A linear iterative unfolding method

András László
Andras.Laszlo@cern.ch
CERN, Geneva
(on leave of absence from RMKI, Budapest)

ACAT-2011
Brunel University, Uxbridge
8 September 2011

Outline

Topic of the talk: introduction to a new linear iterative unfolding algorithm.

- Introduction to the unfolding problem
- Known approaches and shortcomings
- A linear iterative unfolding algorithm
- Test examples
- Physics examples
- Summary

Based on:

[1] A. László: J.Phys. **A39** (2006) 13621.

[2] A. László: in preparation (yet unpublished new results).

Introduction to the unfolding problem

Frequent task in signal processing:

- We want to measure: the spectrum of some quantity.
- Problem: the quantity is smeared by some physical process.
- Task: reconstruct (i.e. *unfold*) the original spectrum.

Probability theory formulation:

- Measured pdf $y \mapsto g(y)$, unknown pdf $x \mapsto f(x)$, known cpdf (*response function*) $(y, x) \mapsto \rho(y|x)$:

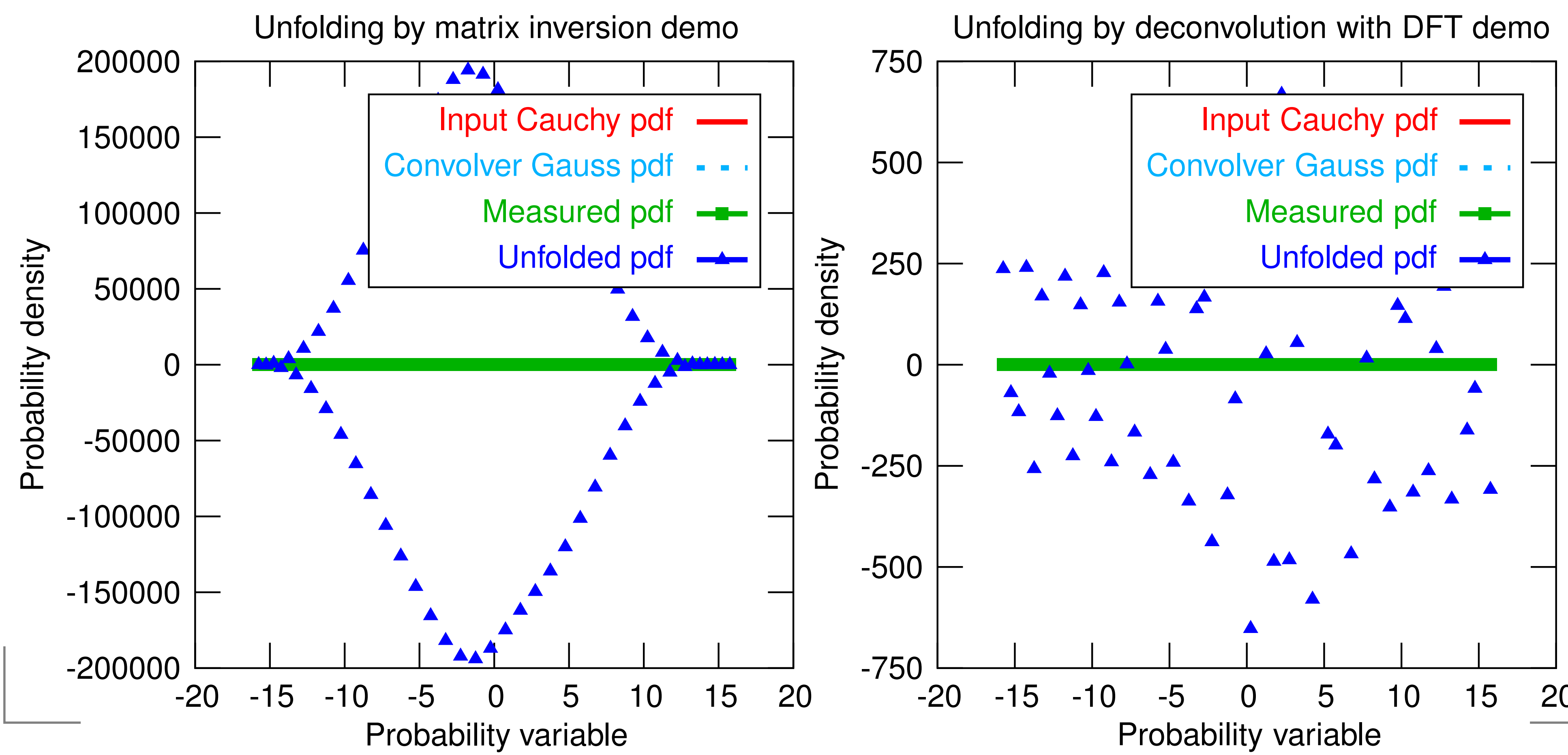
$$g(y) = \int \rho(y|x)\ f(x)\ \mathrm{d}x.$$

