson decay or from gluon decay after $p\bar{p}$ collision. There are lot of events with gluon decay (background) and much fewer events with Higgs decay (signal). Each of these two classes of events have different distribution of the effective mass M_{bb}. The Higgs boson decay is theoretically a Gaussian distribution with mean 120 GeV/c^2 and $\sigma = 15GeV/c^2$, whereas the gluon decay is much broader. The difference between those two distributions can be exploited to decide if Higgs boson decay is present in the data or not. In addition, other physical reasons reject all events in which at least one of the following conditions has been satisfied: at least one jets has $P_T < 15GeV$, at least one jet has pseudorapidity out of the range $(-2.5, 2.5)$, the lepton is electron and $P_T^{lep} < 20GeV$, the lepton is muon and $P_T^{lep} < 6GeV$. All events passing those restrictions form the histogram. Really data do not provide information about the presence of Higgs decay in the event, hence for real data we have available the total distribution of M_{bb} only. For simulated data, we can plot two histograms of M_{bb}, one for background only and the second one for pure signal. Neural networks assume we know the distributions of signal and separated background. So the main idea of how to exploit neural networks to confirm Higgs decay presence is based on filtering events in such a way that percentage of signal will be increased after filtering and at the same time significance $\frac{S_a}{\sqrt{S_a+B_a}}$will stay on the same level.

2.2 Methods

Classification of multivariate data is a problem solved by lot of methods from nearest neighbor method to decission trees, neural networks and genetic algorithms. The problem is generally difficult because of several influences, e.g.

- High problem dimensionality where curse of dimensionality causes excessive grow of processing time. Presence of noise; true data are rarely "pure".
- Multicollinearity, i.e. mutual dependence of individual variables. If variables are originally considered independent, i.e. orthogonal to all others, multicollinearity causes distortion of the space; coordinates are, in fact, not orthogonal.

- Boundary effect. Due to this effect nearest points seem to be rather far and farther points near so that the distance between the nearest and the farthest point of finite data set can be smaller than the distance of the nearest neighbor from the given point.

2.3 Neural network with switching units

Neural network with switching units (NNSU) is a combination of a classical neural network architecture and a decision tree. This network is actually an oriented acyclic graph whose nodes are structures called building blocks. This acyclic graph will be referenced as the outer graph.

Each building block is a neural network consisting of two types of nodes. These nodes are connected in such a way that they again form an acyclic graph but with the restriction that the outputs dimension of the building block is the same for all building blocks in the outer graph. The first type of node, which we refer to as a functional unit, makes a predefined mapping from the input space to the output space of this node. Hence such node can be described by a tuple of integers (input vector dimension and output vector dimension) and by its transfer function. The definition of this transfer function includes parameters of this functional unit (weight vectors, threshold etc. in current neural networks terminology).

The second type of nodes, switching units, collect all outputs from parent functional units, concatenate them together to form one vector, and search for a predefined number of clusters in the set of such input vectors. We use the Forgy-Jancey algorithm which is a non-deterministic clustering procedure.

After clustering each cluster is associated with a corresponding child functional unit and the parameters of this functional unit are adjusted with regard to patterns in the corresponding cluster only. In fact, division of input patterns into two or more disjoint sets, and consecutive learning over these subsets of patterns, put a separation hypersurface into the input space. The type of these hypersurfaces is defined by the type of transfer functions of switching unit parents.

So each building block is learned, the output from each building block is propagated to all children, and the output of the top building block is considered as final output from the neural network.

2.4 Distribution density estimation method

The distribution density estimation method (DDEM, [1]) deals with distances in multidimensional space and simplifies a complex picture of probability distribution of points in this space by mapping functions of one variable. This variable is the distance from the given point (the query point x [3]) in multidimensional space. From it follows that mapping functions are different for different query points and this is cost payed for simplification from n variables in n-dimensional space to one variable.

The distance is basic notion for all approaches dealing with neighbors, especially nearest neighbors. There is a lot of methods of classification based on the nearest neighbors [2]. They estimate the probability density at point x (a query point) of the data space by ratio $\frac{i}{V_i}$ of number i of points of a given class in a suitable ball of volume V_i with center at point x [1].

