Adaptive methods with application to track reconstruction at LHC

A. Strandlie (CERN / Gjøvik University College)

Outline

- **Introduction**
- **Adaptive methods - basic features**
- **Relation to non-HEP applications**
- **Application to track reconstruction at LHC**

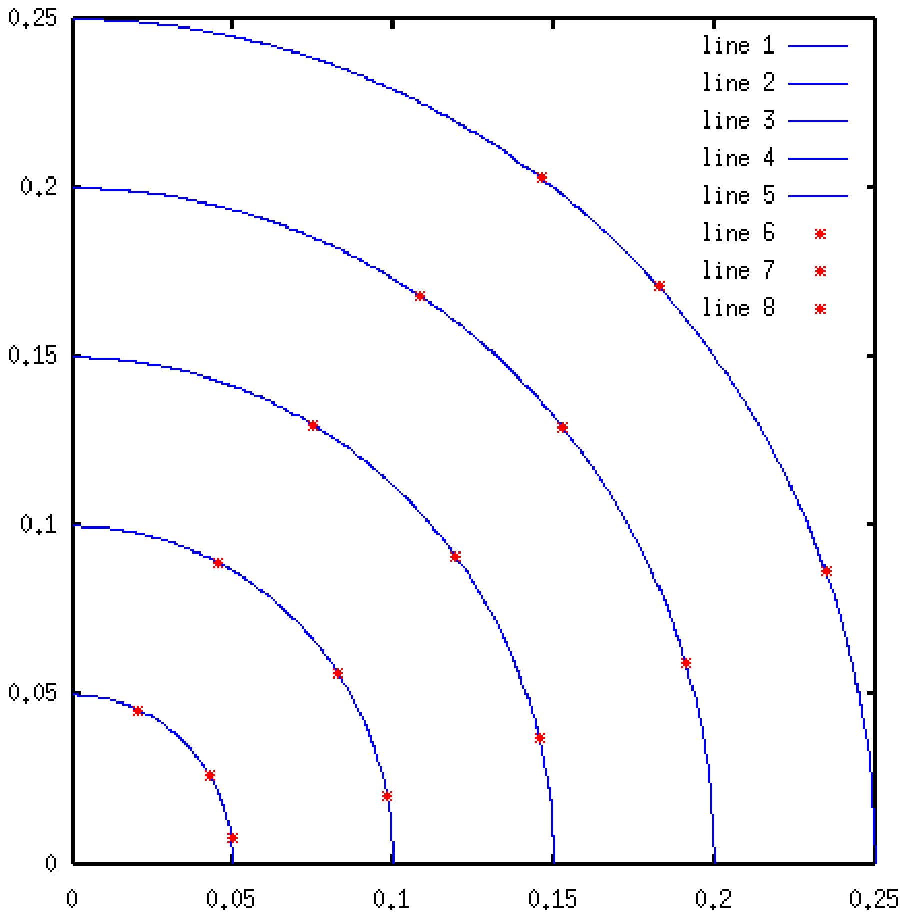
Introduction

- **Track reconstruction is a vital ingredient in the analysis chain of a high-energy physics experiment.**
- **This task is often divided into two different subtasks:**

  1) Track finding/pattern recognition

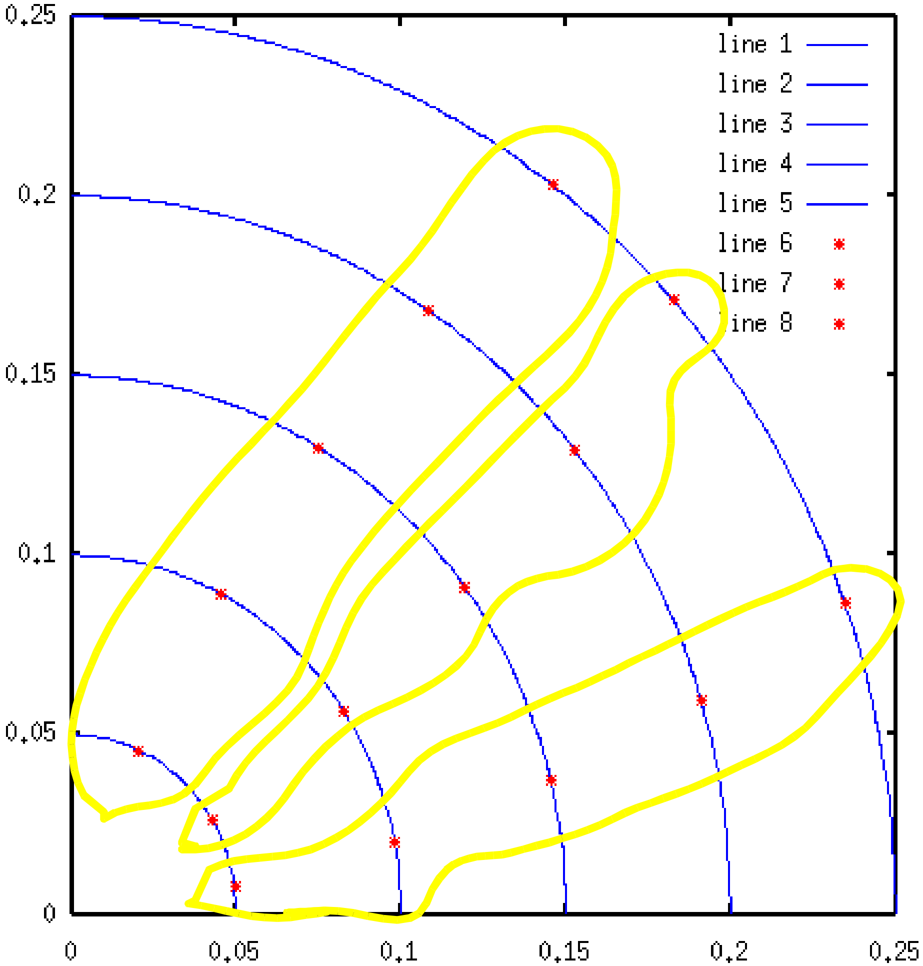
  2) Track fitting/parameter estimation
- **Track finding: starts out with a set of position measurements (provided by a tracking detector).**
- **The aim is to group these measurements together in subsets, each subset containing measurements orginating from one charged particle.**

Introduction



Measurements from three tracks coming from the origin

Introduction

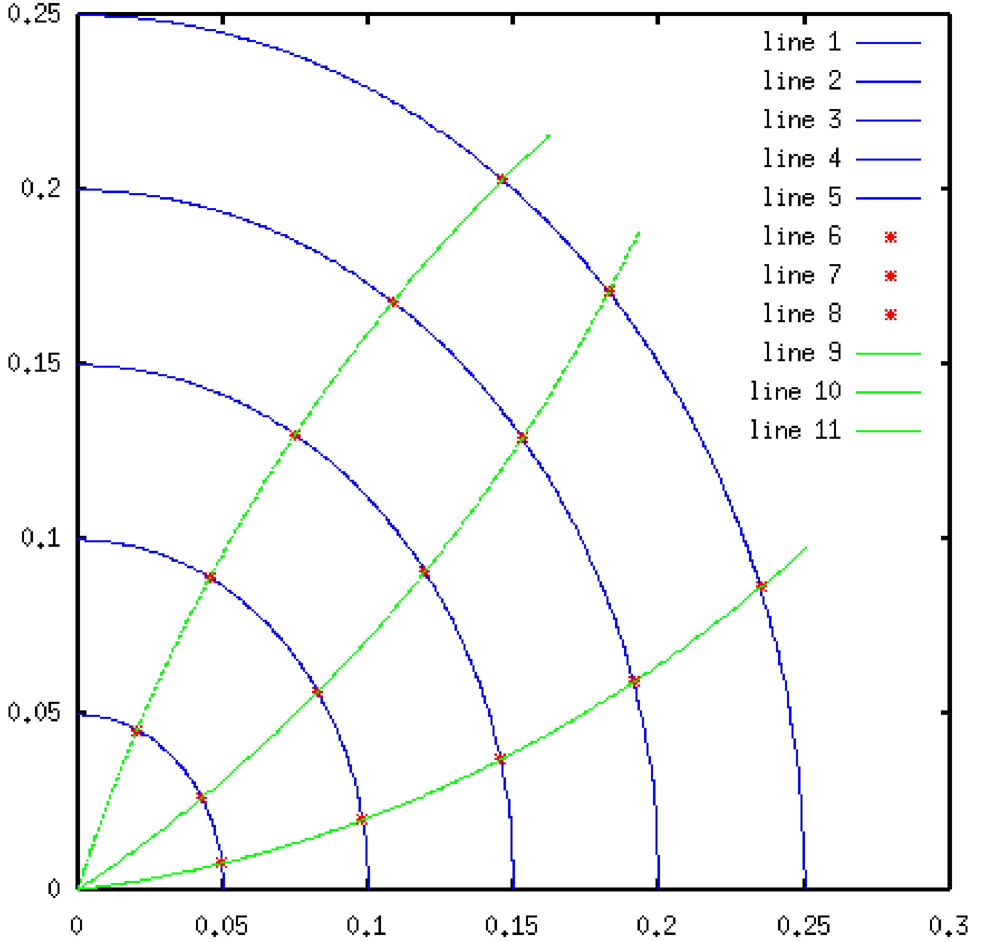


Track finding

Introduction

- **Track fitting: for each of the subsets provided by the track finder, find the optimal estimate of a set of parameters uniquely describing the state of the particle somewhere in the detector.**
- **Example: momentum (absolute value), direction and position at the surface of the detector unit closest to the beam.**
- **The parameters of the tracks are used in higher-level analyses, for instance in vertex reconstruction.**

Introduction



Track fitting

Introduction

- **Track reconstruction - at least track fitting part - has always used least-squares methods in some disguise.**
- **The global least-squares fit - taking all measurements into account simultaneously in the track fit - was dominant before LEP.**
- **Since LEP the Kalman filter has been the most widely used method.**
- **The ability to use the Kalman filter also during track finding enables fast pattern recognition, even usable at a high-level trigger.**

Introduction

- Recently a new class of least-squares track reconstruction methods - adaptive methods - have been developed for the main purpose of coping with the high level of noise and ambiguities at LHC.
- These lectures will give an introduction to the basic features of such methods, point to other applications where similar methods are used, and give examples of their performance in an LHC setting.

Adaptive methods

- **Consider the problem of fitting a set of measurements to a straight line.**
- **Usually this is done by minimizing**

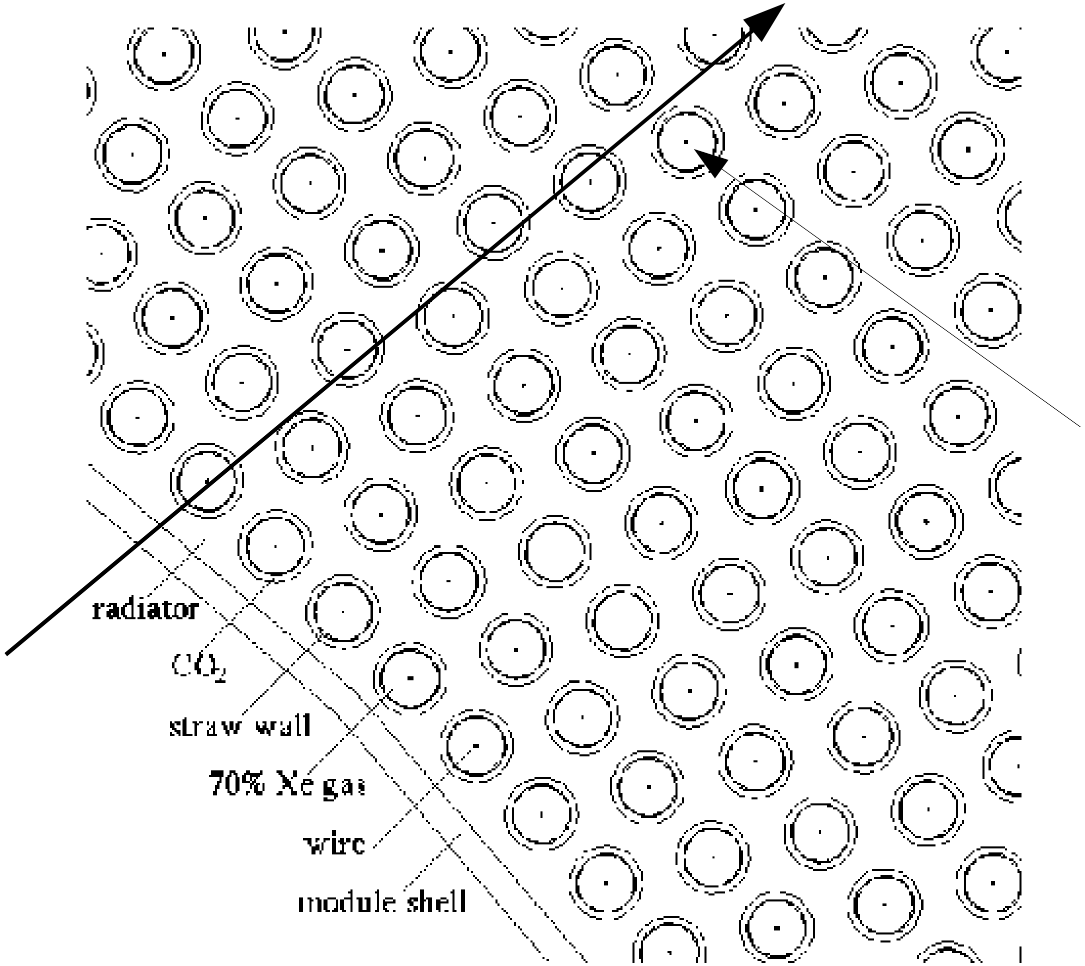
intercept

$$E = \sum_{i=1}^{N} \frac{(y_i - (a_0 + a_1 x_i))^2}{\sigma_i^2}$$

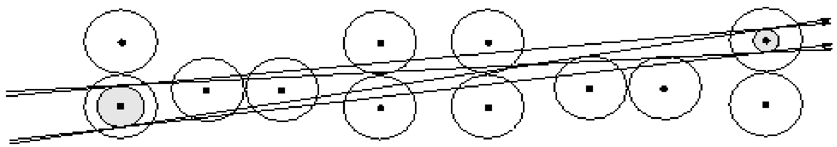
slope

with respect to the parameters of the line.

