Numerical Optimization for Fast Track Finding Based on the Artificial Retina Algorithm

on behalf of the LHCb collaboration

For the 22nd International Conference on Computing in High Energy and Nuclear Physics, CHEP 2016, October 10-14, hosted by SLAC and LBNL

Maxim Borisyak[1,2], Mikhail Belous[2], Andrey Ustyuzhanin[1,2] and Denis Derkach[1,2]

[1] National Research University Higher School of Economics, [2] Yandex School of Data Analysis

Track reconstruction

Track reconstruction

High-energy physics experiments rely on reconstruction of the trajectories of particles produced at the interaction point. This is a challenging task, especially in the high track multiplicity environment generated by p-p collisions at the LHC energies.

Upcoming LHC upgrade

Track reconstruction studies are especially interesting in context of upcoming LHC upgrade as number of particle trajectories will considerably increase.

Artificial Retina Algorithm

Artificial Retina

Given set of hits $\{\mathbf{x}_i\}_{i=1}^N$ and a track model parameterized by θ:

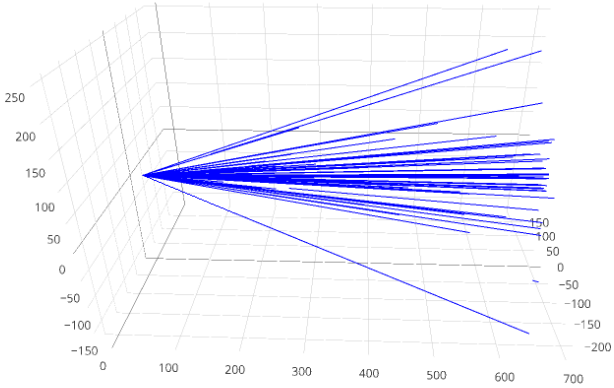
$$R(\theta) = \sum_{i=1}^{N} \exp\left(-\frac{\rho^2(\theta, \mathbf{x}_i)}{\sigma^2}\right)$$

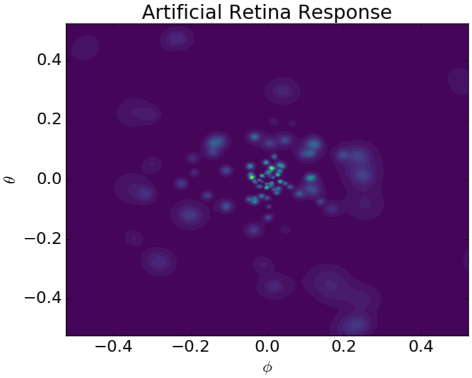
where $\rho(\theta, \mathbf{x}_i)$ — distance function from $\mathbf{x}_i$ to track with parameters θ.

Track reconstruction
Local maxima with $R \gg 1$ are close to true track parameters.

Artificial Retina example

Examples of 3 dimensional event and Artificial Retina response.





Intuition

Interpretation

- approximation of the number of hits that are close to the track;
- conceptually similar to Hough transform.

Features

- Artificial Retina is, essentially, **global optimization problem**;
- the algorithm is defined by the track model and the distance function ρ;
- robust to noisy hits;
- computing response is a well-suited task for massively-parallel processors;

Track reconstruction

Artificial Retina Challenges

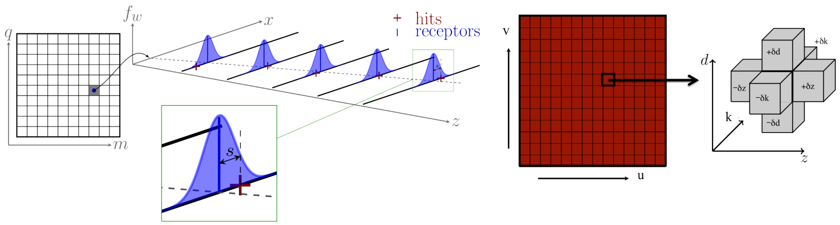
Global Optimization

Grid-search is traditionally used as a global optimization strategy.

- grid-search needs **the whole parameter space** to be explored:
  - computational complexity $\sim (1/\epsilon)^d$, where d — parameter space dimensionality, ϵ — resolution of track reconstruction;
- thus, either:
  - requires a lot of computational power:
    - specialized hardware;
    - massively-parallel processors (e.g. GPU).
  - limits track reconstruction precision.

