Methods for

Non-parametric Density Estimation and Unfolding

Institutsseminar

Institut für Kern- und Teilchenphysik, TU Dresden

25. June 2015

Volker Blobel – Universität Hamburg

The measurement of density functions is the standard task of experimental particle physics. In particle physics experiments a sample of values from a random variable x is extracted from event data, and usually a histogram is generated from the sample as estimate of the density function $f(x)$. More realistic and accurate density estimates are discussed.
Often the true variables x are not directly accessible to the measurement; instead a sample of values from a random variable y is observed, which differs from the variable x because of finite resolution, limited acceptance and perhaps some transformation. Monte Carlo events from a simulation of the measurement process provide information on the transformation of variable x to y, by a sample (X, Y). The inverse ill-posed problem of reconstructing the density $f(x)$ from the statistically limited data sample (Y) and Monte Carlo sample (X, Y), the unfolding problem, is statistically and mathematically complex. Unfolding methodology actually used in HEP is not well-established. Actually used methods and potential new methods are discussed.

---

1. Measurement of densities in Particle Physics
2. $*$ Non-parametric density estimation
3. Singular value decomposition – SVD
4. Iterative unfolding methods
5. Regularized Least Squares
6. $*$ Projection methods
7. Critical review

$*$ = Methods without histograms

(1) Measurement of densities in Particle Physics

(1) Construct density estimate $\widehat{f}(x)$ from a sample of observations.

Observed sample of size N $\quad \{\boldsymbol{X}\}_N \equiv \{X_1, X_2, \ldots X_N\} \quad X_k \in [x_a, x_b]$

improve empirical density function $\quad \widehat{f}(x) = \dfrac{1}{N_{\text{data}}} \sum_{k=1}^{N} \delta\left(x - X_k\right)$

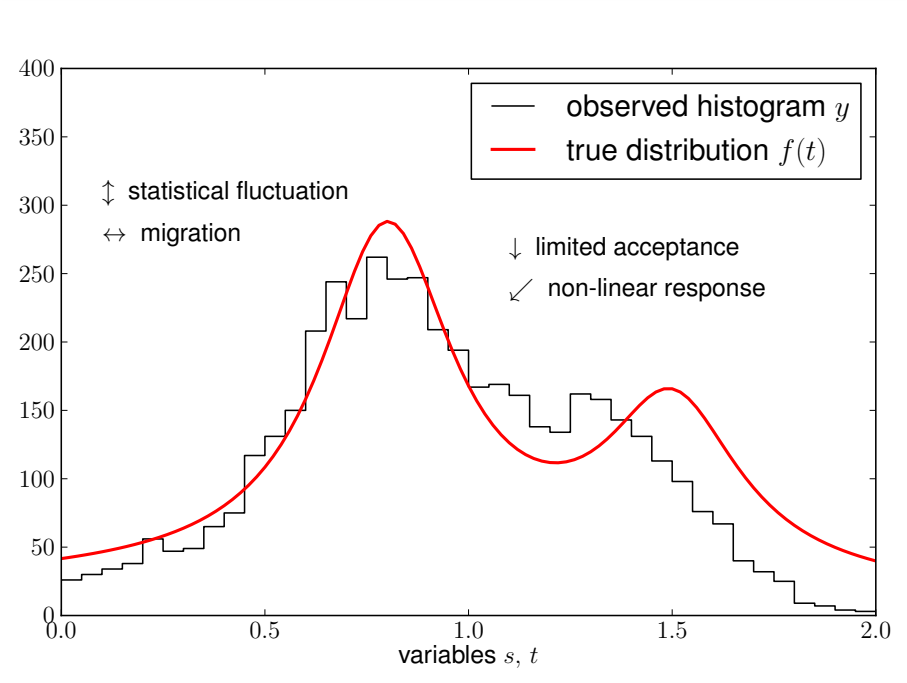
(2) Statistical inverse problem (unfolding): estimate $\widehat{f}(x)$ from measured sample $\{\boldsymbol{Y}\}_N$.

Measurements are recorded by non-perfect detectors: X (true) $\Longrightarrow$ Y (measured)

- smearing by a finite measurement accuracy;
- limited acceptance, causing a loss of part of the sample or a systematic shift of the measured values;
- the trigger and the reconstruction influence the data;
- principally unobservable physical phenomena in the particle collision can occur.

Measured sample of size N $\quad \{\boldsymbol{Y}\}_N \equiv \{Y_1, Y_2, \ldots Y_N\} \quad Y_k \in [y_a, y_b]$

Observed and true distribution



- ↕ Unavoidable random fluctuations: event number/bin follows Poisson distribution
- Measurement process – non-ideal detector

  ↔ migration due to measurement error

  ↓ limited acceptance

  ↙ non-linear response, e.g. loss of energy

Fredholm integral equation

Assumption: the measurement process with the transformation $f \Longrightarrow g$ is described by a *linear* operator $\mathcal{K}$:

$$\mathcal{K} f = g$$

Properties: superposition	$\mathcal{K}(f_1 + f_2) = \mathcal{K} f_1 + \mathcal{K} f_2$
scaling	$\mathcal{K}(\alpha f) = \alpha \mathcal{K} f$

Fredholm integral equation: relation between the density function $f(x)$ and $g(y)$

$$\int_\Omega K(y,x) f(x)\,\mathrm{d}x = g(y)$$

Kernel (or response) function $K(y,x)$ describes the behaviour of the measurement process.

Response function/matrix by MC simulation

In particle physics the measurement process, i.e. the kernel $K(y,x)$ is determined in a MC simulation, based on a certain assumed distribution $f_{MC}(x)$.

$$\text{MC} \qquad \int_\Omega K(y,x) f_{\text{MC}}(x)\,\mathrm{d}x = g_{\text{MC}}(y)$$

Monte Carlo: generate X according to $f_{\text{MC}}(x)$

- Hard interaction: exact matrix element
- QCD bremsstrahlung: parton showers in initial and final state
- Multiple Interactions: modelling beyond factorization
- Hadronization: non-perturbative modelling, and Hadron decays;
- Detector simulation: signal generation, trigger, reconstruction

Result of the Monte Carlo simulation: two statistical samples

$$\begin{aligned} \{(\boldsymbol{X})\}_{MC} &\equiv \{(X)_1, (X)_2, \ldots (X)_{MC}\} && \text{from } f_{MC}(x) \\ \{(\boldsymbol{X},\boldsymbol{Y})\}_{MC'} &\equiv \{(X,Y)_1, (X,Y)_2, \ldots (X,Y)_{MC'}\} && MC' \leq MC \end{aligned}$$

Solution:

$$\int_\Omega K_{\text{prob}}(y,x)\widehat{f}(x)\,\mathrm{d}x \simeq g_{\text{meas}}(y) \qquad \Longrightarrow \text{solution} \quad \widehat{f}(x)$$

$$\text{or} \quad \int_\Omega [K(y,x) f_{\text{MC}}(x)]\,\widehat{f}'(x)\,\mathrm{d}x \simeq g_{\text{meas}}(y) \qquad \Longrightarrow \text{solution} \quad \widehat{f}(x) = f_{\text{MC}}(x)\widehat{f}'(x)$$

Discretization

Unfolding is a **complex statistical and mathematical** problem.

Inverse problems are *ill-posed*: small input data errors or small inaccuracies in the mathematical treatment cause large errors in the solution.

Standard discretization method: Almost all unfolding methods in use in particle physics perform the discretization by histograms, which are the input data to unfolding.

- What is the optimal bin width for the histograms?
  Folklore: bins width show be close to the resolution (one standard deviation).
- How many Monte Carlo events are necessary for the sufficient determination of the response matrix?
  Folklore: a factor of 10 in statistic for Monte Carlo, compared to the data, should be sufficient.

System of linear equations, with probability matrix A and histogram vectors f, g:

$$Af \simeq g$$

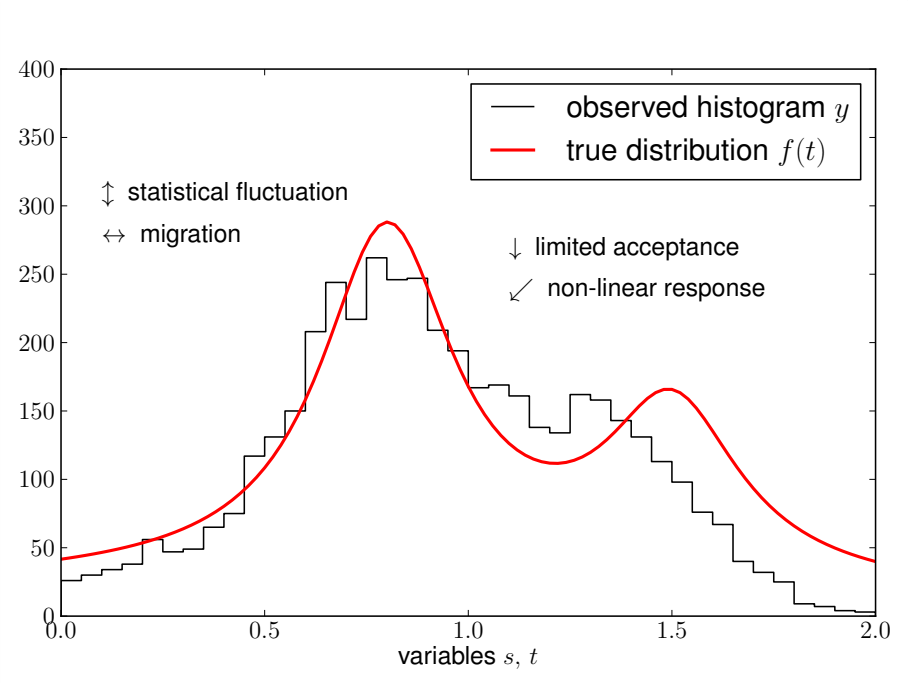
∗ **Alternative methods:** no histograms! Input n-tuples – densities are parametrized by e.g. B-splines or system of orthogonal functions. Avoid mathematical operations that result in unclear or undefined statistical properties of A or g.

Simple bin-by-bin correction factor method

$$\text{bin-correction factor} = C_i = \frac{\text{MC truth-level bin cont}_i}{\text{Reco–level bin cont after event reconstruction and selection}_i}$$

data bin content$_i$ D_i multiplied (*"corrected"*) by correction factor C_i

assumed *"statistical"* standard deviation $\simeq C_i \times \sqrt{D_i}$



- The method is at least rather optimistic; it is essentially a pure *acceptance correction* method.
- In practice non-linear response effects are "corrected" before by a different method.
- "…a HEP-specific heuristic, called bin-by-bin unfolding, which provably accounts for smearing effects *incorrectly* through a multiplicative efficiency correction, is widely used." [V.M. Panaretos]

From a publication: *"…The purity and stability typically exceed* 50%. *If either the purity or the stability is below* 25% *in a bin …, the bin is combined with an adjacent bin …"*

Unfolding in Particle Physics is still "Neuland"

In the past the problem of unfolding was not really discussed in particle physics experiments.

- statement "our data are corrected by Monte Carlo" was assumed sufficient, and accepted as explanation of the applied method.
- bin-by-bin correction: for decades the complex problem of unfolding was "solved" by the "bin-to-bin correction" method, *fixing* the migration between bins by an assumed MC;
- unfolding is often called "data correction", and "histogram" is a synonym for "data"; important data are published as histogram-like binned data.
- even often used unfolding methods are described by wrong or at least unclear terms
  - a certain iterative method (1995) is called by different people either
    - ⋆ Bayesian unfolding, or
    - ⋆ Maximum Likelihood for **Poisson** distributed data, or
    - ⋆ expectation-maximation (EM) algorithm incomplete data (Dempster 1977),
  - but is already known as "Richardson-Lucy method" in astronomy (1972, 1974) and included for image restauration in "Raw Therapee" photo software.
  - a standard method with Tikhonov "second derivative" regularization (1995) is called
    "SVD Approach to Data Unfolding" (and sometimes *a "single"-value decomposition method*).

Phystat 2011 Workshop at CERN

PHYSTAT 2011 workshop (organized by Louis Lyons) in "Statistical Issues . . . " at CERN included one-day on "**unfolding**" (for the first time)

Abstract of introduction by Louis Lyons:

> As a non-expert on unfolding, I wanted to make a few 'obvious' non-controversial remarks about unfolding. It turns out, that even such innocuous comments can become the subject of heated debate.

L.L. invited statistician Victor Panaretos (Lausanne).
The view of Panaretos in a proposal of a new statistical methodology:

> "At present, the unfolding methodology used in LHC analysis is not well-established. . . . these methods suffer from not dealing with two significant issues satisfactorily:
>
> (1) the choice of the regularization strength and
>
> (2) quantification of the uncertainty in the solution."

No real progress in unfolding since 2011.

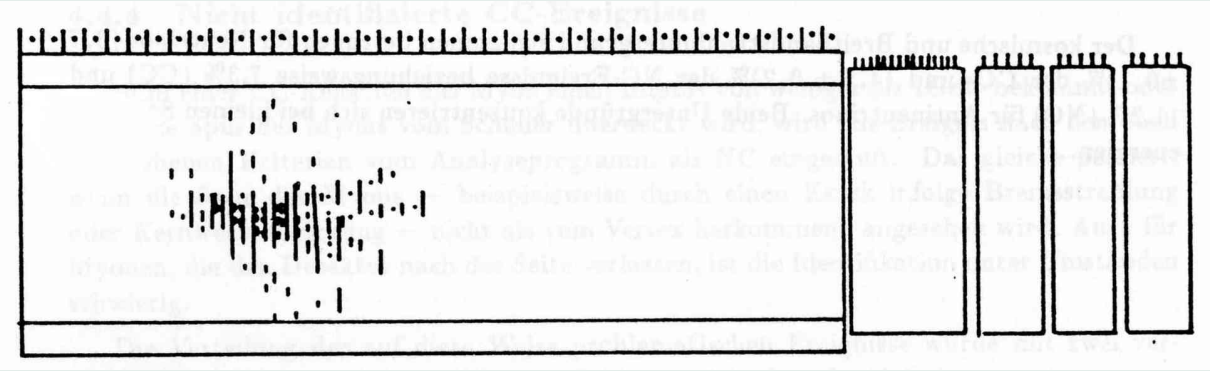
PHYSTAT 2011 Workshop, CERN-2011-006 yellow report, edited by Harrison Prosper, Louis Lyons.

V. Panaretos, M. Kuusela: *Statistical unfolding of elementary particle spectra: Empirical Bayes estimation and bias-corrected uncertainty quantification*, arXiv: 1505.04768

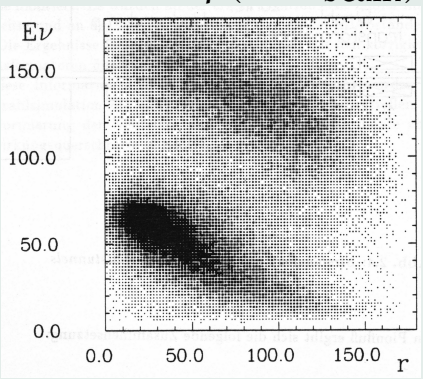
Unfolding is more general than "histogram correction"

Measurement of the differential cross section of inelasticity $y = E_{\text{had}}/E_{\nu,in}$ in neutral current neutrino interactions (narrow band beam): CHARM collaboration

NC: $\nu\, N \to [\text{hadrons}]\, \nu$ $\qquad$ $E_{\nu,in} = E_{\text{had}} +$ **???**



beam flux $\phi(E_\nu, r_{\text{beam}})$



direct problem (by MC) $\quad d\sigma/dy \Longrightarrow$ predicted dist. in E_{had}, r_{beam}

inverse problem (by unfolding) $\quad d\sigma/dy \Longleftarrow$ measured dist. in E_{had}, r_{beam}

Several "unfolding" codes are not able to solve such a problem

.

Unfolding is more general than "histogram data correction" to correct migration effects.

V. Blobel, Regularized Unfolding with $\mathcal{RUN}$ (1979-1984), CERN 1985 Computing School, Aiguablava, Spain

N. Milke, Fortran 77 code translated to C++ (with renaming to $\mathcal{TRUEE}$). TU Dortmund (2012)

Another experiment: "We reconstruct y event-by-event – using the K_ν-peak energy, . . . and later we correct by MC . . . "

∗ (2) Non-parametric density estimation

Task: estimate the density $f(x)$ from a sample $\{\boldsymbol{X}\}_N$ of random observations.

observations, sample of size N $\qquad\qquad \{\boldsymbol{X}\}_N \equiv \{X_1, X_2, \ldots X_N\}$

orthogonal function system $\{\phi_k(x)\}$ $\qquad\qquad \int_a^b \phi_j(x)\phi_k(x)\,\mathrm{d}x = \delta_{jk}$

density function	histogram $\boldsymbol{h}$, discrete sample
$f(x) = \sum_{k=0}^{\infty} f_k \phi_k(x)$	$\widehat{f}(x) = \sum_{k=0}^{p} \widehat{c}_k \phi_k(x)$
$f_k = \int_a^b f(x)\phi_k(x)\,\mathrm{d}x$	$\widehat{c}_k = N^{-1} \sum_{j=1}^{n} h_j \phi_k(\bar{x}_j) \Longrightarrow N^{-1} \sum_{i=1}^{N} \phi_k(X_i)$

- Decay of coefficients f_k fast for a smooth density function $f(x)$;
- histogram with many bins e.g. $n = 1024 = 2^{10}$, as accurate as discrete data;
- property of estimate: $\mathbb{E}\left[\widehat{c}_k\right] = f_k$;
- moments c_k for large k dominated by noise $\rightsquigarrow$ truncation or low-pass filter necessary.

Cosine series expansion

Special basis $\left\{\phi_0(x) = 1,\ \phi_k(x) = \sqrt{2}\cos(\pi k x),\ k = 1, 2, \ldots\right\}$, orthonormal in $[0,1]$:

$$\text{variance of } c_k \quad = \quad V_{kk} = \frac{d_k}{N} \qquad \text{with} \quad d_k \approx 1$$

Moments c_k follow Gaussian distribution (Central limit theorem).

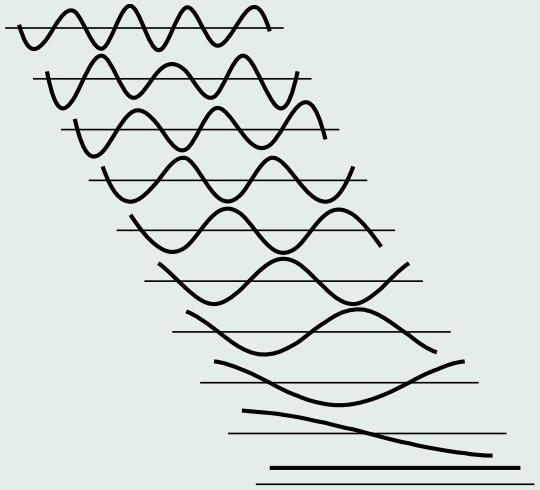
DCT = Discrete Cosine Transformation: in use for compression in JPEG, MPEG, ... and purely real, more efficient than Discrete Fourier Transformation.