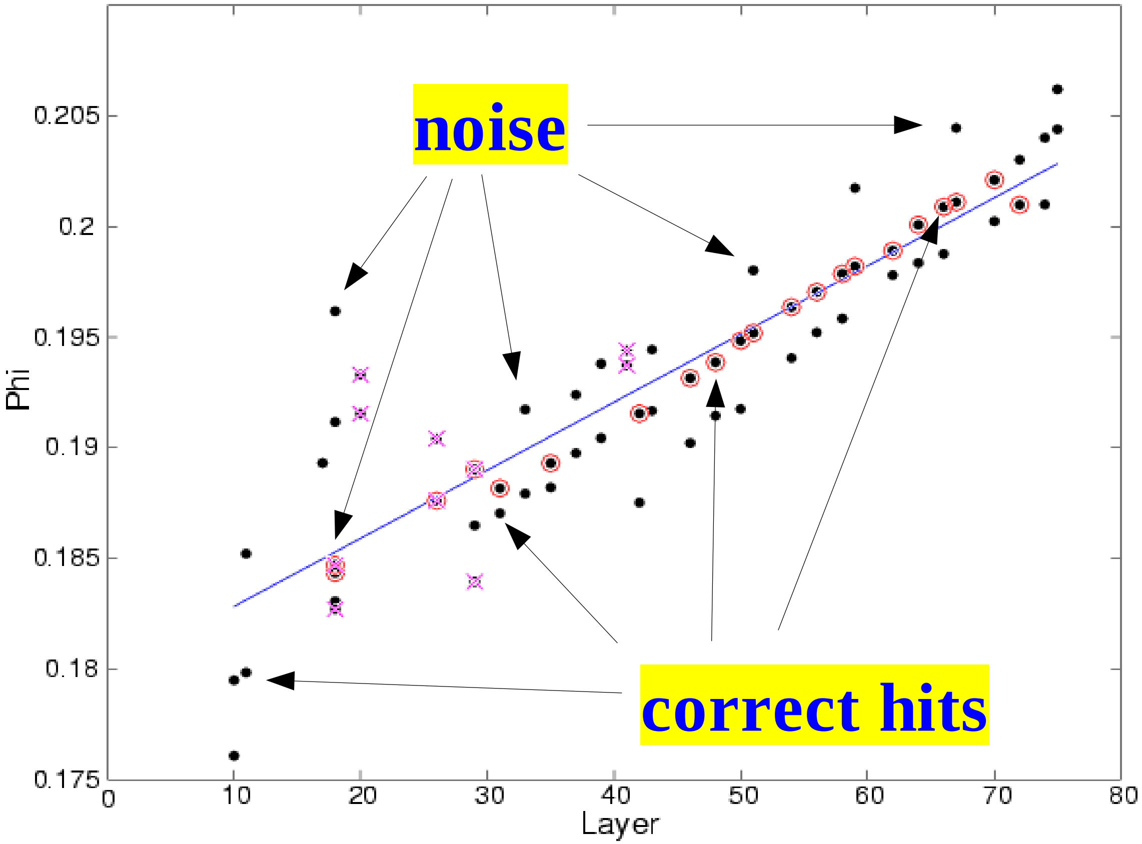
Adaptive methods



Track going through ATLAS TRT



Adaptive methods



- Line shows least-squares fit with equal weight to all measurements
- For most of the correct hits there is an additional noise hit
- Fit is clearly distorted by presence of noise!

Adaptive methods

- **We would of course like to get rid of the influence of the noise hits on the fit.**
- **This can be achieved by minimizing a slightly modified cost or energy function:**

$$E = \sum_{k=1}^{M} \sum_{j=1}^{n_k} s_{jk} \cdot \frac{d_{jk}^2}{\sigma_k^2} \qquad d_{jk} = y_{jk} - (a_0 + a_1 x_k)$$

assignment variables

with respect to the binary assignment variables (taking values 0 and 1, sum of them inside a layer is always one) and the parameters of the line.

Adaptive methods

- **This is a *combinatorial optimization problem*, i.e. it requires exploring a discrete set of combinations during the estimation procedure.**
- **A full exploration requires** $2^M = \exp(M \log 2)$ **combinations to go through**
  - **increases exponentially with the number of layers with measurements!!**
- **With a reasonably large number of layers with measurements, this approach becomes computationally intractable.**

Adaptive methods

- **An adaptive approach would be:**
  1) **Let assignment variables or weights take continuous values between 0 and 1 and initialize them close to one another.**
  2) **Minimize the energy function given previously.**
  3) **Re-calculate weights such that measurements closest to the line tends to get higher weight, whereas other measurements in the same layer get lower weight.**
  4) **Minimize energy and re-calculate weights iteratively until some convergence criterion is fulfilled.**

Adaptive methods

- **Such a procedure hopefully gives correct measurements high weight and wrong measurements low weight.**
- **It is adaptive:**
  - **the weight of a measurement depends on the positions of the other measurements in the same layer competing for inclusion into the fit.**
- **Assume that weights are defined as follows:**

change of notation in continuous case

$$p_{ik} = \frac{\exp\left(-\hat{d}_{ik}^2/T\right)}{\sum_{j=1}^{n_k} \exp\left(-\hat{d}_{jk}^2/T\right)},$$

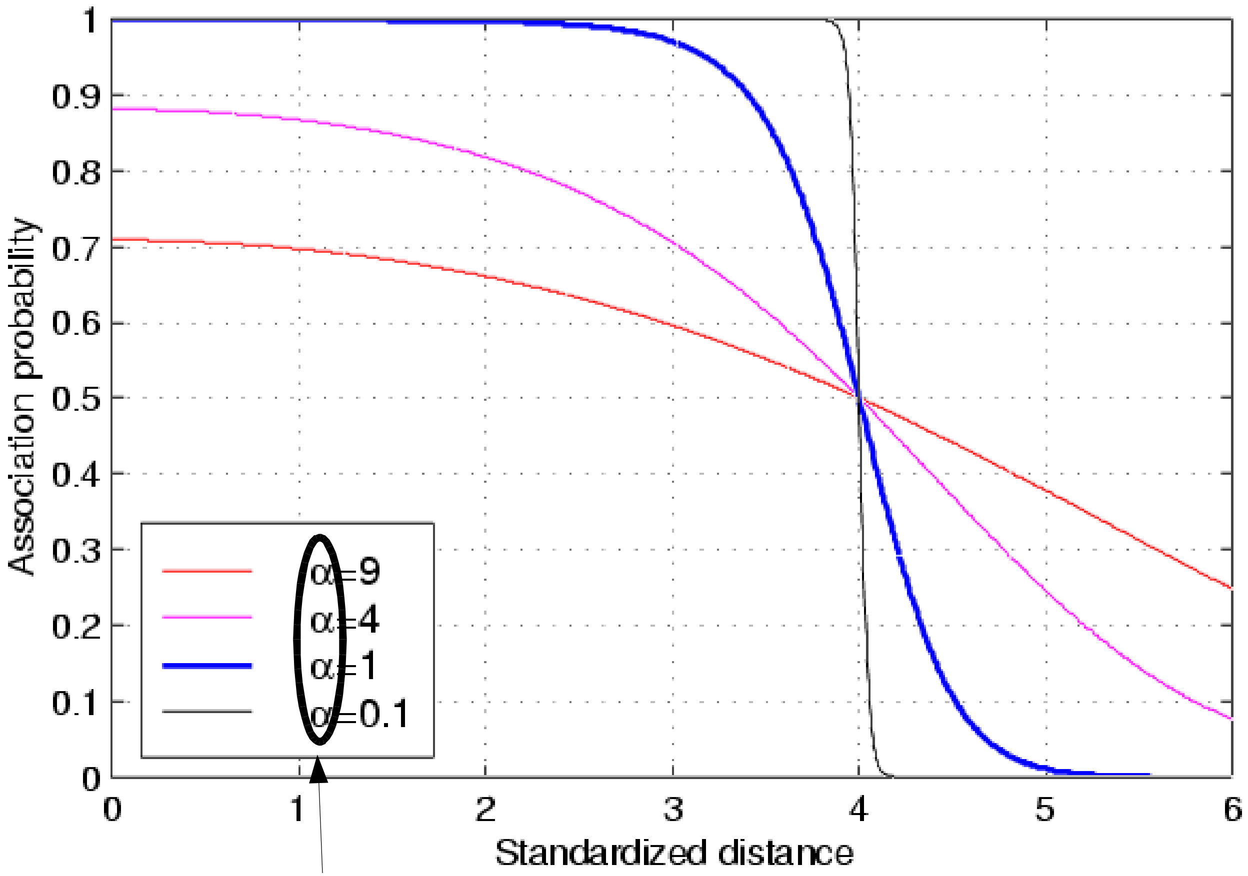
$$\hat{d}_{ik} = d_{ik}/\sigma_k$$

normalized distance

Adaptive methods

- **Consequence:**
  1) **At high values of parameter T all weights are close to one another.**
  2) **At low values of T the closest measurement will have high weight and the others low weight**
     - **the required behaviour is obtained by starting at high T, successively decreasing T during the iterations and ending up at very low T.**
- **Process of decreasing T is called *annealing*, and the set or sequence of temperatures used is called *annealing schedule*.**
- **Option of low weights for all measurements in a layer exist.**

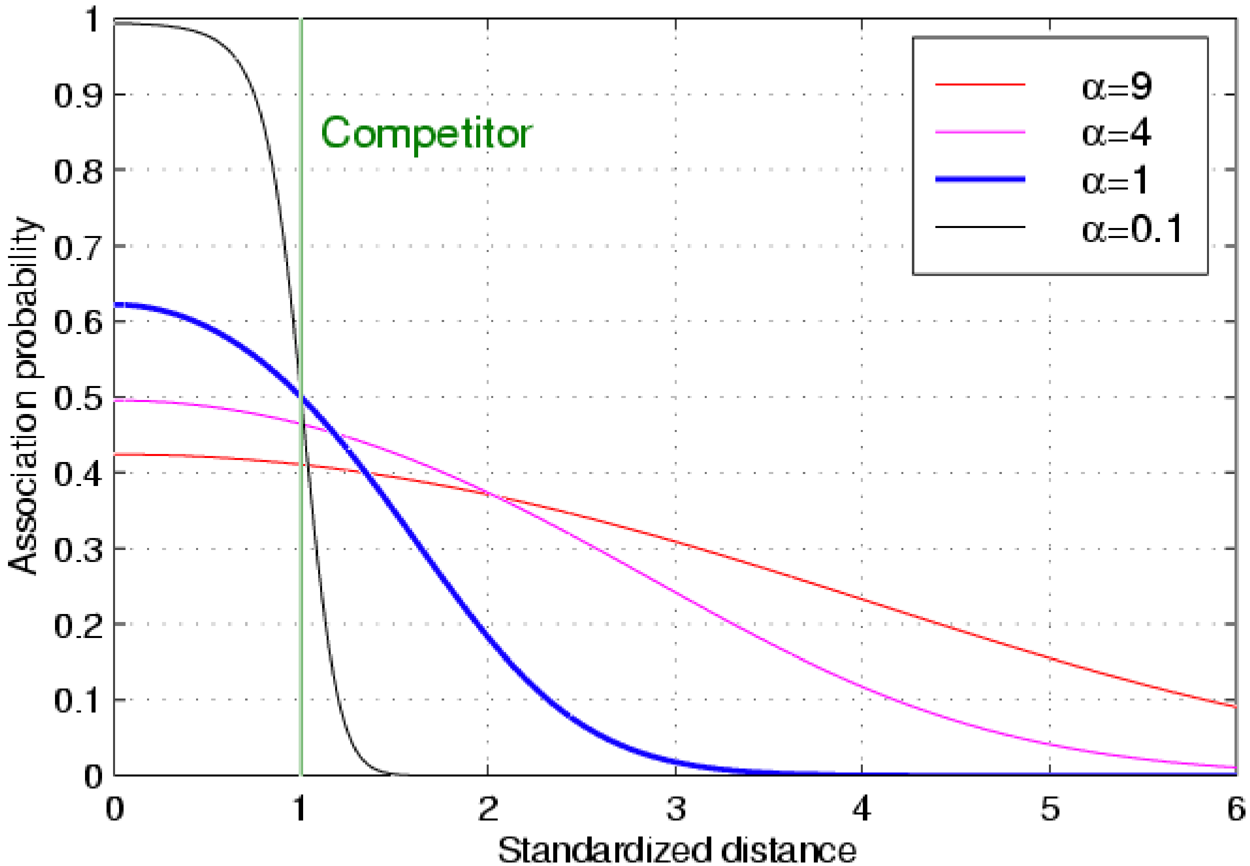
Adaptive methods



- **Weight as function of distance from line**
- **Cutoff at four standard deviations**
- **No competing measurements**

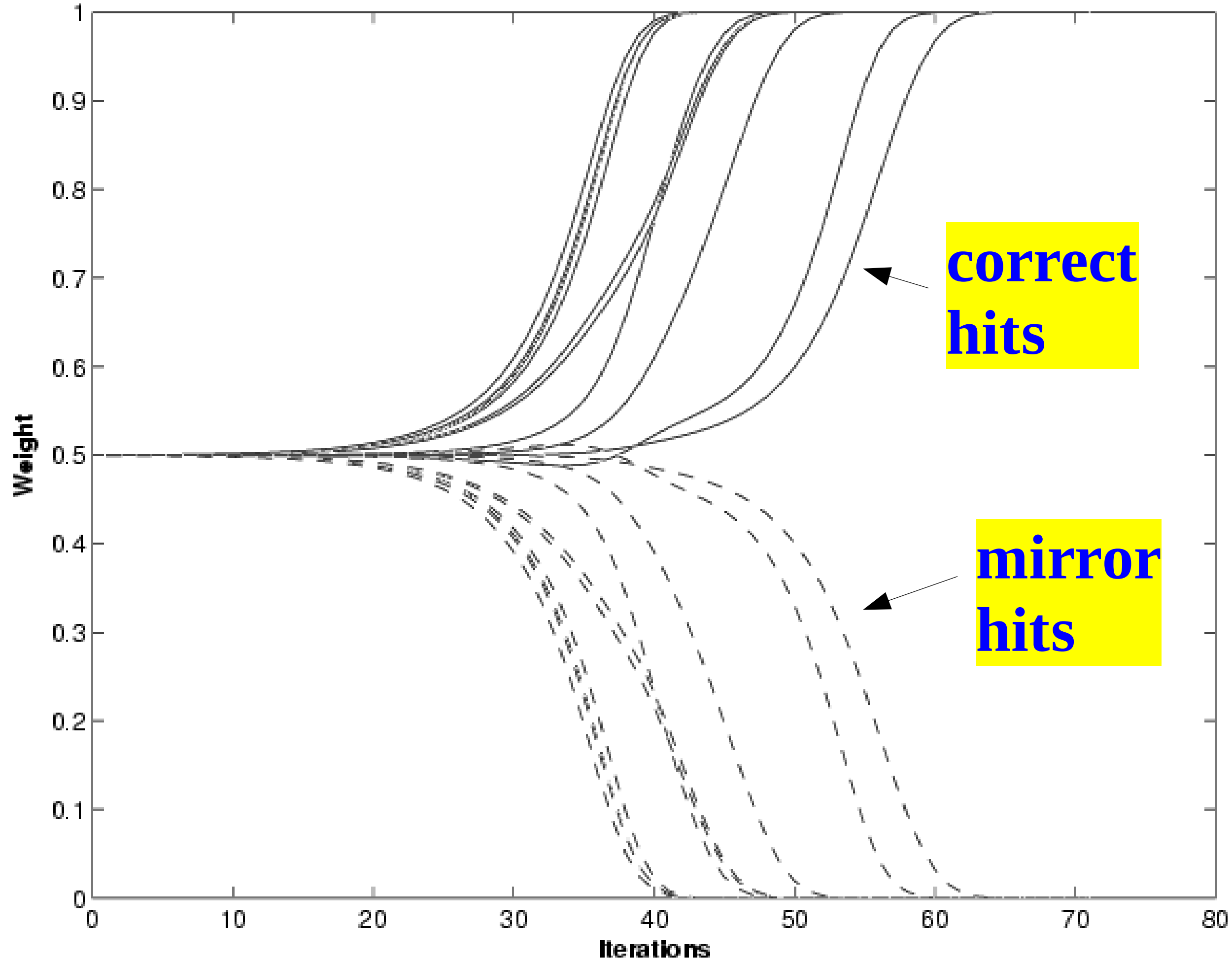
Equiv. to temperature

Adaptive methods



- Competitor at one standard deviation
- Cutoff at four standard deviations
- Behaviour very much affected by presence of competitor!