Inspiration

[Abba et al., 2014] proposes specialized Artificial Retina processor for real-time track reconstruction for CERN LHCb experiment.



Specialized Artificial Retina processor schematic. The processor performs grid-search over the parameter space with two 'major' parameters and three 'minor' ones. Figures from [Abba et al., 2015].

Grid-search requires either **a lot of computational resources** or **specialized hardware.**

Track reconstruction

Numerical optimization study

Numerical optimization study

Our aim

- reduction in **total** computational complexity;
- linear dependency on expected number of tracks;
- linear dependency on track dimensionality.

Main idea

- **exploit gradient information to reduce total computational time;**
- adopt numerical optimization methods for fast local maxima search;

Gradient computation

Intermediate results of $\nabla(R)$ can be reused to compute R:

$$R(\theta) = \sum_i \exp\left(-\frac{\rho^2(\theta, \mathbf{x}_i)}{\sigma^2}\right);$$

$$\nabla R(\theta) = -\frac{2}{\sigma^2} \sum_i \exp\left(-\frac{\rho^2(\theta, \mathbf{x}_i)}{\sigma^2}\right) \rho(\theta, \mathbf{x}_i) \nabla \rho(\theta, \mathbf{x}_i)$$
$$= -\frac{2}{\sigma^2} \sum_i \mathbf{R_i}(\boldsymbol{\theta}) \boldsymbol{\rho}(\boldsymbol{\theta}, \mathbf{x_i}) \nabla \rho(\theta, \mathbf{x}_i)$$

Numerical optimization

Multi-start algorithm general scheme

1. invoke **seed generation procedure** to choose n initial seeds;
2. set initial σ^2;
3. repeat m times in parallel for each seed:
   3.1 perform one step of hill climbing ;
   3.2 decrease σ^2;

- initial σ^2 is to be comparable to mean distance of seeds to true tracks;
- σ^2 should be gradually decreased to the level of desirable precision;
- rate of convergence of optimization method should be taken into account.

Discussion

Total computational complexity:

$$O(nmd) = O\left(\frac{nd}{\epsilon}\right)$$

where:

- n — number of initial seeds;
- $m \sim \frac{1}{\epsilon}$ — number of optimization steps for each processes;
- d — dimensionality of track parameter space;

Complexity for **gradient descent method** and distance function with $O(d)$ computational complexity. Other numerical optimization methods (e.g. Hessian-based) may provide better complexity estimations.

Discussion

- for each seed m step is performed, each includes:
  - response, gradient (and Hessian) components computations;
  - response, gradient (and Hessian) summation.
- limits maximal number of computing units:

$$\text{max}\left(\text{number of computing units}\right) = \text{number of initial seeds} \times \text{number of hits}$$

- hence, less parallelization capacity (relative to grid-search):
  - number of seeds vs. number of steps trade-off;
  - grid-search can be viewed as a special extreme case;
- constant terms in complexity heavily depend on seed generation procedure.

Experiment

Experiment details

Metric

- a track is reconstructed if recovered direction is within ϵ from the true direction;
- $\epsilon = 10^{-3}$ rad, comparable to the maximal error of direction w.r.t. hit errors.

Artificial Retina details

- track parametrized by spherical angles (θ, ϕ);
- distance function — euclidean distance to track.

Implementation details

- Newton Conjugate Gradients method with $m = 3$ steps;
- n initial seeds are randomly picked among hits, so that total amount of computations is α fraction of resources n_{grid} required by plain grid-search to provide ϵ resolution:

$$n = \alpha \cdot n_{\text{grid}} \frac{1}{C_0} \frac{1}{m}$$

  - $C_0 \approx 30$ — normalization constant:

$$C_0 = \frac{\text{time}(\text{optimization step})}{\text{time}(\text{response invokation})}$$

Simplified model of LHCb VELO detector

- tracks - straight lines;
- parameters are motivated by upgrade VELO TDR [LHCb Collaboration, 2013].

