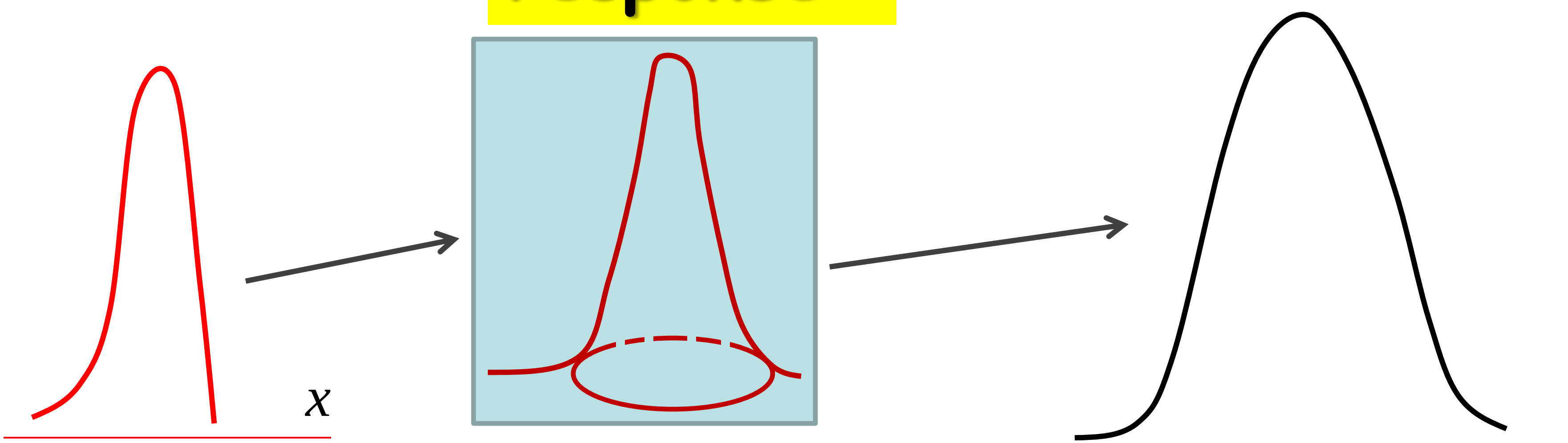
Unfolding Methods

Alberto Rotondi
Università di Pavia

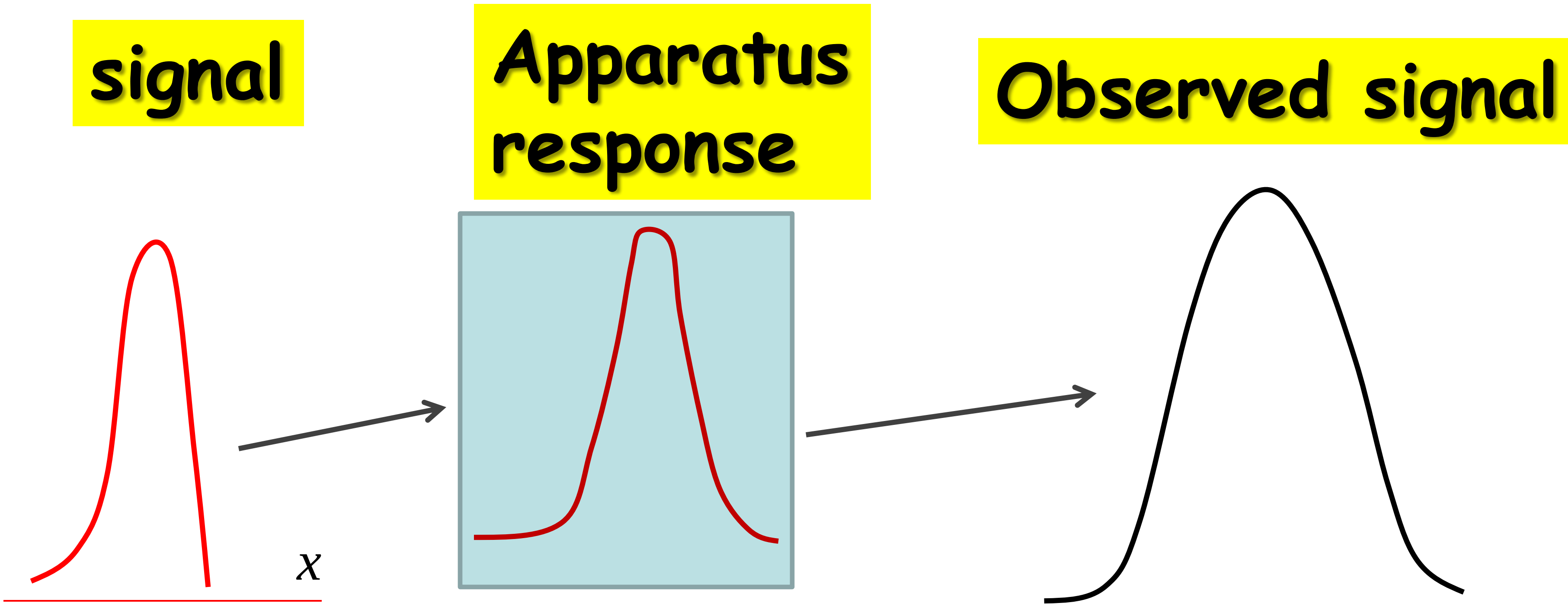
Folding is a common process in physics

signal | **Apparatus response** | **Observed signal**



$f(x)$ $\qquad$ $\delta(x,y)$ $\qquad$ $g(y) = \int f(x)\,\delta(x,y)\,\mathrm{d}x$

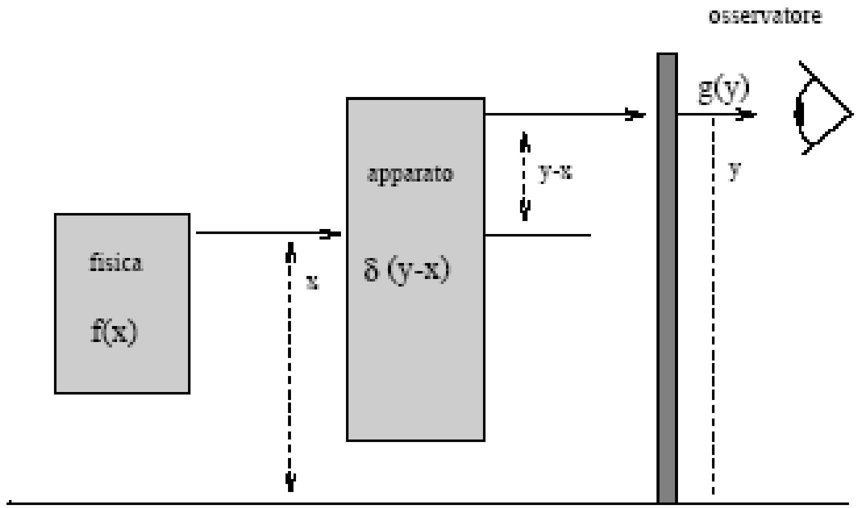
Convolution is a linear folding

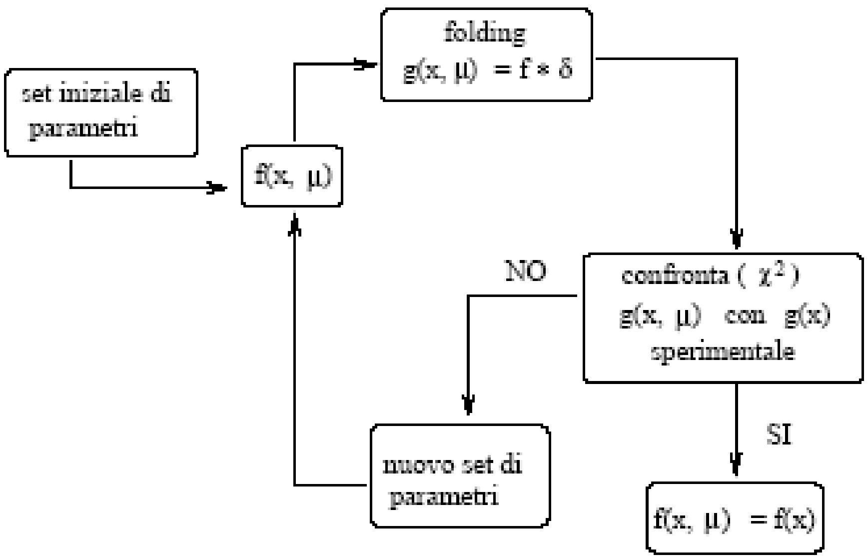


$z = y - x$ $\qquad$ $y = z + x$

$f(x)$ $\qquad$ $\delta(y - x)$ $\qquad$ $g(y) = \int f(x)\,\delta(y - x)\,\mathrm{d}x$

$$g(y) = \int f(x)\,\delta(y - x)\,\mathrm{d}x \ ,$$





Folding theorem

$$Z = f(X_1, X_2)\ .$$

$Z_1 \equiv Z$ e $Z_2 = X_2$ auxiliary variable

$$Z_1 = f(X_1, X_2)\ , \qquad Z_2 = X_2\ .$$

The jacobian is

$$|J| = \begin{vmatrix} \frac{\partial f_1^{-1}}{\partial z_1} & \frac{\partial f_1^{-1}}{\partial z_2} \\ 0 & 1 \end{vmatrix} = \left|\frac{\partial f_1^{-1}}{\partial z_1}\right|\ ,$$