Adaptive methods



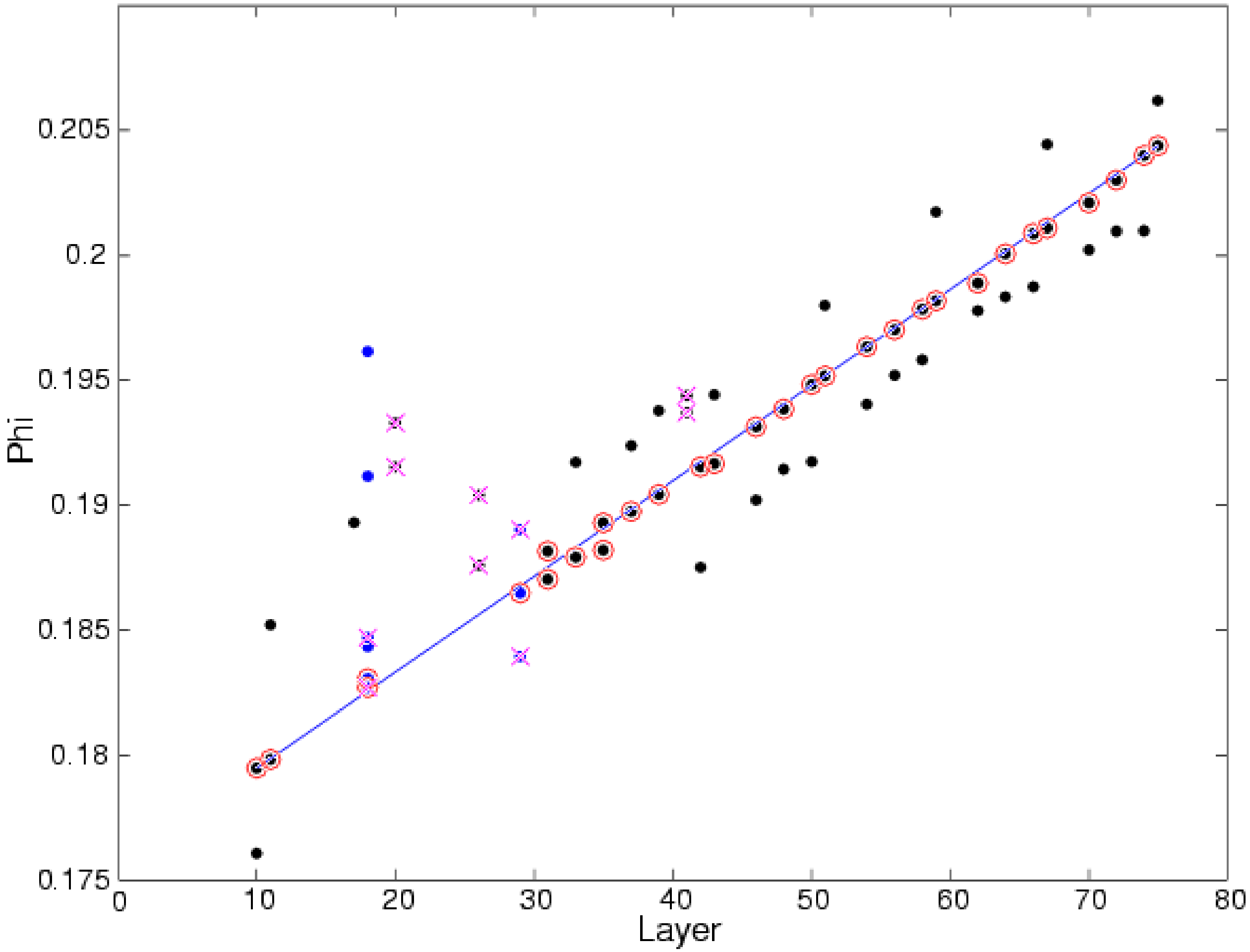
- Evolution of weights in fit as function of iterations
- Weights start out similar at high temperature, end up close to zero/one at low temperature

Adaptive methods

Summary of adaptive algorithm:

- **Initialize weights at starting temperature.**
- **For each of the temperatures in the annealing schedule, repeat until convergence of weights:**
  1) **Obtain parameters of line by minimizing energy function given earlier, regarding weights as fixed quantities.**
  2) **Calculate weights according to current temperature and the parameters of the line.**

Adaptive methods



- Adaptive fit of same set of measurements as shown earlier
- Fit seems to be able to downweight influence of mirror hits and noise and select correct measurements

Adaptive methods

- **For tracking detectors in homogeneous magnetic fields (such as ATLAS Inner Detector and CMS Tracker), the track model is a helix rather than a straight line.**
- **The desired energy function to be minimized is nevertheless basically the same - only the distance measure is different.**
- **To allow for the possibility that none of the measurements in a layer are assigned to the track, the energy function has to be slightly generalized.**

Adaptive methods

- **The expression reads:**

squared cutoff distance

$$E=\sum_{k=1}^{M}\left[S_k\left(\sum_{j=1}^{n_k}s_{jk}\cdot\hat{d}_{jk}^2\right)+\lambda\left(S_k-1\right)^2\right].$$

same meanings as before

also binary (0 or 1), denoting whether all measurements in layer are regarded as noise

Adaptive methods

- **Problem is again minimization of energy function with respect to binary assignment variables and (continuous) parameters of helix.**
- **Also again, this requires combinatorial exploration of all possible configurations of assignment weights**
  - **certainly something to avoid!**
- **Possible way to get rid of combinatorics:**

  1) Require that configurations obey so-called Boltzmann distribution from statistical physics.

  2) Form marginal probability density by summing over all configurations of assignment variables .

Adaptive methods

- **This procedure defines an *effective energy*:**

$$E_{\text{eff}} = -T \cdot \sum_{k} \log\left(n_k \cdot e^{-\lambda/T} + \sum_{j=1}^{n_k} e^{-\tilde{d}_{jk}^2/T} \right).$$

which is seen to be independent of the assignment variables, but implicitely dependent on the track parameters through the distances !!

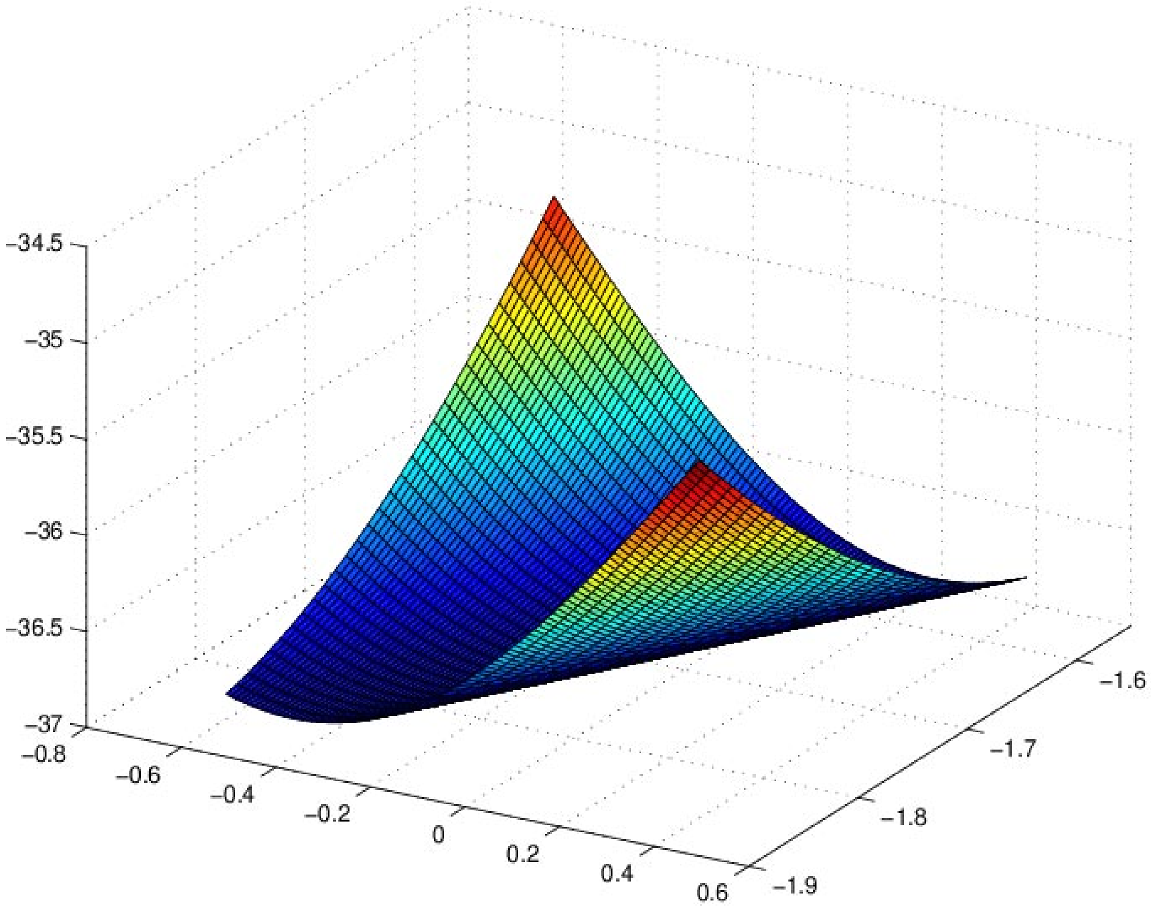
Adaptive methods

- **Strategy is to minimize effective energy at successively lower temperatures and taking zero-temperature limit in the end.**
- **In this limit the effective energy becomes**

$$E_{\text{eff}} = \sum_k \min\left(\{\hat{d}^2_{jk}\}, \lambda\right),$$

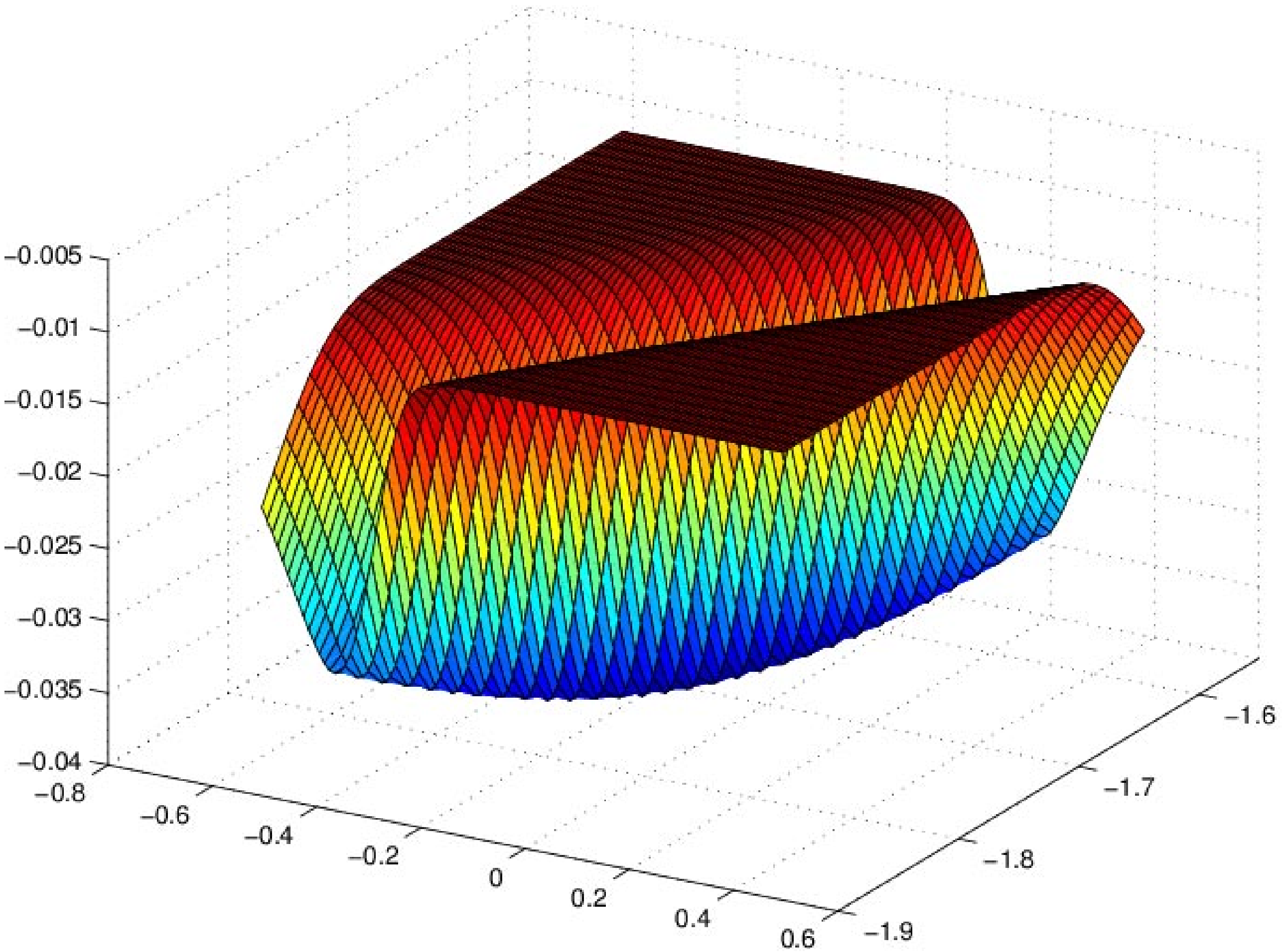
set of squared distances from points in layer k to the track

Adaptive methods

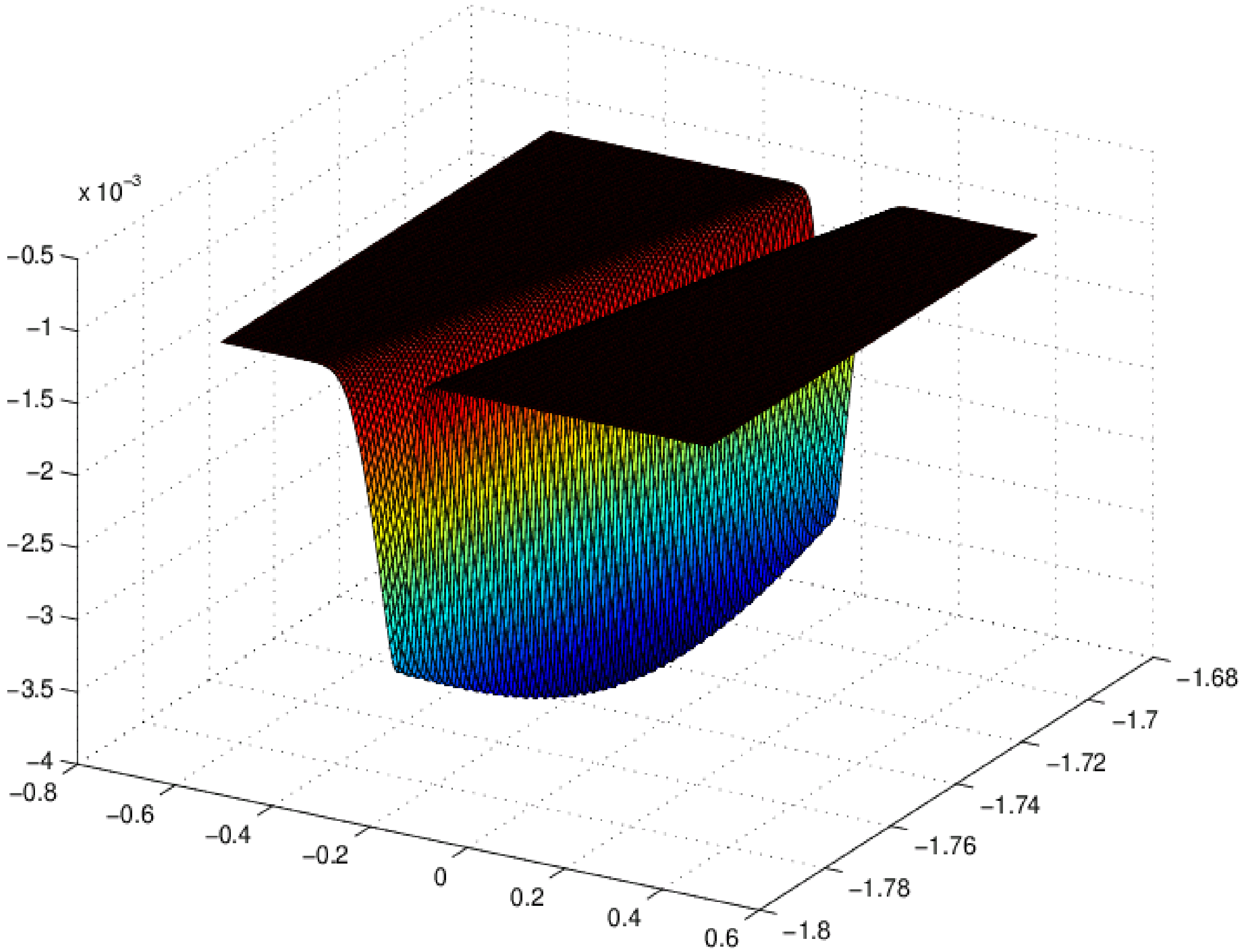


Effective energy landscape for two-parameter case, very high temp. Next slides show evolution as temperature decreases.