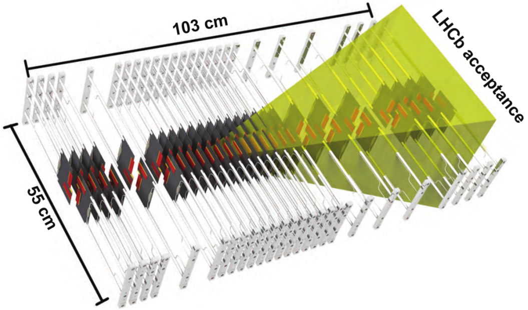
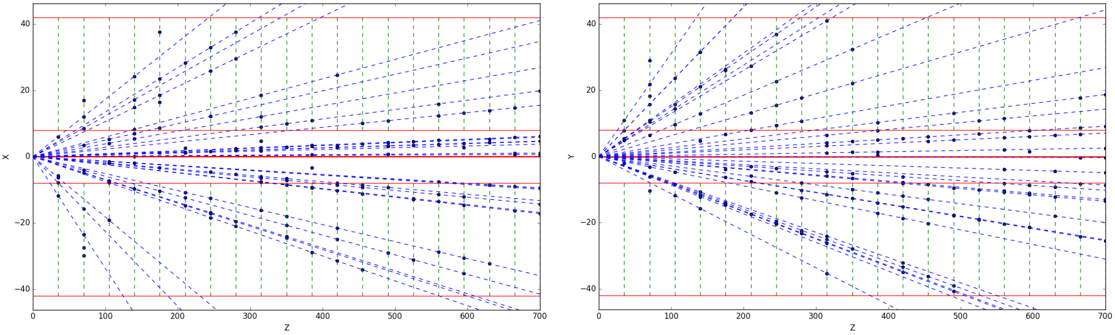


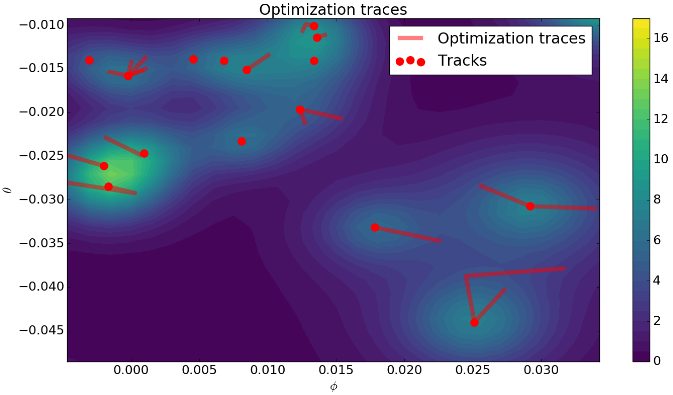
Figure from [LHCb Collaboration, 2013].

Example



An example of low-multiplicity event simulated by simplified VELO model.
Figure shows X-Z (left) and Y-Z (right) projections, blue lines denote tracks, blue dots — hits, green lines — detector planes.

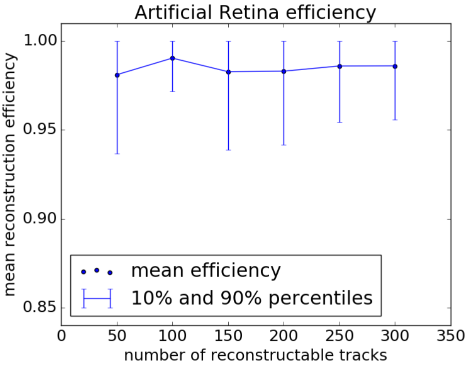
Example



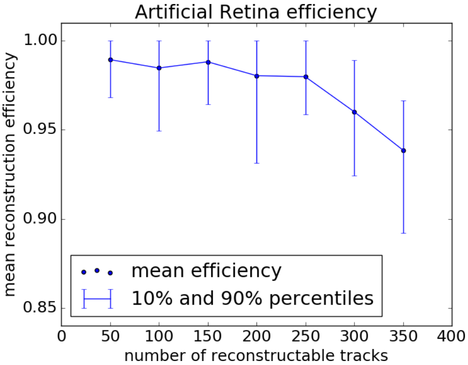
Optimization traces (red lines) for an event in the simplified VELO. Heat map corresponds to the Retina response for the event.