From the general theorem one obtains

$$p_{\mathbf{Z}}(z_1, z_2) = p_{\mathbf{X}}(x_1, x_2) \left|\frac{\partial f_1^{-1}}{\partial z_1}\right|\ .$$

by integrating on the auxiliary variable :

$$p_{Z_1}(z_1) = \int p_{\mathbf{Z}}(z_1, z_2)\, \mathrm{d}z_2\ .$$

hence

$$\begin{aligned} p_Z(z) &= \int p_{\mathbf{X}}(x_1, x_2) \left|\frac{\partial f_1^{-1}}{\partial z}\right| \mathrm{d}x_2 \\ &= \int p_{\mathbf{X}}\left(f_1^{-1}(z, x_2), x_2\right) \left|\frac{\partial f_1^{-1}}{\partial z}\right| \mathrm{d}x_2\ , \end{aligned}$$

which is the probability density

For independent variables:

Convolution theorem

$$p_Z(z) = \int p_{X_1}\left(f_1^{-1}(z, x_2)\right)\, p_{X_2}(x_2) \frac{\partial f_1^{-1}}{\partial z}\, \mathrm{d}x_2 \ .$$

when Z is given by the sum

$$Z = X_1 + X_2 \ ,$$

we have

$$X_1 = f_1^{-1}(Z, X_2) = Z - X_2 \ , \qquad \frac{\partial f_1^{-1}}{\partial z} = 1 \ ,$$

and we obtain

$$p_Z(z) = \int_{-\infty}^{+\infty} p_{\mathbf{X}}(z - x_2, x_2)\, \mathrm{d}x_2 \ .$$

When X_1 and X_2 are independent, we obtain the convolution integral

$$p_Z(z) = \int_{-\infty}^{+\infty} p_{X_1}(z - x_2)\, p_{X_2}(x_2)\, \mathrm{d}x_2 \ ,$$

In physics

(δ instrument function, f signal) :

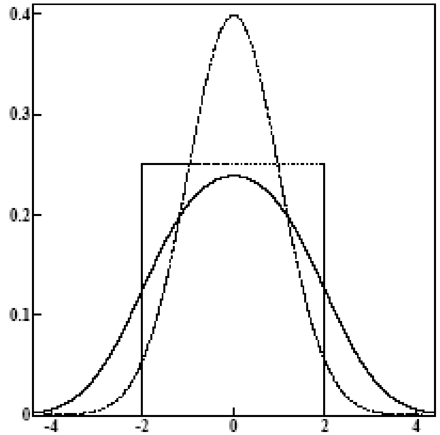
$$g(y) = \int_{-\infty}^{+\infty} f(y - x)\, \delta(x)\, \mathrm{d}x \ ,$$

Uniform*Gaussian

When $Z = X + Y$ **where**
$X \sim N(\mu, \sigma^2)$ e $Y \sim U(a, b)$.
one has immediately

$$p_Z(z) = \frac{1}{b-a} \int_a^b \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(z-y-\mu)^2}{2\sigma^2}\right] \mathrm{d}y$$
$$= \frac{1}{b-a} \int_a^b \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{[y-(z-\mu)]^2}{2\sigma^2}\right] \mathrm{d}y \ .$$

$$p_Z(z) = \frac{1}{b-a}\left[\Phi\left(\frac{b-(z-\mu)}{\sigma}\right) - \Phi\left(\frac{a-(z-\mu)}{\sigma}\right)\right]$$



1D Unfolding

In the reconstruction of an histogram,

- the true histogram (image) where the bin contents are the expected values

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_N) \ , \quad \mu_j = \mu_{\rm tot} p_j = \mu_{\rm tot} \int_{\rm bin\ i} f_t(y)\, {\rm d}y$$

2D Unfolding

The observed $N_{ij}(\exp)$ events have to be compared with the expected values $N_{ij}(\mathrm{th})$ predicted by a model.

$$N_{ij}(\mathrm{th}) = N P_{ij}(\mathrm{obs}) = N \sum_{i'j'} P_{i'j'}(\mathrm{true})\, P_v(\mathrm{obs}_{ij}|\mathrm{true}_{i'j'}) \ , \quad (94)$$

that is, the number of events observed in the ijth-cell is due to the presence into the $i'j'$th-cell, times the probability P_v that the PSF shifts the point from the $i'j'$ to the ij-cell. One has to sum on all the cells near the ij–one.

In eq. (94) the normalization is understood. In practice, from (93, 94), one has

$$P(\mathrm{true}_{ij}|\mathrm{obs}_{i'j'}) = \frac{P_{i'j'}(\mathrm{true})\, P_v(\mathrm{obs}_{ij}|\mathrm{true}_{i'j'})}{\sum_{ij}\left[\sum_{i',j'} P_{i'j'}(\mathrm{true})\, P_v(\mathrm{obs}_{ij}|\mathrm{true}_{i'j'})\right]} \quad (95)$$

In the case of a two dimensional Gaussian point spread function PSF:

$$P_v(\mathrm{obs}_{ij}|\mathrm{true}_{i'j'}) = \frac{1}{2\pi\, \sigma_x \sigma_y} \exp\left[-\frac{(x_{ij} - x_{i'j'})^2}{2\,\sigma_x^2} - \frac{(y_{ij} - y_{i'j'})^2}{2\,\sigma_y^2}\right] , \quad (96)$$

Fourier Techniques

$$f(x) = \int F(t)\, \mathrm{e}^{2\pi i\, xt}\, \mathrm{d}t$$

Convolution:

$$f(x) = \int g(y)\delta(x-y)\,\mathrm{d}y$$

$$\int F(t)\,\mathrm{e}^{2\pi i xt}\,\mathrm{d}t = \int G(t)\,\mathrm{e}^{2\pi i\, yt}\,\Delta(t)\,\mathrm{e}^{2\pi i\,(x-y)t}\,\mathrm{d}t$$

$$\int F(t)\,\mathrm{e}^{2\pi i\, xt}\,\mathrm{d}t = \int G(t)\,\Delta(t)\,\mathrm{e}^{2\pi i\, xt}\,\mathrm{d}t \rightarrow F(t) = G(t)\,\Delta(t)$$

Image Deconvolution

$$D(\boldsymbol{x}) = \int \mathrm{d}\boldsymbol{y} I(\boldsymbol{y})\, \delta(|\boldsymbol{x} - \boldsymbol{y}|)$$

In the absence of noise

$$I = F^{-1}\left[\frac{F(D)}{F(\delta)}\right]$$

where F is the Fourier transform.

For a real image $I(n_1, n_2)$ the Fourier transform is:

$$F(k_1, k_2) = \sum_{n_2=0}^{N_2-1} \sum_{n_1=0}^{N_1-1} \mathrm{e}^{2\pi i k_2 n_2/N_2}\, \mathrm{e}^{2\pi i k_1 n_1/N_1}\; I(n_1, n_2)$$

$$F(k_1, k_2) = FFT_2[FFT_1[I(n_1, n_2)]]$$

For the routines see for example *Numerical Recipes*

The problem with fluctuations

original

Gaussian smearing

Poisson statistics

Fourier (un)restored

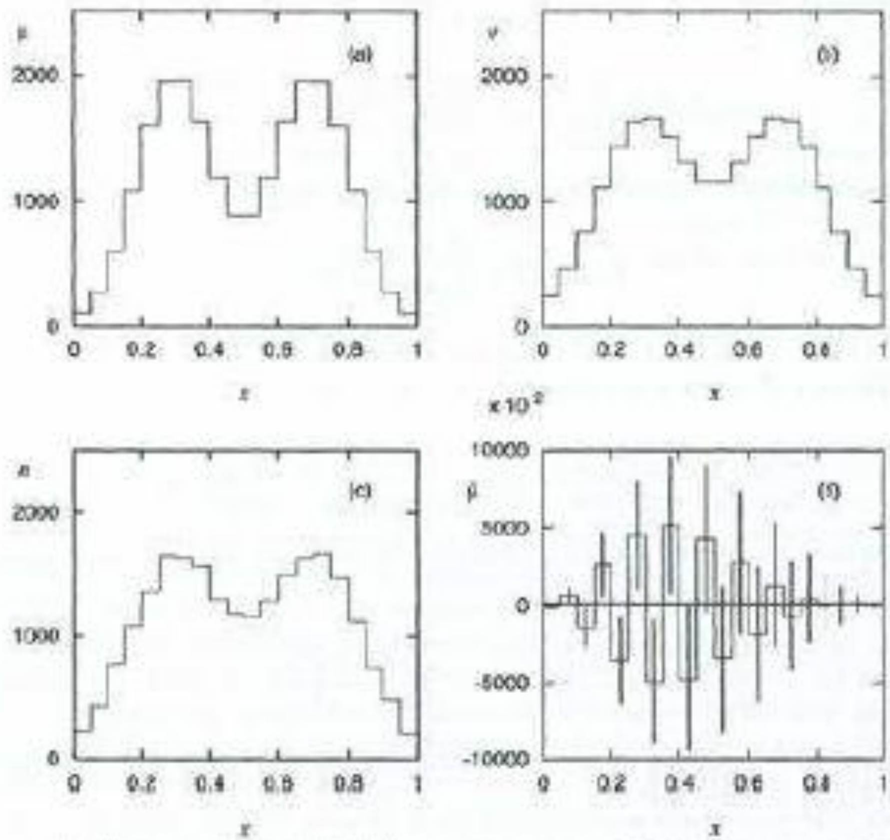


Fig. 11.1 (a) A hypothetical true histogram μ, (b) the histogram of expectation values $\nu = R\mu$, (c) the histogram of observed data n, and (d) the estimators $\hat{\mu}$ obtained from inversion of the response matrix.