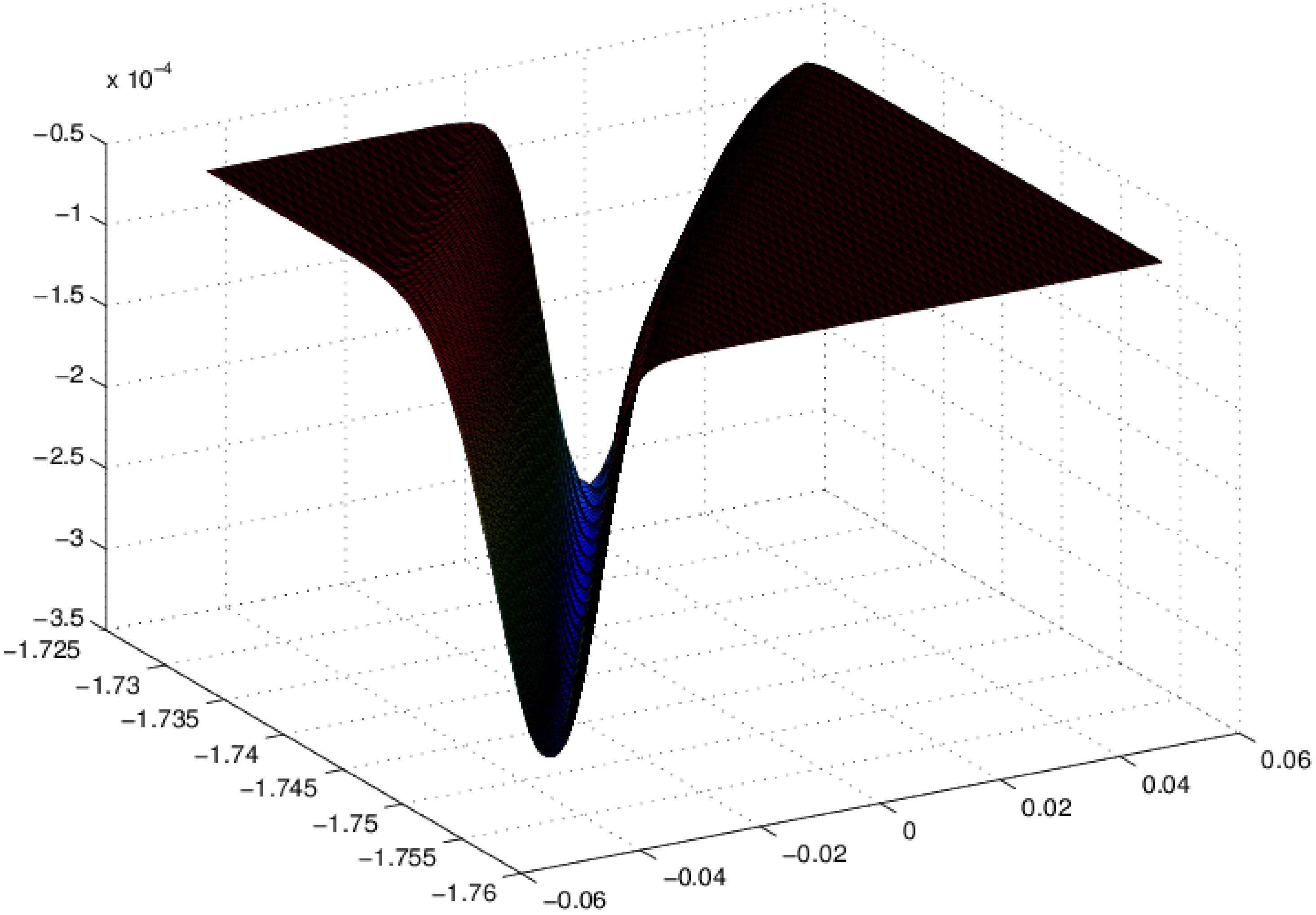
Adaptive methods



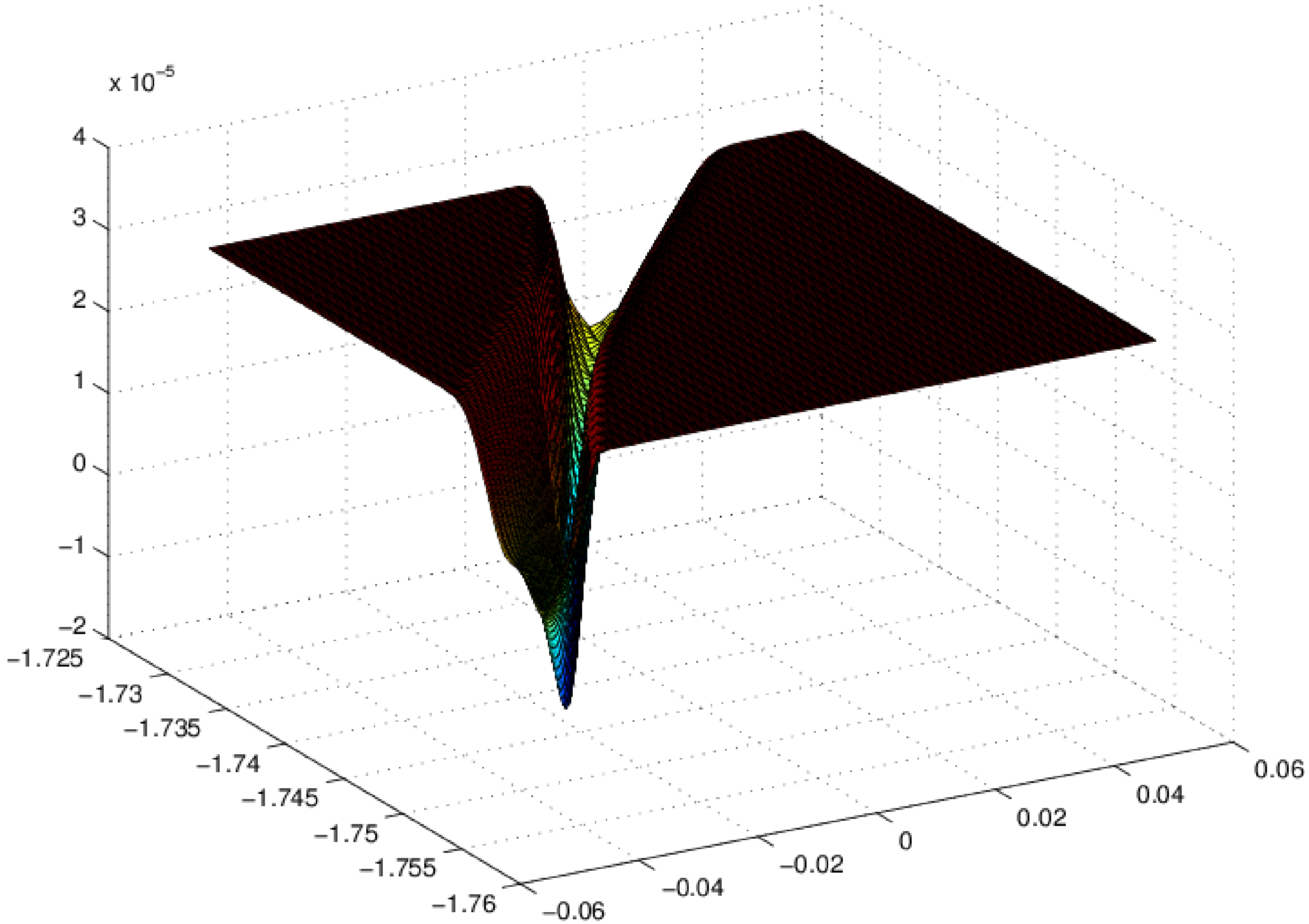
Adaptive methods



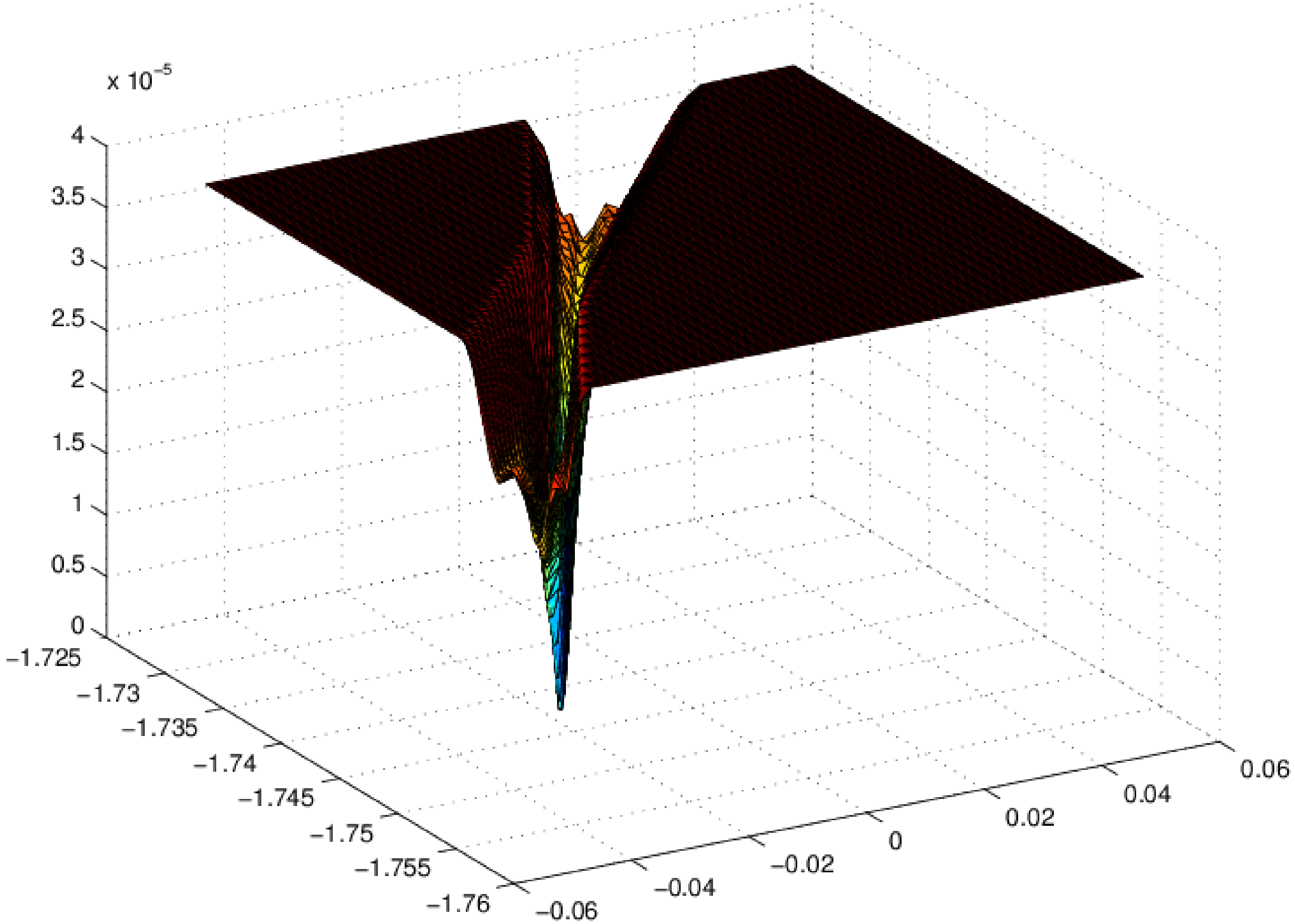
Adaptive methods



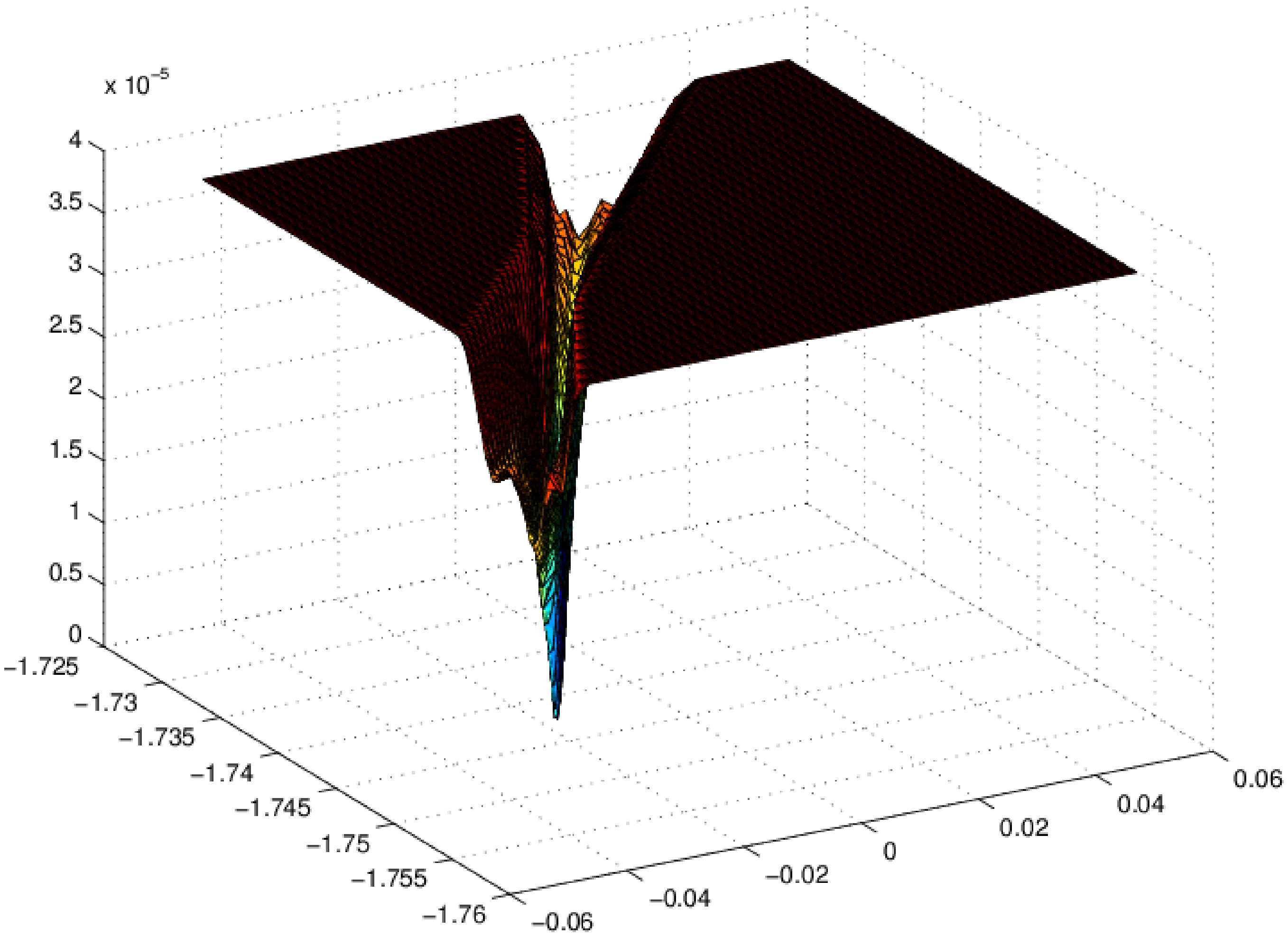
Adaptive methods



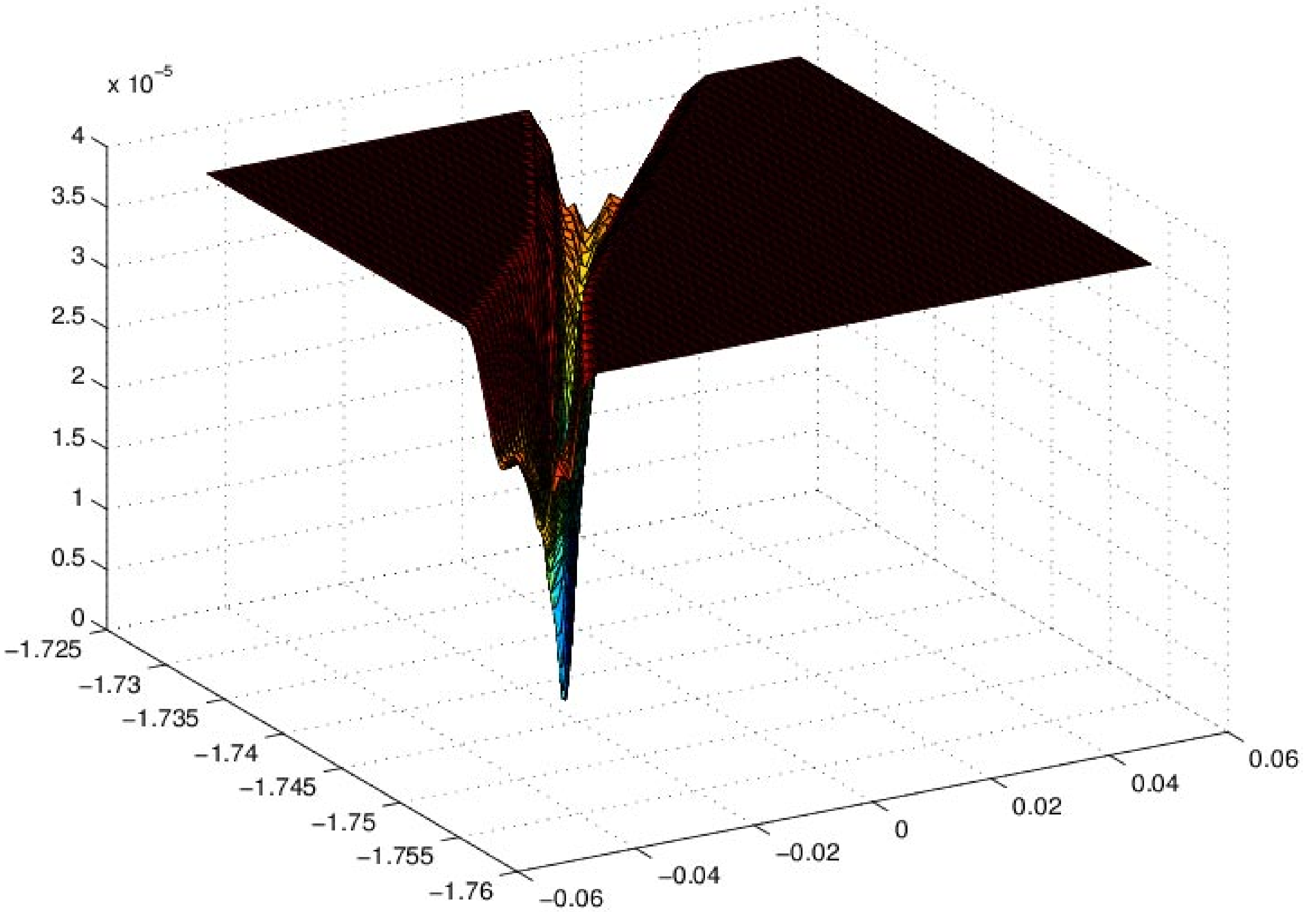
Adaptive methods



Adaptive methods



Adaptive methods



Adaptive methods

- **Non-trivial issue of this algorithm - the *Elastic Arms algorithm* - is the minimization of the effective energy.**
- **The following procedure can be shown to be an alternative way of achieving this goal:**

  1) Calculate the weights according to

$$p_{ik} = \frac{\exp\left(-\hat{d}_{ik}^2/T\right)}{n_k \cdot \exp\left(-\lambda/T\right) + \sum_{j=1}^{n_k} \exp\left(-\hat{d}_{jk}^2/T\right)}.$$

Adaptive methods

2) Minimize

$$Q = \sum_{k=1}^{M} \sum_{j=1}^{n_k} p_{jk} \cdot \hat{d}_{jk}^2$$

with respect to the parameters of the track, regarding the weights as fixed quantities. Repeat until convergence of the weights.

Adaptive methods

- **Key point: minimization of Q is reasonably straightforward, weighted least-squares problem.**
- **Can be done by in principle any least-squares estimator, including the Kalman filter**
  - **Elastic Arms is equivalent to iteratively reweighted Kalman filter with annealing - the *Deterministic Annealing Filter (DAF)*.**
  - **DAF possesses some decisive advantages with respect to standard Elastic Arms (material effects can straightforwardly be taken into account, no elaborate numerical minimization).**

Adaptive methods

- **More general version of Elastic Arms - able to fit several tracks concurrently - can similarly be generalized to a set of Kalman filters with annealing running in parallel**
  - **the *Multitrack Filter (MTF)*.**
- **Crucial feature of MTF is the ability of allowing several tracks to compete for a hit.**