  This is called a *folding*. It is called *convolution*, whenever ρ is translation invariant: $\forall\ x, y, z\ :\ \ \rho(y + z|x) = \rho(y|x - z)$. In that special case, ρ can be expressed by a single pdf: $\rho(y|x) = \rho(y - x|0)$, and we write $g = \rho(\cdot|0) \star f$.
- Task: have to invert this linear integral operator ("Fredholm"-like operator, although the kernel function ρ does not satisfy the regularity conditions of being Fredholm).
- In practice, the measured pdf is also contaminated by (e.g. a statistical) noise term.

Known approaches and shortcomings

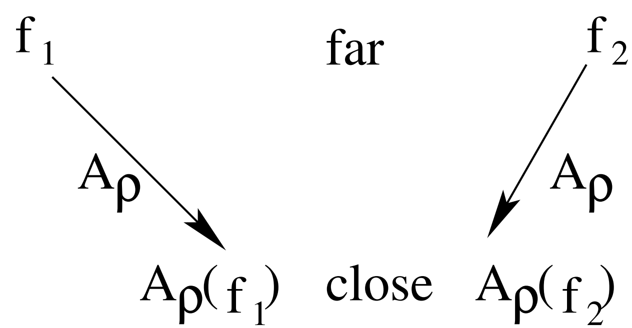
- A naive approach is to discretize the problem and invert the discrete matrix.
- A naive approach (in case of convolution) is to use the convolution theorem:
  $F(\rho(\cdot|0) \star f) = F(\rho(\cdot|0))F(f) \quad \Rightarrow \quad f = F^{-1}\left(F(g)/F(\rho(\cdot|0))\right)$ (F: Fourier transf.)

Does not work. Example with Gauss $\star$ Cauchy convolution:



The reason for these issues:

- The inverse operator (in the generic case) can be proved not to be $L^1 \to L^1$ continuous.
  $\Rightarrow$ Numerical instability, as g has a *noise* term in practice.
- This means that the image of two far pdfs (f_1, f_2) by the folding operator (A_ρ) can also be close.



- I.e. it can be difficult to distinguish initial pdfs after folding (ill-posed problem).
- This also means that if we discretize the problem, the matrix norm of the inverse operator tends to infinity as we reach continuum limit. I.e. infinitely magnifies certain undesired modes.
- Very difficult to regularize without a priori knowledge about f.

Possible treatments:

- Using naive deconvolution (in case of convolutions) with frequency regularization to suppress highly oscillating modes. This is involves kind of cooking.
- Using naive matrix inversion, but with suppressing low eigenvalue modes. Similar drawbacks as for frequency regulatization.
- Using an Ansatz for f, and fit parameters, so that $A_\rho(f)$ gets close to g. This is of course biased if we are not 100% sure about the Ansatz. The χ^2 can eventually be insensitive to the details of the true f.
- Bin-by-bin fitting of f so that $A_\rho(f)$ gets close to g. This becomes unstable as naive inversion, except if fitting is performed by adding penalty function to χ^2 which penalizes large local gradients. This can be slightly insensitive to the details of the true f, and solves a slightly different problem.
- Using method of convergent weights (Bayesian iteration scheme) of Kondor-Mülthei-Schorr-d'Agostini. Looks more promising as it preserves the positivity and the integral of the pdf. Regularization achieved by stopping the iteration at a finite order. No proof of convergence is available. When to stop? (Error propagation?)