Transformation of histogram $\boldsymbol{h}$ either

- by product with matrix $\boldsymbol{U}_{\text{DCT}}$ and inverse by $\boldsymbol{U}^{\text{T}}_{\text{DCT}}$:

$$\boldsymbol{U}^{\text{T}}_{\text{DCT}}\boldsymbol{h} = \boldsymbol{b} \qquad\qquad \boldsymbol{U}_{\text{DCT}}\boldsymbol{b} = \boldsymbol{h}$$

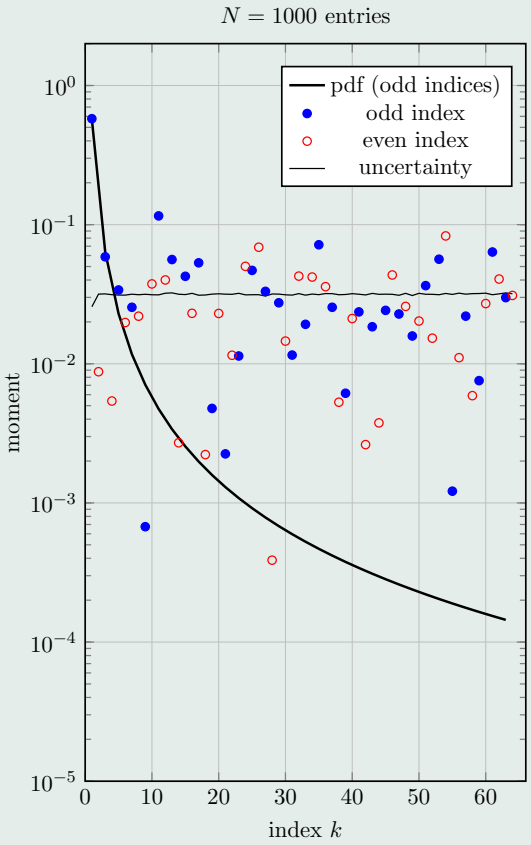
- or by "Fast Cosine Transform", cpu-time $\mathcal{O}(n \times \log n)$.
- generalization to multidimensional data: e.g. 2D-DCT by two 1D-DCT sequences.

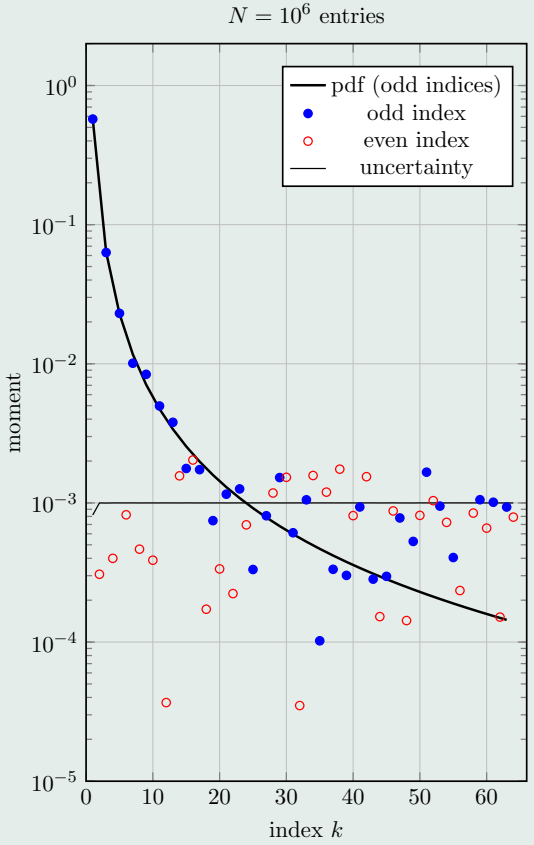


N. Ahmed, T. Natarajan, K.R. Rao, *Discrete Cosine Transform*, IEEE Transactions on Computers, p. 90 – 93, (1974)

Moments from histogram data by DCT

- Fast decay of "true" moments from probability density function;
- uncertainty from histogram moments $\propto 1/\sqrt{N}$;
- moments reach noise-level after $p = 8$ moments (10^3 entries) and after $p = 25$ moments (10^6 entries).

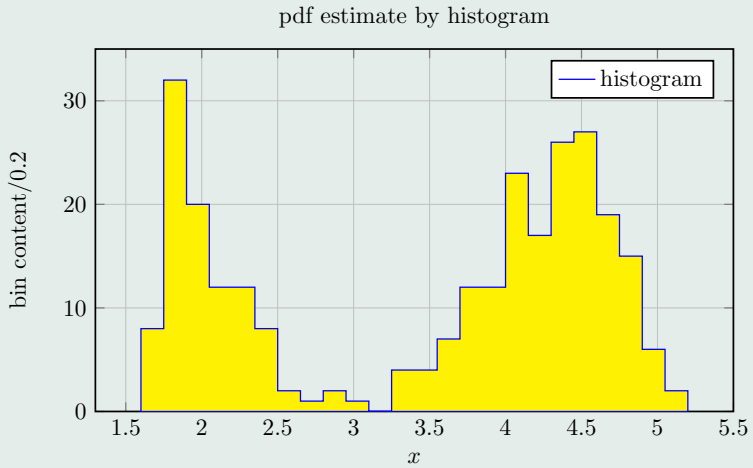


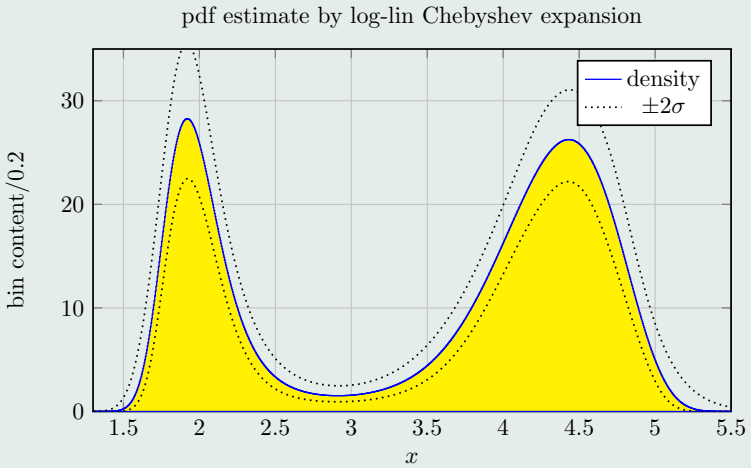


Density estimates I

Example: Geyser-data sample $\{X\}_{272}$ from Yellowstone National park, from the web.
http://www.stat.cmu.edu/~larry/all−of−statistics/=data/faithful.dat

Duration of geyser eruption (in mins)





\+ no correlations between different bins;

− discontinuous (5-bit) density estimate,

− required parameters: nr of bins, origin

− bad/no estimate in low-density regions

− Poisson uncertainty from bin content.

− non-zero correl. between different points;

\+ continuous accurate density estimate,

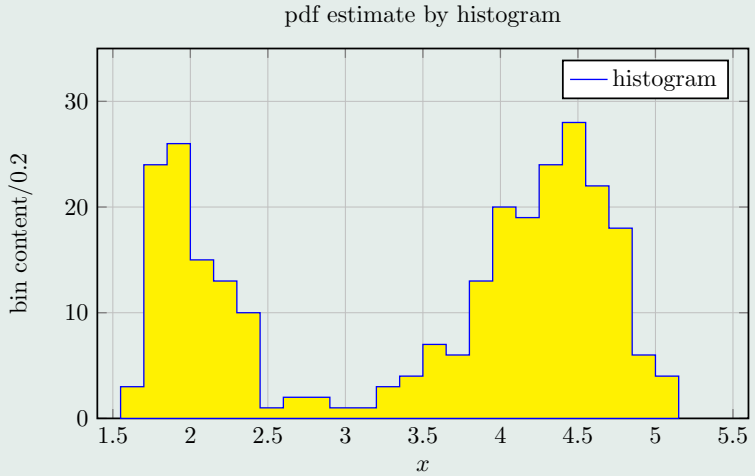
\+ **no** parameters required,

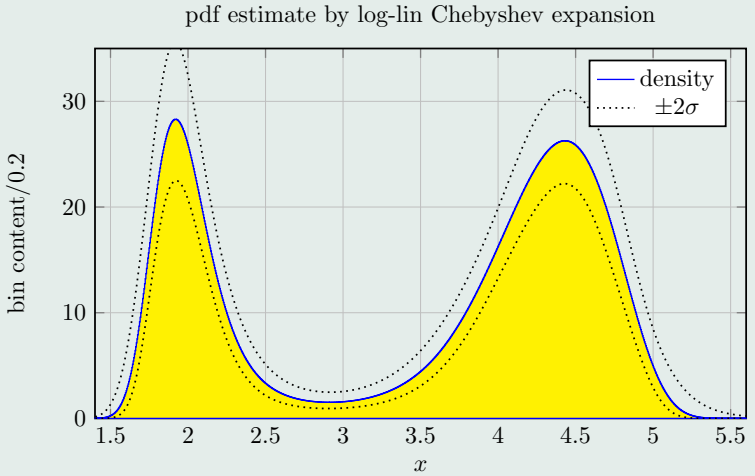
\+ estimate > 0 in whole region,

\+ uncertainty available (log-normal).

Example: Geyser-data sample $\{X\}_{272}$ from Yellowstone National park, from the web.
http://www.stat.cmu.edu/~larry/all−of−statistics/=data/faithful.dat

Duration of geyser eruption (in mins)





\+ no correlations between different bins;

− discontinuous (5-bit) density estimate,

− required parameters: nr of bins, origin

− bad/no estimate in low-density regions

− Poisson uncertainty from bin content.

− non-zero correl. between different points;

\+ continuous accurate density estimate,

\+ **no** parameters required,

\+ estimate > 0 in whole region,

\+ uncertainty available (log-normal).

Log-linear models

Task: construct estimate $\widehat{f}(x)$ of pdf, if a limited number of moments c_k of $f(x)$ is known?

$$\widehat{f}(x) = \sum_{k=0}^{p} \widehat{c}_k \phi_k(x) \Longrightarrow \exp\left[\sum_{k=0}^{p} \gamma_k \phi_k(x)\right]$$

Log-normal: Intuitivly one could expect a **log-normal distribution** of estimated density values, which by definition are restricted to positive density values.

Hiroshi Sekimoto, *An Unfolding Method Leading to a Positive Solution Only*, Nucl. Instrum. Methods Phys. Res. A **228**, pp. 129 – 132, 1984.

Generalized linear models: log-linear models for counts (Poisson)
linear model for predictor $\boldsymbol{\eta}$, related to solution $\widehat{f}(x)$ by logarithmic link function (for Poisson data):

$$\widehat{f}(x) = \exp\left(\eta(x)\right) \qquad \eta(x) = \sum_{k=0}^{p} \gamma_k \phi_k(x)$$

J.A.Nelder and R.W.M.Wedderburn, *Generalized linear models*, J.R.Statist. Soc A **135**, pp. 370–384, 1972

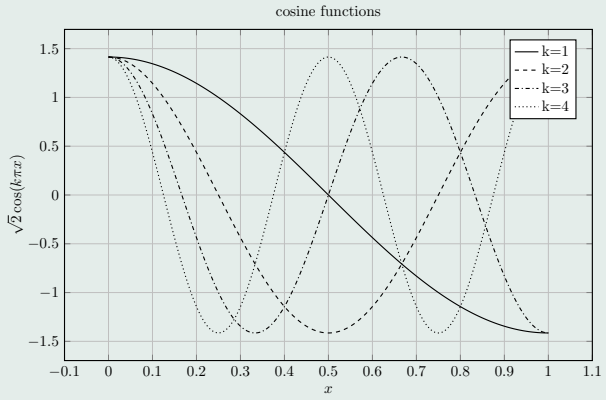
Maximum-entropy approach: The maximum entropy method is based on the concept that the distribution that *maximizes the information entropy* is the statistically most likely to occur. Require *equality constraints for p moments* c_k and ...

maximization of the entropy functional $\qquad S = -\int_a^b \left[f(x) \ln f(x)\right] \, \mathrm{d}x$

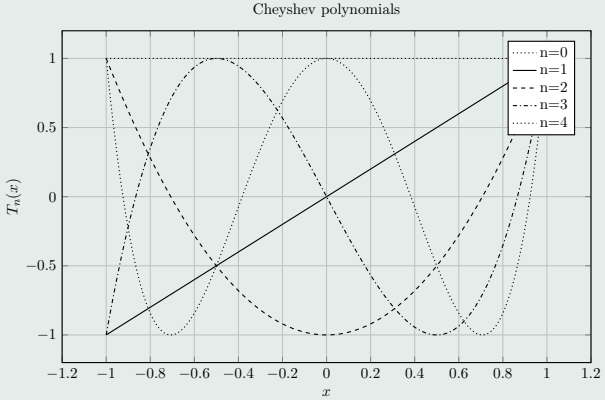
E. T. Jaynes, *Information Theory and Statistical Mechanics*, Physical Review **106** pp. 620–630 (1957)

Cosine and Chebyshev polynomial expansion

$$\cos(k\pi t)$$



$$T_n(x) \equiv \cos(n \arccos(x))$$

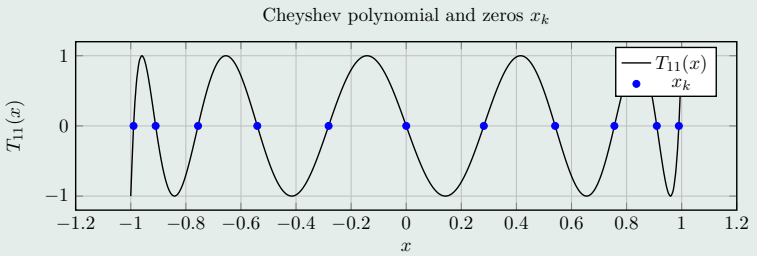


$$\int_{-1}^{+1} \frac{T_n(x)T_m(x)}{\pi\sqrt{1-x^2}}\mathrm{d}x = \begin{cases} 1 & : n = m = 0 \\ 1/2 & : n = m \neq 0 \\ 0 & : n \neq m \end{cases}$$

$$c_n = 2\int_{-1}^{+1} \frac{T_n(x)f(x)}{\pi\sqrt{1-x^2}}\mathrm{d}x \quad n = 1, 2\ldots$$

Gauss-Lobato zeros for $k = 0, 1 \ldots n-1$:

$$x_k = \cos(\pi(2k+1)/(2n))$$
$$1/\pi \arccos(x_k) = t_k = (2k+1)/(2n)$$



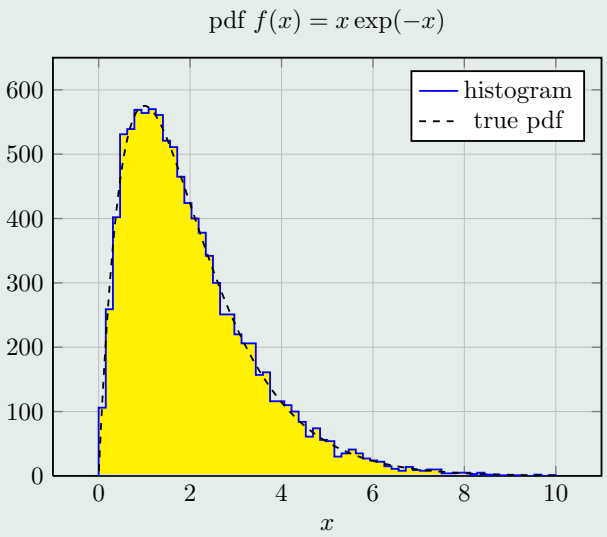
$$\frac{1}{m}\sum_{k=0}^{m-1} T_i(x_k)T_j(x_k) = \begin{cases} 1 & i = j = 0 \\ 1/2 & i = j \neq 0 \\ 0 & i \neq j \end{cases}$$

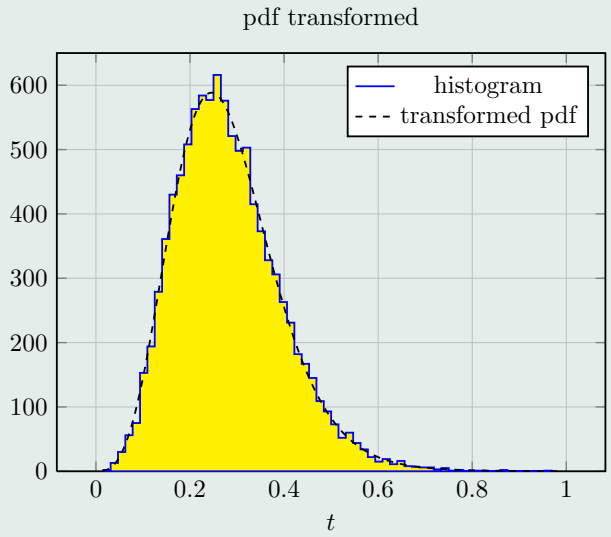
$$c_j = \frac{2}{n}\sum_{k=0}^{n-1} f(x_k)T_j(x_k) \quad j = 1, 2\ldots$$

Transformation $x \Longrightarrow t$ by arccos

Only cos-terms after transformation:

$$x = -\cos(\pi t) \in [-1, +1] \qquad t = \frac{1}{\pi} \arccos(-x) \in [0, 1]$$





Chebyshev expansion of $\widetilde{f}(x)$ by cosine expansion of $\widetilde{g}(t)$

$$h_i = N \times \frac{2}{Z_2} \exp\left[\sum_{k=1}^{m} \gamma_k T_k(\bar{x}_i)\right] \qquad g_i = N \times \frac{\pi \sin(\pi \bar{t}_i)}{Z_1} \exp\left[\sum_{k=1}^{m} \gamma_k \cos(\pi k \bar{t}_i)\right]$$

with same coefficients γ_k for cosine and Chebyshev expansion.

(3) Singular value decomposition – SVD

SVD . . . a new way to see into the heart of a matrix.
GILBERT STRANG

The SVD in least-squares problems allows insight into the structure of the matrix $\boldsymbol{A}$ and the LS solution of the matrix equation $\boldsymbol{A}\boldsymbol{f} \simeq \boldsymbol{g}$

Singular value matrix decomposition $\quad \boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}} = \sum_{i=1}^{n} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^{\mathrm{T}} \qquad \boldsymbol{\Sigma} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{V}$

$\boldsymbol{\Sigma} = \mathbf{diag}\,\{\sigma_1, \ldots, \sigma_n\}$ with ordered *singular values* $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$.

$$\begin{pmatrix} & {\scriptstyle (n)} \\ {\scriptstyle (m)} & \boldsymbol{A} \\ & \end{pmatrix} = \left(\begin{array}{c|c|c|c|c} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_{n-1} & \boldsymbol{u}_n \end{array}\right) \begin{pmatrix} \sigma_1 & & & & \mathbf{0} \\ & \sigma_2 & & & \\ & & \ddots & & \\ & & & \sigma_{n-1} & \\ \mathbf{0} & & & & \sigma_n \end{pmatrix} \left(\begin{array}{c} \boldsymbol{v}_1^{\mathrm{T}} \\ \hline \boldsymbol{v}_2^{\mathrm{T}} \\ \hline \cdots \\ \hline \boldsymbol{v}_{n-1}^{\mathrm{T}} \\ \hline \boldsymbol{v}_n^{\mathrm{T}} \end{array}\right)$$

Matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ with orthonormal columns: the m-vectors $\boldsymbol{u}_i$ and the n-vectors $\boldsymbol{v}_i$ are called *left* and *right singular vectors* of matrix $\boldsymbol{A}$. They have, with increasing index an increasing number of *sign-changes* in their elements, corresponding to *higher frequencies*.