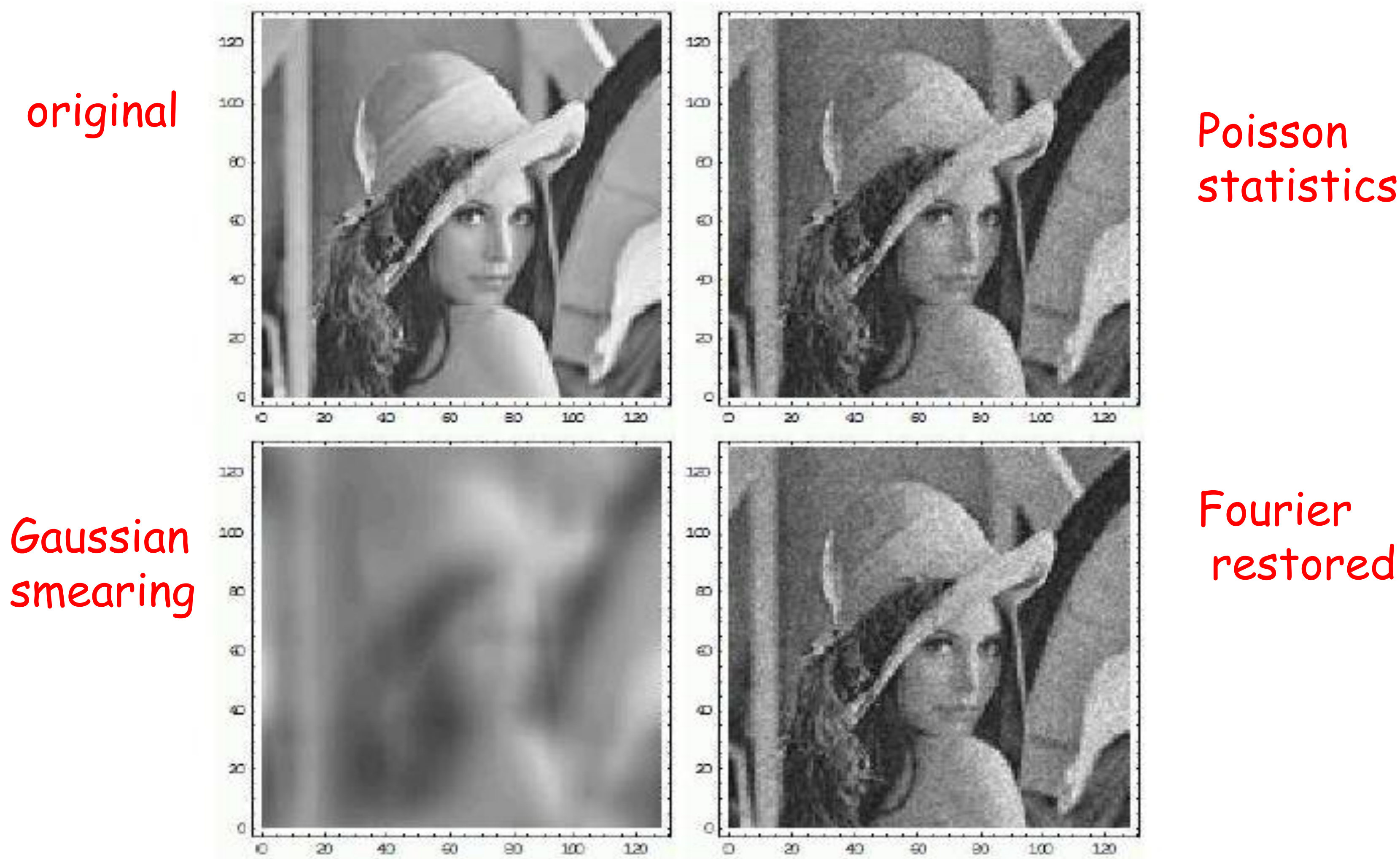
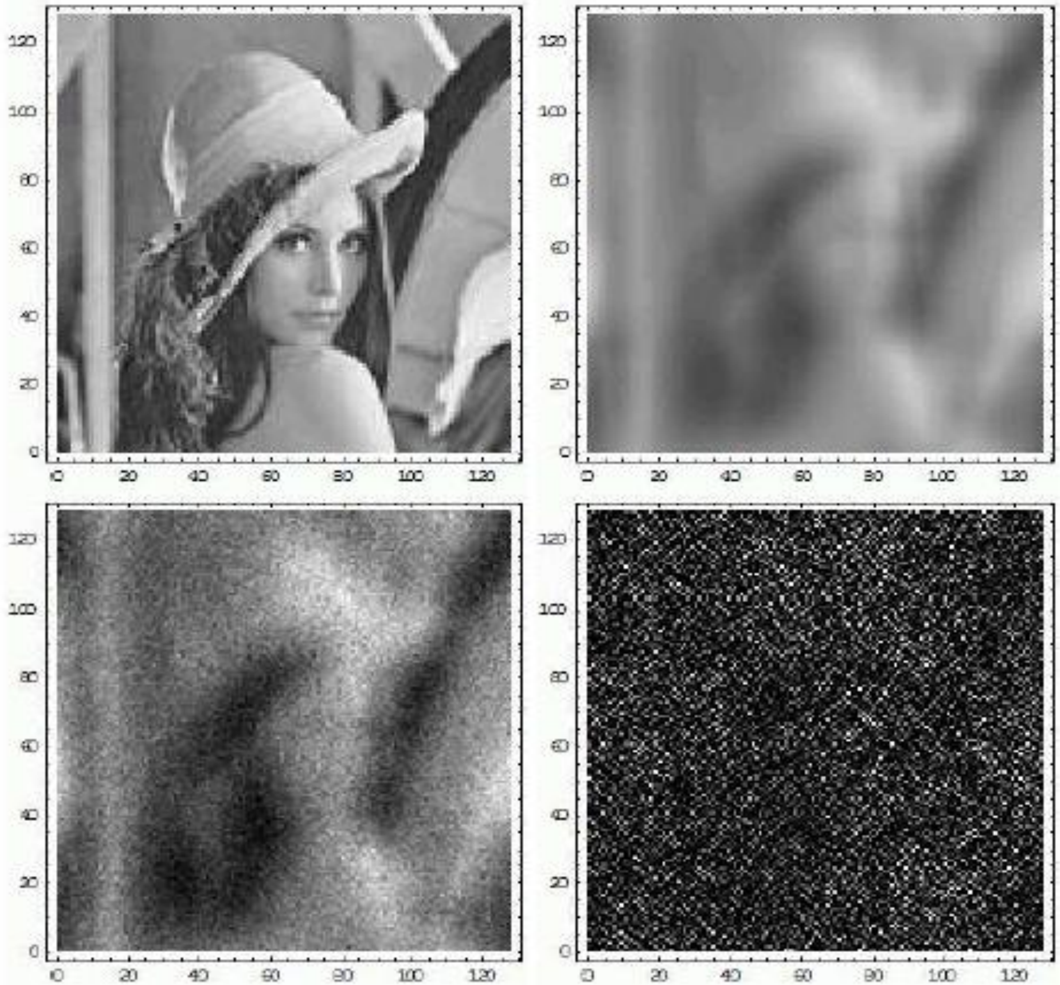


Figure 11: **Lena restored by FFT: The original image (top left) is sampled with Poisson statistics (top right) and smeared with a 2D 10-bins Gaussian PSF (bottom left): the Fourier restored image (bottom right) is similar to the Poisson sampled image. In this case the noise term N is neglected.**

original

Gaussian smearing



Poisson statistics

Fourier (un)restored

Figure 12: **Lena not restored by FFT: In this case the noise term N is not ignored: the original image (top left) is smeared with a 2D 10-bins Gaussian PSF (top right) and the result is sampled with Poisson statistics (bottom left): the Fourier restored image (bottom right) cannot recover the information lost in the noise. Another approach, statistical in nature, is required.**