A linear iterative unfolding algorithm

- Let us chose a pdf η which is a nowhere zero Schwarz function, having nowhere zero Fourier transform (e.g. a Gaussian). Let A_η be the convolution by η ("smoother").
- Let us define the normalization factor

$$K_{\rho,\eta} = \max_x \int\int (A_\eta\rho)(y|z)\,(A_\eta\rho)(y|x)\,\mathrm{d}y\,\mathrm{d}z.$$

  ($K_{\rho,\eta} = 1$ holds automatically if A_ρ is a convolution.)
- Take the iteration scheme in search for f (with $g = A_\rho f$, and $f \in L^1 \cap L^2$):
  - $f_0 = K_{\rho,\eta}^{-1}\,(A_\eta A_\rho)^T\,A_\eta g$,
  - $f_{N+1} = f_N + \Big(f_0 - K_{\rho,\eta}^{-1}\,(A_\eta A_\rho)^T\,A_\eta A_\rho f_N\Big)$,
    where $(A_\eta A_\rho)^T$ is the transpose folding by $A_\eta A_\rho$.
    (Whenever $\forall\, x\; :\; \rho(\cdot|x) \in L^1 \cap L^2$, then one may drop A_η by setting it to $A_\eta = I$.)
- Setwise convergence can be proved whenever $K_{\rho,\eta} < \infty$. Proof based on Riesz-Thorin theorem $+$ spectral representation of positive operator over L^2 space (*to be published*). Convergence holds e.g. for convolution, calorimeter response etc.
- Iteration scheme motivated by the "Neuman series" known in functional analysis.
- Resembles to "Landweber iteration" (`Am.J.Math.` **73** `(1951) 615`), but that result is proved with too restrictive regularity requirements (Fredholmness) for our problem.

Upper estimate to the residual term at finite iteration order approximation:

- In practical applications, besides fact of convergence, one also may want to know the deviation from limiting pdf at given finite iteration order N.
- Let S be a compact set on the domain of f.
- Then,

$$\frac{1}{\mathrm{Volume}(S)} \int_S \left| f - P_{\mathrm{Ker}(A_\rho)} f - f_N \right| (y) \, \mathrm{d}y \leq \frac{1}{\sqrt{\mathrm{Volume(S)}}} \, \|f\|_{L^2} \, \frac{1}{N+2},$$

  $P_{\mathrm{Ker}(A_\rho)}$ being the orthogonal projection to the kernel of A_ρ.
- $f - P_{\mathrm{Ker}(A_\rho)} f$ contains all the information about pdf f which is not lost upon the folding.
- I.e. residual deviation of f_N averaged to the set S decreases with $\frac{1}{\sqrt{\mathrm{Volume(S)}}}$ and decreases as $\frac{1}{N+2}$.

Properties of statistical error propagation:

- As the method is linear, exact propagation of covariance matrix of initial pdf is possible.
- Let C be the covariance matrix of the measured pdf g. Since covariance matrix is positive definite, one can always define (not uniquely) error matrix E for which $C = E\,E^T$ holds. (If C is diagonal, construction of such an E is just trivial.)
- Perform the iteration:
  - $f_0 = K_{\rho,\eta}^{-1}\,(A_\eta A_\rho)^T\,A_\eta g,$
    $E_0 = K_{\rho,\eta}^{-1}\,(A_\eta A_\rho)^T\,A_\eta E,$
  - $f_{N+1} = f_N + \Big(f_0 - K_{\rho,\eta}^{-1}\,(A_\eta A_\rho)^T\,A_\eta A_\rho f_N\Big),$
    $E_{N+1} = E_N + \Big(E_0 - K_{\rho,\eta}^{-1}\,(A_\eta A_\rho)^T\,A_\eta A_\rho E_N\Big).$
- Then, the statistical covariance matrix in each step will be $C_N = E_N\,E_N^T$.
- Iteration stop when L^1 norm of statistical error content exceeds predefined limit.