- **It works such that the track closest to the hit has the largest probability of getting the hit assigned.**
- **Any single-track algorithm assigns hits to the track under consideration irrespective of the presence of other, nearby tracks.**

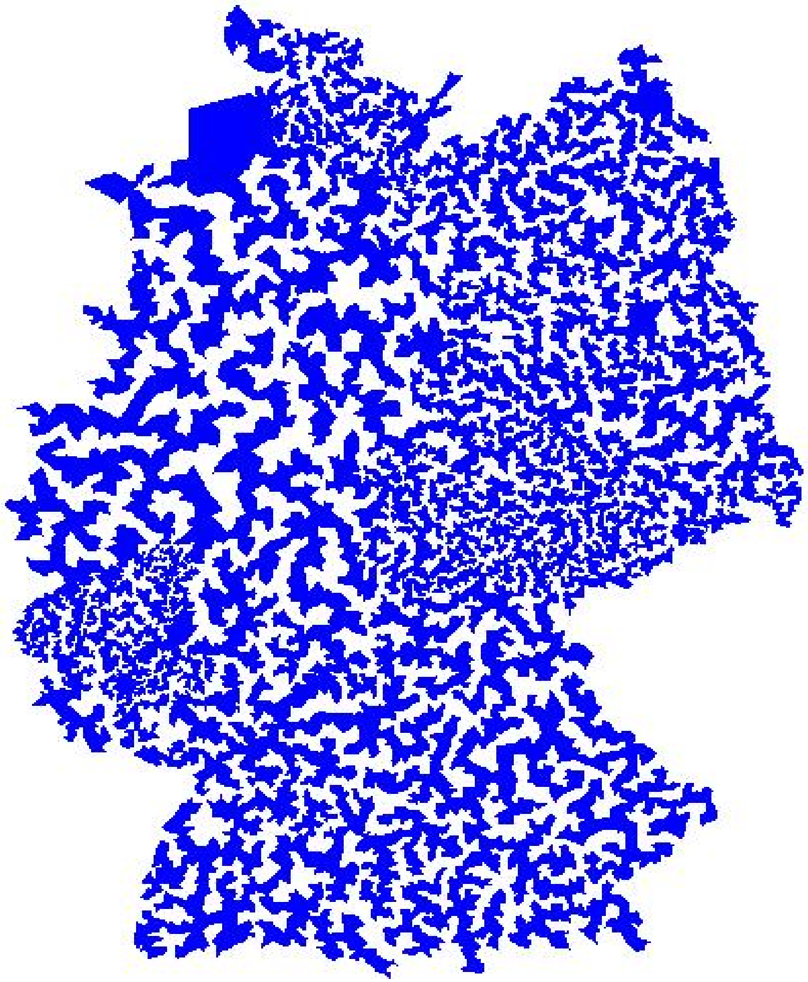
Other applications

- **Track reconstruction in a noisy environment is a *combinatorial optimization problem*.**
- **Combinatorial optimization is a field of research of its own with a large range of applications.**
- **Solving such problems by constructing and minimizing energy functions is in widespread use.**
- **A few examples of this will be mentioned here.**

Other applications

- **The *Travelling Salesman Problem* is maybe the the most thoroughly studied problem in combinatorial optimization.**
- **Assume that a salesman is going to visit N cities - not any city more than once - such that the overall distance is minimized.**
- **Since the total number of paths is N!, a full exploration of all paths is impossible for large N.**

Other applications

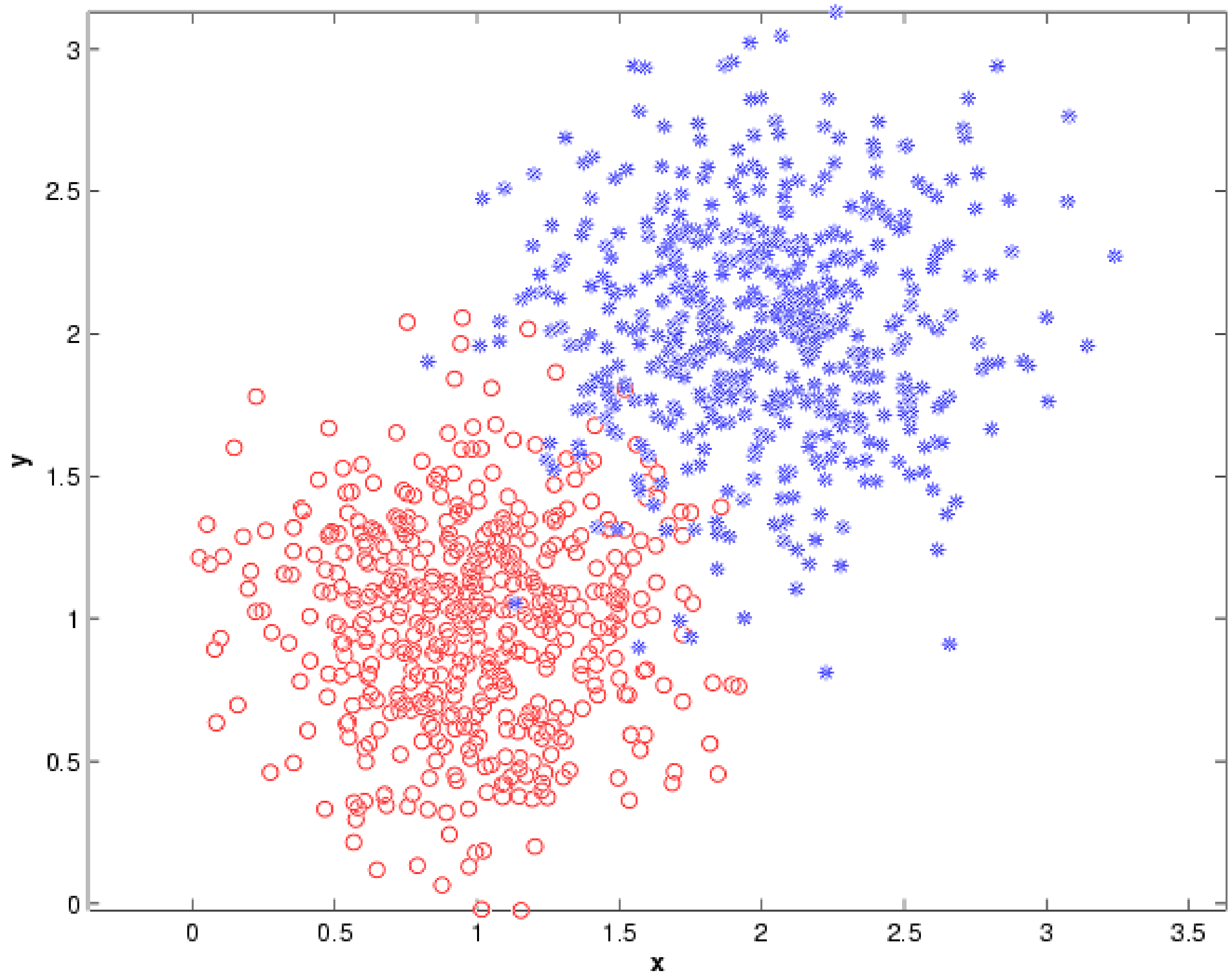


Optimal tour through 15 112 cities in Germany. Computing took 22 years scaled to a 500 Mhz CPU!

Other applications

- **One approach due to Durbin and Willshaw - the *Elastic Net method* - is to construct an energy function and minimize it a successively lower temperatures.**
- **Their approach has been shown to yield very good solutions with acceptable computing cost.**
- ***Clustering* starts out with a set of measured quantities and tries to group them into a number of subsets which naturally belong together.**
- **Finding a representative quantity of the subsets, for instance the mean values, is also desired.**

Other applications



- Example of two clusters
- Two-dimensional position measurements with different true values
- What are roughly the correct values in this case?