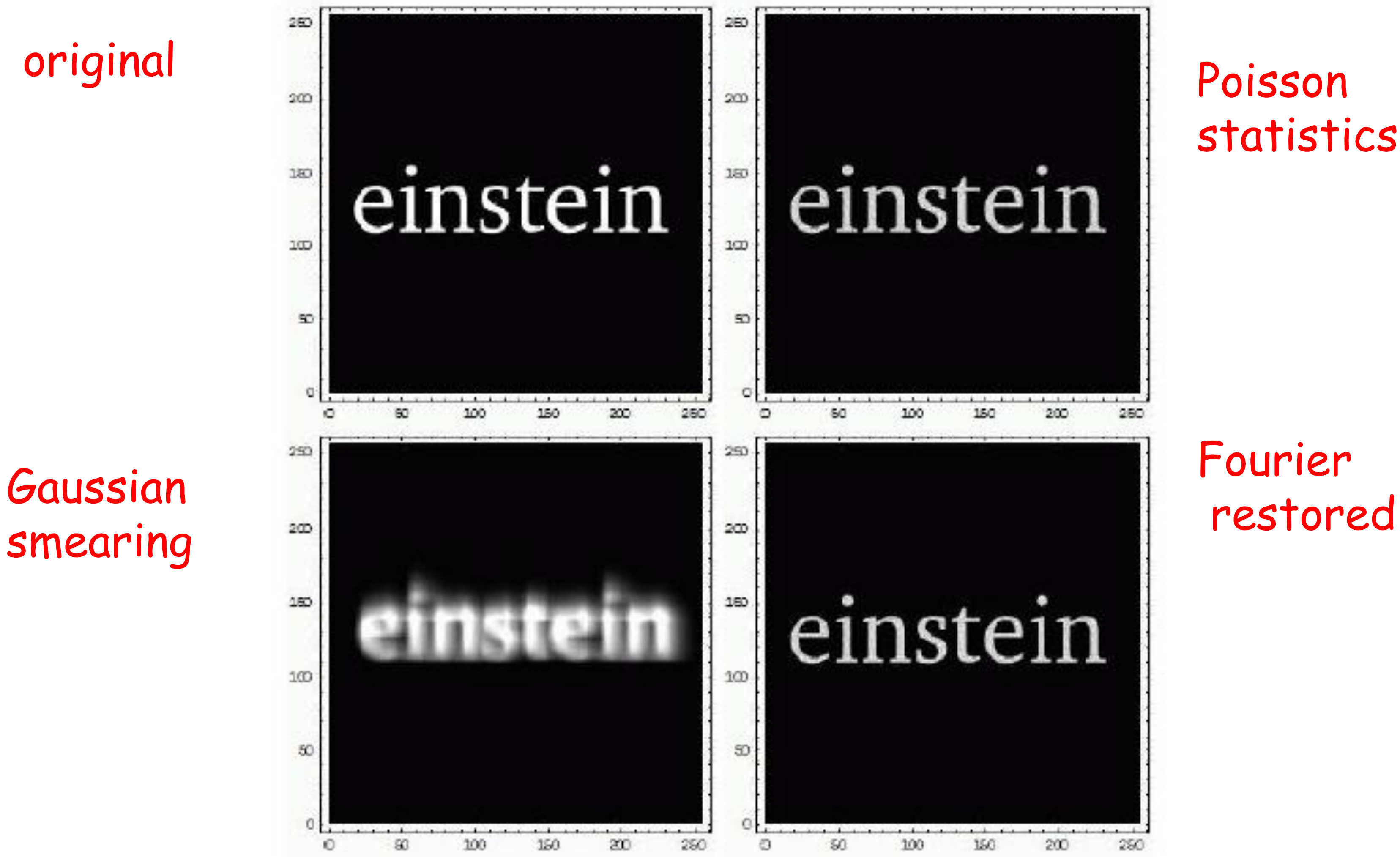


Figure 13: **Einstein restored by FFT: explanation as in Figure 1.**

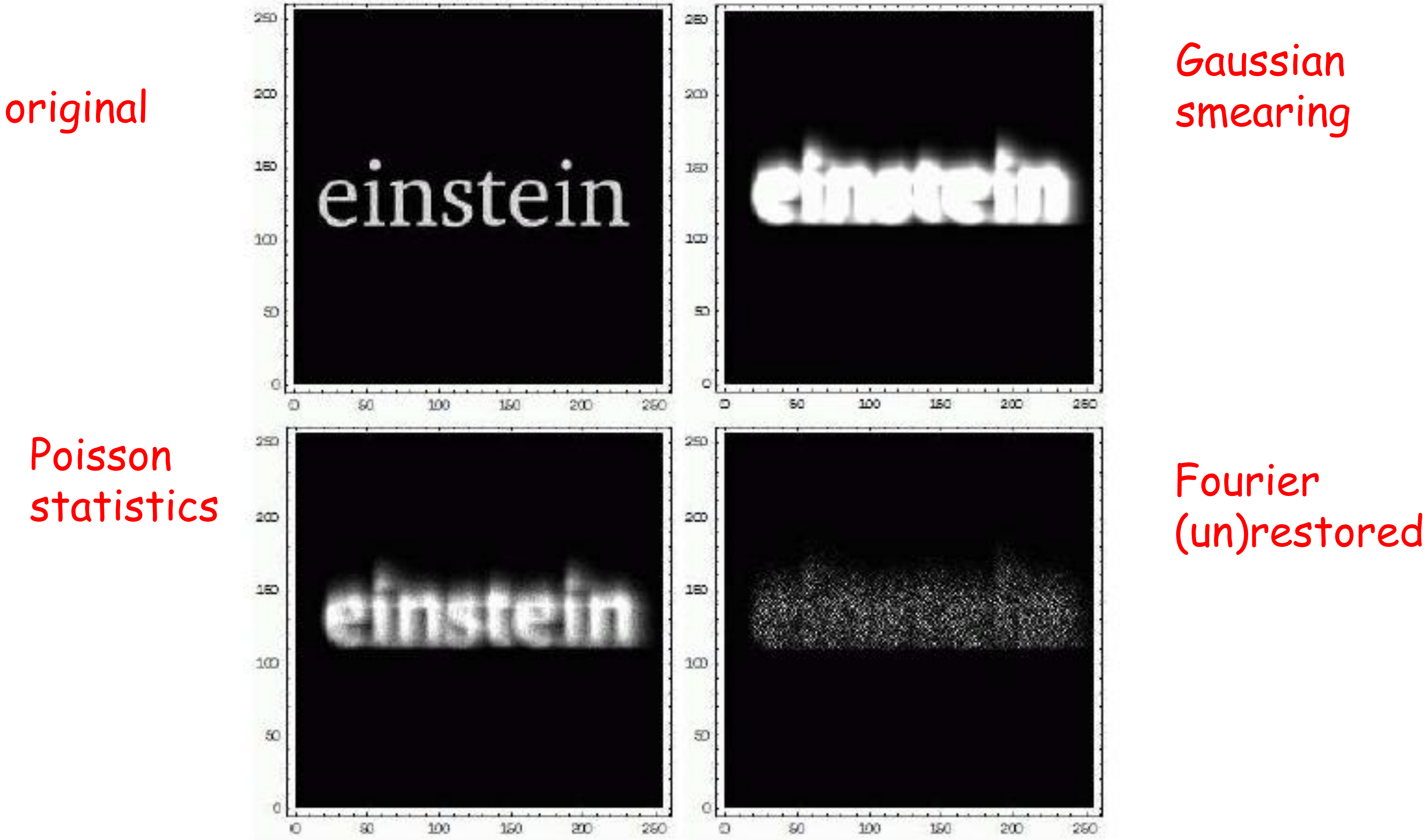
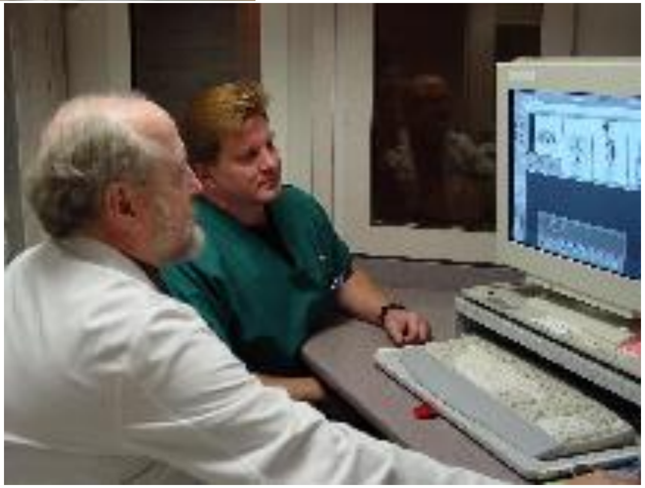
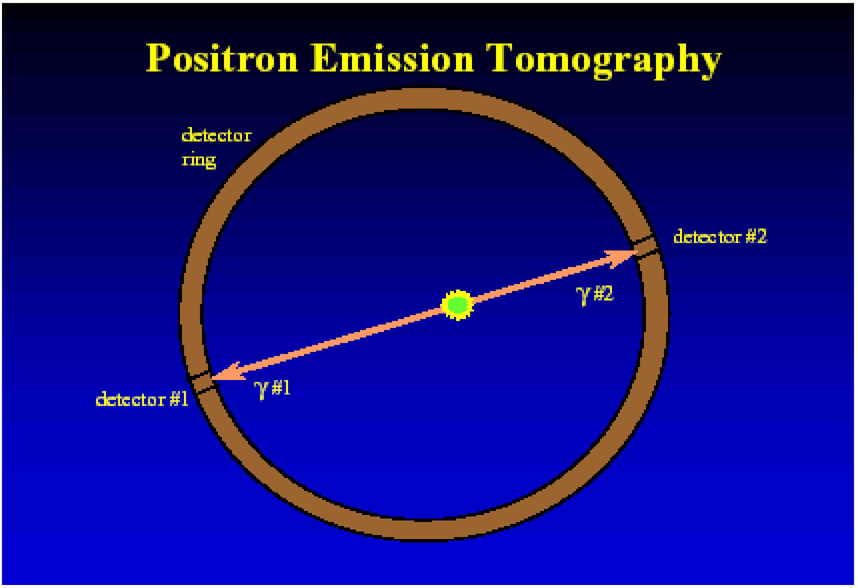
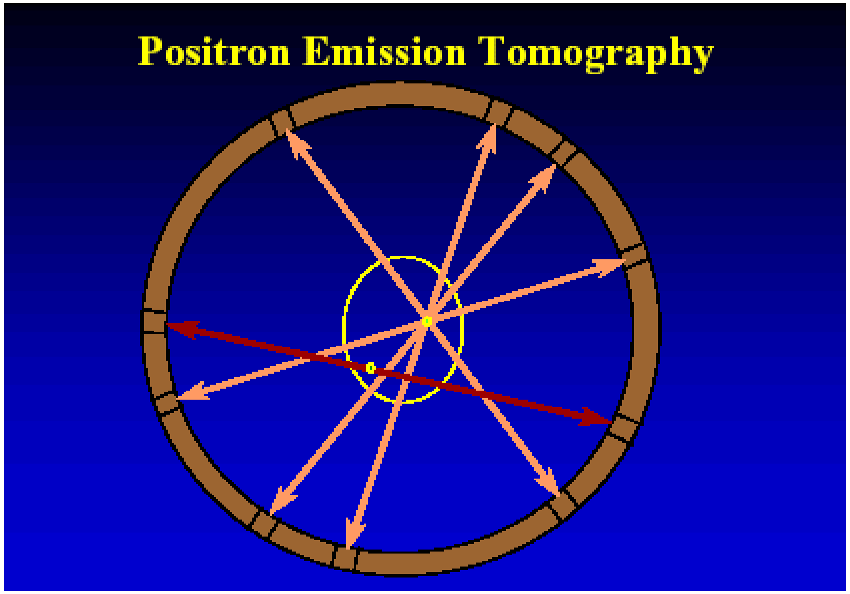