The method used is also based on distances of the training set samples x_s, $s = 1, 2, \ldots k$ from point x. It can be shown that the sum of reciprocals of q-th power of these distances, where q is a suitable number, is convergent and can be used as a probability density estimate. From the fact of high power of distances in multidimensional Euclidean space ($q > 1$, usually between 1 and n), fast convergence, i.e. small influence of distant samples, follows. The speed of convergence is the better the higher dimensionality and the larger q.

Using distances, i.e. a simple transformation from n-dimensional Euclidean space E_n to one-dimensional Euclidean space E_1, and no iterations, the curse of dimensionality is straightforwardly eliminated. The method can be also considered as a variant of the kernel method, based on a probability density estimator but using a much simpler metric and does not satisfy some mathematical conditions. The value q is something like effective dimensionality of problem including boudary effect and other effects mentioned above.

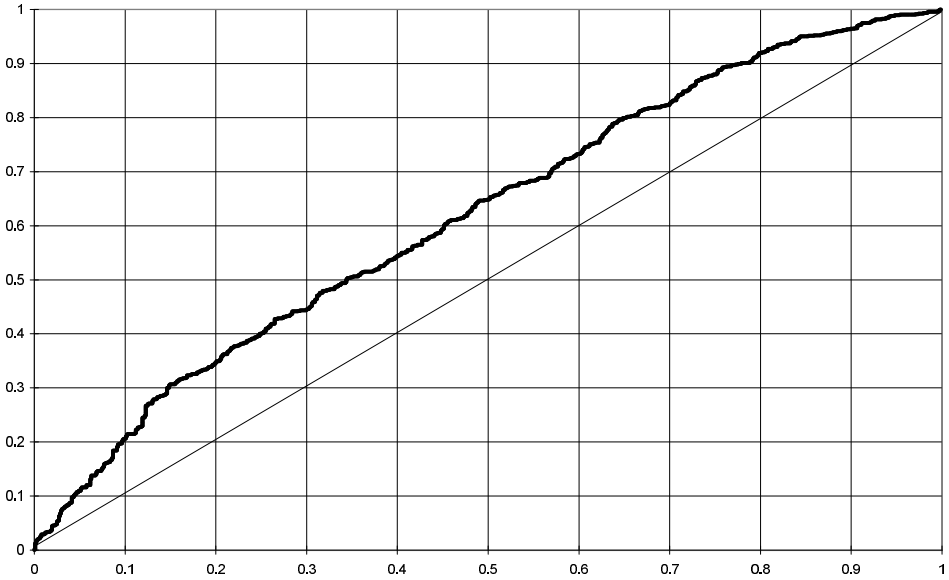

Figure 2.1: In this figure plot of signal efficiency (vertical axis) vs. background error (horizontal axis) is shown. It can be seen that for cut supressing 90% of background it remains 22% of the signal. Then enrichment (or – say – enhancement) is in this case equal to 2.2 .

The method uses normalized data, i.e. the individual coordinates of samples of the learning set are normalized to zero mean and unit variance and the same normalization constants (empirical mean and empirical variance) are applied to all other (testing and of unknown class) data. This transformation does not mean any change in form of the distribution, i.e. uniform distribution remains uniform, exponential distribution remains exponential (with $\sigma = 1$ and shifted by 1 to the left), etc. This normalization is a part of program SFSloc7a we use for practical computations.

2.5 Direct application of NNSU neural network and DDEM for separation enhancement

We want mention that all work has been done with "fair" data sets, i.e. the learning set and the testing set are disjoint. It means that after the learning the testing data were "newer seen before", thus simulating similar but unknown future reality.

Note that on horizontal axis in Fig. 2.2 there are values of the DDEM output which is, in fact, its estimation of probability of data item (sample, event) being a signal. The same is valid for horizontal axis of Fig. 2.3.

In this figure the dependence of the enrichment factor on the DDEM output is shown. The enrichmnet factor is, in fact, the ratio by which the percentage of signal in the mixture of signal and background is enlarged after the cut, i.e. after the filter.

2.6 Testing quality of reconstruction

Using the same data we tested also how difficult is to recognize the well reconstruted events from badly reconstructed ones. In Fig. 2.5 it is seen that peeks of histograms are better separated than when we try separate Higgs/noHiggs from the same data. The curve in bottom part of Fig 2.5 also shows beter behavior – compared with Fig. 2.1.