Properties of systematic error propagation:

- An upper estimation can be proved using spectral representation of operators.
- Let δg be the systematic error on the measured pdf g. We also can include here systematic error $\delta\rho$ of the response function: this gives a systematic error contribution $A_{\delta\rho} f$ to δg.
- Let S be a compact set on the domain of f.
- Then,

$$\frac{1}{\mathrm{Volume}(S)} \int_S |\delta f_N|\,(y)\;\mathrm{d}y$$

$$\leq \frac{1}{\sqrt{\mathrm{Volume(S)}}} \left\| K_{\rho,\eta}^{-1} \left(A_\eta A_\rho\right)^T A_\eta \delta g \right\|_{L^2} \left(\Psi(N+2)+\gamma\right)$$

$$\approx \frac{1}{\sqrt{\mathrm{Volume(S)}}} \left\| K_{\rho,\eta}^{-1} \left(A_\eta A_\rho\right)^T A_\eta \delta g \right\|_{L^2} \left(1+\ln(N+1)\right)$$

(Ψ being digamma function, γ being Euler's constant).

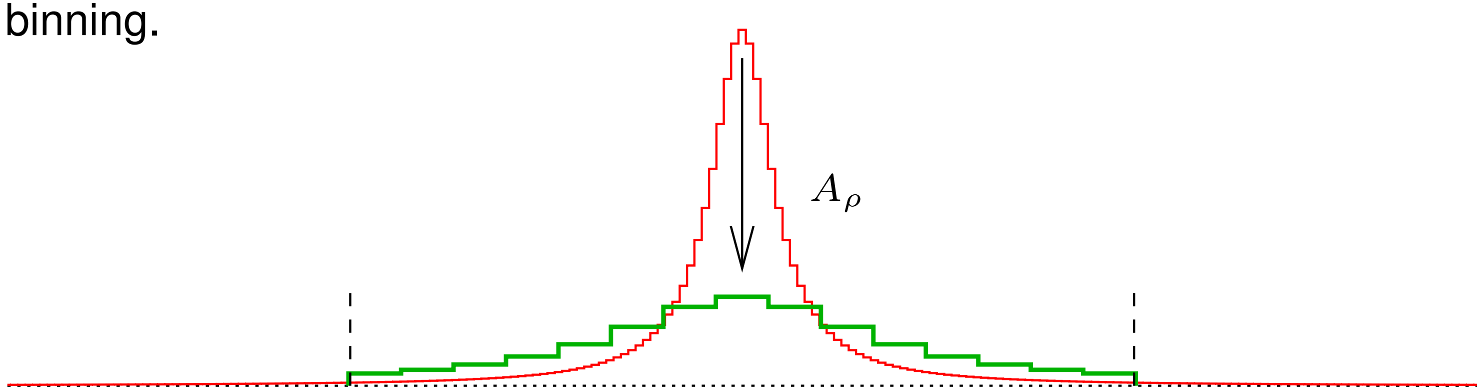
- I.e. systematic error of f_N averaged to the set S decreases with $\frac{1}{\sqrt{\mathrm{Volume(S)}}}$ and grows logarithmically with iteration order N.

The method does not require initial binning of pdfs:

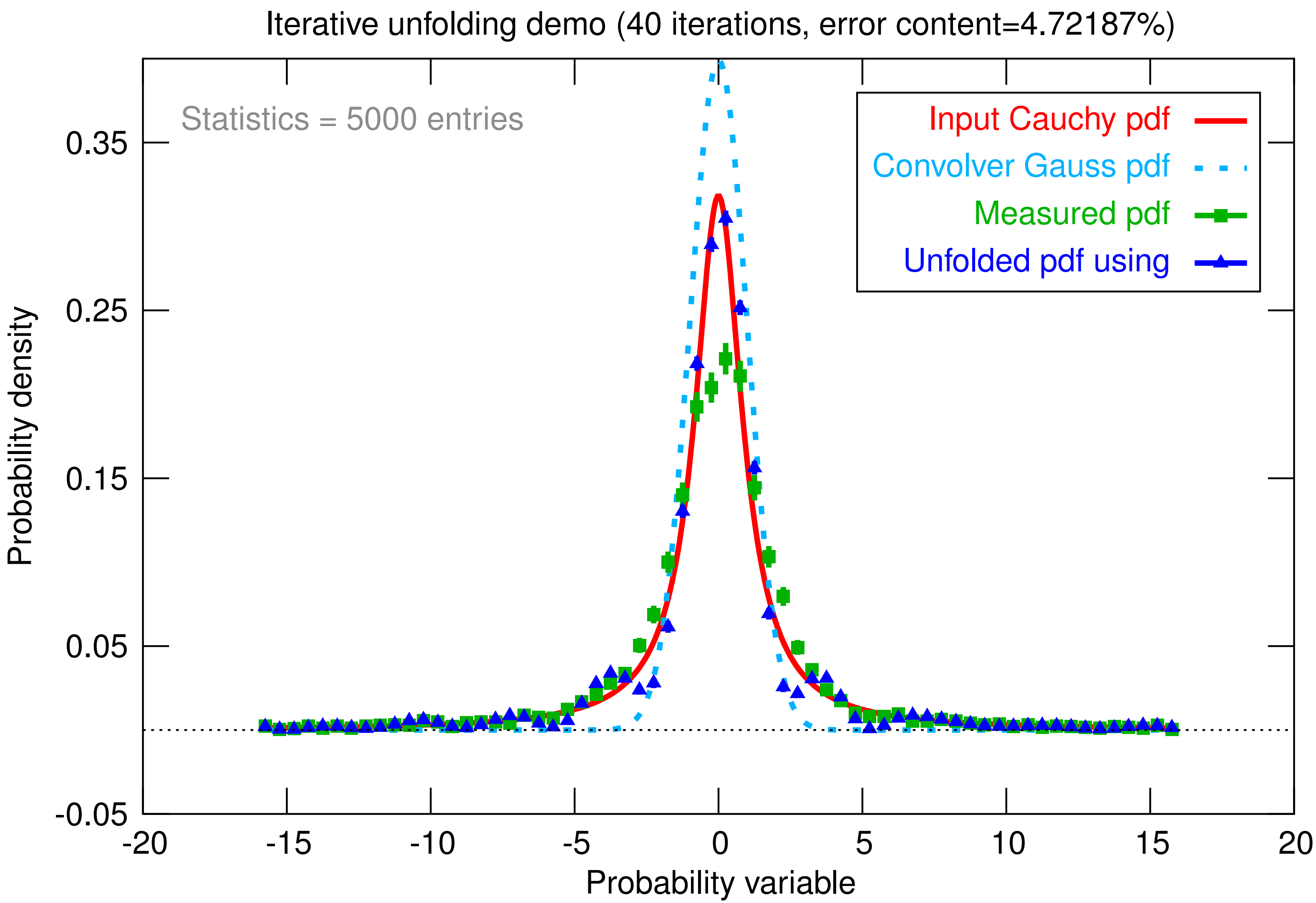
- As the convergence theorem is proved in continuum limit, discretization (e.g. with histograming) is not necessary. It may be, however, used to numerically model the pdf.
- One may use as well unbinned empirical distribution function instead of histograming.

The procedure of histograming is as well a folding:

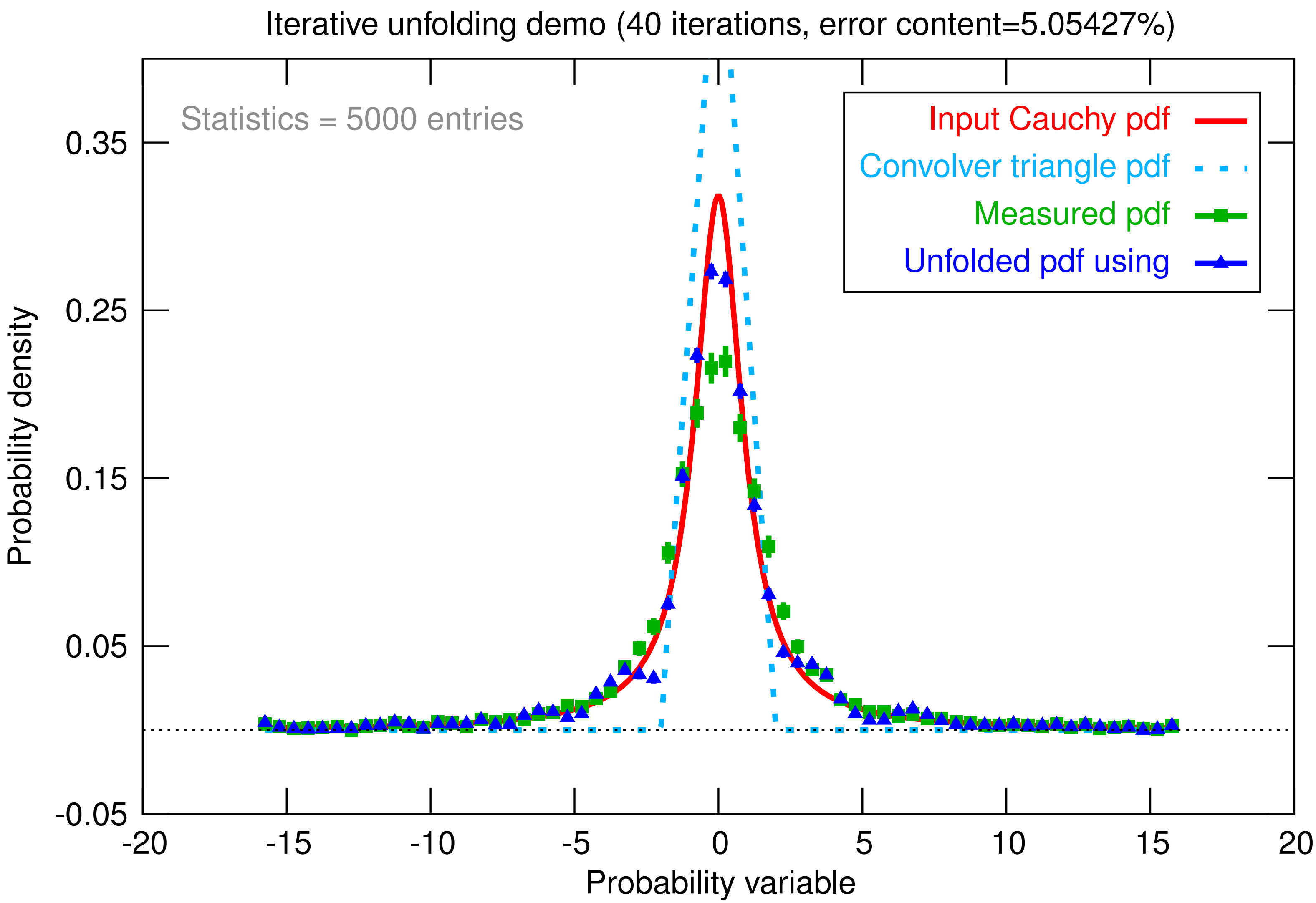
- As the histograming process is also itself a folding operator, one may include it in the folding A_ρ.
- One may assume better binning density and wider histograming domain for the initial (unknown) pdf when implementing A_ρ and A_ρ^T. I.e. the initial pdf can be assumed to better approximate the continuum limit. In that case, one can unfold as well the effect of binning.