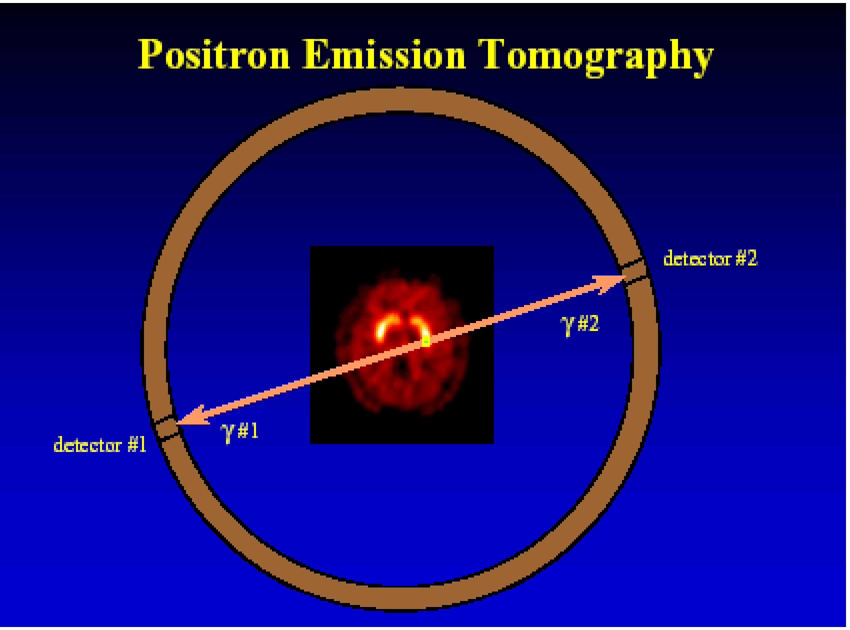
Figure 14: **Einstein not restored by FFT: explanation as in Figure 2.**

PET: positron emission thomography




Positron Emission Tomography
detector
ring
detector #2
γ #2
γ #1
detector #1


Positron Emission Tomography


Positron Emission Tomography
detector #2
γ #2
γ #1
detector #1

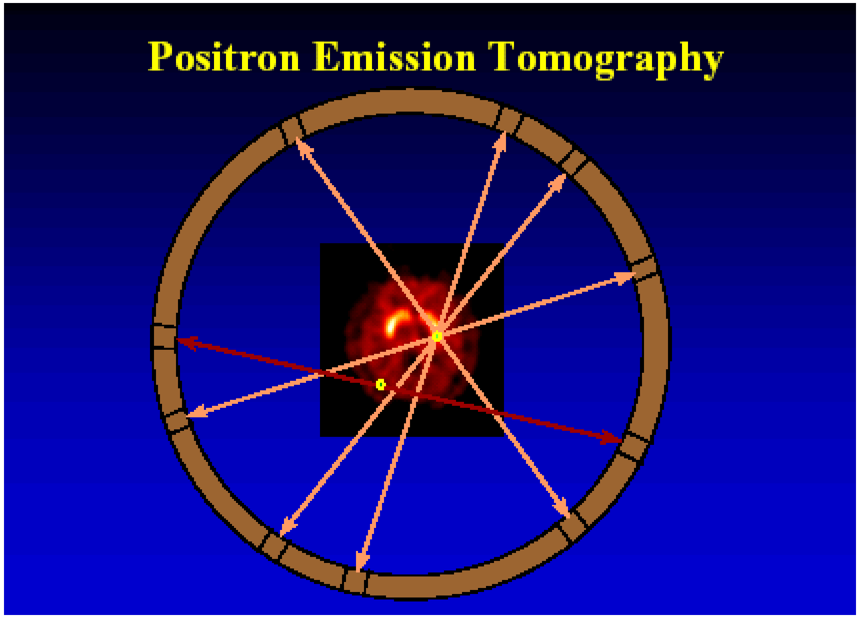

Positron Emission Tomography

Image restoration

When true $\rightarrow$ obs $\equiv \mu \rightarrow \nu$ deterministic methods (FFT) can be used

$$\boldsymbol{\nu} = R * \boldsymbol{\mu} \tag{19}$$

When true $\rightarrow$ smeared $\rightarrow$ obs $\equiv \mu \rightarrow \nu \rightarrow n$ statistical methods must be used

$$\boldsymbol{n} = \boldsymbol{\nu} + \boldsymbol{\rho} = R * \boldsymbol{\mu} + \boldsymbol{\rho}$$

In the poissonian or binomial case we have to minimize:

$$-\ln L(\boldsymbol{\mu}) = -\sum_i \ln P(n_i, \nu_i)$$

In the gaussian case we must minimize

$$\chi^2(\boldsymbol{\mu}) = \sum_{ij} (\nu_i - n_i)(V^{-1})_{ij}(\nu_j - n_j)$$

In 2-D, when $N_{ij}(\mathrm{exp})$ contains fluctuations, we have to minimize:

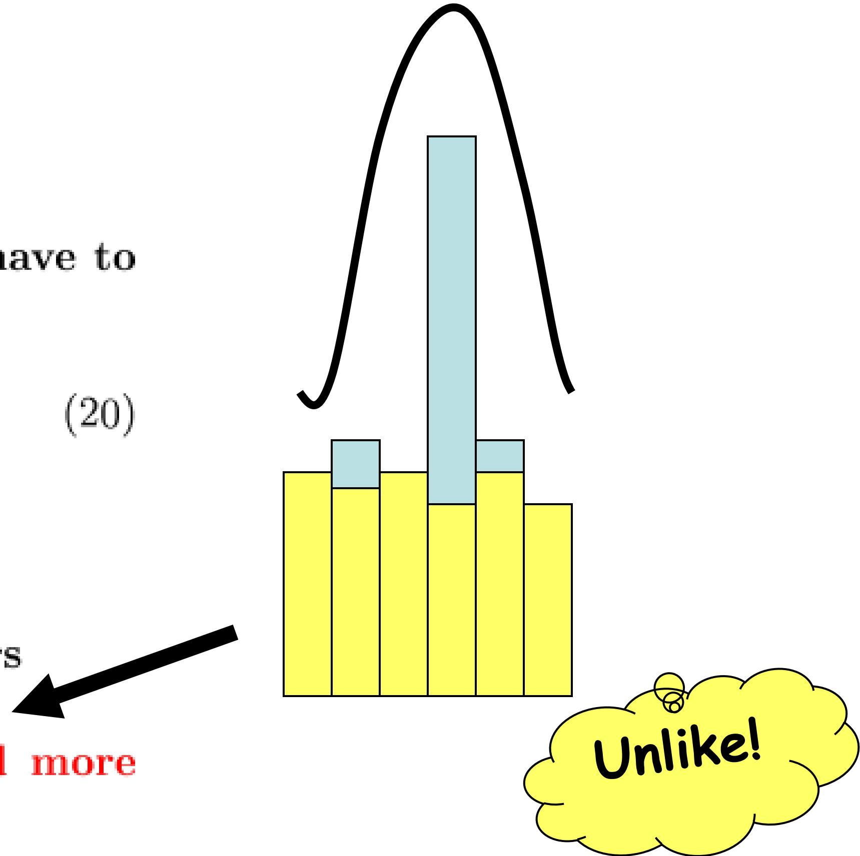
$$-2 \ln L(\boldsymbol{n}|\boldsymbol{\nu}, \boldsymbol{\mu}) \simeq \chi^2 = \sum_{ij} \frac{[N_{ij}(\mathrm{exp}) - N P_{ij}(\mathrm{obs})]^2}{N P_{ij}(\mathrm{obs})} \tag{20}$$

where

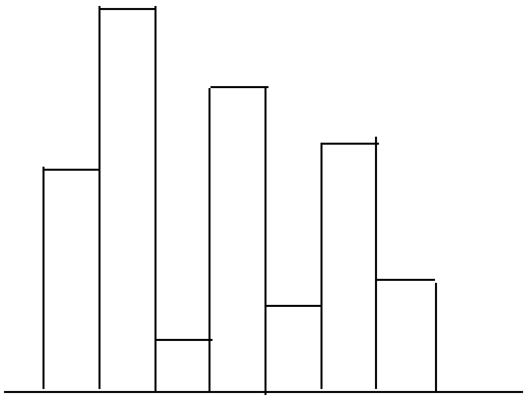
$$P(\mathrm{obs}) = P(\nu|\mu)\, P(\mu)$$

If all the pixel contents $\boldsymbol{\mu}$ are the free parameters to be determined the problem has zero DoF

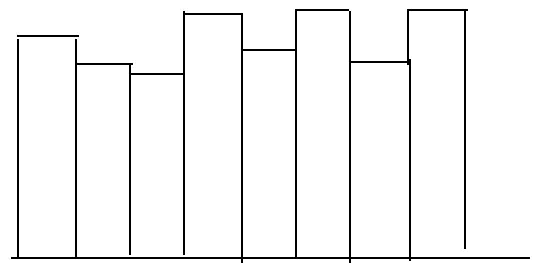
This is a ILL-POSED problem with many (and more probable) unrealistic solution!!



Explanation:



spike solution
HIGLY PROBABLE



smooth solution
UNLIKE

The smeared distributions of two input distributions cannot be distinguished if they agree on a large scale of x but differ by oscillations on a "microscopic" scale much smaller than the experimental resolution

many solutions give a good χ^2

the spike ones are more probable!