Results

$\alpha = 1/3$



$\alpha = 1/10$



Efficiency depending on the number of initial seeds (set by restricting computational resources to α fraction of resources for plain grid-search). Ghost rate is strictly zero in all cases. Multiple results for the same track are merged within ϵ-radius.

Summary

Summary

Numerical optimization for Artificial Retina

- numerical optimization for local maxima search.
- gradient information can be utilized to reduce total computational complexity;
- gradient and Hessian can be efficiently computed;
- asymptotic complexity linear in number of dimensions;

Experiment

- results comparable to grid-search Artificial Retina methods;
- 0.98 average reconstruction efficiency with $1/3$ of plain grid-search resources.

References I

Abba, A., Bedeschi, F., Citterio, M., Caponio, F., Cusimano, A., Geraci, A., Marino, P., Morello, M., Neri, N., Punzi, G., et al. (2015).
Simulation and performance of an artificial retina for 40 mhz track reconstruction.
Journal of Instrumentation, 10(03):C03008.

Abba, A., Punzi, G., Spinella, F., Marino, P., Tonelli, D., Stracka, S., Lionetto, F., Ninci, D., Petruzzo, M., Cusimano, A., et al. (2014).
A specialized track processor for the lhcb upgrade.
Technical report.

LHCb Collaboration (2013).
LHCb VELO Upgrade Technical Design Report.
Technical Report CERN-LHCC-2013-021. LHCB-TDR-013.

Backup

Track parametrization

Track parametrization:

- pseudo-rapidity η and angle in the traverse plane ϕ;
- track direction $\mathbf{n} = (n_x, n_y, n_z)$;
- angle t and offset x_0 (2 dimensional case).

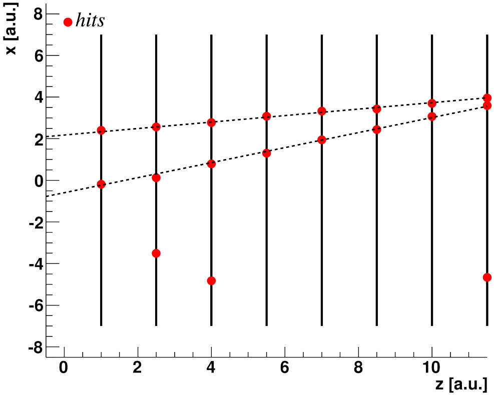


Figure from [Abba et al., 2014].

Distance function

Distance functions:

- projection error:
  $\rho(\mathbf{n}, \mathbf{x}_i) = \|\mathbf{x}_i - \mathbf{n}(\mathbf{n} \cdot \mathbf{x}_i)\|_2$;
- projection error in the corresponding detector plane ($z = \mathrm{const}$).

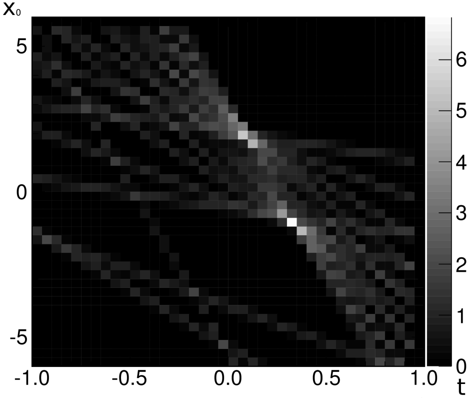


Figure from [Abba et al., 2014].

Simplified model of LHCb VELO detector

Simplified model of LCHb VELO was simulated:

- tracks - straight lines;
- simplified 'VELO' parameters:
  - 20 disc layers with inner $r = 8$ mm, outer - $R = 42$ mm;
  - length: $L = 700$ mm;
  - probability of a particle interacting with a layer: $P_{\mathrm{int}} = \frac{1}{2}$;
  - hit error: $\epsilon \sim \mathcal{N}(0, 10^{-3})$ mm;
  - number of noisy hits: $N' \sim \mathrm{Poisson}(250)$;
- $\eta \sim \mathrm{Uniform}[1, 5]$;
- $\phi \sim \mathrm{Uniform}[0, 2\pi]$.

Parameters are motivated by upgrade VELO TDR.