SVD developed in 19. century by geometers; stable algorithm by Golub and Reinsch (1970 ALGOL).

Simple model:

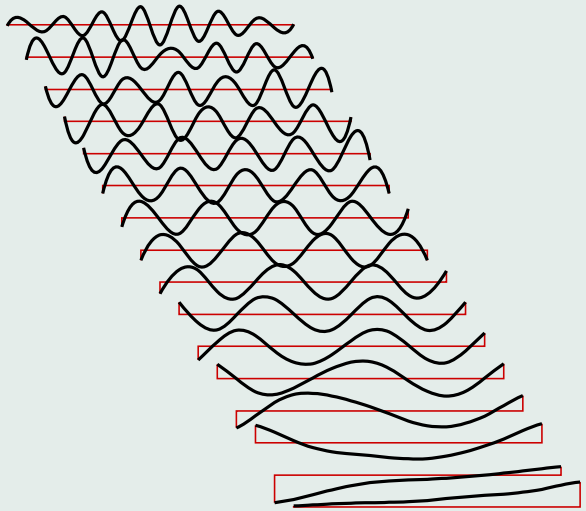
$$\underbrace{A}_{\text{response matrix}} \times \overbrace{f}^{\text{solution}} \simeq \underbrace{(g + \mathcal{N}_m)}_{\text{measurement}}$$

Direct (folding): calculate expected measured distribution

$$g = Af = U\Sigma V^T f = \sum_{j=1}^{n} \sigma_j \left(v_j^{\mathrm{T}} f \right) u_j$$

The "true" vector f is decomposed into components $\left(v_j^{\mathrm{T}} f \right)$, and the expected "measured" vector g is a superposition of the vectors u_j, **weighted with singular values** σ_j.

Components with $\sigma_j = 0$ or $\sigma_j \ll 1$ *disappear* in the real measured vector $g + \mathcal{N}$ (noise).



Singular vectors

...and the (2) inverse problem

Naive Least Squares: minimize $\|Af - g\|^2 \quad \rightsquigarrow \quad \widehat{f} = A^\dagger g \qquad A^\dagger = \left(A^T A\right)^{-1} A^T$

Inverse (unfolding): SVD solution identical to LS solution using generalized inverse $A^\dagger$

$$\widehat{f} = A^\dagger g = V\Sigma^{-1}\left(U^T g\right) = \sum_{j=1}^{n} \frac{1}{\sigma_j}\left(u_j^T g\right) v_j = \sum_{j=1}^{n} \frac{1}{\sigma_j} c_j v_j$$

The estimated "true" vector $\widehat{f}$ is a superposition of the vectors v_j, with 'measured" Fourier coefficients $c_j = \left(u_j^T g\right)$, and **weighted with the inverse singular values** $1/\sigma_j$.

$$V_f = A^\dagger A^{\dagger T} = V\Sigma^{-2}V^T = \sum_{j=1}^{n} \left(\frac{1}{\sigma_j^2}\right) v_j v_j^T$$

Note: building blocks of the solution are not single bins, but whole distributions v_j

Properties of the naive LS solution

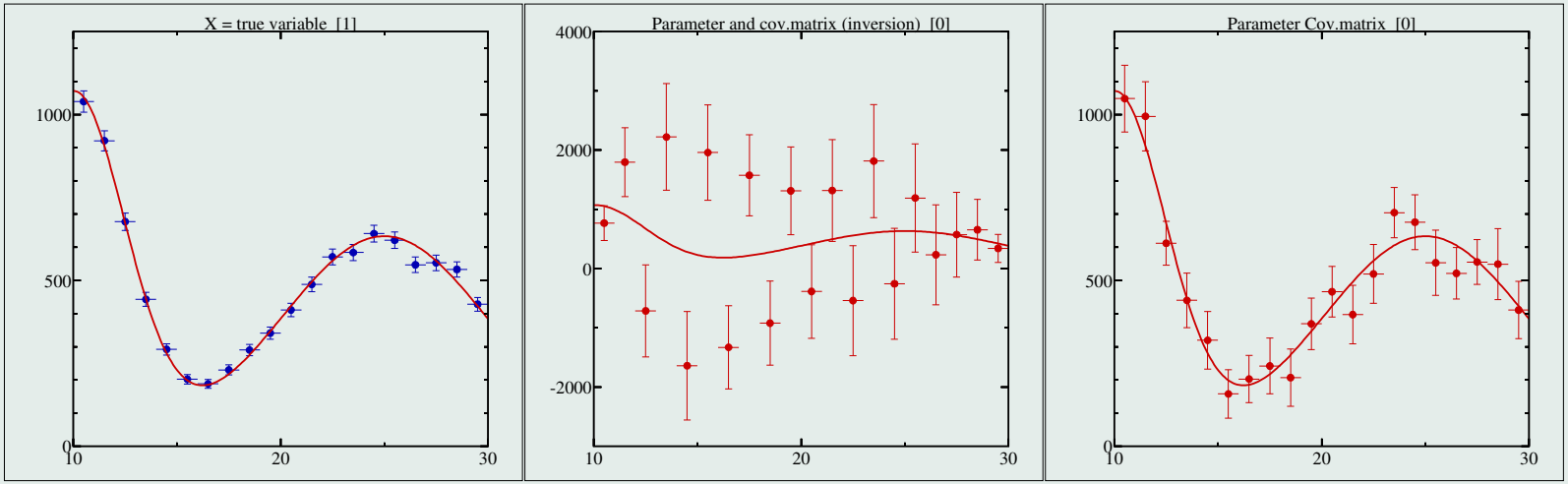
(1) Check of a potential bias in the solution:

$$\text{estimator} \qquad \widehat{f} = A^{+} g \quad \text{with} \quad E[g] = A f_{\text{exact}}$$

$$E\left[\widehat{f}\right] = A^{+} E[g] = \left(A^{+} A\right) f_{\text{exact}} = f_{\text{exact}} \qquad \widehat{f} \text{ unbiased}$$

(2) Variance: V_f = lower bound, given by the Rao-Cramér-Frechet (RCF) inequality. Gauss-Markov theorem: the least square estimate is unbiased and has smallest possible variance.

But: the result $\widehat{f}$ will often show large, unacceptable fluctuations!



10000 entries, no smearing | unfolding by inversion | truncation – $p = 15$ terms

Strategies

Naive unfolding results will be wildly fluctuating and are not acceptable.

Wider bins – reduce number n of bins of estimated "true" vector $\hat{x}$; this will avoid very small singular values σ_j.

Cut/truncation – ignore insignificant Fourier coefficients c_j (value $\approx$ uncertainty) with small values σ_j:

$$\hat{f} = A^{+} g = \sum_{j=1}^{n} \frac{1}{\sigma_j} c_j v_j \qquad \Rightarrow \qquad \sum_{j=1}^{p} \frac{1}{\sigma_j} c_j v_j \quad \text{with} \quad p < n$$

A sharp cut-off may result in Gibbs oscillations.

Regularization – add e.g. 'smoothing" term $\|Lx\|^2$ to least squares condition:

$$\underset{a}{\text{minimize}} \ \|Af - g\|^2 + \tau f^{\mathrm{T}} \left(L^{\mathrm{T}} L \right) f$$

Result is a smooth cut-off, that avoids Gibbs oscillations.

Parametrized unfolding – if a well-known parametrization $f(t; a)$ has to be tested, this parametrization can be used directly, without the need for regularization.

Effect of a Gaussian resolution: convolution ...

Approximation of an even function $f(t)$ with period 1:

$$f(t) \approx a_0 + \sum_{k=1}^{n-1} a_k \cos(\pi k t) \qquad g(s) \approx \alpha_0 + \sum_{k=1}^{n-1} \alpha_k \cos(\pi k s)$$

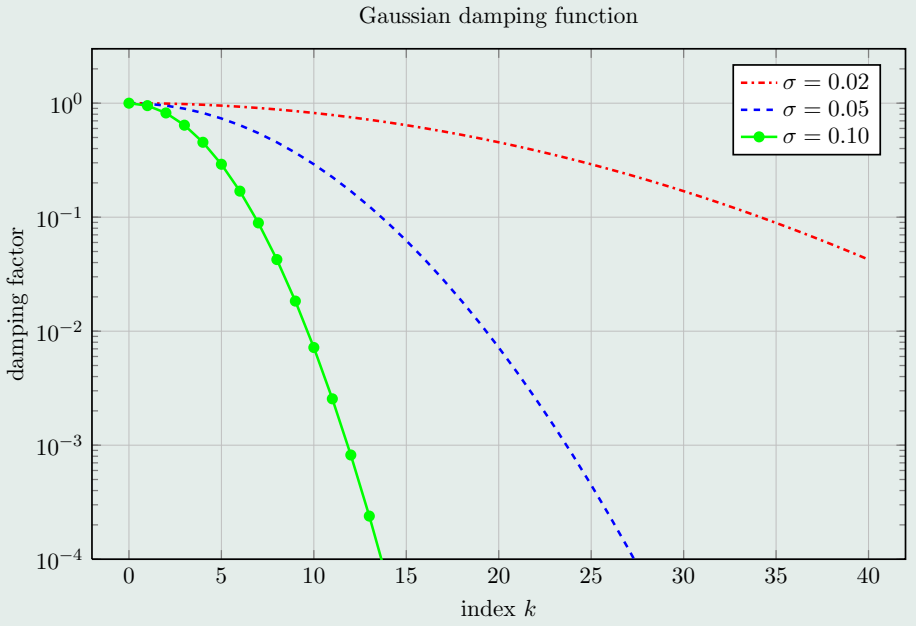
Convolution: Effect of a Gaussian kernel with standard deviation σ:

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(s-t)^2}{2\sigma^2}\right) \times \cos(\pi k t)\,\mathrm{d}t = \exp\left(-\frac{(\pi k \sigma)^2}{2}\right) \times \cos(\pi k s)\,,$$

blue dashed curve:
$\sigma = 0.05$, i.e. full range $\equiv 20\,\sigma$

Damping factor becomes

$$\approx 10^{-2} \text{ for } k = 19$$
$$\approx 10^{-4} \text{ for } k = 27$$



The exponential factor $\exp\left(-\frac{\pi^2 k^2 \sigma^2}{2}\right)$ corresponds to the singular value σ_k.

Deconvolution: correct coefficients

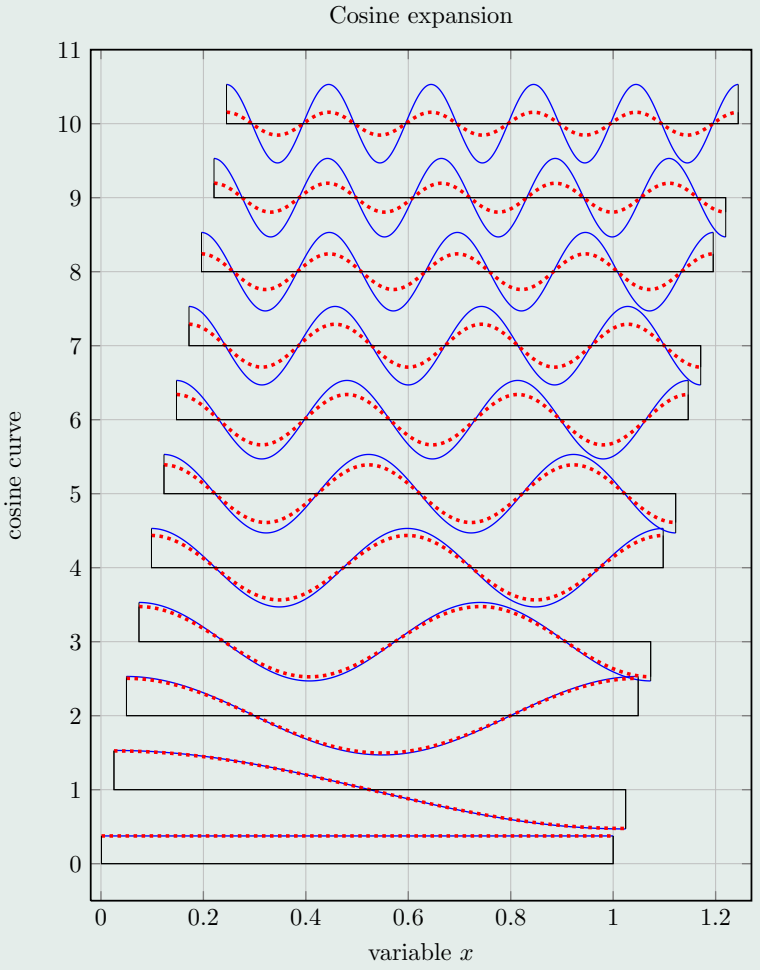
for small k :

$$\widehat{a}_k = \alpha_k \times \exp\left(\frac{\pi^2 k^2 \sigma^2}{2}\right)$$

for large k :

$$\widehat{a}_k = (\text{noise level}) \times \exp\left(\frac{\pi^2 k^2 \sigma^2}{2}\right)$$

Blue noise:
with k^2 exponentially increasing noise!

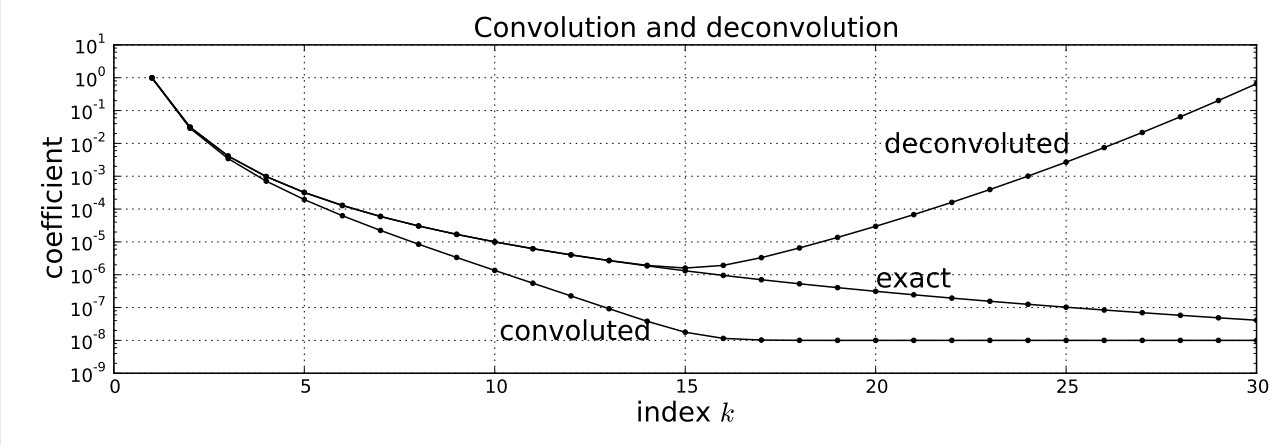


. . . and deconvolution

Deconvolution: expand convoluted ("measured") function $g(s)$ to obtain coefficients α_k:

Correct coefficients back: $\quad \widehat{a}_k = \exp\left(\frac{\pi^2 k^2 \sigma^2}{2}\right) \times \alpha_k \qquad$ for small k

$$= \exp\left(\frac{\pi^2 k^2 \sigma^2}{2}\right) \times \sqrt{\alpha_k^2 + \epsilon^2} \xrightarrow[k\to\infty]{} \epsilon \times \exp\left(\frac{\pi^2 k^2 \sigma^2}{2}\right)$$



- Fourier coefficients α_k are below the exact ones a_k and reach for $k = 15$ a level of about 10^{-8}, due to **round-of-errors** ϵ.
- The deconvoluted Fourier coefficients are correct only up to $k = 15$.
- Deconvolution with $k \gg 15$ will have result dominated by noise.

Singular values and MC statistics

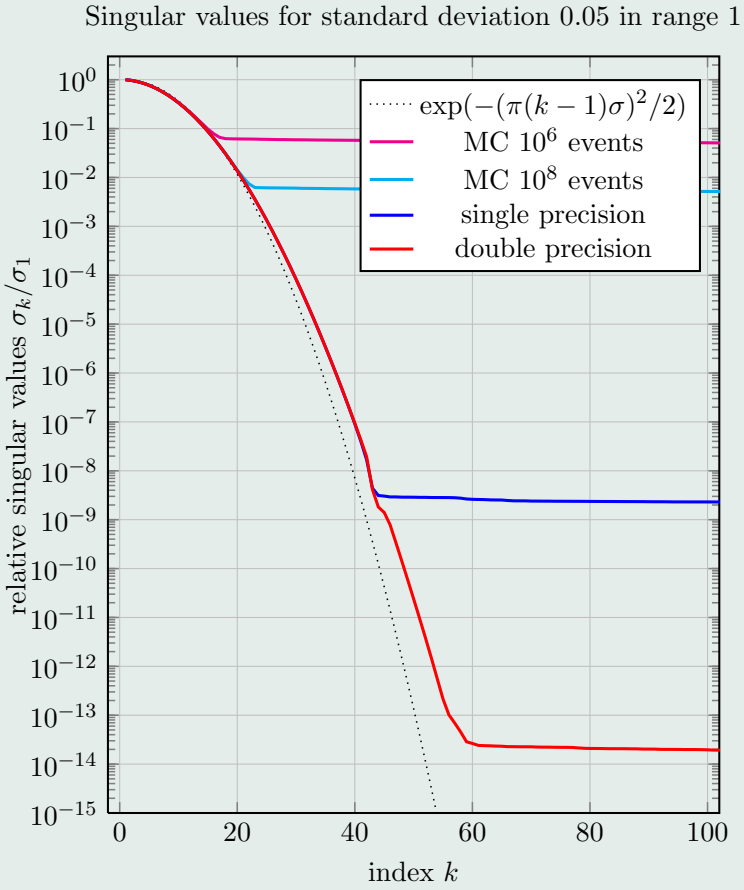
How many singular values and vectors can be determined by MC simulation with limited number of events?

Four versions of response matrix:

- MC 10^6 and 10^8 events
- Single and double precision analytical computation

Limitation by statistical fluctuations and roundoff-errors!