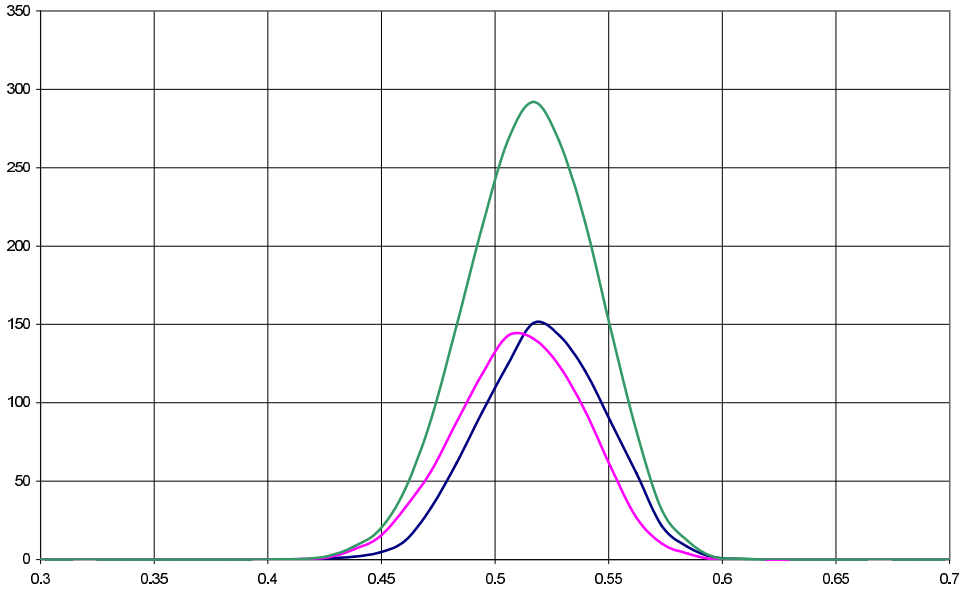

Figure 2.2: Histograms of background (lovest peak), signal (middle peak), and sum of both (the largest peak). Horizontal axis: probability of event being a signal (DDEM output).

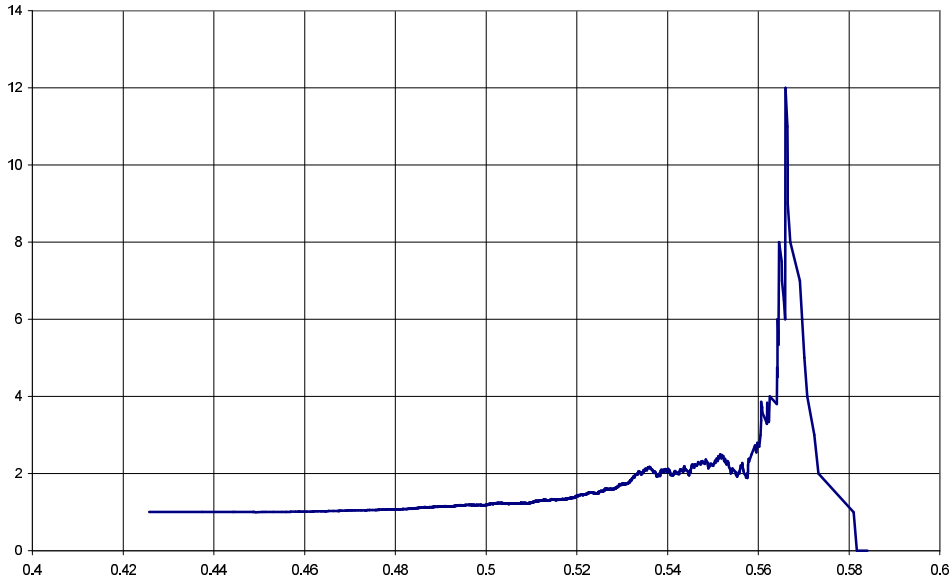

Figure 2.3: Enrichment factor vs. probability (output of DDEM).