Test examples

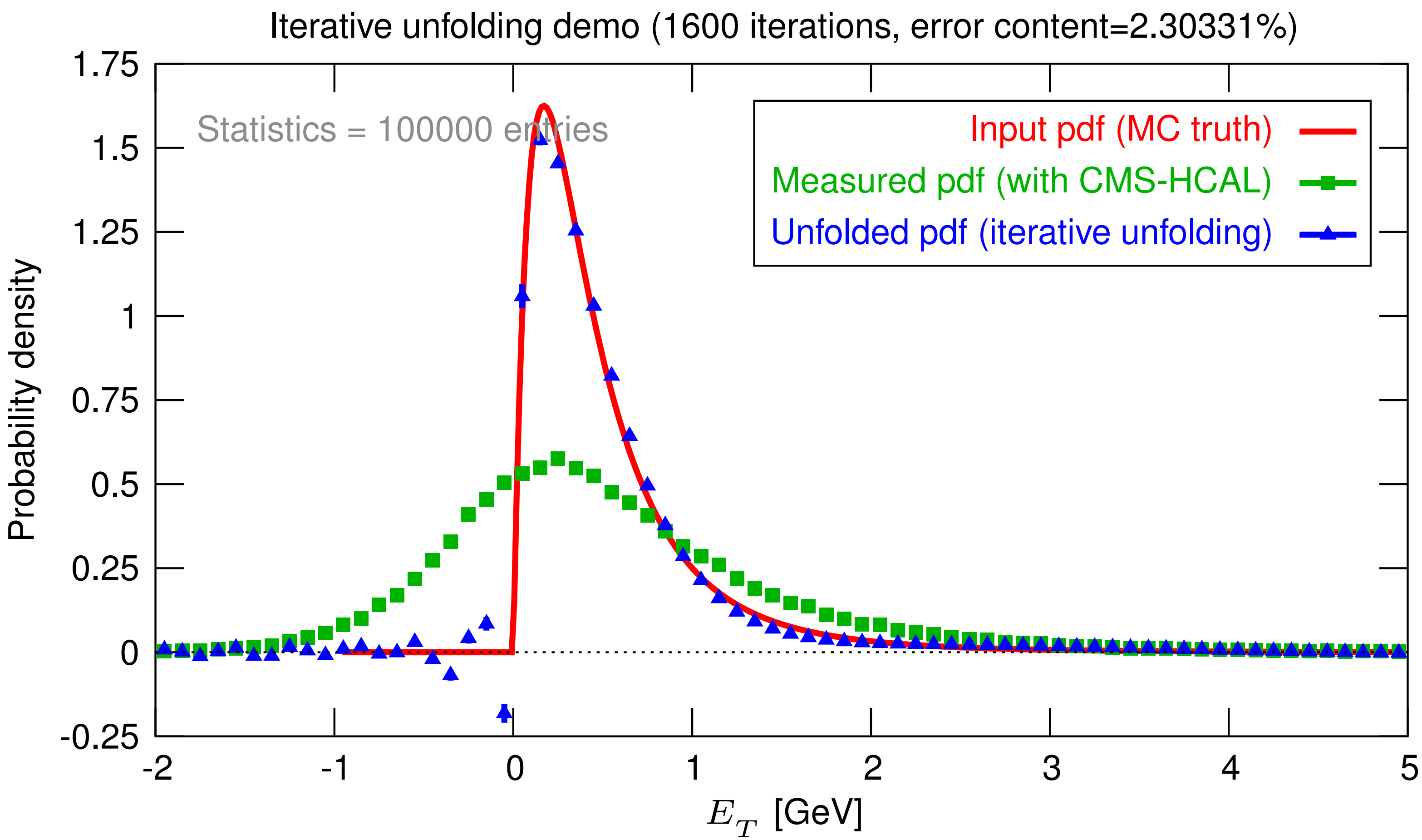


Gauss $\star$ Cauchy distribution + binning demo

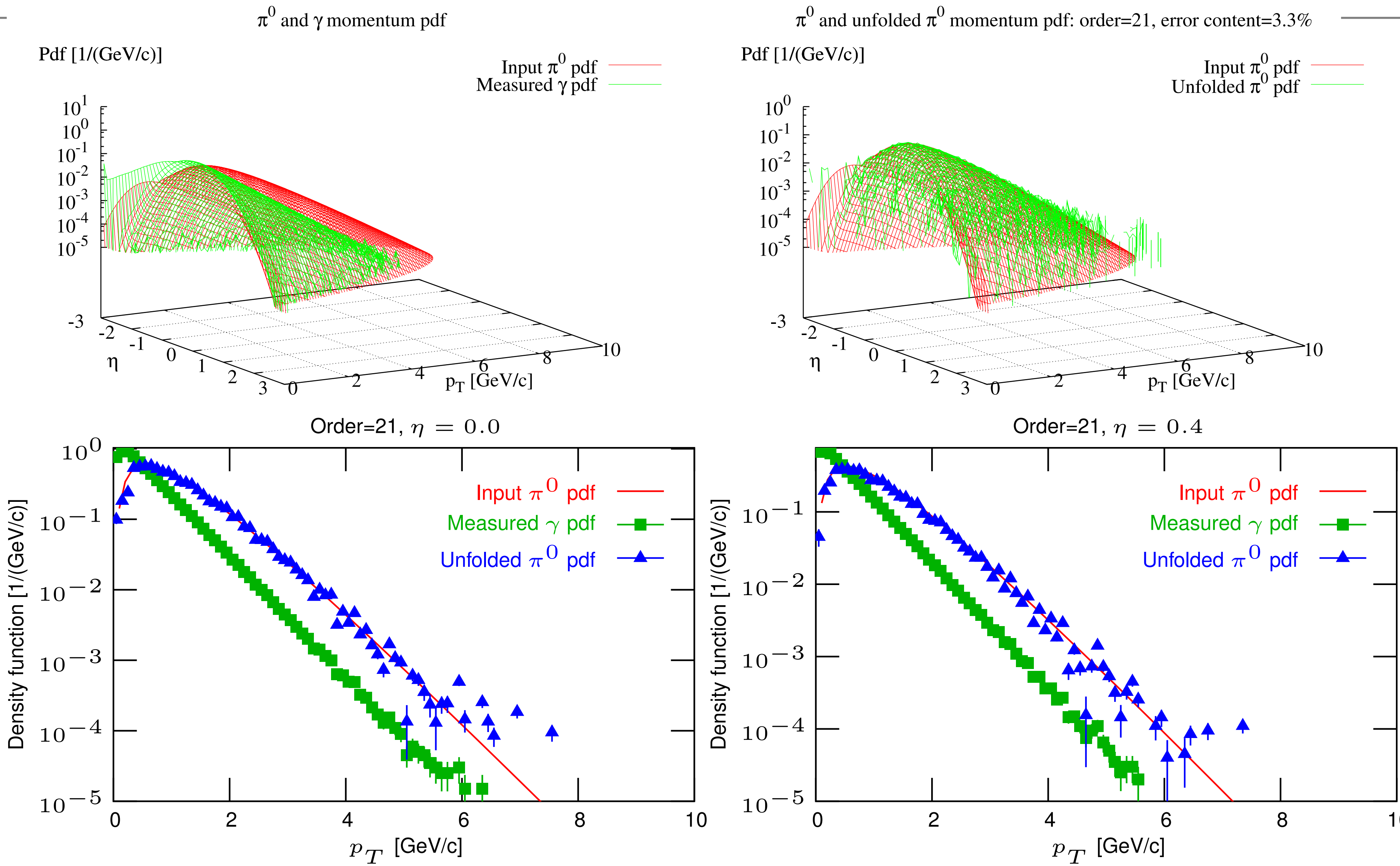


Triangle ⋆ Cauchy distribution + binning demo

Physics examples



Charged hadron E_T spectrum at midrapidity, $\sqrt{s} = 7\,\mathrm{TeV}$ (PRL **105** (2010) 022002), when measured with CMS hadronic calorimeter (JPCS **160** (2009) 012056) + binning.



π^0 single particle spectrum from γ single particle spectrum ($\pi^0 \to \gamma + \gamma$ decay).

Summary

- A linear unfolding iteration algorithm is proposed. Can be binning-free.
- Convergence to the undestroyed part of the initial pdf is proved under certain conditions (which hold e.g. for convolution, calorimeter response etc.).
- Exact propagation of statistical errors (due to linearity).
- Upper estimates for propagated systematic errors of measured pdf and resp function.
- Positivity of pdf is not strictly conserved (due to linearity).
- Available application library in C:
  `http://www.rmki.kfki.hu/~laszloa/downloads/libunfold.tar.gz`
  Should be included into RooUnfold?
- Interesting open questions:
  - The folding operator can be evaluated by Monte Carlo without a priori sampling (free of binning effect). Can this be done to the transpose folding?
  - Proof is still not known for the convergence of the Bayesian unfolding of Kondor-Mülthei-Schorr-d'Agostini. Worth to investigate as that method preserves positivity and integral of pdf $\Rightarrow$ well suited for probability theory problems.
- Thanks to: prof. Günter Zech for valuable discussions.