	$p =$ nr of σ_i	σ_p
10^6 MC	15	0.06
10^8 MC	21	0.006
single prec.	43	3×10^{-9}
double prec.	56	3×10^{-14}



(4) Iterative unfolding methods

Iterative methods to solve the equation $Af = g$ for the vector f access the (unaltered) response matrix only by matrix-vector multiplication with A and A^T and generate a sequence of iteration vectors $f^{[k]}, k = 1, 2, \ldots$ They allow to solve **large-scale inverse problems**, where factorization methods are infeasible (response often point spread function).

"...need iterative schemes with the intrinsic property that they ...pick up those SVD components corresponding to the largest singular values ... (semi-convergence)" [Hansen]

Methods

Landweber method (1951): additive corrections determined iteratively.
Corresponds to Least Squares solution with **Gaussian** distributed data.

L.Landweber(1951): An iteration formula for Fredholm integral equations of the first kind. Amer. J. Math. 73, 615–624

Richardson-Lucy method (1972, 1974): multiplicative corrections determined iteratively, used for restoration of Hubble ST images
Corresponds to Maximum Likelihood solution with **Poisson** distributed data.

W.H. Richardson,(1972) *Bayesian-Based Iterative Method of Image Restoration*, Journal of the Optical Society of America A, 62 (1): 55 – 59. L.B.Lucy(1974), *An iterative technique for the rectification of observed distributions*, Astronomical Journal 79 (6): 745 – 754.

Landweber method

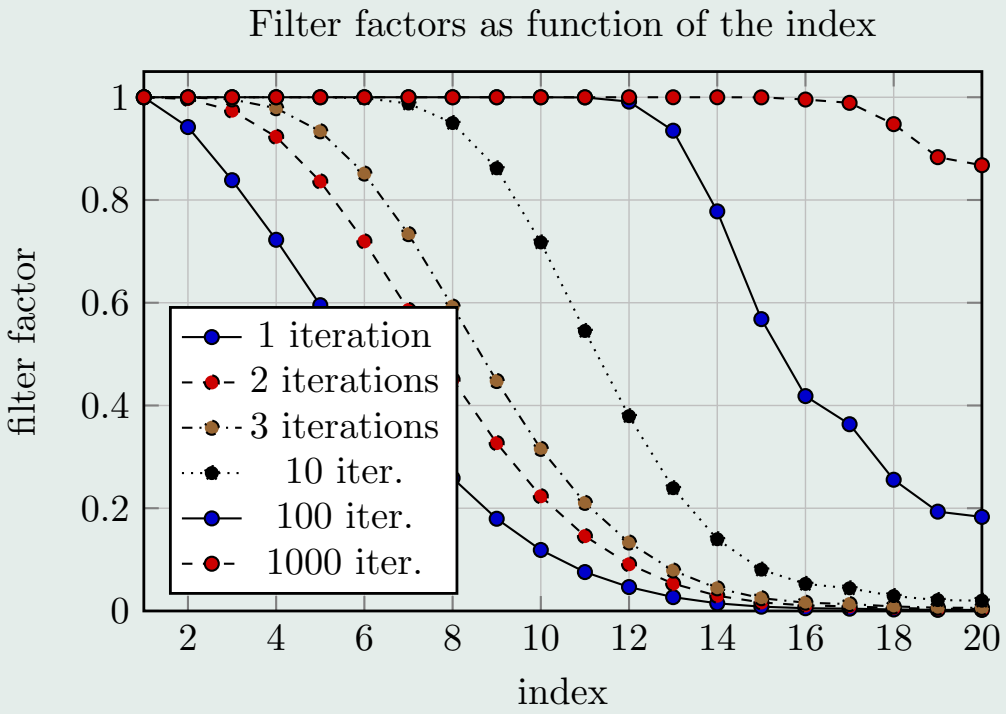
Starting (iteration $k = 0$) from distribution $\widehat{f}_j^{(0)}$ (e.g. uniform):

$$\text{iterate:} \quad \widehat{\boldsymbol{f}}^{(k+1)} = \widehat{\boldsymbol{f}}^{(k)} + \frac{\omega}{\sigma_1^2} \boldsymbol{A}^{\mathrm{T}} \left(\boldsymbol{g} - \boldsymbol{A}\widehat{\boldsymbol{f}}^{(k)} \right) \qquad 0 < \omega < 2$$

Filter factors for $\widehat{\boldsymbol{f}}^{(k)}$ ($\omega = 1$):

$$G_i^{[k]} = 1 - \left(1 - \frac{\sigma_i^2}{\sigma_1^2} \right)^k$$

$$i = 1, 2, \ldots n$$



Richardson-Lucy (RL) unfolding

Starting (iteration $k = 0$) from distribution $\widehat{f}_j^{(0)}$ (e.g. uniform) iterate:

$$\widehat{f}_j^{(k+1)} = \left(\sum_i A_{ij} \frac{g_i}{g_i^{(k)}} \right) \cdot \frac{\widehat{f}_j^{(k)}}{\alpha_j} \qquad \text{with} \quad g_i^{(k)} = \sum_j A_{ij} \widehat{f}_j^{(k)}$$

The algorithm shows semi-convergence: often a smooth solution $\widehat{f}$ is obtained after few iterations, and then oscillations occur corresponding to the *naive* maximum likelihood solution. [Shepp, L. A.; Vardi, Y. (1982), *Maximum Likelihood Reconstruction for Emission Tomography*, IEEE Transactions on Medical Imaging 1: 113]

Identical to algorithm called "Bayesian unfolding" in HEP:

G. D'Agostini, A multidimensional unfolding method based on Bayes' theorem, Meth. in Phys. Res. A362 (1995) 487

The RL-algorithm has the following properties

- Estimates $\widehat{f}_j$ are always positive (starting values $\widehat{f}_j^{[0]}$ modified by positive *factors*);
- result is ML solution for Poisson distributed bin entries;
- no straightforward determination of the covariance matrix;
- regularization (damping of oscillations) dependent on iteration number (and by histogram binning and start distribution): no prescription for *blind* unfolding; sometimes intermediate smoothing recommended to suppress developing oscillations (in HEP);
- typical application is picture deblurring using simple point-spread-function (PSF).

Opinions about the RL algorithm

- "…the iteration sequence converges quickly to a reasonable unfolded histogram, sometimes after a single iteration."
- "…one can realize that in most of the cases a good agreement is reached after a few iterations."
- "In these cases where the observed distribution indicates that there are sharp structures in the true distribution, the iterative method permits to implement these in the input distribution. In this way the number of iterations is reduced and oscillations are avoided."
- "…solution obtained by stopping the iteration does depend on the starting distribution. We may choose it according to our expectation."
- "Unfolding is not an entirely objective procedure."

…studies from Particle Physics

- "convergence can be extremely slow"
- "The slow convergence of the method is sometimes argued to be a positive feature of the algorithm, since a fast program would bring us quickly close to the minimum norm solution that is usually nonsense …"
- "…need iterative schemes with the intrinsic property that they …pick up those SVD components corresponding to the largest singular values …"
- "…if the system of equations has many solutions, then the algorithm will converge to the solution that is closest to the …starting distribution."
- "**Contributions corresponding to small singular values** show an extremely slow convergence, and are, after a few iterations, still **biased to the initial assumption**."

…from the statistical literature.

(5) Regularized Least Squares

More realistic with $\mathcal{N}$ = noise = statistical fluctuations plus systematic deviations

$$\underbrace{(\boldsymbol{A}+\mathcal{N}_{MC})}_{\text{response matrix}} \times \overbrace{\boldsymbol{f}}^{\text{solution}} + \underbrace{(\boldsymbol{b}+\mathcal{N}_b)}_{\text{background}} = \underbrace{(\boldsymbol{g}+\mathcal{N}_m)}_{\text{measurement}}$$

Noise contributions combined with small singular values (response matrix) will result in fluctuating/oscillating result.

- Optimization: regularization term (penalty) and χ^2-expression

Regularization	minimize	$\|\boldsymbol{Af}-\boldsymbol{g}\|^2 + \tau \times$ regularization term

- or (better) maximum likelihood (Poisson) with (instead χ^2) deviance

$$D(\boldsymbol{g};\boldsymbol{g}^{\text{fit}}) = 2\sum_{i=1}^{n}\left[g_i \log g_i - g_i \log g_i^{\text{fit}} - (g_i - g_i^{\text{fit}})\right]$$

- Reduce, by regularization, influence of noise (in all its forms) on the solution, but balance between variance and bias.
- Regularization based on norm, squared second derivatives or entropy.

Norm regularization

Norm regularization with diagonal penalty term:

$$\text{minimize} \quad \|Af - g\|^2 + \tau \times f^T f$$

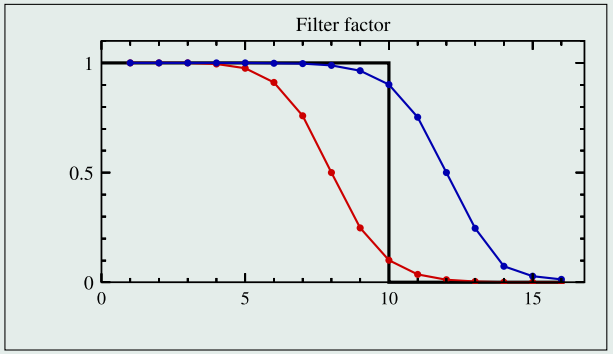
Solution by SVD of response matrix A recommended (but not necessary).

$$\text{Solution:} \quad \hat{f} = \sum_{k=1}^{n} \varphi_k \times \frac{c_k}{\sigma_k} v_k \qquad \text{with} \quad \text{Fourier coefficients } c_k$$

filter factor $\varphi_j = \dfrac{\sigma_j^2}{\sigma_j^2 + \tau}$

$= 1 \quad \ldots \quad 0.5 \quad \ldots \quad 0$

$\varphi_k = \quad 0.5 \qquad \text{for} \quad \tau = \sigma_k^2$



Caution — the shape of the filter curve is fixed!

- Significant terms should not be reduced, in order to avoid a bias;
- but insignificant contributions should be damped, to reduce influence of noise;
- no standard procedure to determine optimal value of τ.
- Standard Tikhonov regularization with a single regularization parameter τ may be too simple under the conditions of HEP experiments (several noise sources).

Squared second derivative regularization

Regularization based on square of second derivatives $\approx -f_{i-1} + 2f_i - f_{i+1}$.

$$\text{minimize} \quad \|\boldsymbol{A}\boldsymbol{f} - \boldsymbol{g}\|^2 + \tau \times \boldsymbol{f}^{\mathrm{T}} \left(\boldsymbol{L}^{\mathrm{T}}\boldsymbol{L}\right) \boldsymbol{f}$$

$$\text{second der.} \quad \boldsymbol{L} = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{pmatrix} \in \mathbb{R}^{n \times n}$$

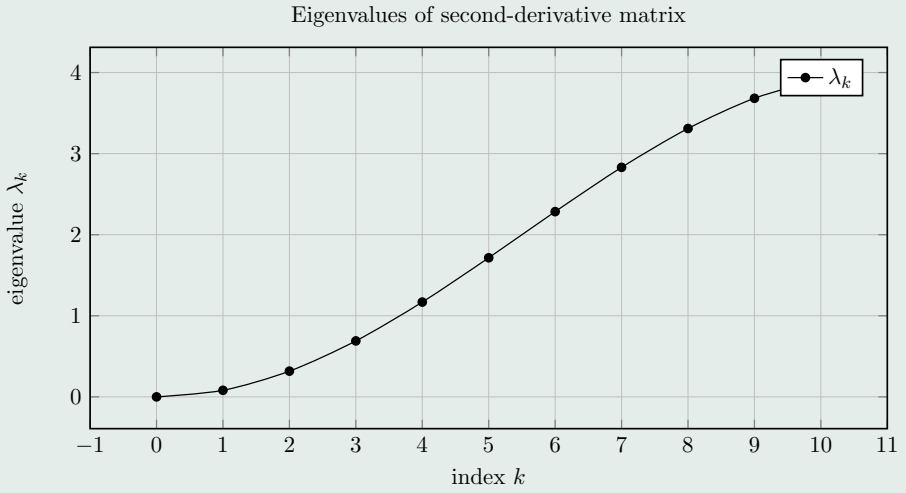
Requires **GSVD** = generalized SVD (two non-diagonal matrices) (not in CERN libraries).

Decomposition analytically known:

$$\boldsymbol{C} = \left(\boldsymbol{L}^{\mathrm{T}}\boldsymbol{L}\right) = \boldsymbol{U}_{\mathrm{DCT}}\, \boldsymbol{\Lambda}^2\, \boldsymbol{U}_{\mathrm{DCT}}^{\mathrm{T}}$$

Do not try to invert $\boldsymbol{C} = (\boldsymbol{L}^{\mathrm{T}}\boldsymbol{L})$!

$$\textbf{filter factor} \quad \varphi_j = \frac{\gamma_j^2}{\gamma_j^2 + \tau}$$



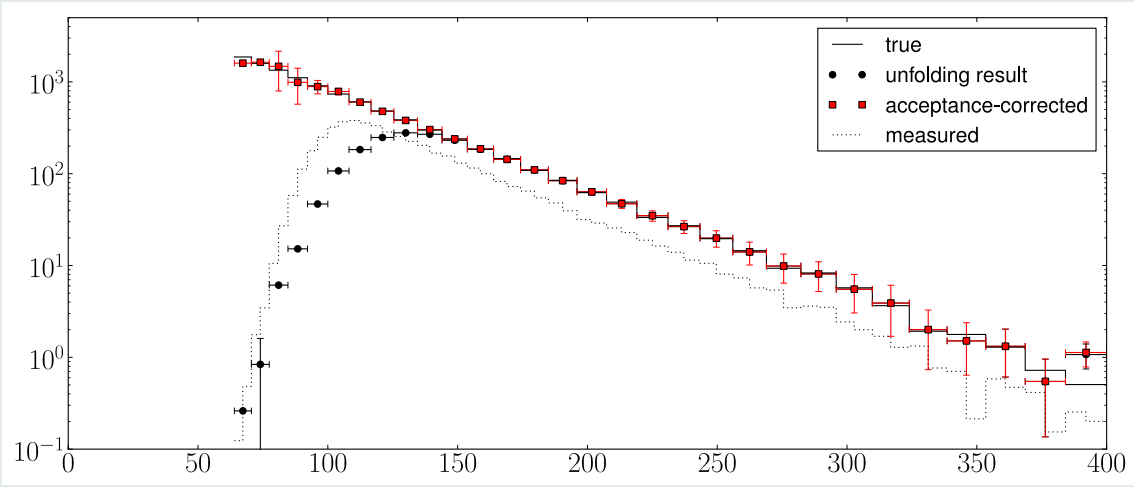
G. Golub and C.F.Van Loan, *Matrix Computation*. C.F.Van Loan, *Generalizing the Singular Value Decomposition* (1976)

Example: Unfolding of inclusive jet cross section as a function of p_T

Measurement:

- p_T, measured by calorimeter, systematically underestimated (bias);
- region of small p_T unmeasurable: uncertainties increase for region $p_T \to 0$

jet transverse momentum p_T distribution



MC simulation

FNAL-publication: Correction in two separate steps, for (1) bias, and for (2) limited resolution (smearing) using bin-by-bin CF method (standard procedure).

Unfolding (see figure): in one step, taking into account bias, finite resolution (smearing), and limited acceptance;

- allows the consistent determination of the covariance matrix,
- steeply falling distribution: intermediate transformation to $\sqrt{p_T} \Rightarrow$ variable bin size and constant standard deviation.

Unfolding with Tikhonov regularization, using square of second derivatives.

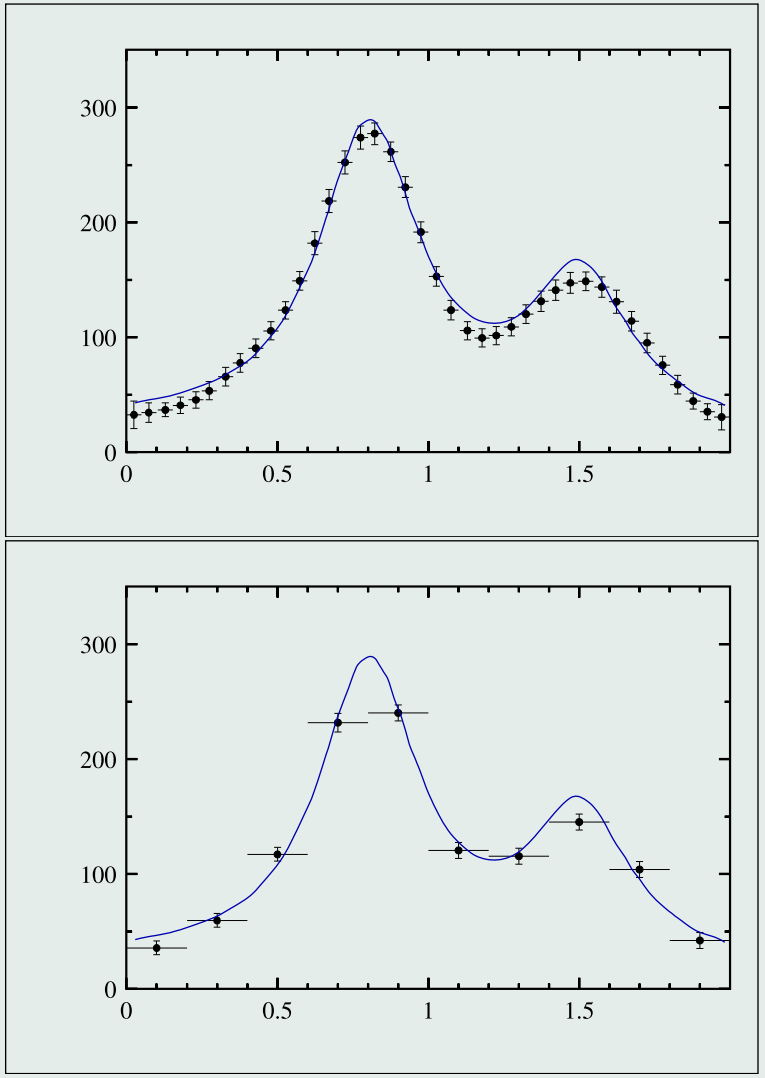
"Histogram"-like presentation of result.
(from A. Hoecker and V. Kartvelishvili: NIM A 372)

Result with 40 data points:

- almost like a "band" representing result
- constructed from 10 significant parameters, with 40-by-40 covariance matrix, singular with rank 10;
- large positive correlations, therefore few sign-changes of the residuals to MC input distribution.