Cure: to add to χ^2 an empirical regularization term $C[p]$.

$$\chi^2 \to \alpha\,\chi^2 + C[P(\text{true})]$$

or

$$\chi^2 \to \chi^2 + \alpha C[P(\text{true})]$$

or
to increase the DoF by using a parametric model

$$P(\nu \mid \mu)P(\mu) \to P(\nu \mid \mu')$$

Remember

$$\mu_i \stackrel{PSF}{\longrightarrow} \nu_i \stackrel{random}{\longrightarrow} n_i$$

and consider (95) as a form of the Bayes theorem

$$P(\text{true}_{ij}|\text{obs}) \propto \sum_{i',j'} P_{i'j'}(\text{true})\, P_v(\text{obs}_{ij}|\text{true}_{i'j'}) = L(\boldsymbol{n}|\boldsymbol{\mu})P(\boldsymbol{\mu})$$

Bayesians say: posterior = likelihood × prior One maximizes $P(\text{true}_{ij}|\text{obs}) \equiv F(\boldsymbol{\mu})$ (or minimize $-F(\boldsymbol{\mu})$):

$$F(\boldsymbol{\mu}) = \ln L(\boldsymbol{n}|\boldsymbol{\mu}) + \ln P(\boldsymbol{\mu}) \qquad (99)$$

The frequentist assumes

$$P(\mu) = 1$$

following the Maximum Likelihood (ML) principle. The practical (no Bayesian) experimentalist introduces an empirical regularization parameter α and considers the prior $P(\boldsymbol{\mu})$ as a regularization function $C(\boldsymbol{\mu})$:

$$F(\boldsymbol{\mu}) = \alpha \ln L(\boldsymbol{n}|\boldsymbol{\mu}) + C(\boldsymbol{\mu}) \qquad (100)$$

By keeping fixed the normalization:

$$\nu_T = \sum_i \sum_j R_{ij}\hat{\mu}_j + \rho_i = n_T$$

the objective function is

$$F(\boldsymbol{\mu}) = \alpha \ln L(\boldsymbol{n}|\boldsymbol{\mu}) + C(\boldsymbol{\mu}) + \lambda(n_T - \sum_i \nu_i) \qquad (101)$$

where λ is a Lagrange multiplier

$$\frac{\partial F}{\partial \lambda} = 0 \rightarrow \sum_i n_i = n_T$$

The objective function to be minimized is

$$-F(\boldsymbol{\mu}) = -2\ln L(\boldsymbol{n}|\boldsymbol{\mu}) - \alpha\, C(\boldsymbol{\mu}) + \lambda(n_T - \sum_i \nu_i) \quad (23)$$

$$\mu_j = \mu_{\text{tot}} p_j = \mu_{\text{tot}} \int_{\text{bin j}} f_t(y)\,\text{d}y$$

where $\alpha > 0$. Some regularization terms:

Regularization terms

- minimum second derivative (Tichonov)

$$C(\boldsymbol{\mu}) = -\int [f_t''(y)]^2\,\text{d}y \simeq -\sum_{i=1}^{M-2} [-\mu_i + 2\mu_{i+1} - \mu_{i+2}]^2$$

- minimum variance:

$$C(\boldsymbol{\mu}) = -\text{Var}[\boldsymbol{\mu}] \equiv ||C\boldsymbol{\mu}||^2 = -\sum_i \mu_i^2$$

- maximum entropy (MaxEnt)

$$C(\boldsymbol{\mu}) = -\sum_i p_i \ln p_i = -\sum_i \frac{\mu_i}{\mu_T} \ln \frac{\mu_i}{\mu_T}$$

- cross-entropy

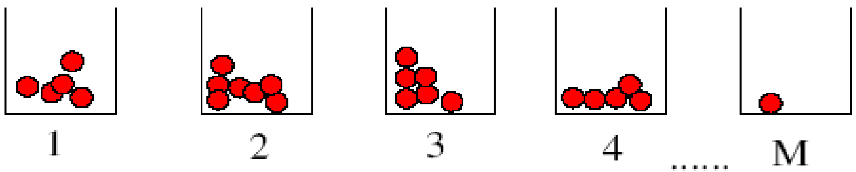
$$C(\boldsymbol{\mu}) = -\sum_i p_i \ln \frac{p_i}{q_i} = -\sum_i \frac{\mu_i}{\mu_T} \ln \frac{\mu_i}{\mu_T\, q_i}$$

where $\boldsymbol{q} = (q_1, q_2, \ldots, q_n)$ is the most likely a priori shape for the true distribution μ_i.

We have M boxes and a monkey that throws N ball randomly into them.
What is the box-balls configuration of highest probability? Probability of a configuration:

$$p = \frac{1}{M^N} \frac{N!}{n_1! n_2! \dots n_N!} = \mathrm{e}^{\ln p}$$

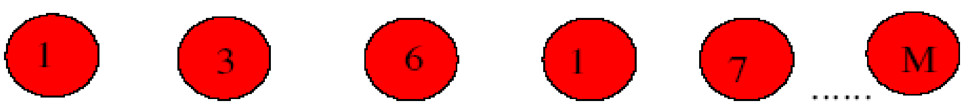
What is MaxEnt ???



M boxes and N balls

equal to

N balls labelled randomly from 1 to M



$$\ln p = -N \ln M + \ln N! - \sum_i \ln(n_i!)$$

Stirling formula: $n! = \sqrt{2\pi n}\, n^n \mathrm{e}^{-n} \rightarrow \ln n! \simeq n \ln n - n$

$$\ln p = -N \ln M + N \ln N - N + \sum n_i - \sum_i n_i \ln n_i$$

$$\sum p_i = 1 \ , \qquad p_i = \frac{n_i}{N} \ , \ \ln p = -N \ln M - N \sum_i p_i \ln p_i$$

The most probable configuration means to maximize

$$\ln p(\boldsymbol{\mu}) = S = -\sum_i p_i \ln p_i$$

The iterative principle

$$N_{ij}(\text{th}) = N P_{ij}(\text{obs}) = N \sum_{i'j'} P_{i'j'}(\text{true})\, P_v(\text{obs}_{ij}|\text{true}_{i'j'}) \quad (25)$$

We write this equation considering the operator R:

$$n = R * \mu$$

The iterative method (Van Cittert 1930) adds with a weight β the residual r_k to the current solution

$$\mu_{k+1} = \mu_k + \beta[n - R * \mu_k] \quad (26)$$

The method is based on the known equation

$$\sum_{i=0}^{k} q^n = \frac{1 - q^{k+1}}{1 - q} \quad (27)$$

for $k \to \infty$ the series converges if $|q < 1|$.
By applying iteratively (26)

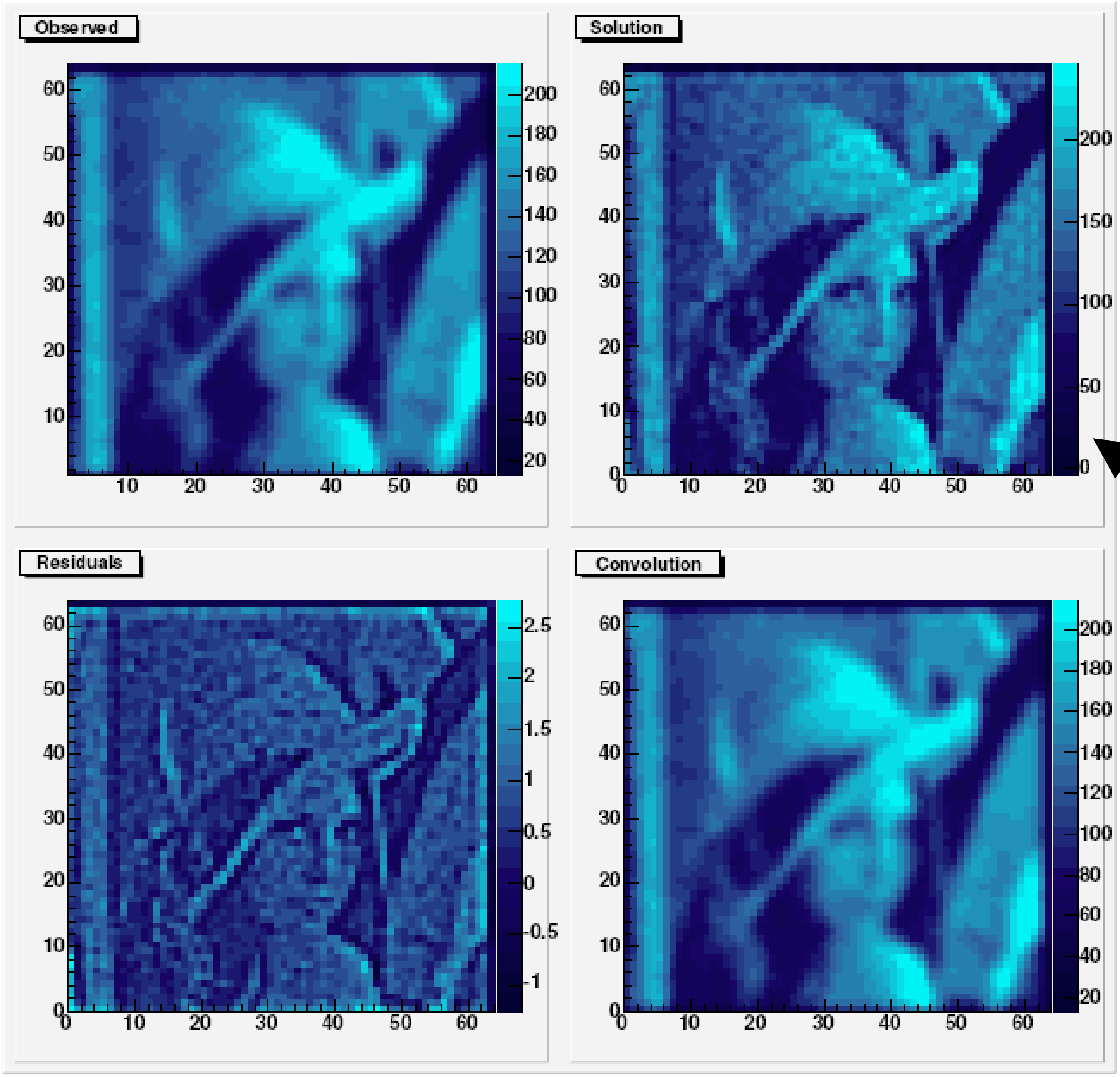
$$\begin{aligned}
\mu_{k+1} &= \beta n + (1 - \beta R)\mu_k = \beta n + (1 - \beta R)(\beta n + (1 - \beta R)\mu_{k-1}) \\
&= \beta n + \beta(1 - \beta R)n + (1 - \beta R)^2 \mu_{k-1} \\
&= \beta n + \beta(1 - \beta R)n + \beta(1 - \beta R)^2 n + (1 - \beta R)^3 \mu_{k-2} \ldots \\
&= \sum_{i=0}^{k} \beta(1 - \beta R)^i n \;.
\end{aligned}$$