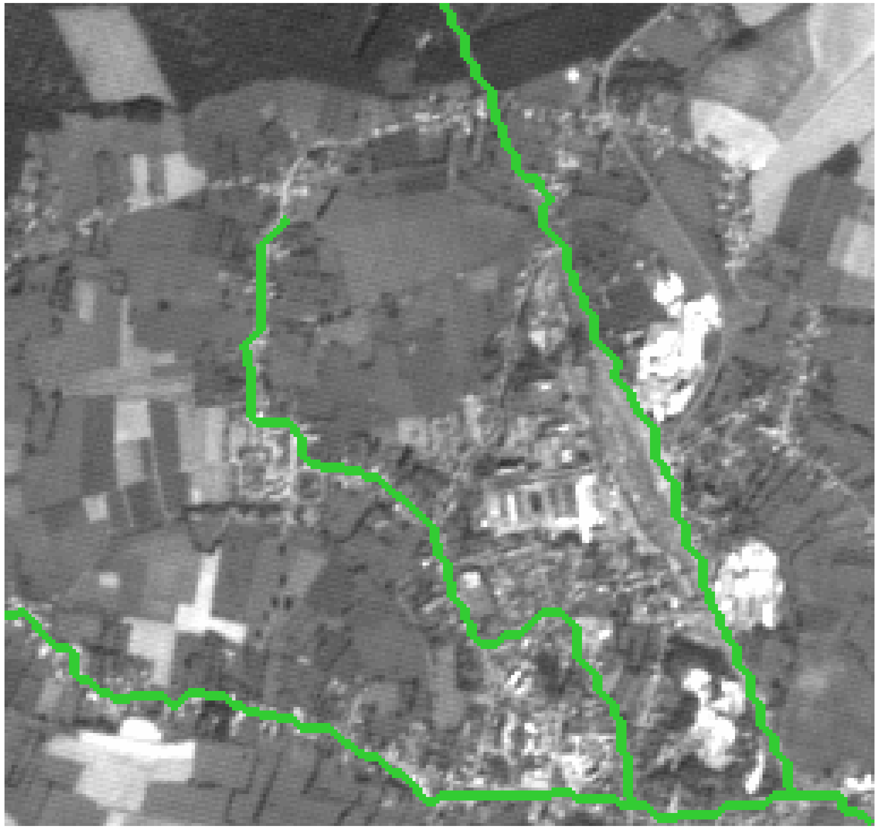
Other applications

- **An algorithm based on the minimization of an energy function with exactly the same structure as the Elastic Arms algorithm has been developed.**
- **In situations with high level of noise, i.e. for a high fraction of measurements not coming from any of the relevant subsets one wants to cluster, this algorithm has been shown to be superior to conventional algorithms.**

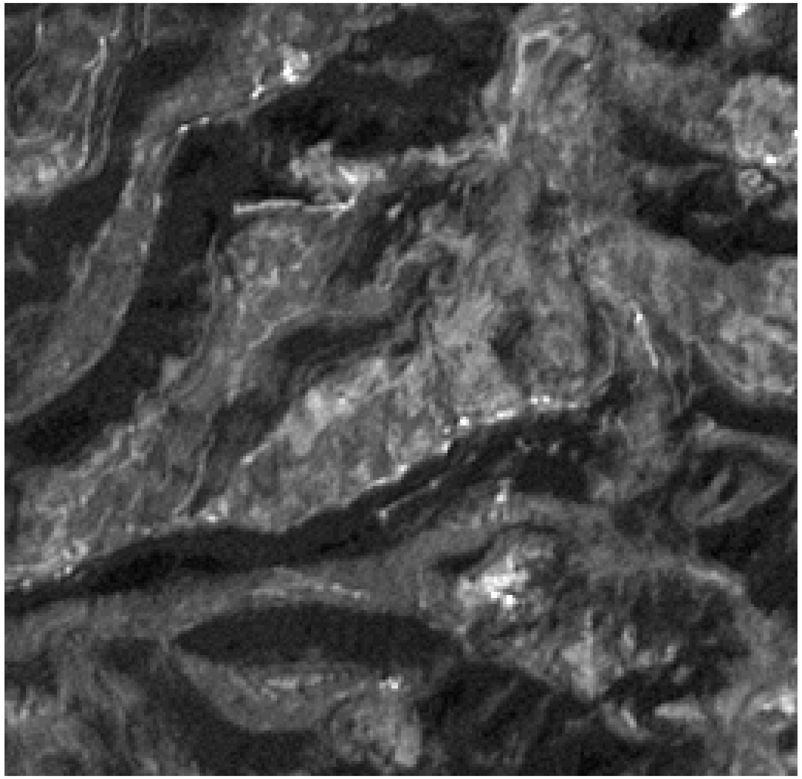
Other applications

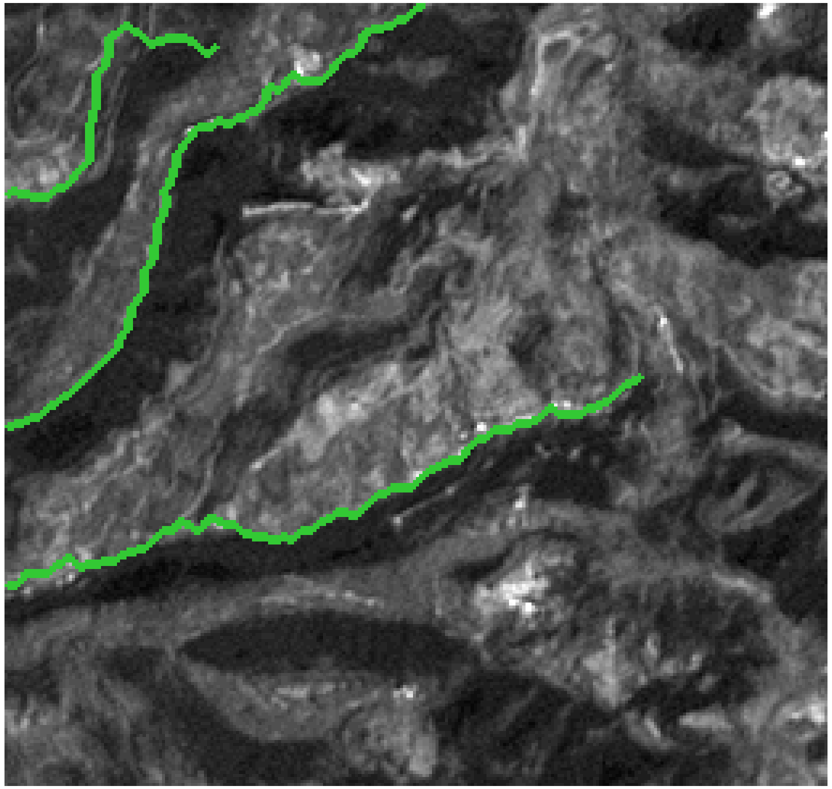
- ***Image processing*** **deals with using computers for extracting information from (digital) images.**
- **Detecting lines and edges in images are important preliminary steps for various high-level vision processes.**
- **These are combinatorial optimization problems, since they basically consist of assigning a binary value (e.g. 0 = no edge, 1 = edge) to every pixel in an image.**
- **Approaches based on minimizing energy functions have been shown to be efficient.**

Other applications



Other applications

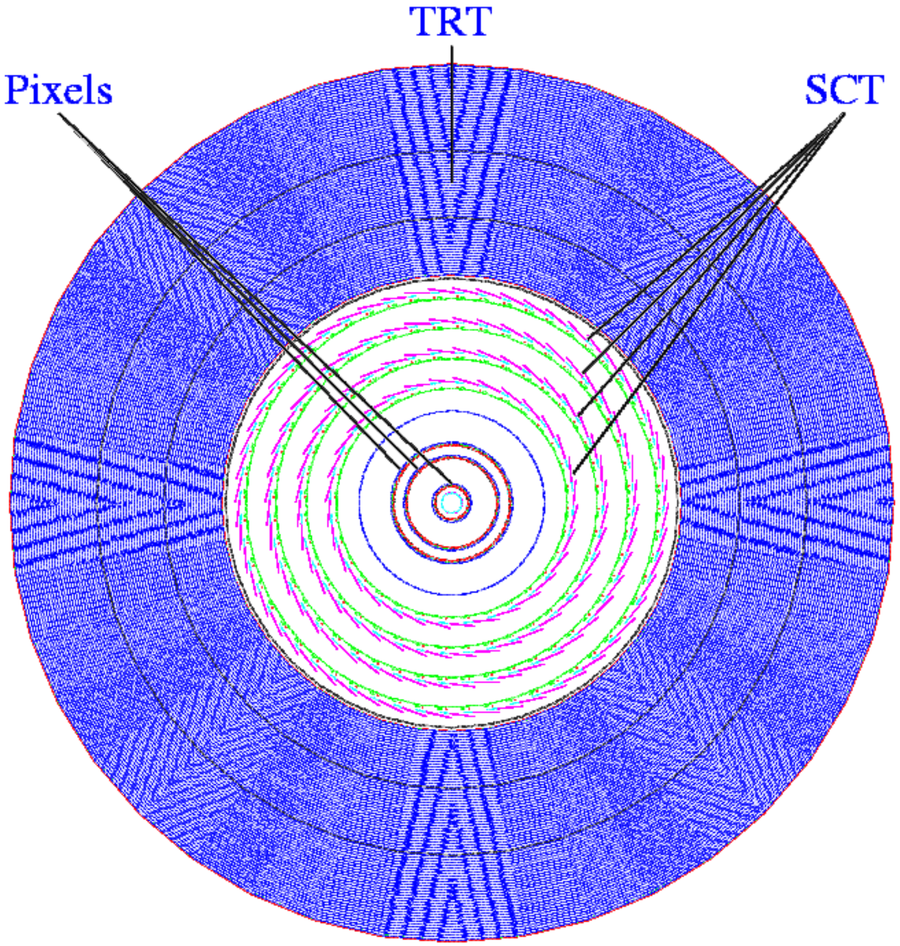




Other applications

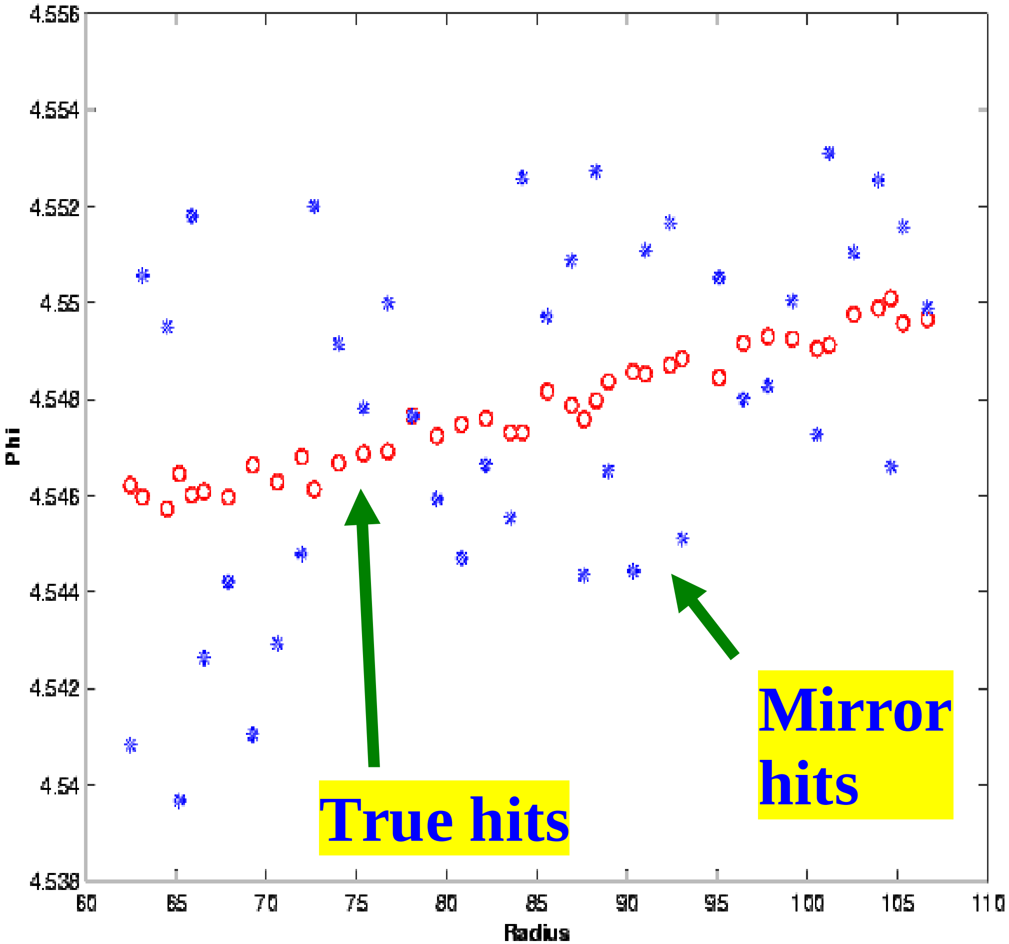
- As for the Elastic Arms algorithm, the strategy is to sum up the effect of the assignment variables in an effective energy and minimize this.
- Introducing a temperature parameter and an annealing schedule turns again out to be vital.

Study - ATLAS TRT



- Drift-tube detector-> left/right ambiguities
- About 35 measurements per track in barrel part of TRT
- Radius of barrel extending from about 50 to 100 cm from the beam

Study - ATLAS TRT



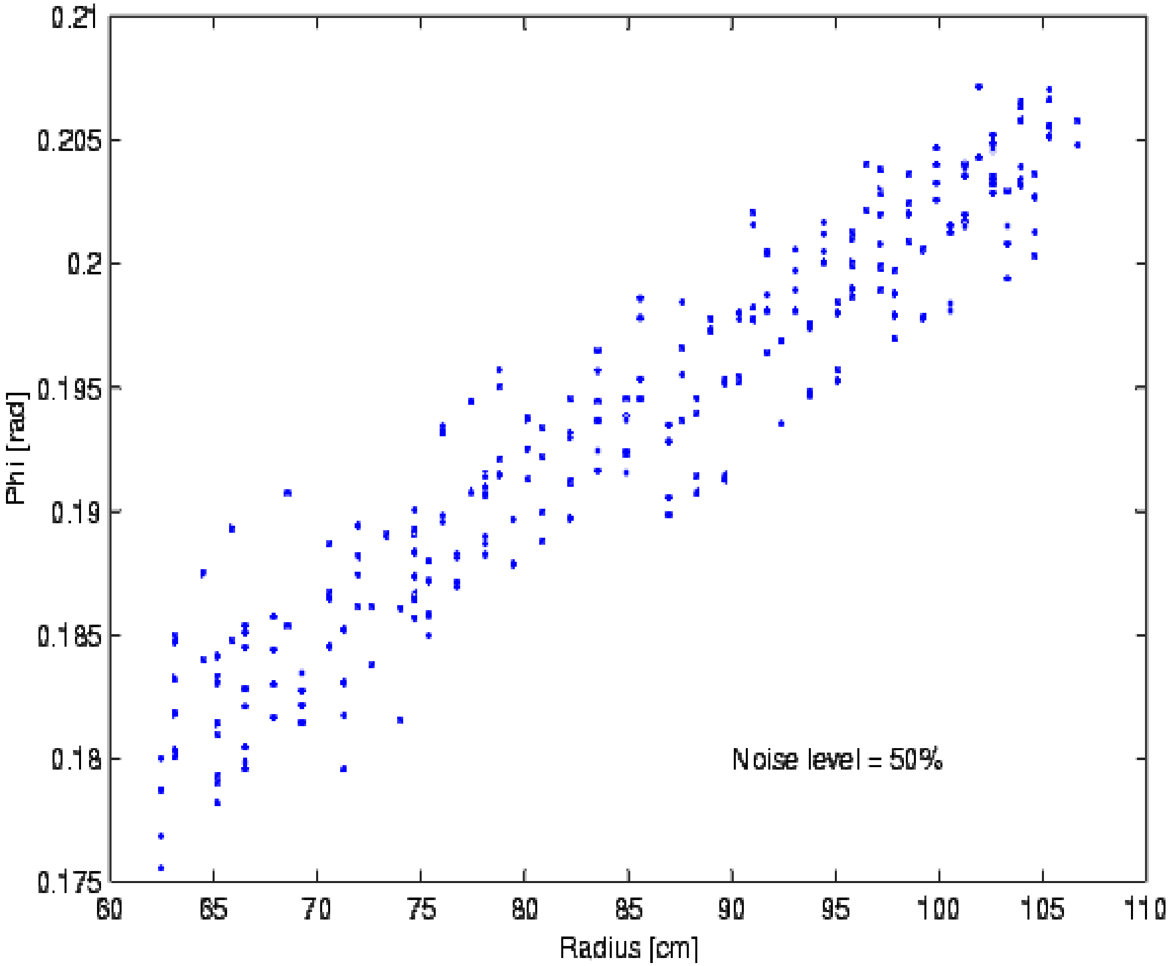
- Evaluating the abilities of the algorithms to solve left/right ambiguities
- Assuming separate pattern recognition has been done first
- Initialization by least-squares fit to straight line in R-Phi-plane

Study - ATLAS TRT

Method	V_{rel}	t_{rel}
DAF nominal	1.54	1.21
DAF frozen	1.74	1.41
GSF all	1.59	7.04
GSF best	1.78	7.04
EAA nominal	1.56	2.12
EAA frozen	1.71	2.44
KF	~ 1500	0.08