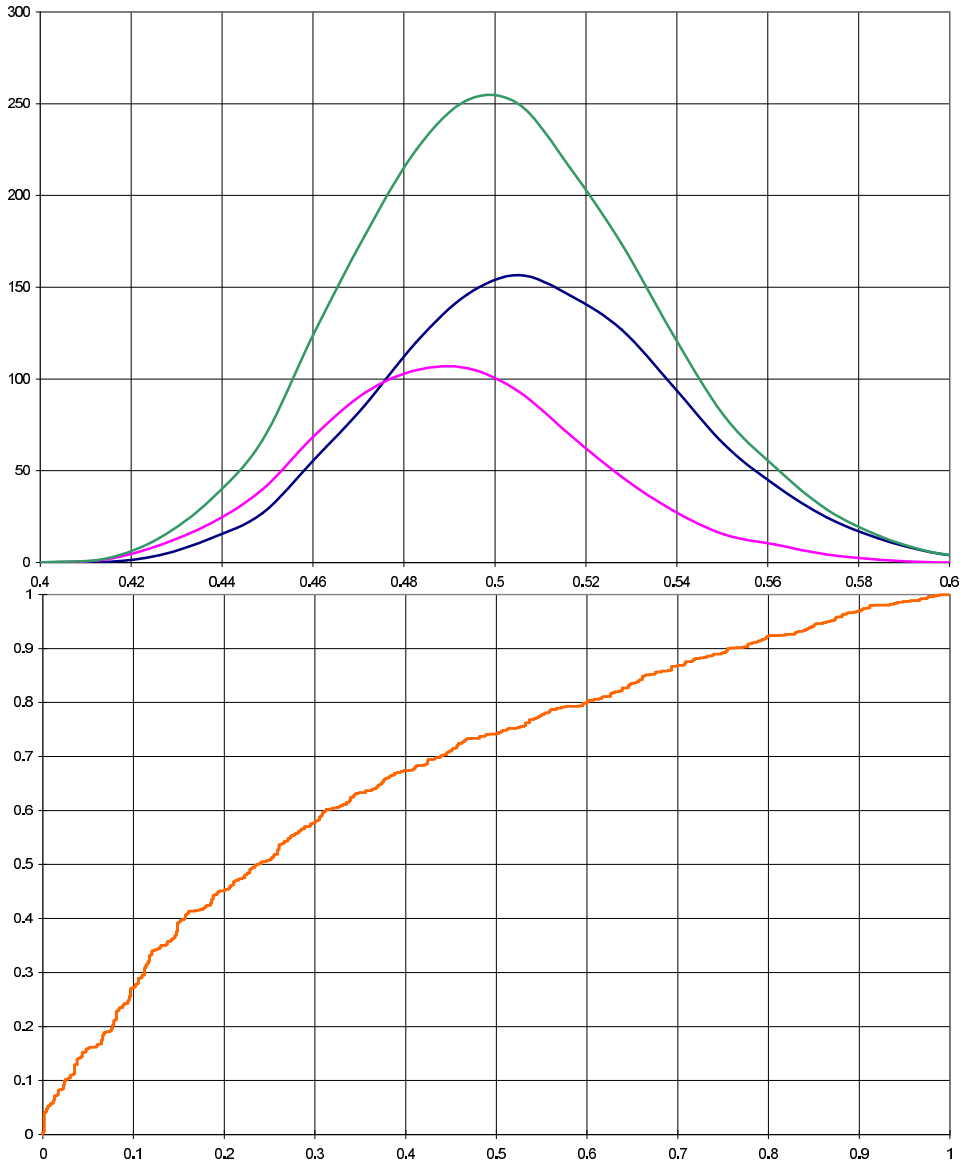

Figure 2.4:
UPPER: Histogram of background (lovest), signal (middle), and sum (upper) for reconstruction quality estimation.
BOTTOM: Signal efficiency vs. Background error (horizontal) for reconstruction quality estimation.
In both of these figures it is seen that estimation of the reconstruction quality is easier than classification of the Higgs/noHiggs events. At the same time, it is not enough as it is seen below.

2.7 Well recostructed data for Higgs/noHiggs separation with NNSU and DDEM

We made two experiments:

First, learning data were split to two equal parts, part of better reconstructed events and part of worse (bad) reconstructed events. Because of simulated data, the quality of reconstruction can be known. Originally reconstruction quality is given by numbers Qr from 0 to 11. The better half consist of events with Qr=0, 1, 2, and approx. 30% of events with Qr=3. The bad half is from all remaining events. In the same way also testing data were split. Corresponding data and figures and lines on them are denoted simply "well", sometimes with letter "w".

Second, only events with Qr=0, and 1 were used. Corresponding data and figures and lines on them are denoted simply "01".