From (27):

$$\mu_{k+1} = \frac{I - (I - \beta R)^{k+1}}{\beta R} \beta n \to R^{-1} n = \mu \;, \quad \text{for} \quad k \to \infty \;.$$

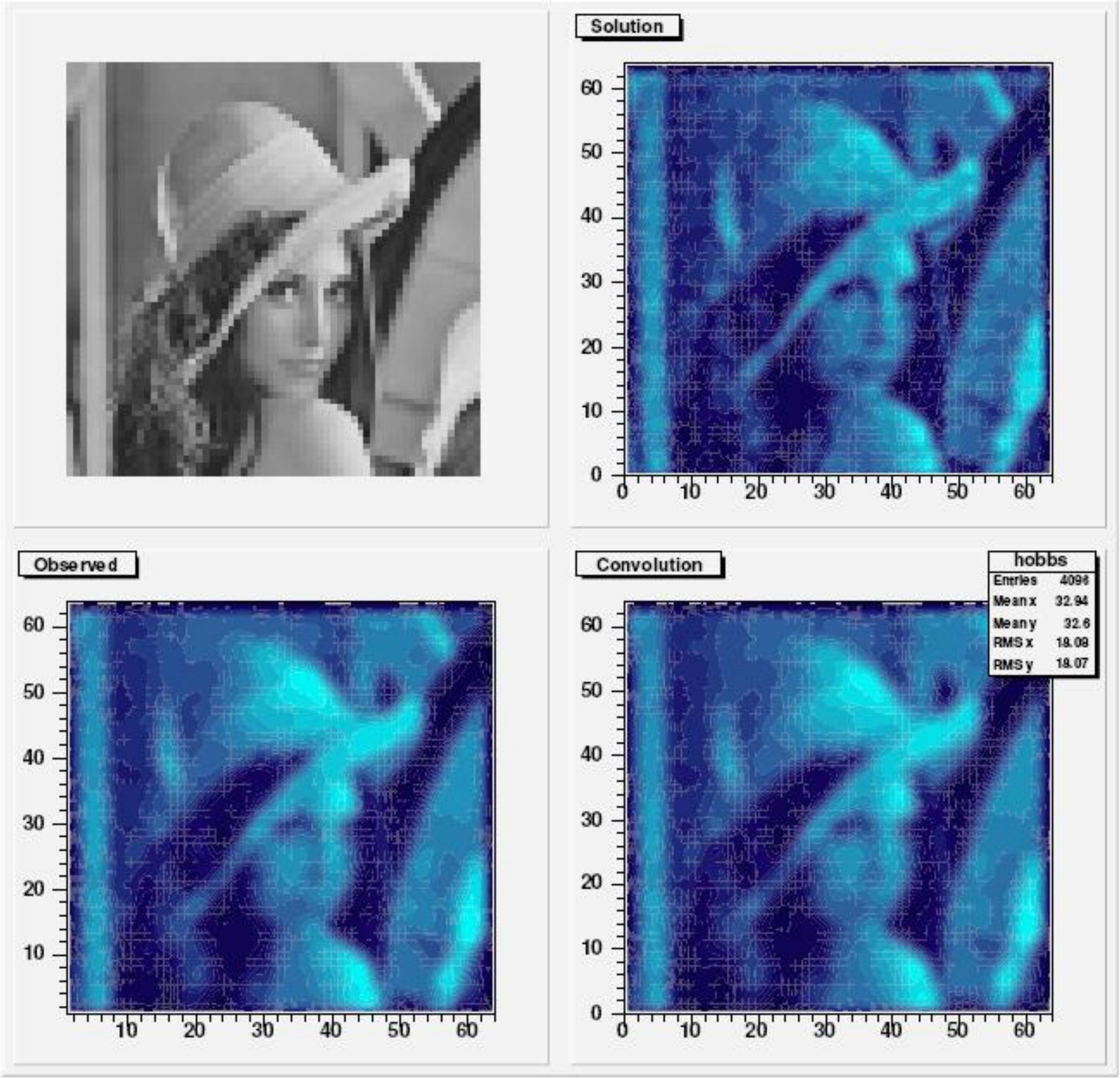
if $|I - \beta R| < 1$

Without β $\mu_{k+1} = \mu_k + [n - R * \mu_k]$

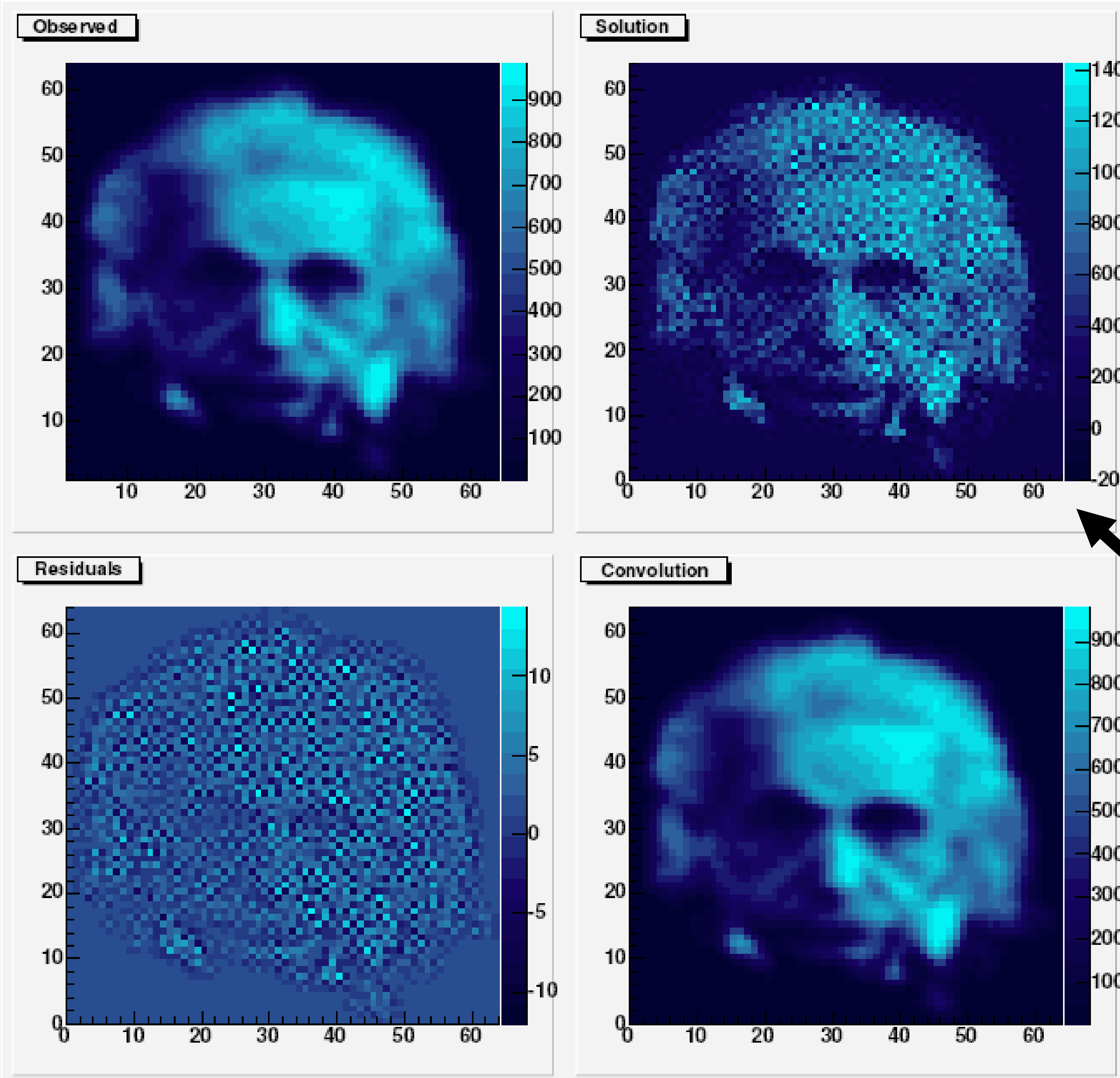


The iterative Principle without best fit

Good!



The iterative Principle without best fit + smoothing



The iterative Principle without best fit

Bad!