(Same) result with 10 data points:

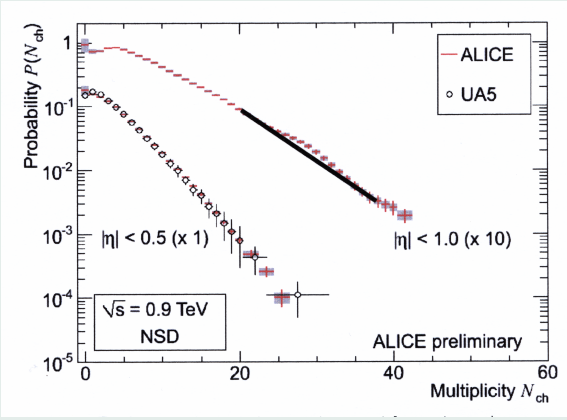
- each point represents a bin average of the result
- 10-by-10 covariance matrix non-singular, inverse is weight matrix.
- small and negligible correlations.



Example for positive correlation

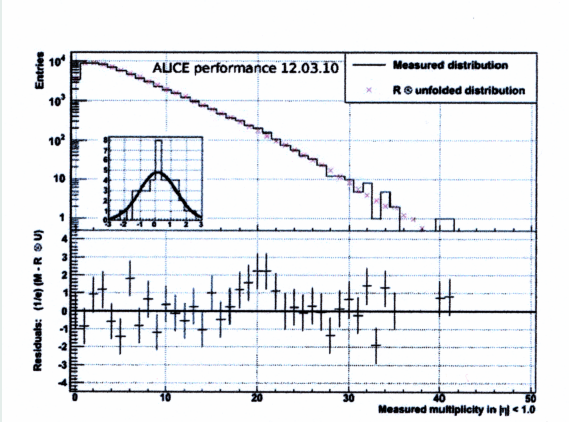
Unfolding of charged multiplicity distribution in ALICE:

Hump observed around $N_{\text{ch}} = 30 \pm 5$ $\Rightarrow$ new physics?



Charged Multiplicity distribution

Residuals of measured distribution



Small upward fluctuation observed in ≈ 4 bins around $N_{\text{ch}} = 20$.

Bins of the unfolded distribution are (positively) correlated over a large range. The origin of the 10-bins hump at $N_{\text{ch}} = 30$ is a 4-bin fluctuation at observed $N_{\text{ch}} = 20$.

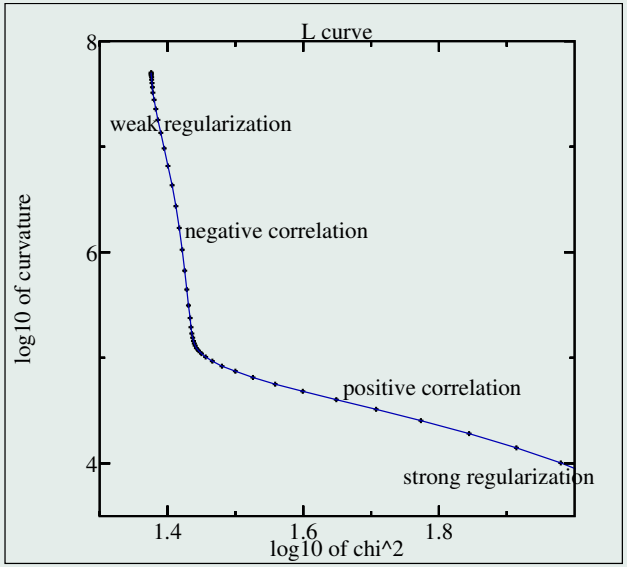
Note: $n = 40$ bins are unfolded here from $m = 30$ measured bins – covariance matrix must have rank defect > 10.

From: Jan Fiete Grosse-Oetringhaus: Comments on Unfolding Methods in ALICE, PHYSTAT 2011

Discrepancy principle (Morozov): choose τ such that $\left\| \boldsymbol{A}\boldsymbol{f}_{\text{reg}} - \boldsymbol{g} \right\|_2^2 = n_{df}$

L-curve method: plot $\left\| \boldsymbol{L}\boldsymbol{f}_{\text{reg}} \right\|_2$ versus $\left\| \boldsymbol{A}\boldsymbol{f}_{\text{reg}} - \boldsymbol{g} \right\|_2$ for a set of τ-values $\Rightarrow$ dependence has shape of an L with a distinct corner $\rightsquigarrow$ optimal value of τ.

L-curve



Effective number of degrees of freedom: ($\mathcal{RUN}$) determine n_{df} from spectrum of Fourier coefficients c_j, and determine τ such that sum of filter factors

$$\sum_{j=1}^{p} \frac{\sigma_j^2}{\sigma_j^2 + \tau} = n_{df}$$

Minimum of global correlation: minimum mean value of global correlation coefficients

Definition: $$\rho_j = \sqrt{1 - \left[\left(\boldsymbol{V}_f \right)_{jj} \cdot \left(\boldsymbol{V}_f^{-1} \right)_{jj} \right]^{-1}} \quad \text{with} \quad 0 \leq \rho_j \leq 1$$

∗ (6) Projection methods ... without histograms

Histograms with few bins (e.g. $32 = 2^5$) deteriorate the resolution.

Alternative method:

- Avoid the use of histograms with low number of bins;
- use unbinned data or histogram with many bins, e.g. $1024 = 2^{10}$ bins and response array with 1024×1024 bins;
- transform all vector- and matrix arrays using well-defined DCT matrix $\boldsymbol{U}_{\mathrm{DCT}}$ after arccos transformation to *frequency space*.

measured data	$\widetilde{g} = \boldsymbol{U}_{\mathrm{DCT}}^{\mathrm{T}}\, g$
MC input	$\widetilde{f}_{MC} = \boldsymbol{U}_{\mathrm{DCT}}^{\mathrm{T}}\, f_{MC}$
response matrix	$\widetilde{A} = \boldsymbol{U}_{\mathrm{DCT}}^{\mathrm{T}}\, A \boldsymbol{U}_{\mathrm{DCT}}$

The solution $\widetilde{f}$ is obtained in the space of frequencies:

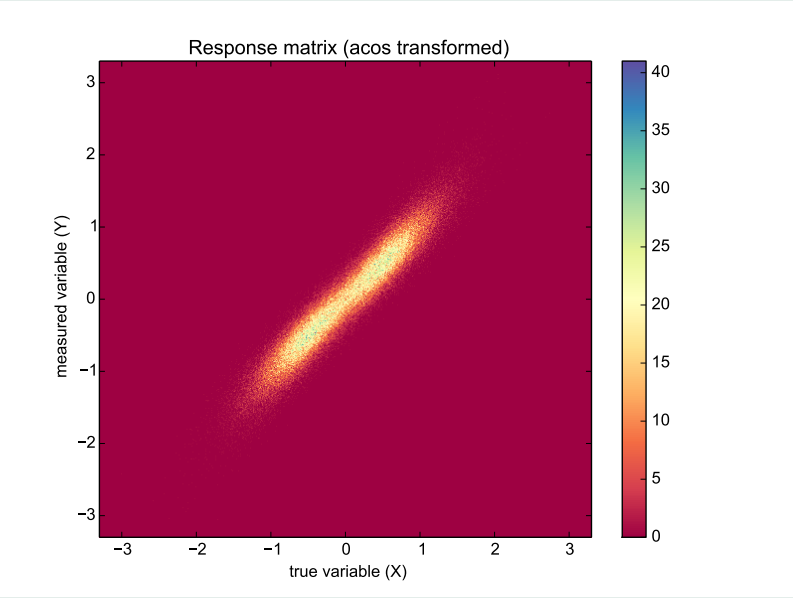
$$\widetilde{A}\widetilde{f} = \widetilde{g}$$

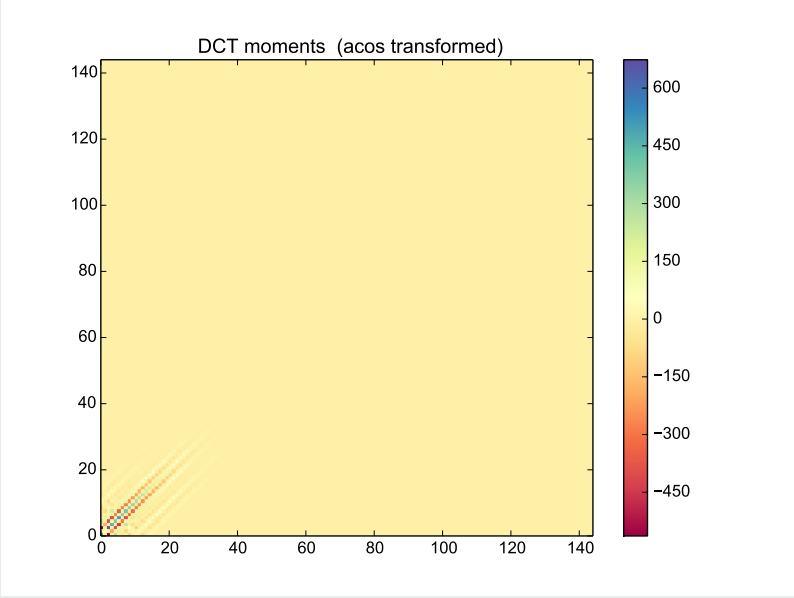
and is finally expressed as log-parametrization $\widetilde{f}(x)$ in terms of Chebyshev polynomials.

Transformation of 2D-histogram matrix A by DCT:

$$U_{\mathrm{DCT}}^{\mathrm{T}} A U_{\mathrm{DCT}} = \widetilde{A}$$

in analogy to SVD: $U^{\mathrm{T}} A V = \Sigma$





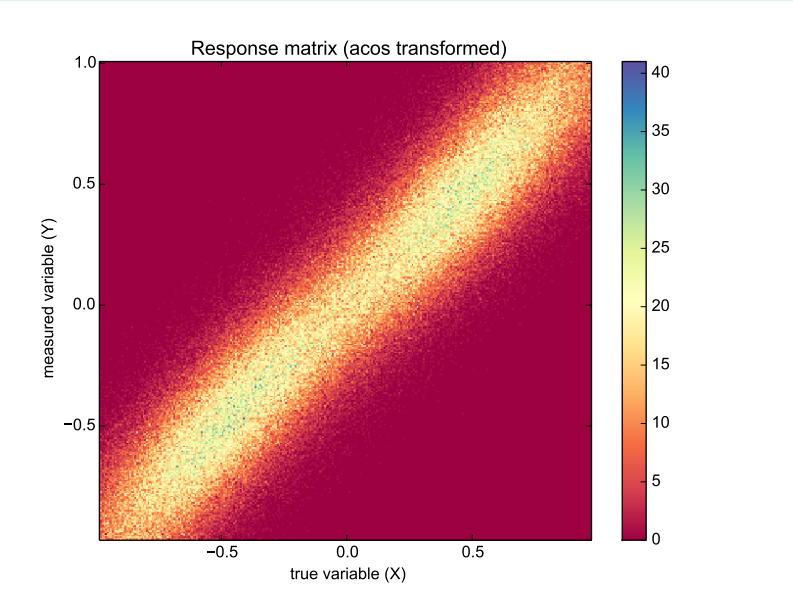
1024×1024 bin response histogram from Monte Carlo simulation.

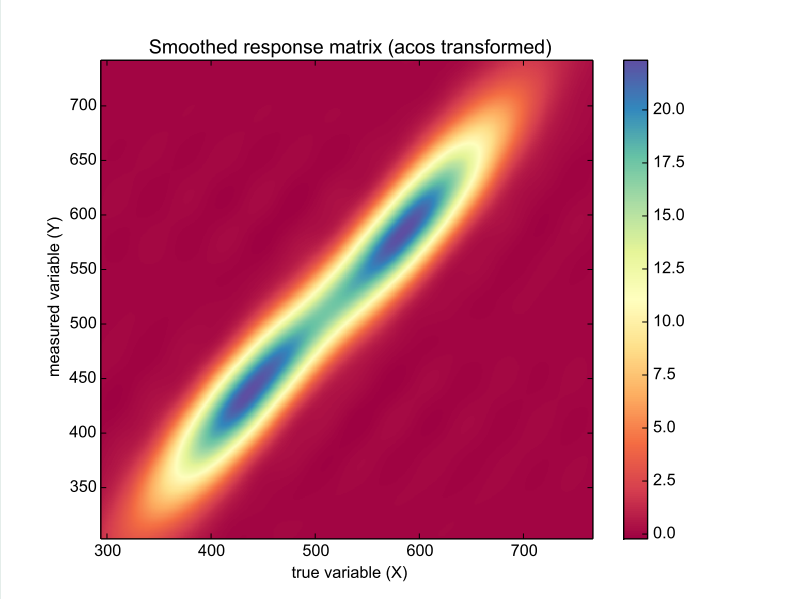
compression: almost all 1024×1024 moments are compatible with zero

The SVD is applied to the lower-left corner of the moment array;

$$\widetilde{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$$

with the diagonal matrix of singular values. The inverse singular values are used to construct the final moments, i.e. the elements of $\widetilde{f}$ from the elements of $\widetilde{g}$.

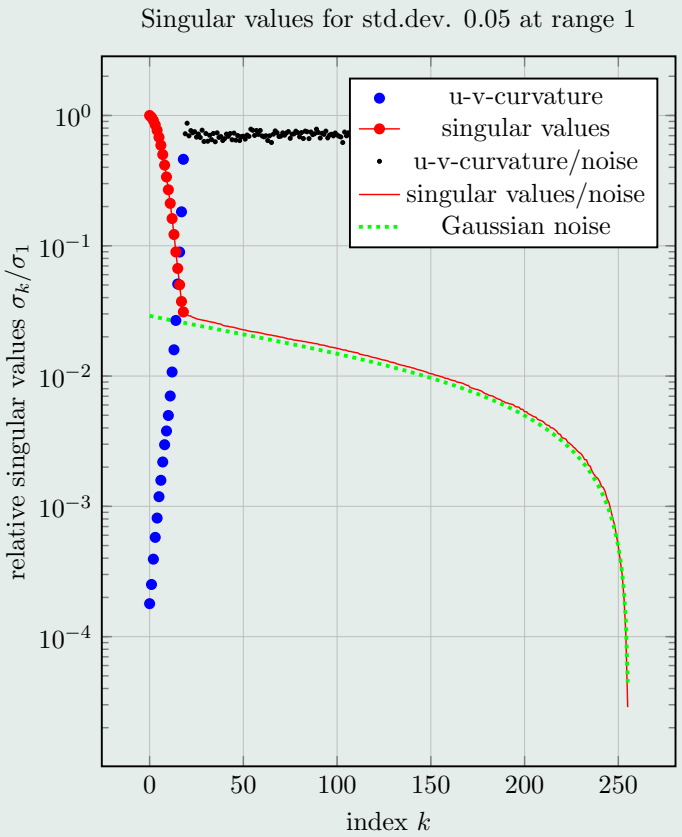




Only a fraction of the singular value is significantly different from zero; the corresponding colums vectors of $\boldsymbol{U}$ and $\boldsymbol{V}$, the left- and right-singular vectors are well-defined.

- Green dashed curve assumes Gaussian random value of σ_i with mean zero.
- All singular values σ_i have the same uncertainty!
- Singular vectors for insignificant σ_i are random vectors with $\approx$ same curvature.

$\rightsquigarrow$ **sharp index limit** on well-defined response!



- Use all the significant moments and singular values/vectors, without regularization;
- Measured moments c_k from $g(y)$ are multiplied by inverse singular values $\Longrightarrow$ reconstructed moments of $f(x)$;
- Log-linear determination of $\widehat{f}(x) = \sum_{k=0}^{p} \widehat{\gamma}_k T_k(x)$, from equality constraints for reconstructed moments.

Is the linear Fredholm integral equation the correct model?

- Is the problem described by a linear integral equation?
- How to take into account the migration *in and out* the measurement region?
- Even unfolding a 1D-distribution can require a more-dimensional measurement!
- How to discretize statistically correct the 2D-, 3D- ... distributions?

$\rightsquigarrow$

Use a realistic Monte Carlo input function $f_{\mathrm{MC}}(x)$ close to the expected result in order to avoid distorting effects from a potentially non-linearity and *in/out* migration, without introducing a significant bias towards the expectation.

Criticisms

- "Each year, the experimental collaborations working with LHC data publish dozens of papers using . . . unsatisfactory unfolding techniques."
- "Furthermore, . . . methods suffer from not dealing with two significant issues satisfactorily: (1) the choice of the regularization strength and (2) quantification of the uncertainty in the solution."
- "The delicate problem of choosing the regularization strength is handled in most LHC analyses using non-standard heuristics or, in the worst case scenario, by simply fixing a certain value "by hand". When quantifying the uncertainty of the unfolded spectrum, the analyses rarely attempt to take into account the uncertainty related to the choice of this regularization strength."
- ". . . ignores the Poisson nature of the observations and does not enforce the positivity of the solution."
- "Almost all unfolding methods in particle physics use directly histograms for the observed distribution and the response matrix . . . although histograms may cause severe problems."

Two opposition opinions

- A fundamental problem in unfolding is the discrepancy, that, on one hand,
  - unfolding itself is a complex statistical and mathematical procedure, and on the other hand,
  - the procecure should be used, more or less blindly, by physicists, which are experts in particle physics, but have only limited knowledge in the theory of unfolding.
- This discrepancy can be resolved by the development of unfolding procedures, which require *a minimum of steering parameters* for unfolding.

versus

- Unfolding is a complex data analysis task that involves several assumptions and approximations. It is crucial to understand the ingredients that go into an unfolding procedure.
  Unfolding algorithms should never be used as black boxes!

THE END

Complexity and simplicity

"For every complex problem there is an answer
that is clear, simple, and wrong."

H. L. MENCKEN

"Everything should be made as simple as possible, but not simpler."

ALBERT EINSTEIN

"Using a simple tool to solve a complex problem does not result in a simple solution."

JOHN DOUGLAS PORTER

"Seek simplicity and distrust it."

ALFRED NORTH WHITEHEAD

"You can get it wrong and still you think it's all right"

JOHN LENNON and PAUL MCCARTNEY, from

We can work it out

The credibility of a quotation is increased substantially if it can be ascribed to a widely-recognized genius such as Albert Einstein.

Literature

R.C.Aster, B.Borchers, C.H.Thurber,
Parameter Estimation and Inverse Problems, Elsevier (2013)

Per Christian Hansen.
Rank-deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion,
Volume 160 of *SIAM monographs on mathematical modeling and computation*. Society for Industrial and Applied Mathematics, Philadelphia, 1997.

Discrete Inverse Problems – Insight and Algorithms,
Volume 160 of *Fundamentals of Algorithms*. Society for Industrial and Applied Mathematics, Philadelphia, 2010.

Jari Kaipio and Erkki Somersalo.
Statistical and Computational Inverse Problems,
Volume 160 of *Applied Mathematical Science*. Springer, 2004.

Statistical inverse problems: Discretization, model reduction and inverse crimes,
Journal of Computational and Applied Mathematics, 198:493 – 504, 2007.

Curtis R. Vogel.
Computational Methods for Inverse Problems,
Volume 160 of *SIAM Frontiers in Applied mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, 2002.

Andreas Rieder.
Keine Probleme mit Inversen Problemen.
Vieweg, 2003.