There are shown, that

- original results on Higgs/noHiggs separation using DDEM (black line). This is basis for all other comparisons
- results using data on "well" reconstructed events (the better half of the data set) as the learning set, and
  - testing with data on well reconstructed events from the testing set (again the better half of all testing set)
  - testing with all the testing set with well as well as badly reconstructed events.
- results using data on "01", i.e. the best reconstructed events from the learning set, and
  - testing with data on "01", i.e. the best reconstructed events from the testing set (upper line)
  - testing with all the testing set with well as well as badly reconstructed events (bottom line).

It is seen from figures that the best case is the upper line corresponding to "01" learning as well "01" testing data. It is therefore clear that reconstruction quality heavily influences Higgs/noHiggs separation. The problem is to estimate the reconstruction quality or organize computation so that better reconstructed data can be used.

2.8 Conclusion to Higgs boson search

We found that quality of reconstruction, especially quality of reconstruction of the testing set, i.e. data simulating real data from the experiment, especially influence Higgs/noHiggs recognition. We have shown that some improvement is possible but problem remains in way how to recognize truly well reconstructed events.

There are 120 possible combinations from which only one is true. In data there is a combination which seems to be true best. This "true" combination comes from reconstruction procedure as fulfilling best the reconstruction criteria. But it can differ from physical facts. We plan to use part of reconstruction procedure into consideration and deal with more data samples for each event. Each event leads to one of 120 possibilities and all these possibilities can be considered. This would be rather extensive approach. A better way would be to use during reconstruction procedure in each step both (or all) possibilities which fulfill softer criterion than to use the one best only. Thus we can get several samples for each event but not all 120. Worst combinations will be excluded and number of samples reduced from 120 to, say, 5 or 10 or so. All these selected combinations can be tested using our methods from the point of view of reconstruction quality and the one best really used for Higgs/noHiggs recognition enhancement.