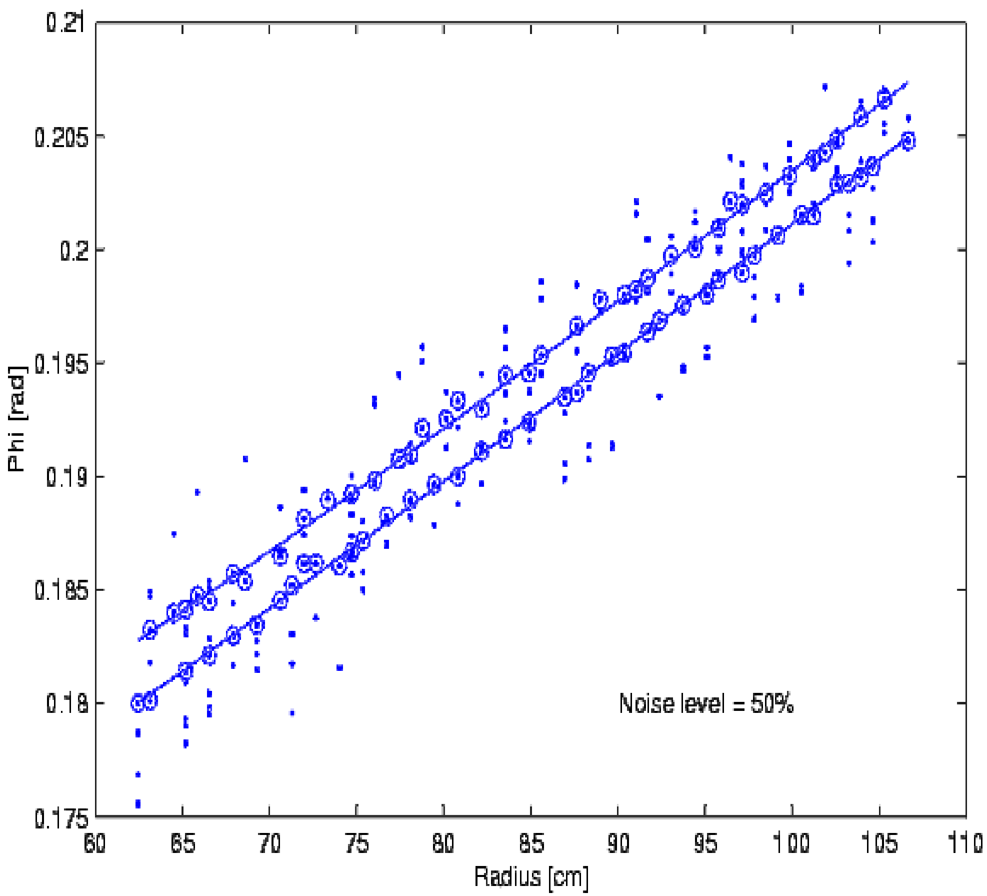
- **Relative generalized variance and relative time consumption for different methods**
- **Baseline is least-squares fit to only true hits**
- **DAF, GSF and EA equally precise (MS turned off), DAF fastest**
- **Zero-temperature limit not optimal**

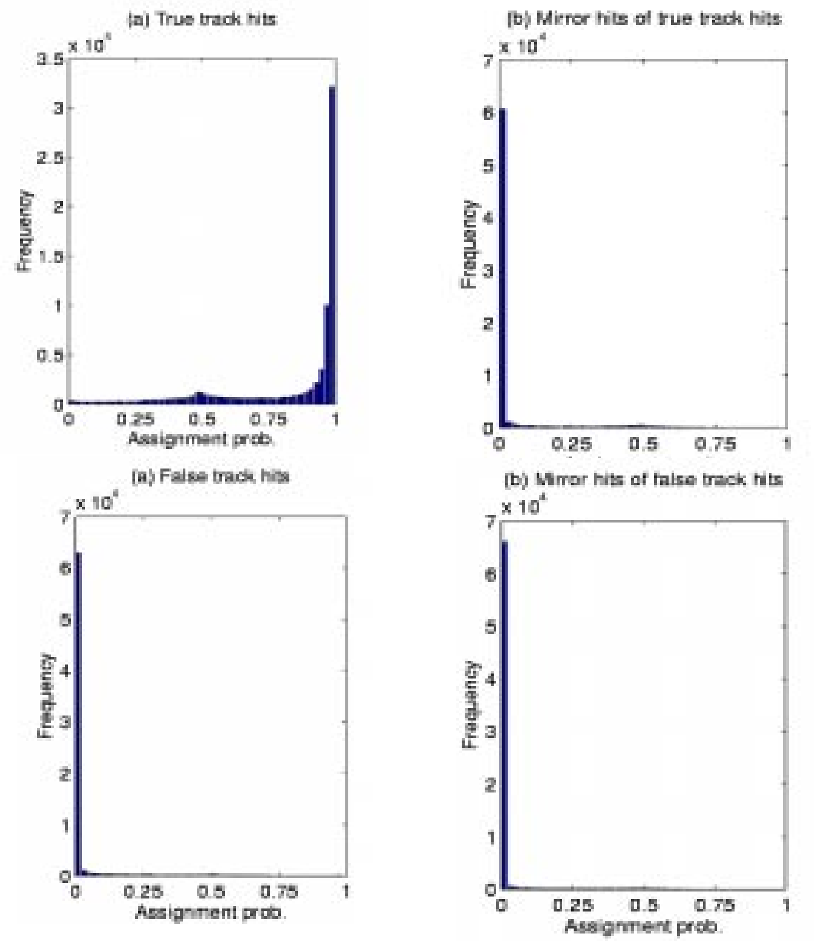
Study - ATLAS TRT



- **Next, simulation experiment of two nearby tracks**
- **Comparing MTF and DAF**
- **Assuming measure-ments of track pair to be found by track finder**
- **Initializing tracks close to true values**

Study - ATLAS TRT



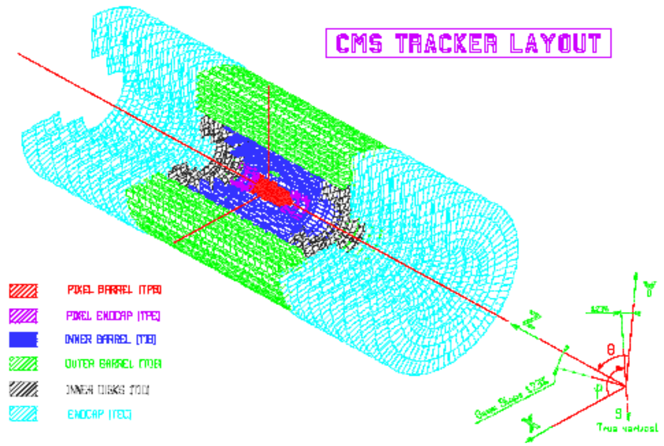


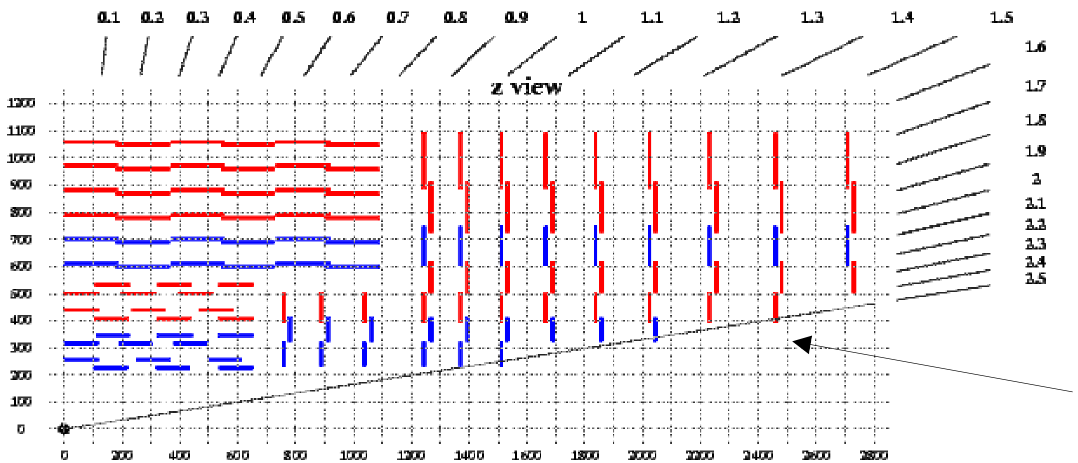
Study - ATLAS TRT

Noise level	Method	
	DAF	MTF
0%	281	4.52
10%	270	5.35
20%	388	6.26
30%	358	7.19
40%	409	9.50
50%	653	11.66

- **DAF is for sure confused by the nearby track**
- **MTF yields very good results, also at high levels of noise**
- **Baseline is generalized variance of fit to each track separately, no noise and no mirror hits**

Study - CMS tracker





- **All-silicon based tracker with pixels closest to the beam and silicon strip detectors outside**

- **Radial extension: ~110 cm**
  Longitudinal: ~2*270 cm

R-z view of one quadrant of CMS tracker

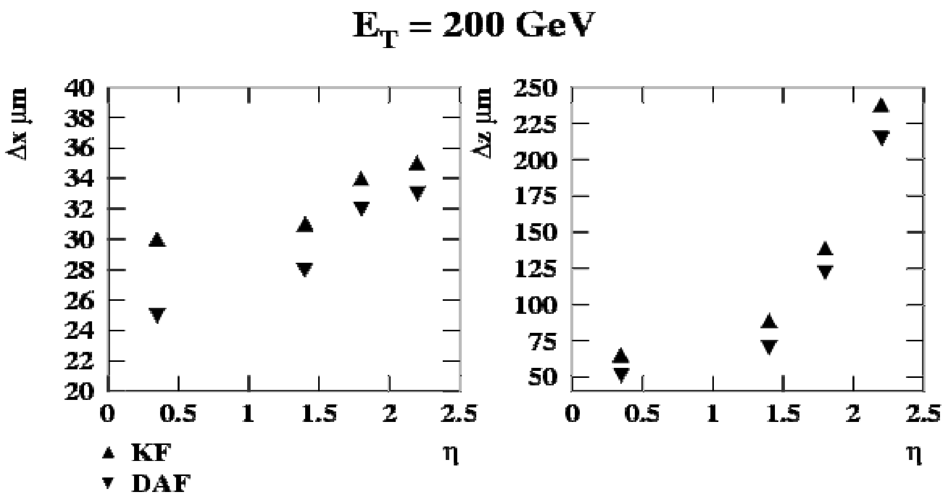
Study - CMS tracker

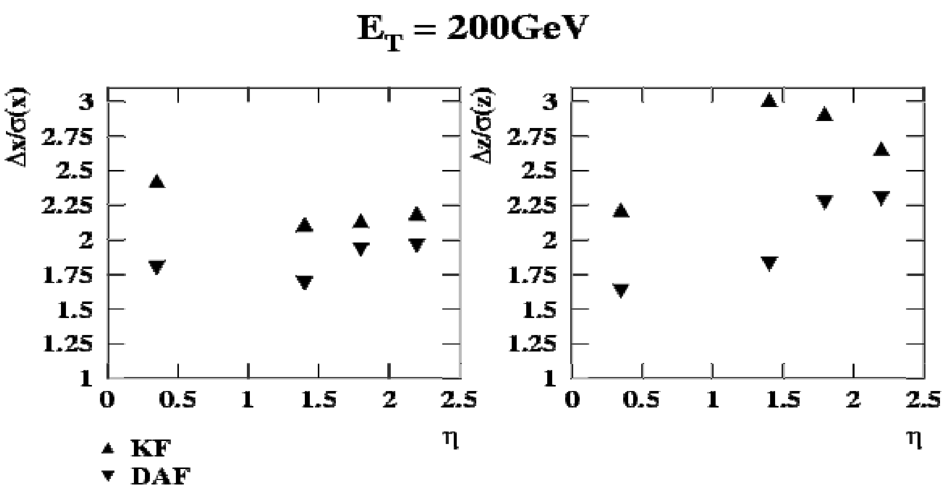
- Both DAF and MTF have been implemented in the official reconstruction framework of CMS - ORCA (M. Winkler, PhD thesis 2002).
- Performance of methods has been thoroughly evaluated and systematically compared to the Kalman filter.
- Track finding (initialization) done by combinatorial Kalman filter, smoothing either by KF, DAF or MTF.
- Will focus on two channels:

  1) 200 GeV transverse energy b-jets.

  2) 3-prong tau jets from 500 GeV SUSY Higgs decay.

Study - CMS tracker





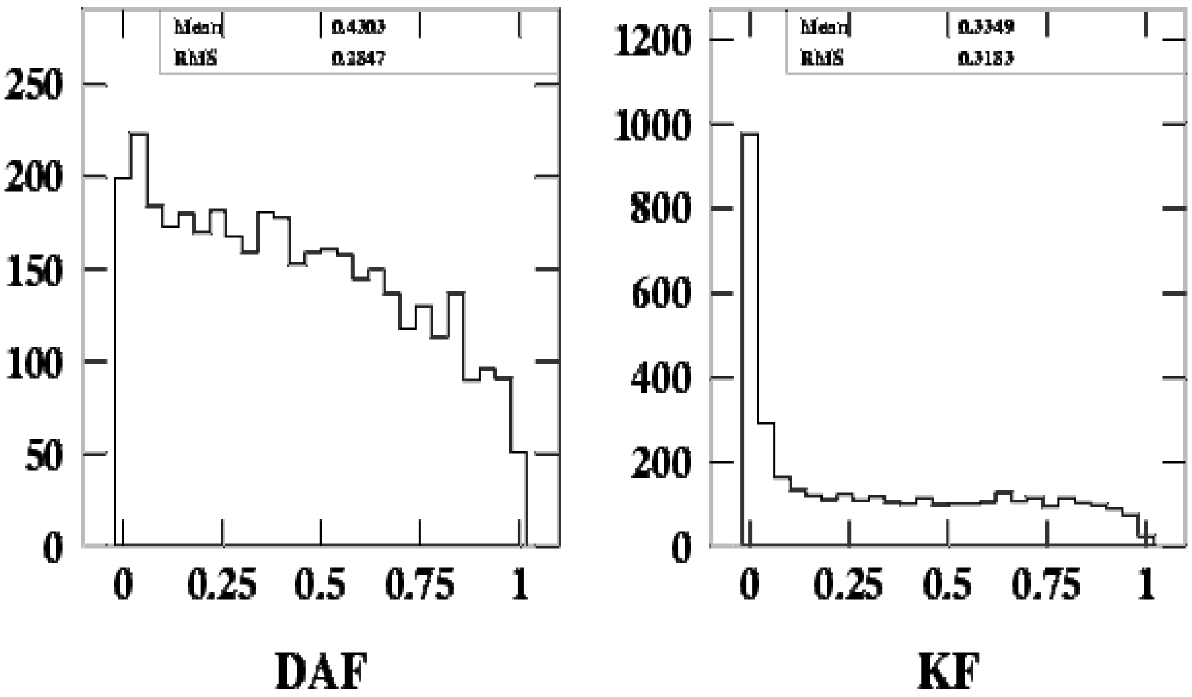
Impact parameter resolutions and pulls of tracks (transverse momentum larger than 15 GeV) in 200 GeV b-jets