Ake Bjork.
Numerical Methods for Least Squares Problems.
Society for Industrial and Applied Mathematics, Philadelphia, 1996.

homepage www.desy.de/~blobel

Unfolding codes

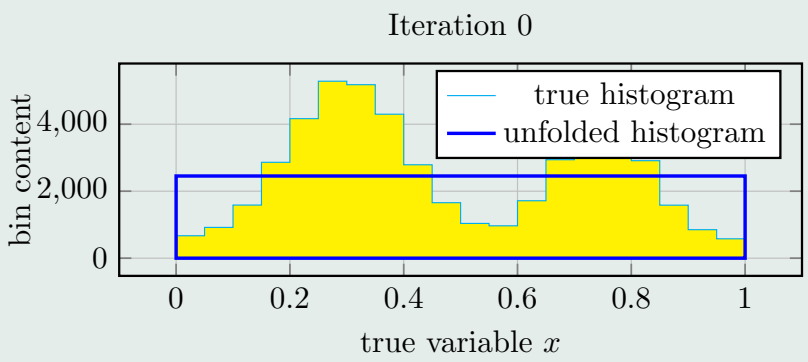
GURU (Fortran, SVD, by Andreas Hoecker and Vato Kartvelishveli)
TUNFOLD, *an algorithm for correcting migration effects in high energy physics*, by Stefan Schmitt, DESY
Both in RooUnfold (ROOT Unfolding Framework, by Tim Adye et al.)

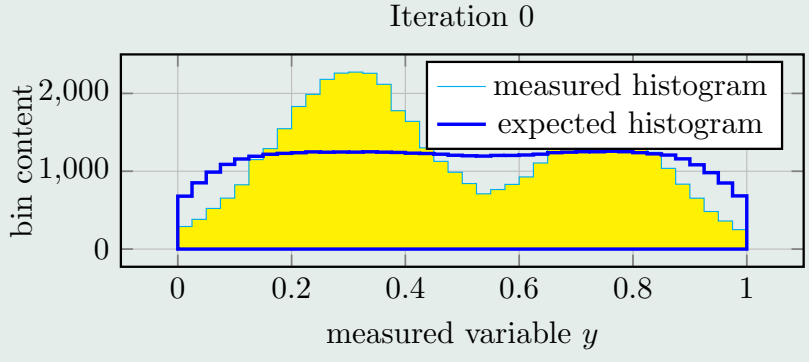
$\mathcal{RUN}$ (Fortran), converted to C++ by Natalie Milke (Uni Dortmund)

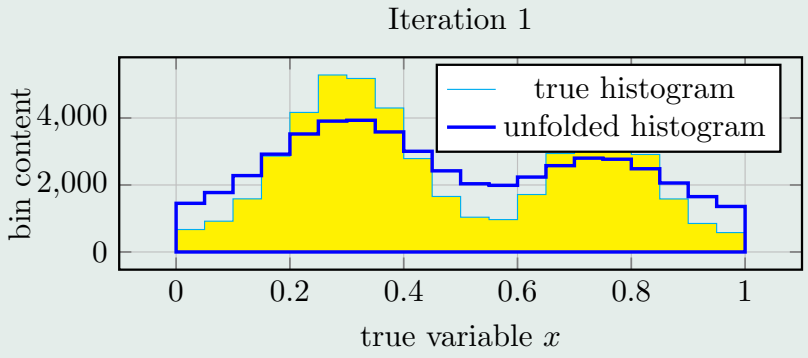
Method	$\mathcal{RUN}$	GURU	Tunfold	Iterative
Input: matrix		✔	✔	✔
Input: n-tuple	✔			
Orthogonalization	✔	✔		
Input errors	✔	✔	✔	
Least squares		✔	✔	?
MaxLik (Poisson)	✔			
Regularization	✔	✔	✔	implicit
iterative				✔
automatic binning	✔			
Cov.mat. by prop.	✔	✔	✔	
MC re-weighting	✔			