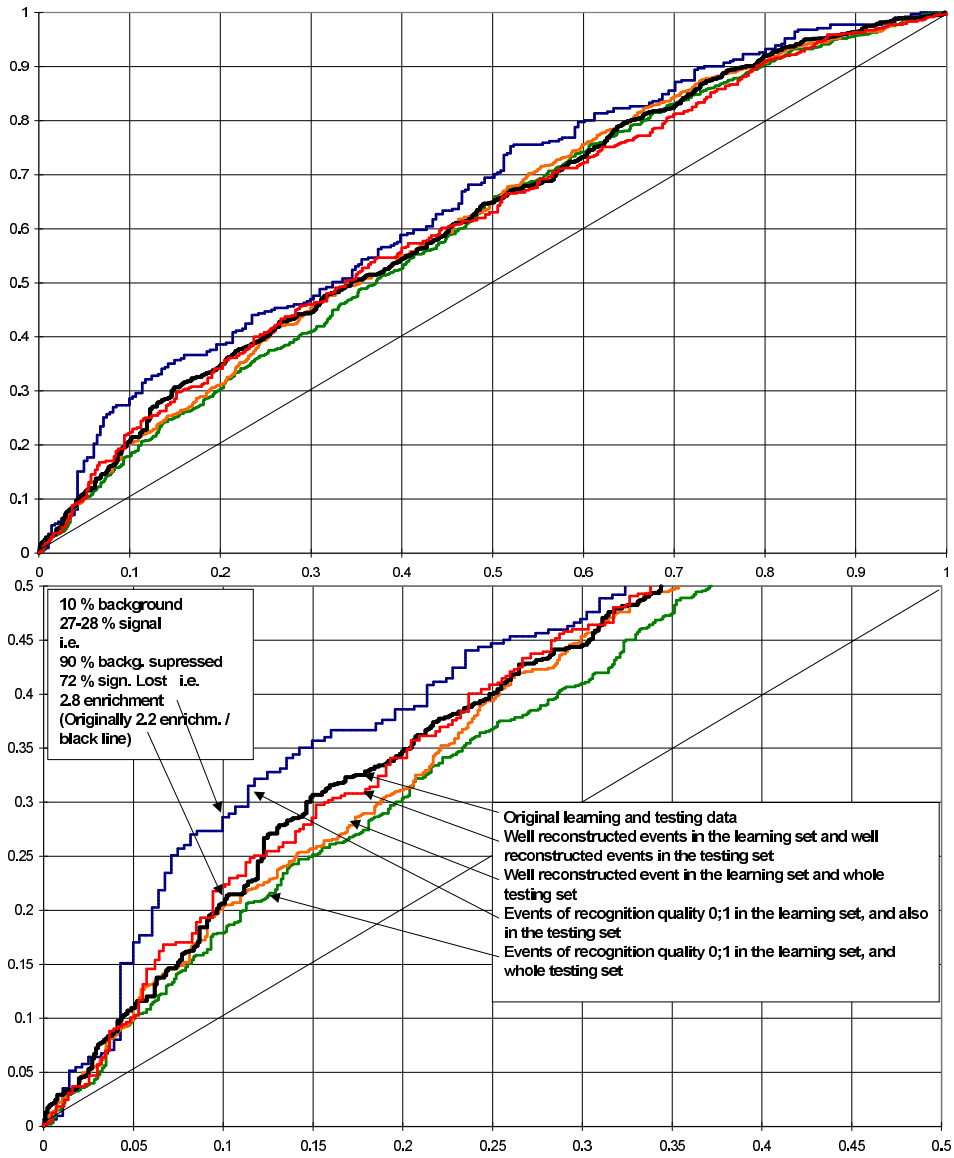

Figure 2.5: Signal efficiency vs. Background Error (horizontal) for Higgs recognition (bottom figure is a detail of upper figure).
In both of these figures it is seen that quality of classification of the Higgs/noHiggs events heavily depends on quality of reconstruction.

2.9 Acknowledgement

Authors are deeply indebted to Dr. Elsbieta Richter-Was for her kind understanding that we do not undestand physics, for providing data for our experiments and explanation of their contents as well as reconstruction procedure, and her fruitful discussions over partial and often not too splendid results.

Bibliography

[1] Jiřina M., "Classification of Multivariate Data Using Distribution Mapping Exponent." Proc. of the Symposium in Memoriam John von Neumann, Budapest, Hungary, Dec. 12, 2003.

[2] Silverman, B. W., "Density Estimation for Statistics and data Analysis." Chapman and Hall, London, 1986.

[3] Hinnenburg A., Aggarwal C.C., Keim D.A."What is the nearest neighbor in high dimensional spaces ? " Proc. of the 26th VLDB Conf. , Cairo, Egypt, 2000, pp 506-515.

[4] Duda R., Hart P. Stork, D.G.: Pattern Classification. John Wiley and Sons, 2000.

[5] ATLAS Technical Proposal for a General-Purpose pp Experiment at the Large Hadron Collider at CERN. CERN/LHCC 94-43 LHCC/P2 15 Dec. 1994, Chaps. 11.11, 3.4.

[6] Gavrilenko, I.: Description of Global Pattern Recognition Program (XKalman). ATLAS Note-INDET-97-165 (ATL-I-PN-165), 25 Apr. 1997.

[7] Billoir, Pierre: Progressive Track Recognition with a Kalman-like Fitting Procedure. Computer Physics Commnications 57 (1989), pp. 390-394.

[8] Frühwirth, R.: Applicaton of Kalman Filtering to Track and Vertex Fitting. Nucl. Instr. and Meth. in Phys res. A262 (1987), pp. 444-450.

[9] Billor, P., Qian, S.: Simultaneous Pattern Recognition and Track Fitting by the Kalman Filtering Method. Nucl. Instr. and Meth. in Phys res. A294 (1990), pp. 219-228.

[10] Badalá, A., Barbera, R., Lo Re, G., Palmeri, A., Pappalardo, G. S., Pulvirenti, A., Riggi, F.: Tracking inside the ALICE Inner Tracking System. CERN-ALI-2001-034; CERN-ALICE-PUB-2001-034. Geneva, CERN, 30 Jul 2001, 11 p.

[11] Jiřina, M., Hakl, F.: The Wave Algorithm for Track Searching. Technical Report No. 868, Institute of Computer Science, Academy of Sciences of the Czech Republic, July 2002, pp. 22-28.

[12] Andl, J.: Mathematical Statistics (in Czech), SNTL/ALFA, Prague, 1985.

[13] Leamer, E. E.: Specification searches. Ad hoc inference with nonexperimental data. John Wiley and Sons, New York, 1978.