DAF is significantly more precise and has better pulls

Tails of residual distributions much smaller with DAF

Study - CMS tracker

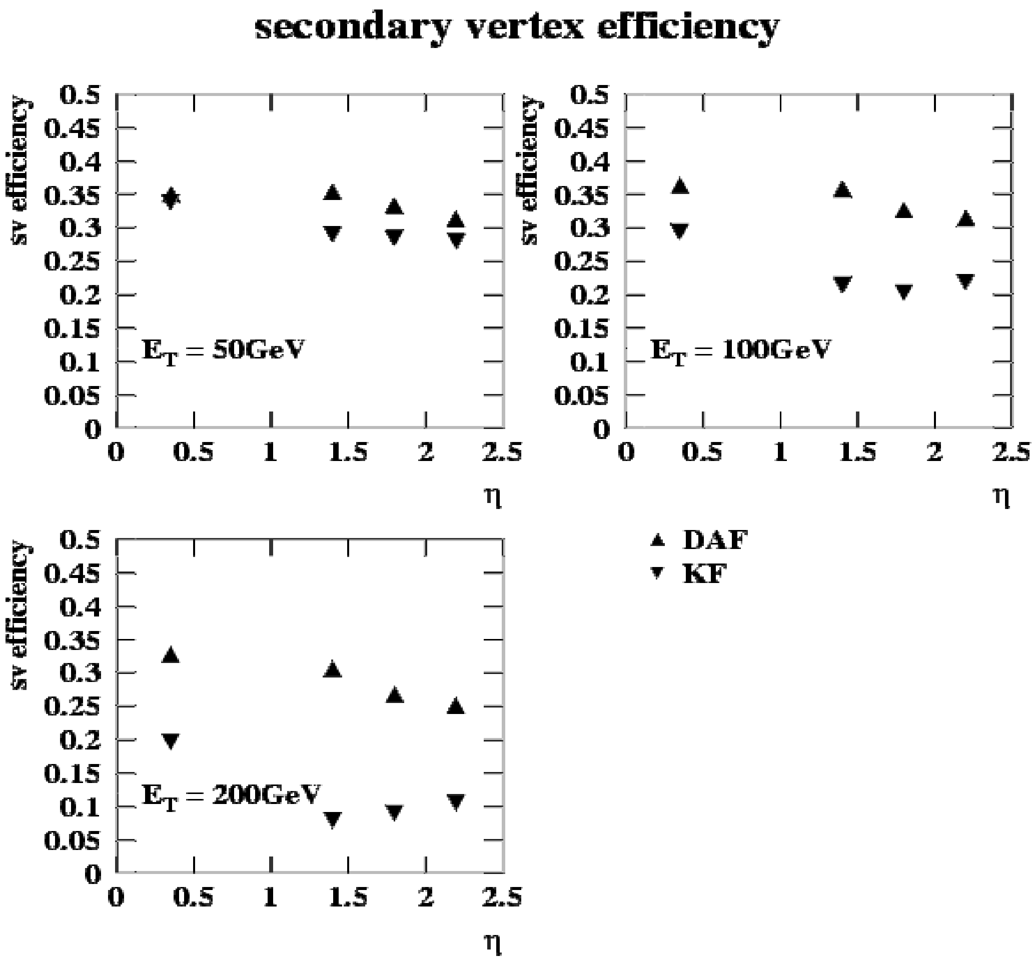
χ^2 probability



Consequently, quality of error estimation is better with the DAF

Histogram not flat -> still room for improvement !!

Study - CMS tracker



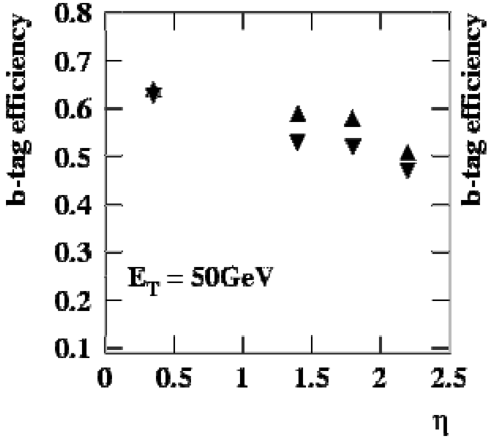
SV finding efficiency for KF depends strongly on jet energy

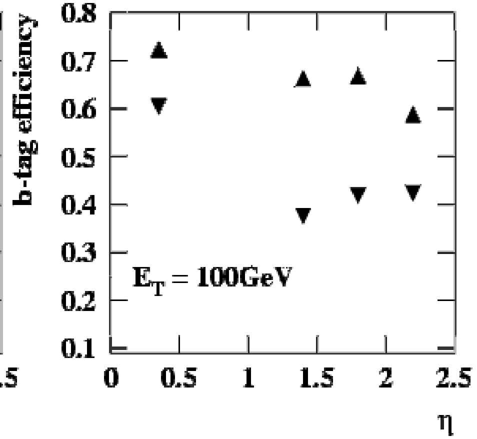
For DAF: efficiency virtually independent of energy

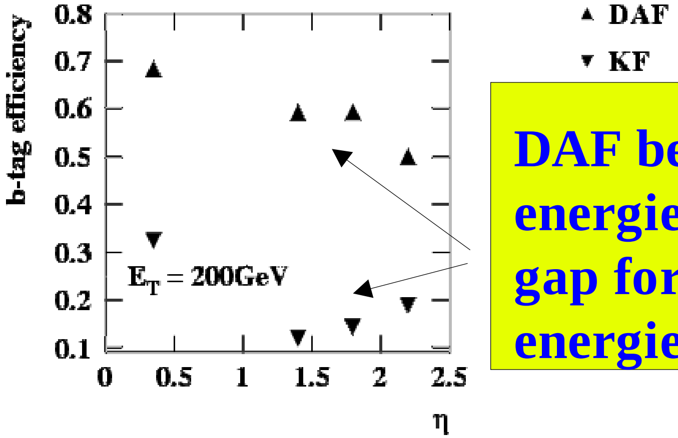
DAF significantly better, particularly at high energies

Study - CMS tracker

b-tagging efficiency

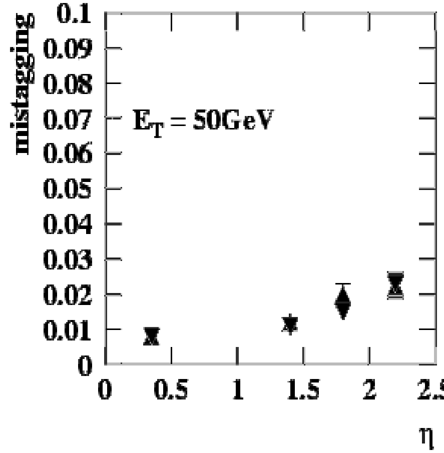


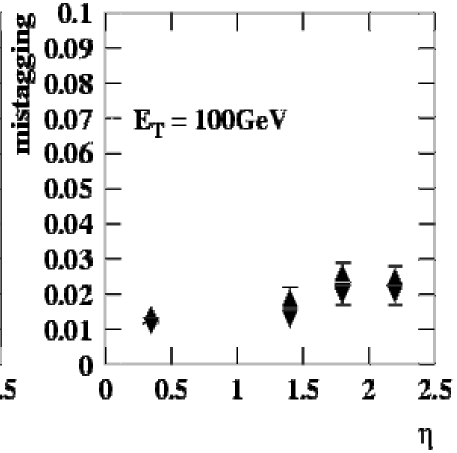


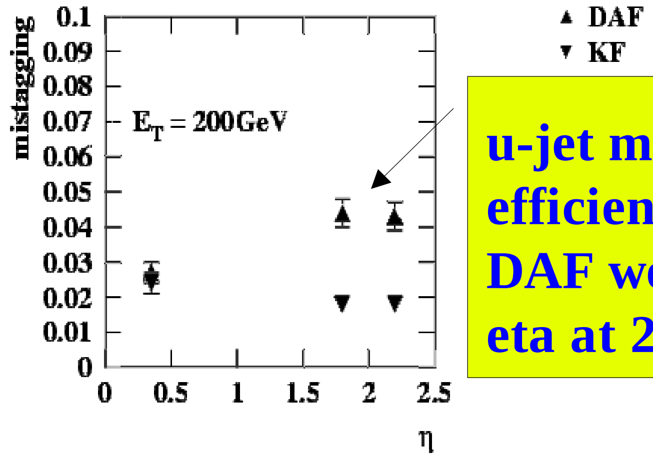


DAF better at all energies, largest gap for high energies

mistagging

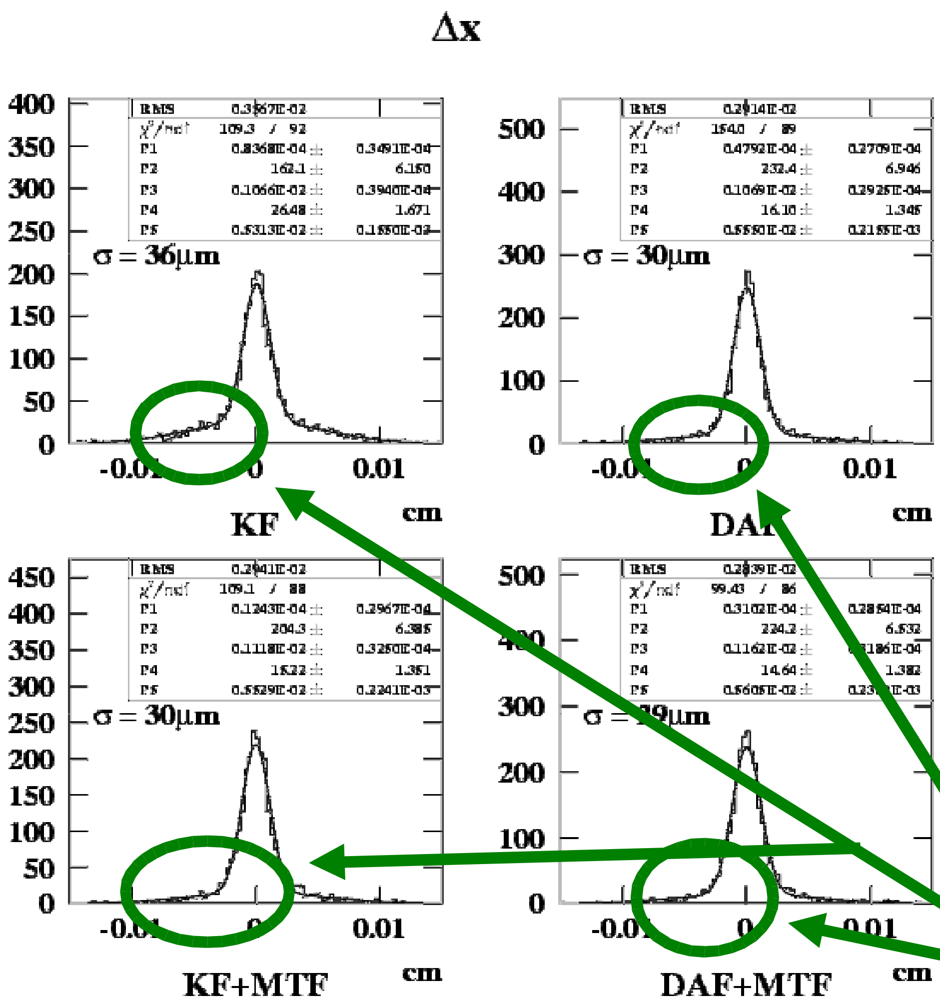






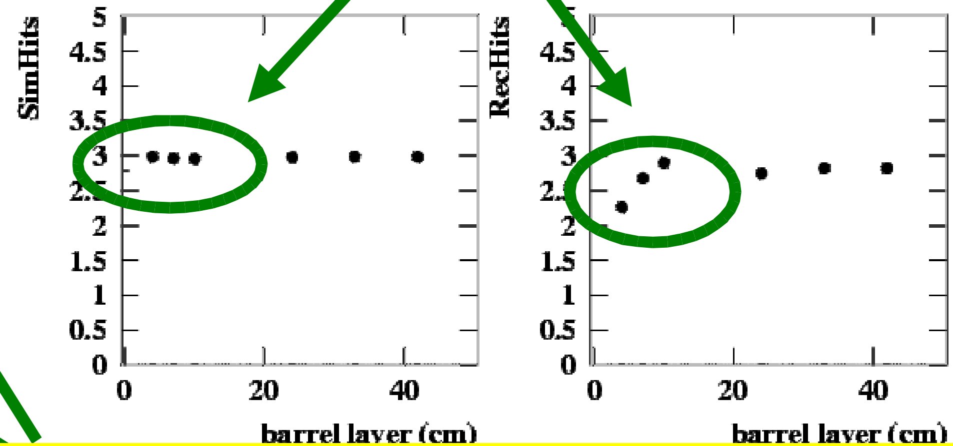
u-jet mistagging efficiency similar, DAF worse for high eta at 200 GeV

Study - CMS tracker



3-prong tau decays from heavy Higgs

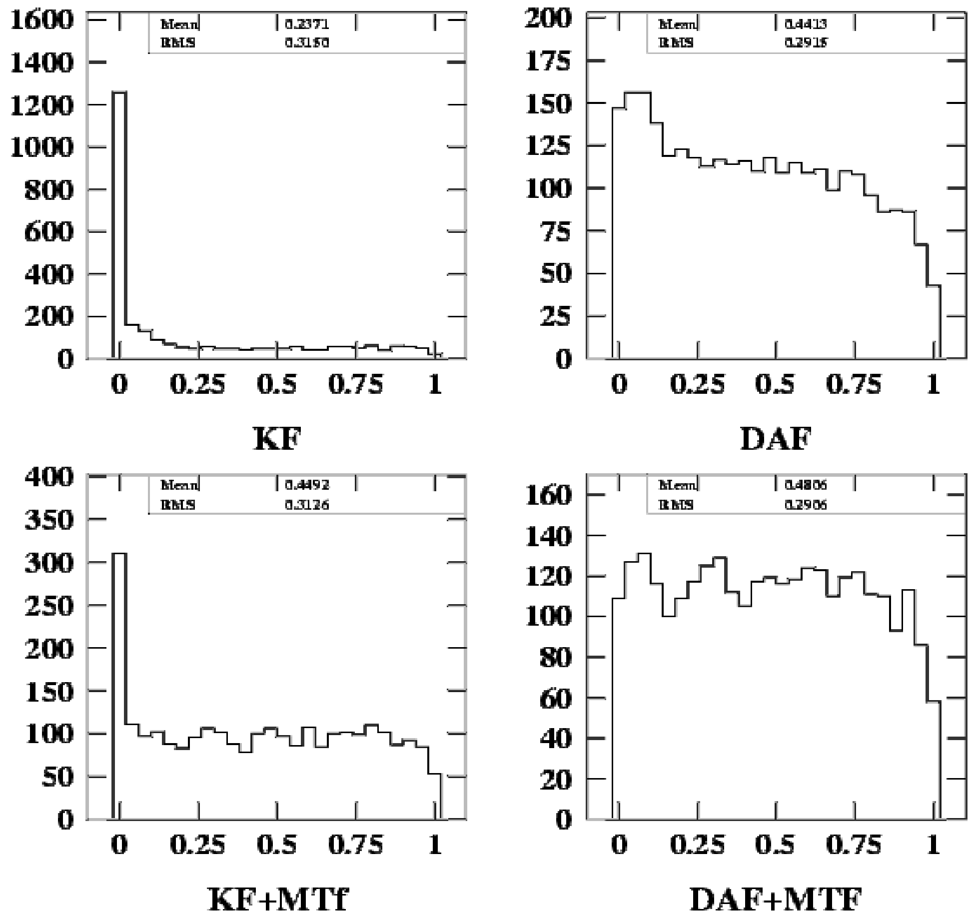
No improvement in resolution with MTF, environment not hostile enough!



Tails again reduced significantly

Study - CMS tracker

χ^2 probability



However, due to different and more correct structure of assignment weights, quality of error estimate is better with the MTF

Could expect small improvement in tagging efficiency with respect to DAF (not studied yet)

Conclusions

- The basic features of adaptive methods have been introduced through an example of a line fit.
- Starting from the Elastic Arms algorithm, state-of-the-art adaptive algorithms for track reconstruction such as the Deterministic Annealing Filter and the Multitrack Filter have been derived.
- Relations to other combinatorial optimization problems, such as clustering and certain image analysis tasks, have been mentioned.

Conclusions

- The performance of the adaptive algorithms on reconstruction problems in the CERN LHC detectors ATLAS TRT and CMS tracker has been discussed.
- Adaptive algorithms show no gain with respect to standard algorithms in clean environments (little noise, no ambiguities, good track separation).
- Significant improvements can be achieved under harsh conditions, such as very high energy b-jets and narrow jets from tau decays.