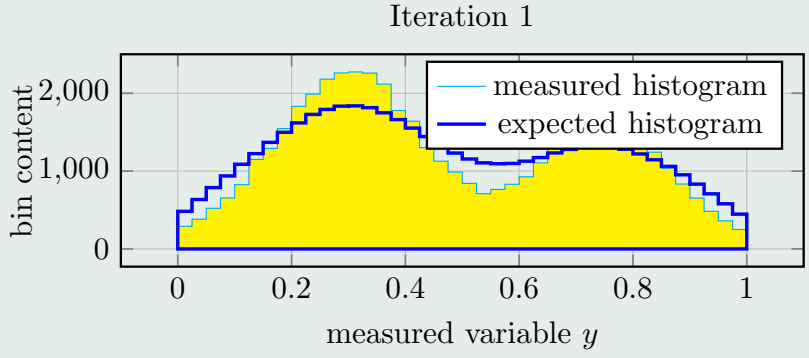
RL: uniform start distribution I

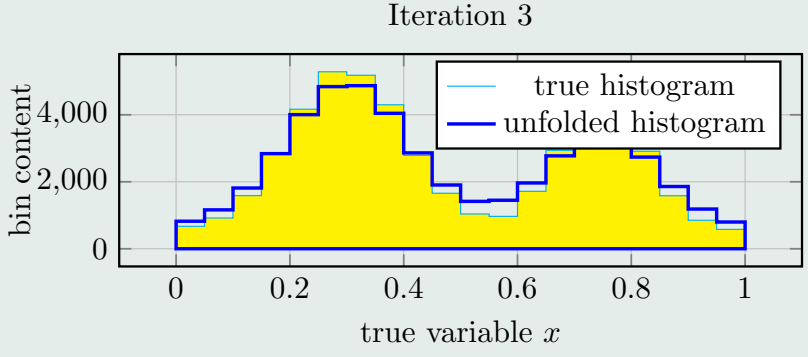
0 . . . 3 iterations

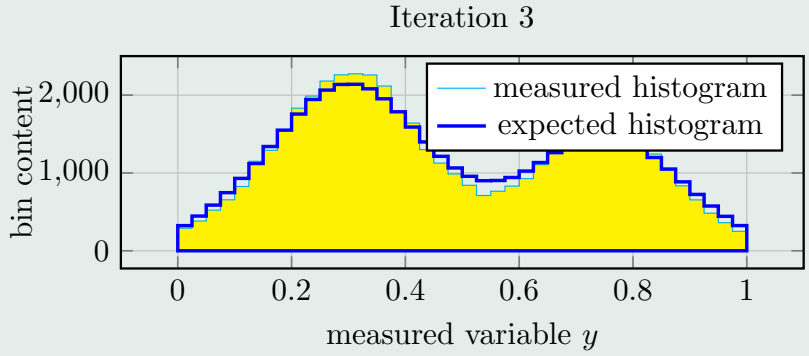






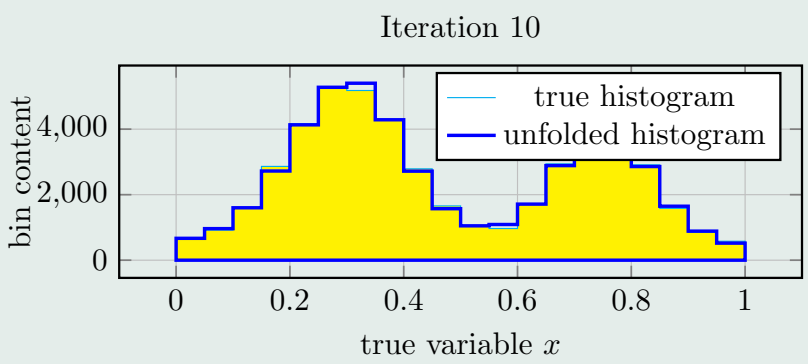


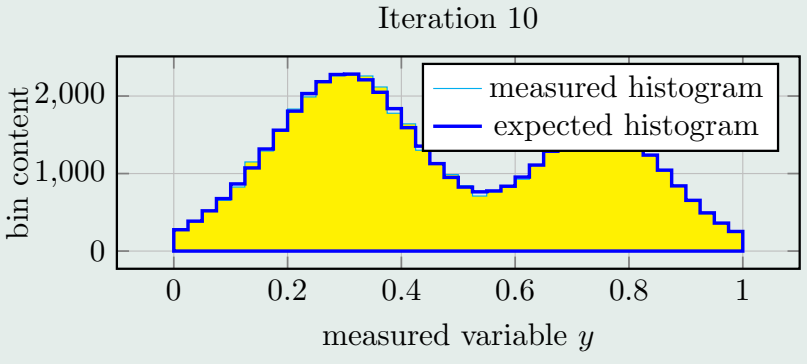


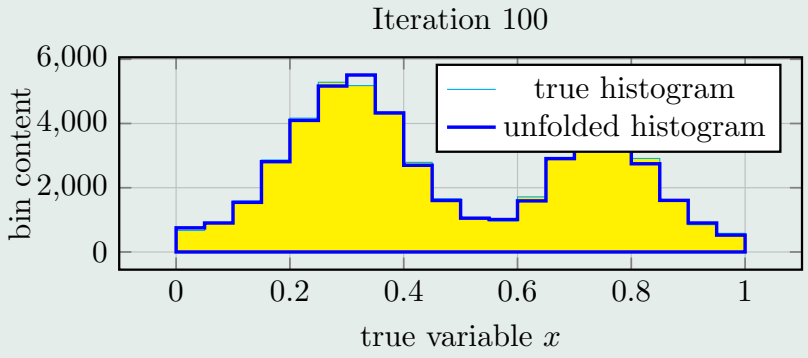


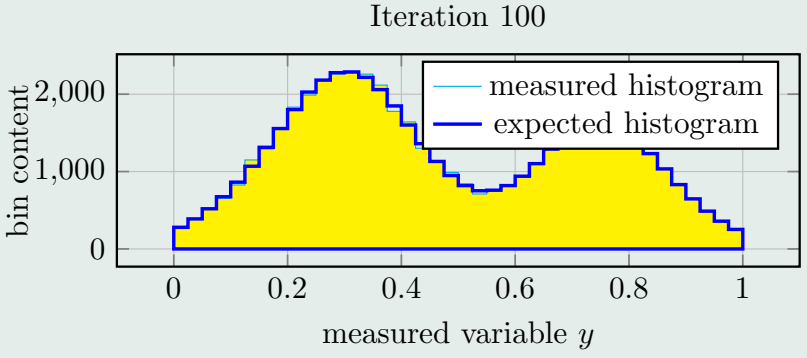
RL: uniform start distribution II

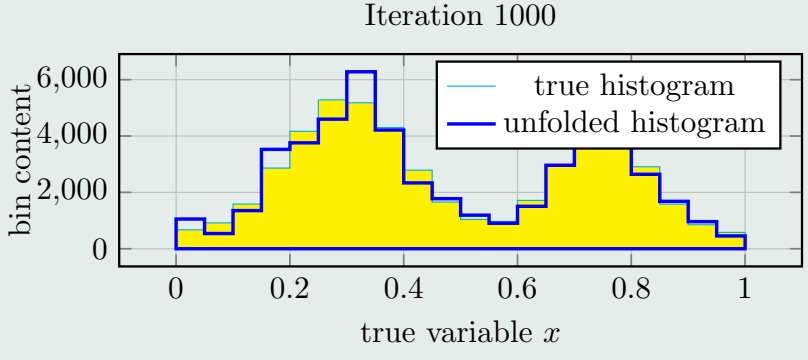
10 . . . 100 iterations

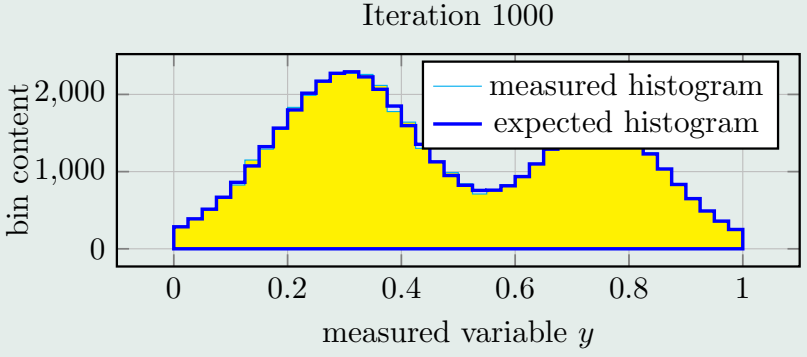






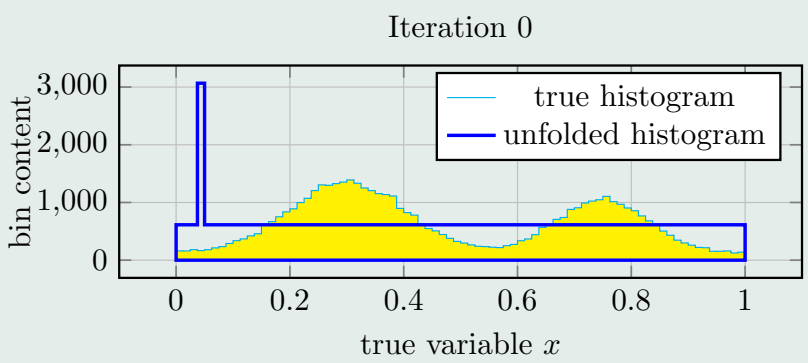


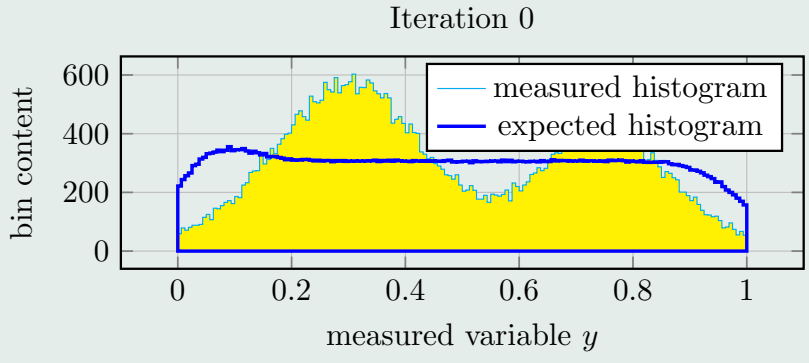


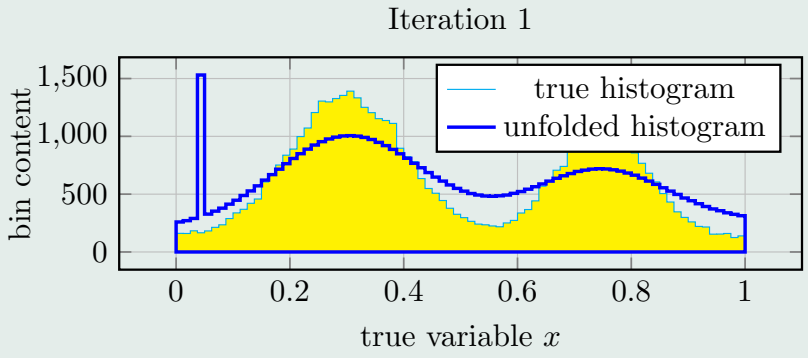


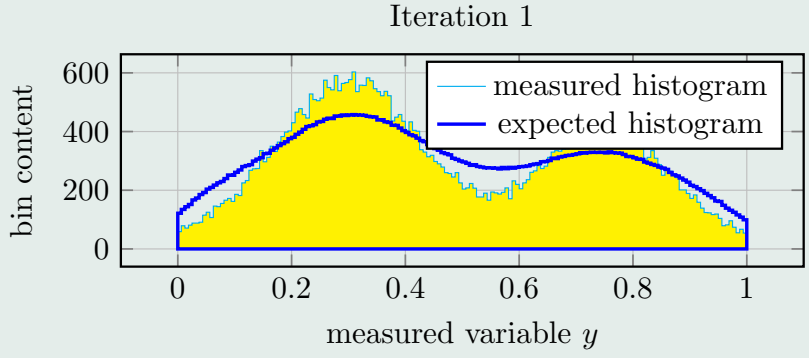
RL: uniform dist + spike I

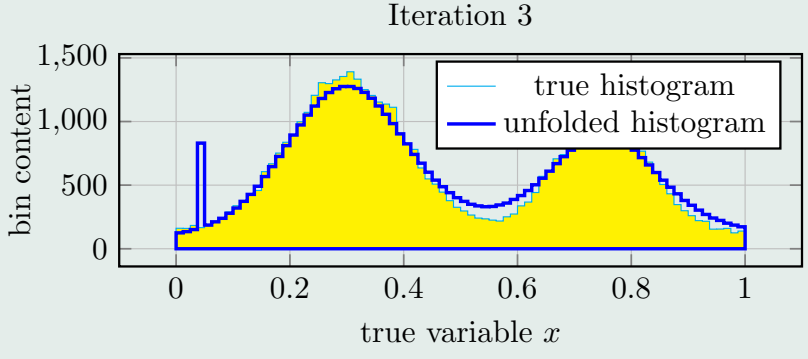
0 . . . 3 iterations

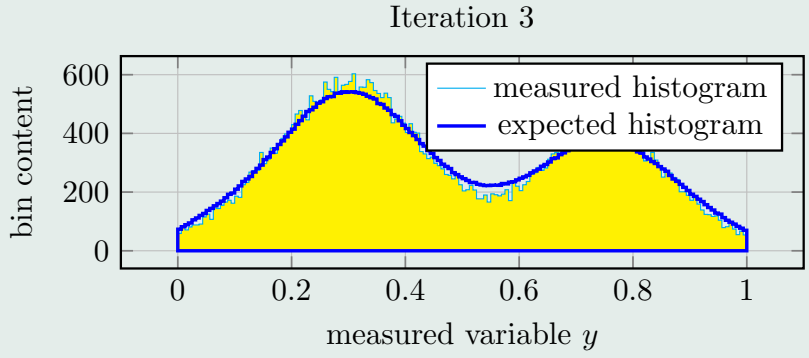


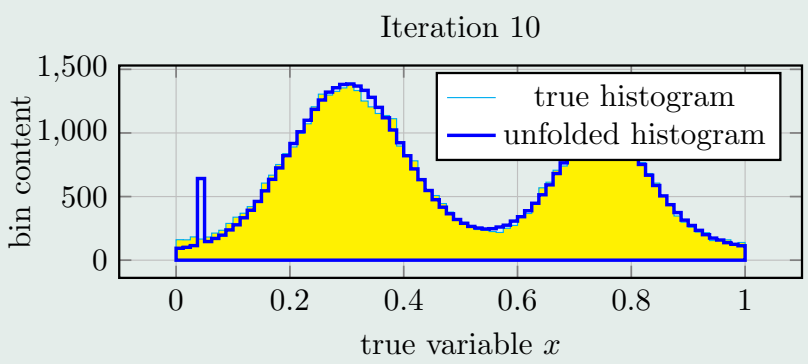


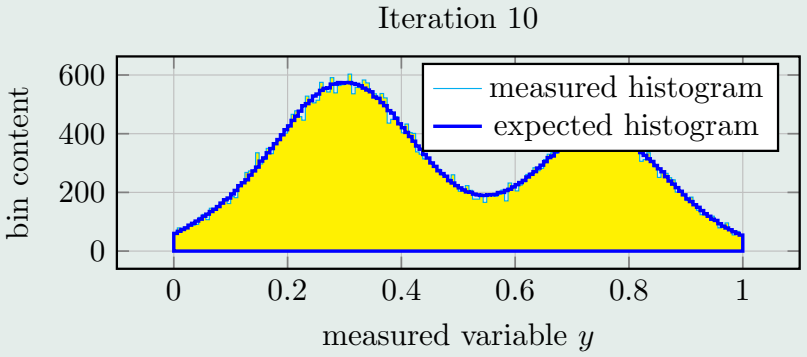


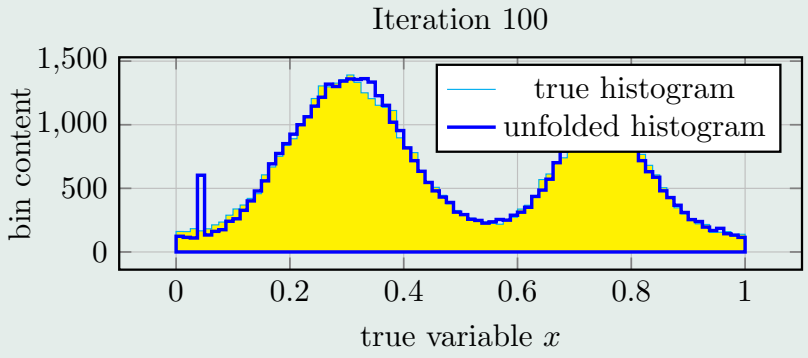


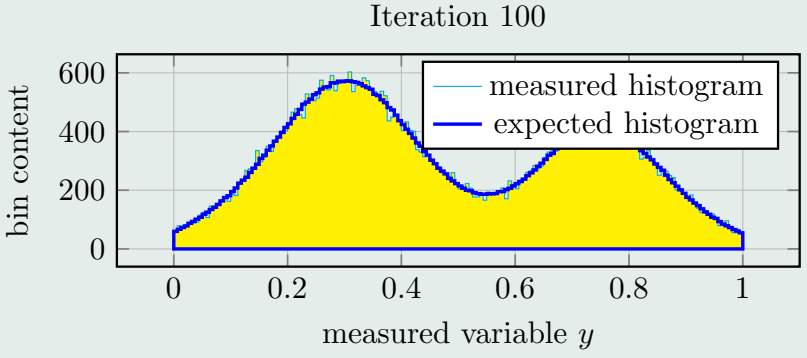


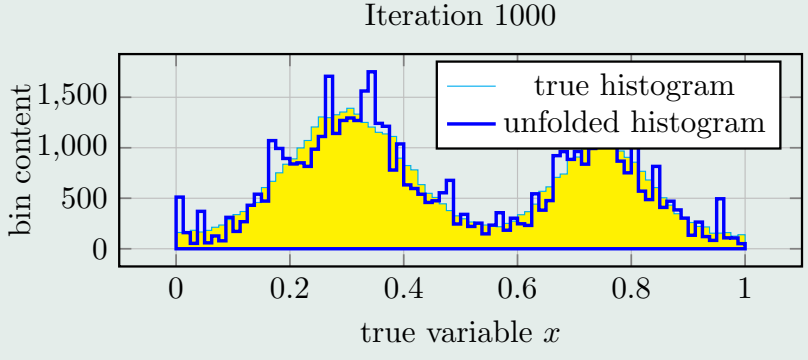


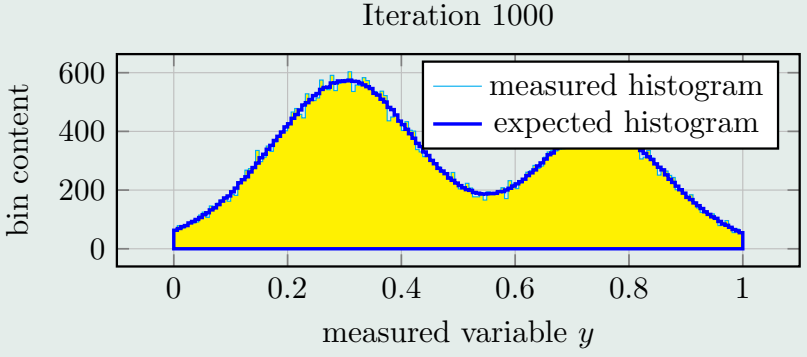

Iteration 10
true histogram
unfolded histogram
bin content
1,500
1,000
500
0
0
0.2
0.4
0.6
0.8
1
true variable x


Iteration 10
measured histogram
expected histogram
bin content
600
400
200
0
0
0.2
0.4
0.6
0.8
1
measured variable y


Iteration 100
true histogram
unfolded histogram
bin content
1,500
1,000
500
0
0
0.2
0.4
0.6
0.8
1
true variable x


Iteration 100
measured histogram
expected histogram
bin content
600
400
200
0
0
0.2
0.4
0.6
0.8
1
measured variable y


Iteration 1000
true histogram
unfolded histogram
bin content
1,500
1,000
500
0
0
0.2
0.4
0.6
0.8
1
true variable x

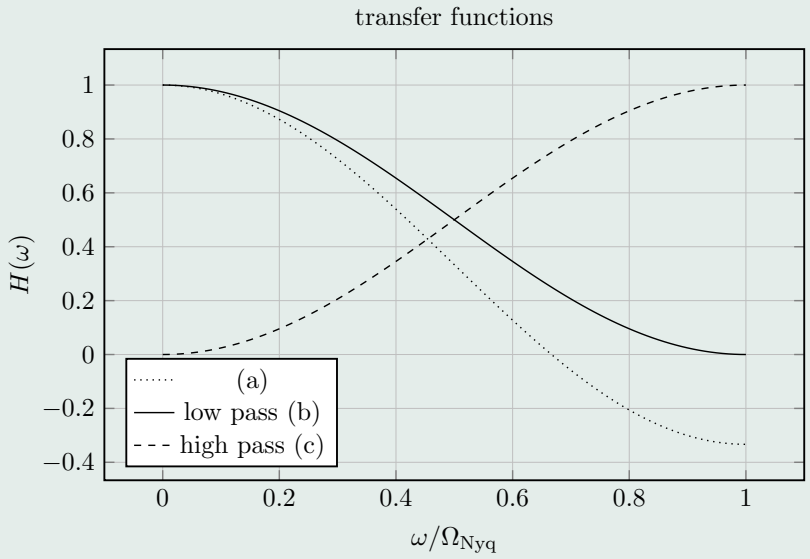

Iteration 1000
measured histogram
expected histogram
bin content
600
400
200
0
0
0.2
0.4
0.6
0.8
1
measured variable y

Smoothing

$$\text{smooth } f \Longrightarrow y$$

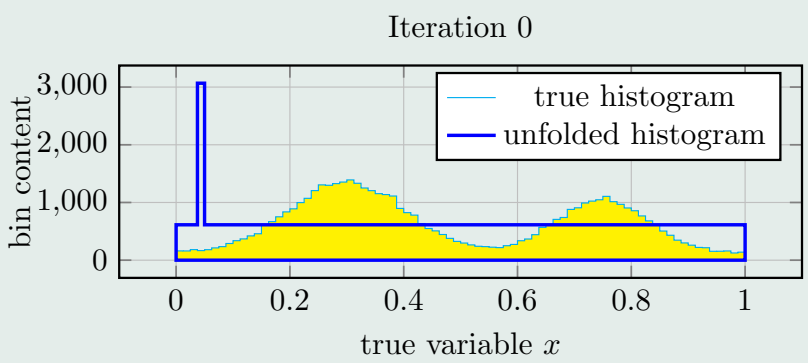
	algorithm		transfer function
(a)	$y_k = \frac{1}{3}(f_{k-1} + f_k + f_{k+1})$		$H(\omega) = \frac{1}{3}\left(1 + 2\cos\pi\frac{\omega}{\Omega_{\text{Nyq}}}\right)$
(b)	$y_k = \frac{1}{4}(f_{k-1} + 2f_k + f_{k+1})$	low-pass filter	$H(\omega) = \cos^2\frac{\pi}{2}\frac{\omega}{\Omega_{\text{Nyq}}}$

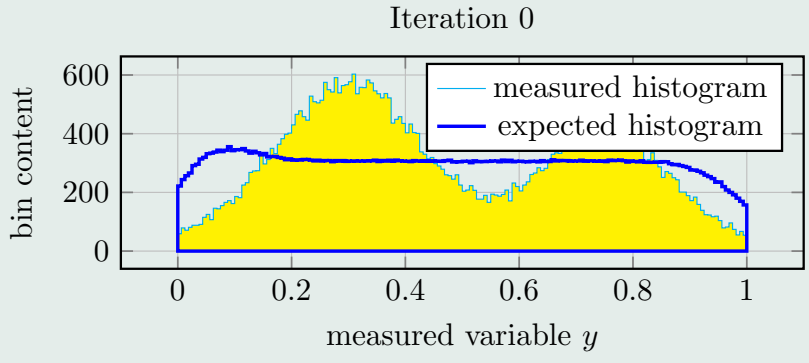
The Nyquist frequency is 1/2 of the sampling rate of a discrete signal processing.

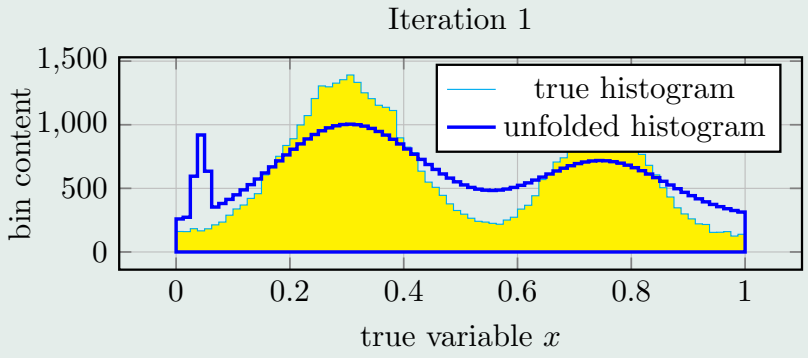


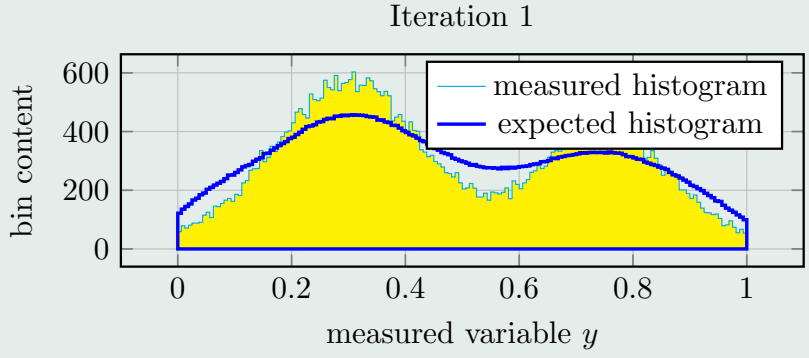
Non-consistent algorithm (a): non-optimal smoothing, zero transfer for $\omega = 2/3\Omega_{\text{Nyq}}$ and negative (!) for higher frequencies.

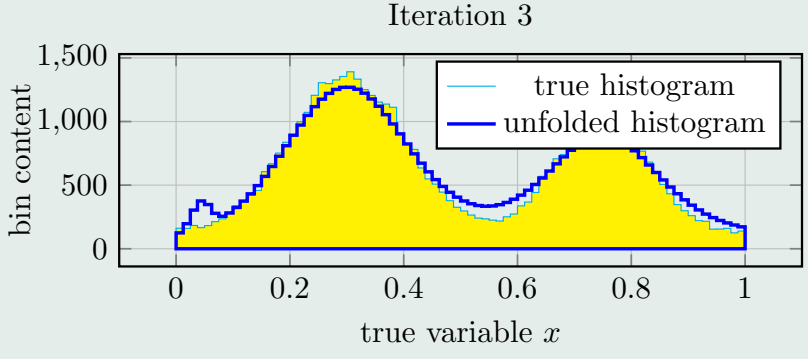
Correct algorithm (b): damping of higher frequencies without zero-transfer and without sign-change. Transfer 1/2 for $\omega = 1/2\Omega_{\text{Nyq}}$.

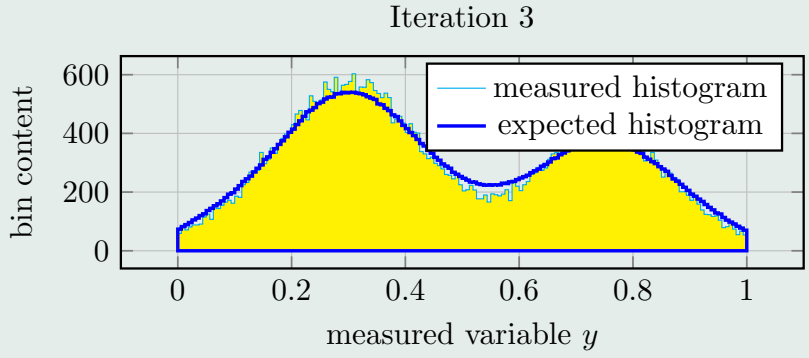

Iteration 0
bin content
3,000
2,000
1,000
0
true histogram
unfolded histogram
0
0.2
0.4
0.6
0.8
1
true variable x

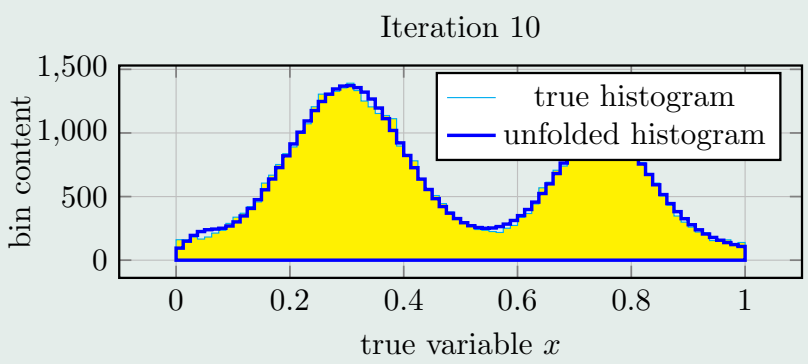

Iteration 0
bin content
600
400
200
0
measured histogram
expected histogram
0
0.2
0.4
0.6
0.8
1
measured variable y

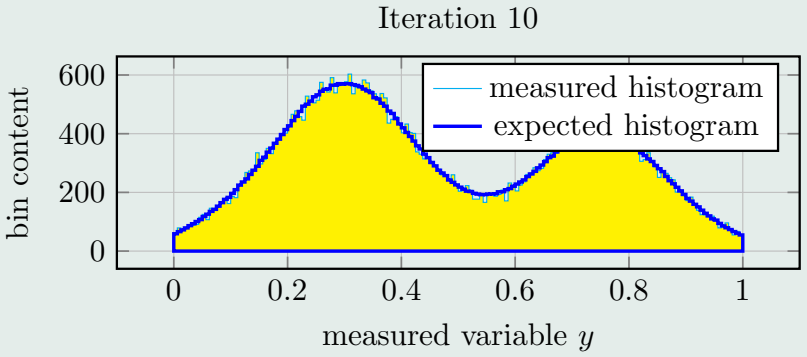

Iteration 1
bin content
1,500
1,000
500
0
true histogram
unfolded histogram
0
0.2
0.4
0.6
0.8
1
true variable x

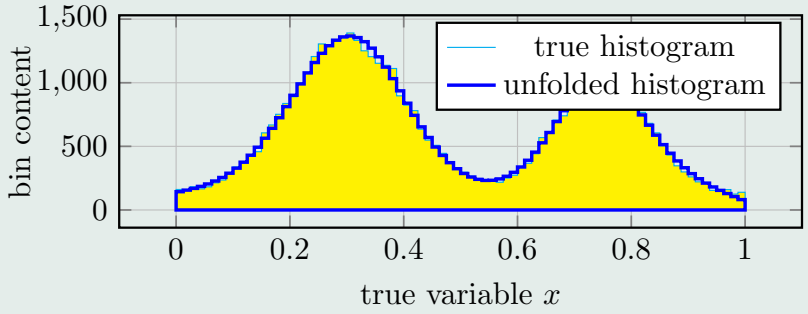

Iteration 1
bin content
600
400
200
0
measured histogram
expected histogram
0
0.2
0.4
0.6
0.8
1
measured variable y

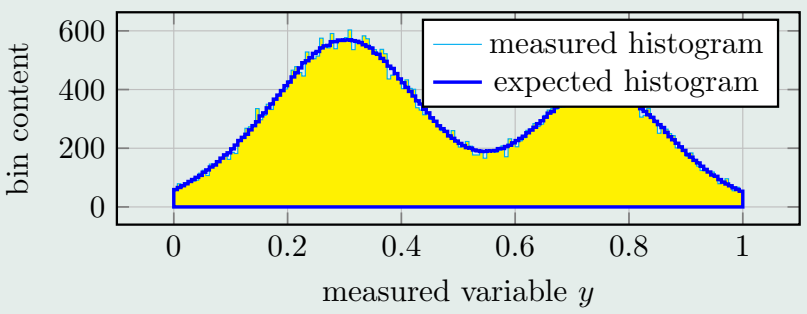

Iteration 3
bin content
1,500
1,000
500
0
true histogram
unfolded histogram
0
0.2
0.4
0.6
0.8
1
true variable x

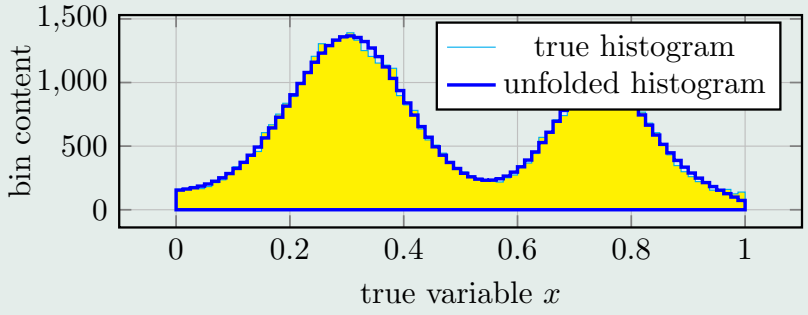

Iteration 3
bin content
600
400
200
0
measured histogram
expected histogram
0
0.2
0.4
0.6
0.8
1
measured variable y

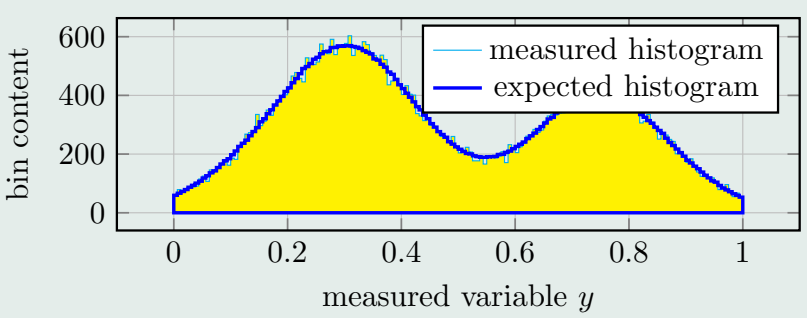

Iteration 10
true histogram
unfolded histogram
bin content
1,500
1,000
500
0
0
0.2
0.4
0.6
0.8
1
true variable x


Iteration 10
measured histogram
expected histogram
bin content
600
400
200
0
0
0.2
0.4
0.6
0.8
1
measured variable y


Iteration 100
true histogram
unfolded histogram
bin content
1,500
1,000
500
0
0
0.2
0.4
0.6
0.8
1
true variable x

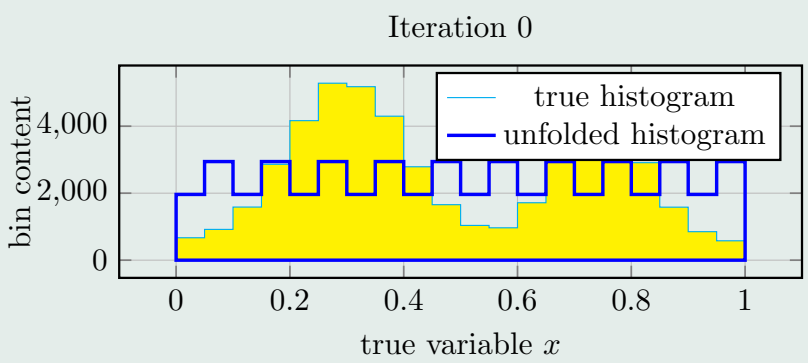

Iteration 100
measured histogram
expected histogram
bin content
600
400
200
0
0
0.2
0.4
0.6
0.8
1
measured variable y

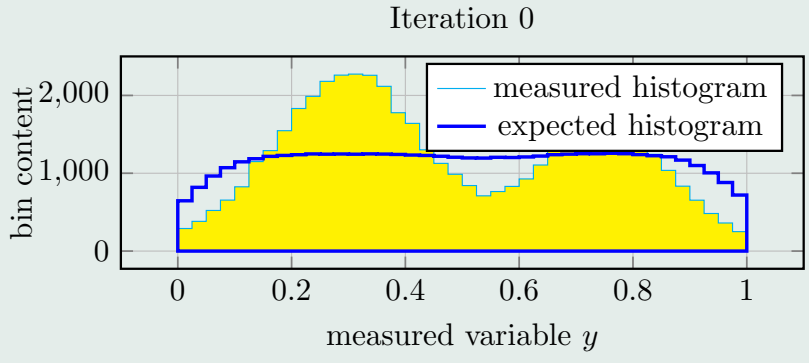

Iteration 1000
true histogram
unfolded histogram
bin content
1,500
1,000
500
0
0
0.2
0.4
0.6
0.8
1
true variable x

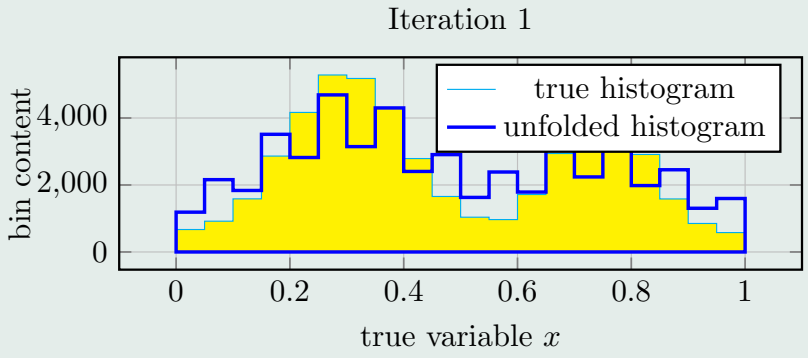

Iteration 1000
measured histogram
expected histogram
bin content
600
400
200
0
0
0.2
0.4
0.6
0.8
1
measured variable y

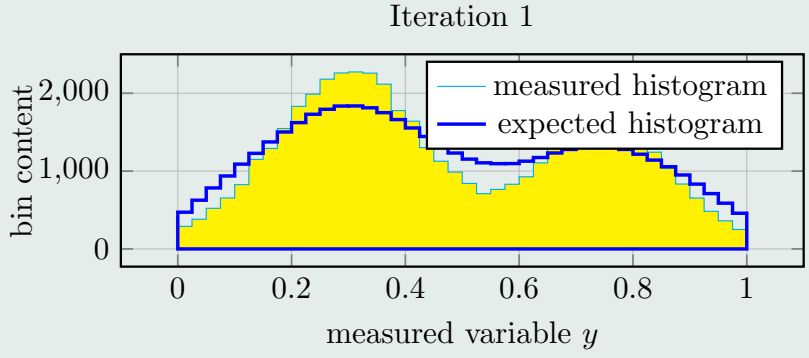
RL: hi-freq start distribution I

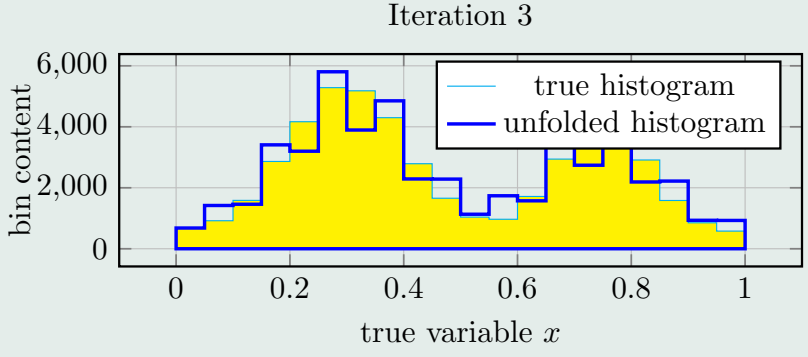
0 . . . 3 iterations

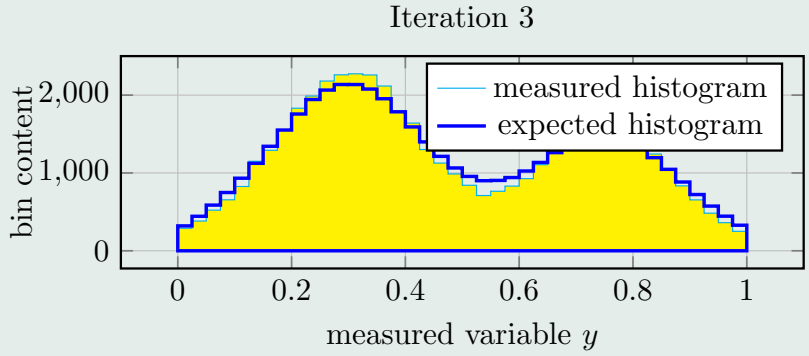






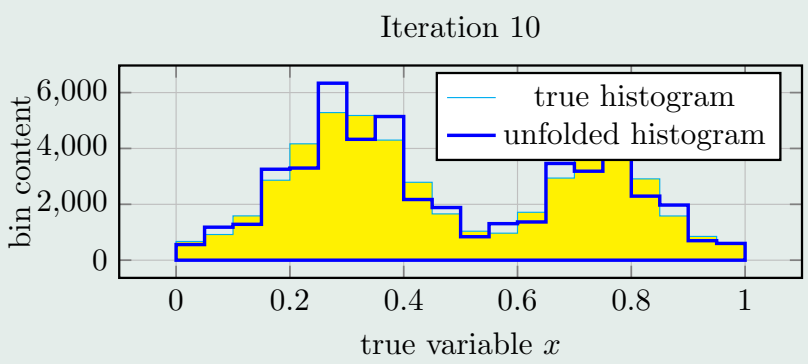


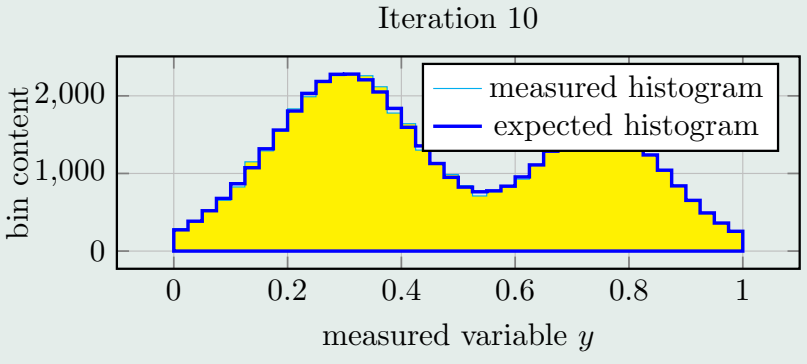


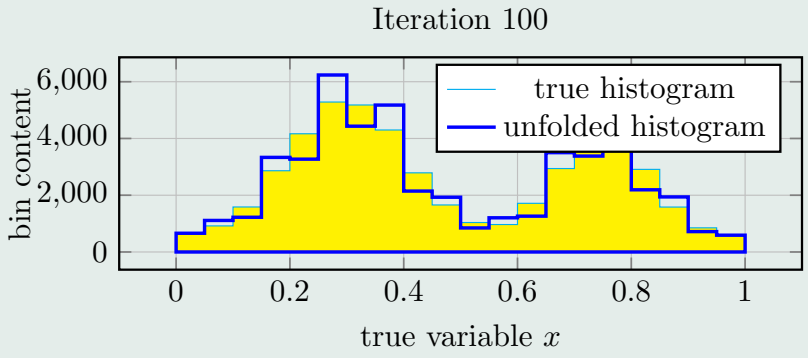


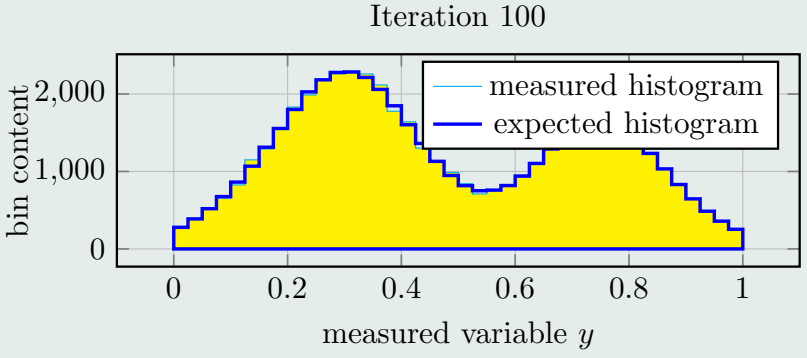
RL: hi-freq start distribution II

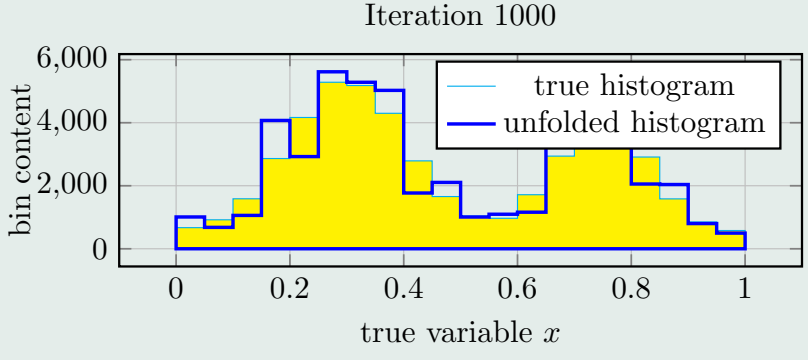
10 . . . 100 iterations

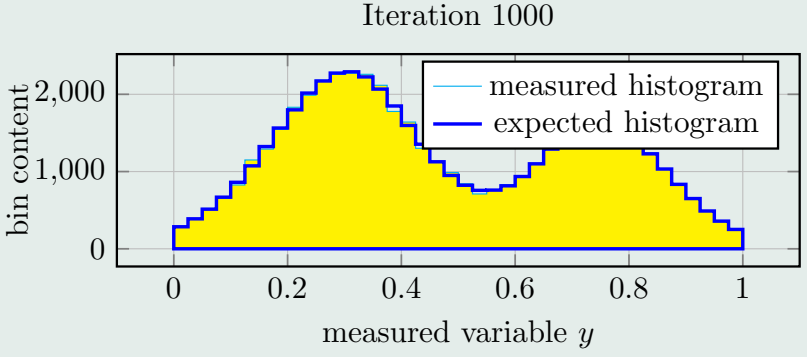












Frequentist and Bayesian probability

Two fundamental approaches to statistical analysis: Frequentism and Bayesianism

Method	Frequentist	Bayesian
Meaning of probability	Frequentist	Degree of belief
Probability for parameters?	No, no, no	Yes

Frequentists: $P(A)$ = number of times A occurs, divided by nr of trials (objective; the probability of Quantum mechanics; can only be applied to repeatable phenomena – most scientific work).

Bayes theorem: Probability of both A and B being true:

$$P(A \text{ and } B) = P(A|B)P(B) = P(B|A)P(A)$$

$$\text{which implies} \qquad P(B|A) = \frac{P(A|B)P(B)}{P(A)} \qquad \text{Bayes theorem}$$

Bayes theorem itself is not controversial between frequentists and Bayesians.

Bayesian use of Bayes theorem: A = observed data, B = theoretical parameter

$$\underbrace{P(\text{parameter}|\text{data})}_{\textit{posterior density}} \propto \underbrace{P(\text{data}|\text{parameter})}_{\textit{likelihood function}} \times \underbrace{P(\text{parameter})}_{\textit{prior density}}$$

"The resolution in each bin is checked using a Monte Carlo simulation. Two variables are calculated for this purpose, the purity $P = N_{rec,gen}/N_{rec}$ and the stability $S = N_{rec,gen}/N_{gen}$, where N_{rec} (N_{gen}) is the total number of reconstructed (generated) Monte Carlo events in the bin and $N_{rec,gen}$ is the number of events which are both generated and reconstructed in the same bin. The purity and stability are calculated for both the electron and the Σ methods. For the cross section measurement the method with the higher purity is used. *. . . The purity and stability typically exceed* 50%. *If either the purity or the stability is below* 25% *in a bin for the chosen reconstruction method, the bin is combined with an adjacent bin.*

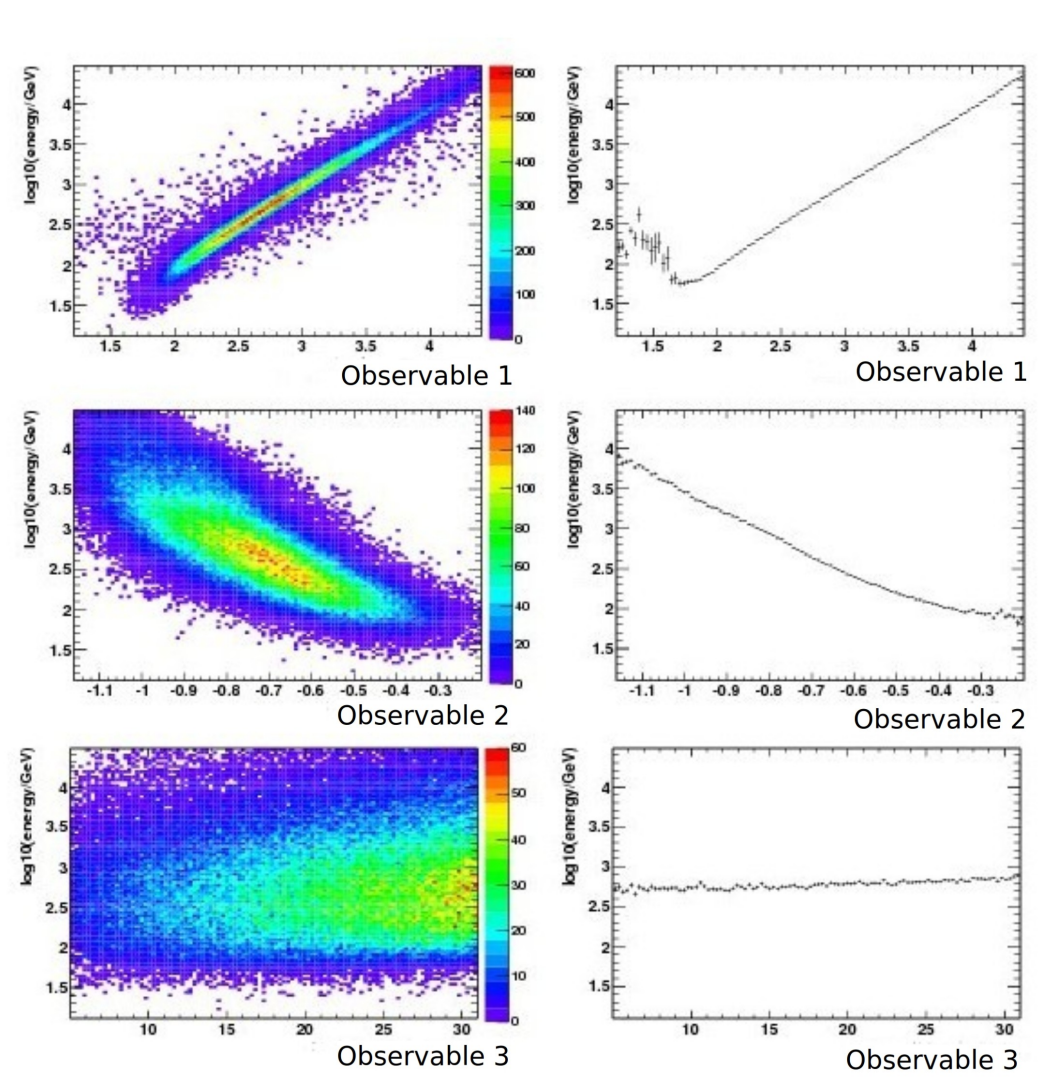
Events are only taken into account from bins which pass the stability and purity criteria and are covered by the chosen method.

The radiative and bin centre corrections can be determined using the Monte Carlo simulation.

The correction for the detector acceptance using Monte Carlo modelling requires the cross section model used in the simulation to be sufficiently close to the data, such that migrations between the bins are well reproduced.

In practice, this is achieved using an iterative MC event reweighting procedure which converges after one iteration for the measurement region."

Measurement: flux of cosmic gamma-rays, up to tens of TeV, by system of two Cherenkov telescopes. **Vertical** scale = $\log_{10}$ of energy/GeV **horizontal** scale: observables



Observable 1:

energy estimate – good correlation with energy;

Observable 2:

light-distribution parameter, has some correlation with energy;

Observable 3:

angle, no direct correlation with energy, but energy estimates differ for different angles.

3-dim distribution of observables used to unfold (reconstruct) the 1-dim energy flux.

Contents