Consider the case with statistical fluctuations

$$\mu_i \overset{PSF}{\rightarrow} \nu_i \overset{random}{\rightarrow} n_i$$

$$\mathbf{n} = \mathbf{R}{*}\mu + \mathbf{r}$$

To have regular solutions and to make the method robust, we must search for an iterative solution which minimize the χ^2:

$$\chi^2 = ||\mathbf{R} * \mu - \mathbf{n}||^2 = \frac{1}{2}\sum_{\mathbf{ik}}(\sum_{\mathbf{mn}} \mathbf{R}_{\mathbf{i-m,k-n}}\,\mu_{\mathbf{mn}} - \mathbf{n}_{\mathbf{ik}})^2 \ .$$

Note that is the case in which
the PSF depends on the pixel difference only (translational invariance)
In this case we consider R as an operator and we can work with symmetric $M \times N$ matrices.
Minimum χ^2 w.r.t μ_{ik} gives the equations:

$$\frac{\partial \chi^2}{\partial \mu_{mn}} = \sum_{ik}(\sum_{rs} R_{i-r,k-s}\,\mu_{rs} - n_{ik})\, R_{i-m,k-n} = 0$$

$$0 < m < M \ , \quad 0 < n < N \ .$$

Hence :

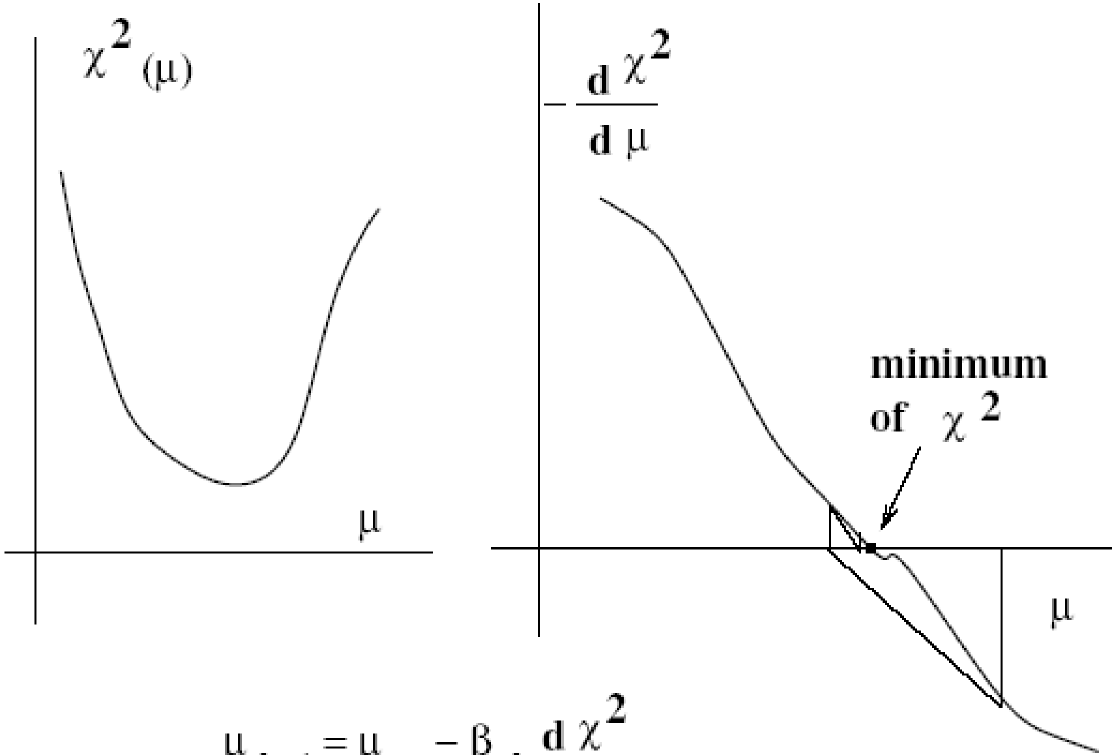
$$\frac{\partial \chi^2}{\partial \mu} = R * [R * \mu - n] = 0 \qquad (29)$$

The iterative solution toward the μ's that minimize χ^2 is obtained with the Robbins-Monro algorithm (1951).

The iterative algorithm + best fit

The iterative algorithm + Best fit

It belongs to the family of the Newton-Raphson methods From (29):



$$\mu_{k+1} = \mu_k - \beta_k \frac{d\chi^2}{d\mu}$$

$$\mu_{k+1} = \mu_k + \beta_k R * [n - R * \mu] \qquad (30)$$

This formula minimizes χ^2. For example when $\beta_k = \beta$:

$$\begin{aligned}
\mu_{k+1} &= \mu_k + \beta R * [n - R * \mu_k] \\
&= \beta R * n + (I - \beta R^2) * \mu_k \\
&= \beta R * n + \beta(I - \beta R^2) * R * n + (I - \beta R^2)^2 * \mu_{k-1} \\
&= \sum_{i=0}^{k} \beta(I - \beta R^2)^i * R * n = \frac{I - (I - \beta R^2)^{k+1}}{\beta R^2} \beta R * n \to R^{-1} n
\end{aligned}$$

$$\mu_{k+1} = \mu_k + \beta_k R * [n - R * \mu_k] \tag{31}$$

Previous method converges if

$$||I - \beta R * R|| < 1$$

when β is independent of k. In this case

$$0 < \beta < \frac{2}{\text{max eigenvalue of } (R * R * \mu) * \mu^{-1}}$$

When β_k depends on k convergence is assured if (Robbins and Munro 1951)

$$\lim_{N\to\infty} \beta_N = 0 \ , \quad \sum_{N=1}^{\infty} \beta_N = \infty \ , \quad \sum_{N=1}^{\infty} \beta_N^2 < \infty \ ,$$

Next, the method must be regularized by adding a term to the χ^2.

$$\chi^2 = ||R * \mu - n||^2 + \alpha||C * \mu||^2 \tag{32}$$

For example,

$$||C * \mu||^2 = \sum_{ik} \mu_{ik}^2$$

$$||C * \mu||^2 = |\sum_{ik} \mu_{ik} \ln \mu_{ik}/\mu_{\text{tot}}| \ , \qquad \alpha < 0$$

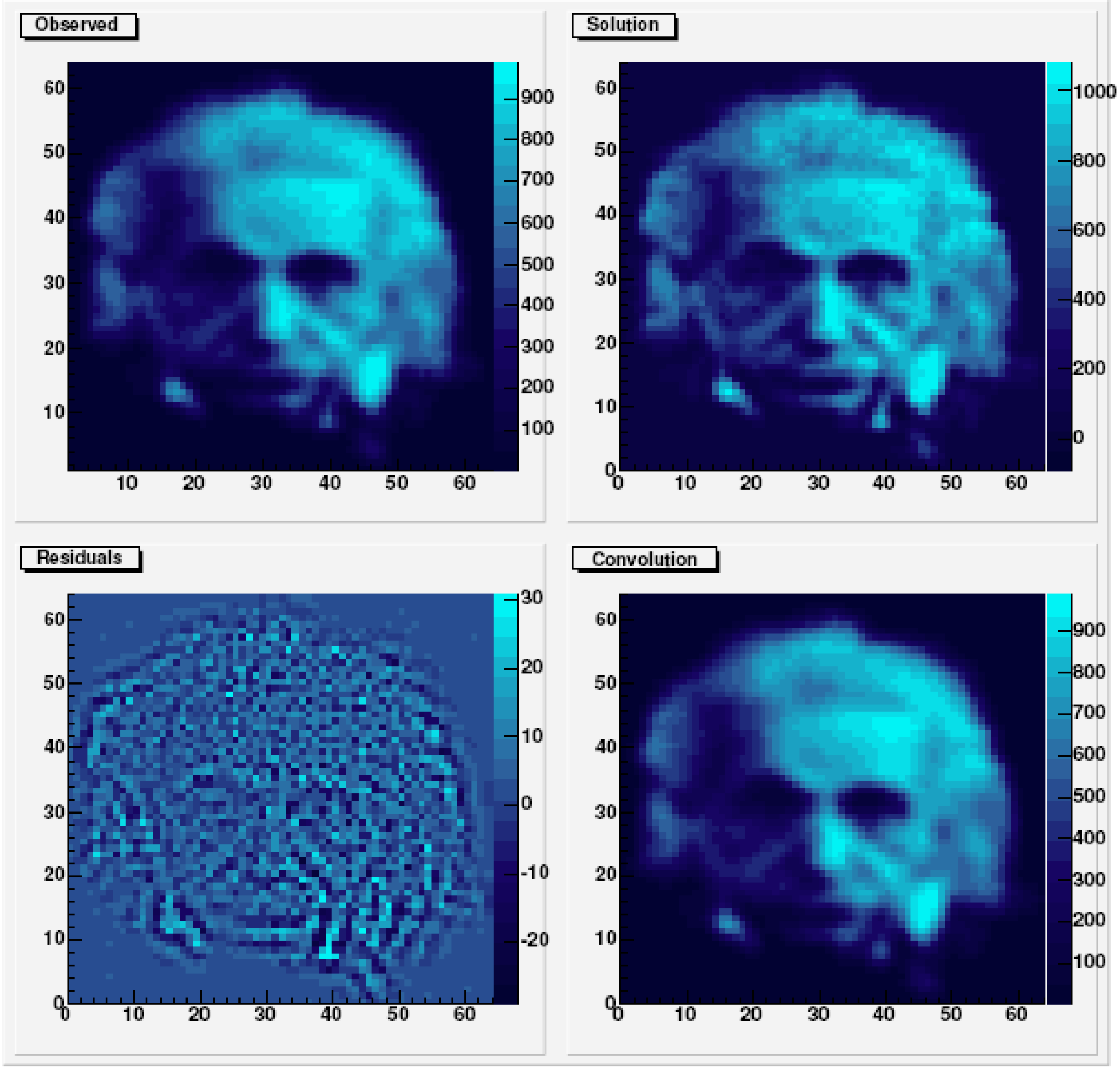
The iterative solution becomes

$$\mu_{k+1} = \mu_k + \beta_k [R * n - (R * R + \alpha C * C)\mu_k] \tag{33}$$

The iterative algorithm + best fit + regularization

$$\mu_{k+1} = \mu_k + \beta_k[R * n - [R * R * \mu_k + \alpha(\ln \mu_k/\mu_T + I)]$$

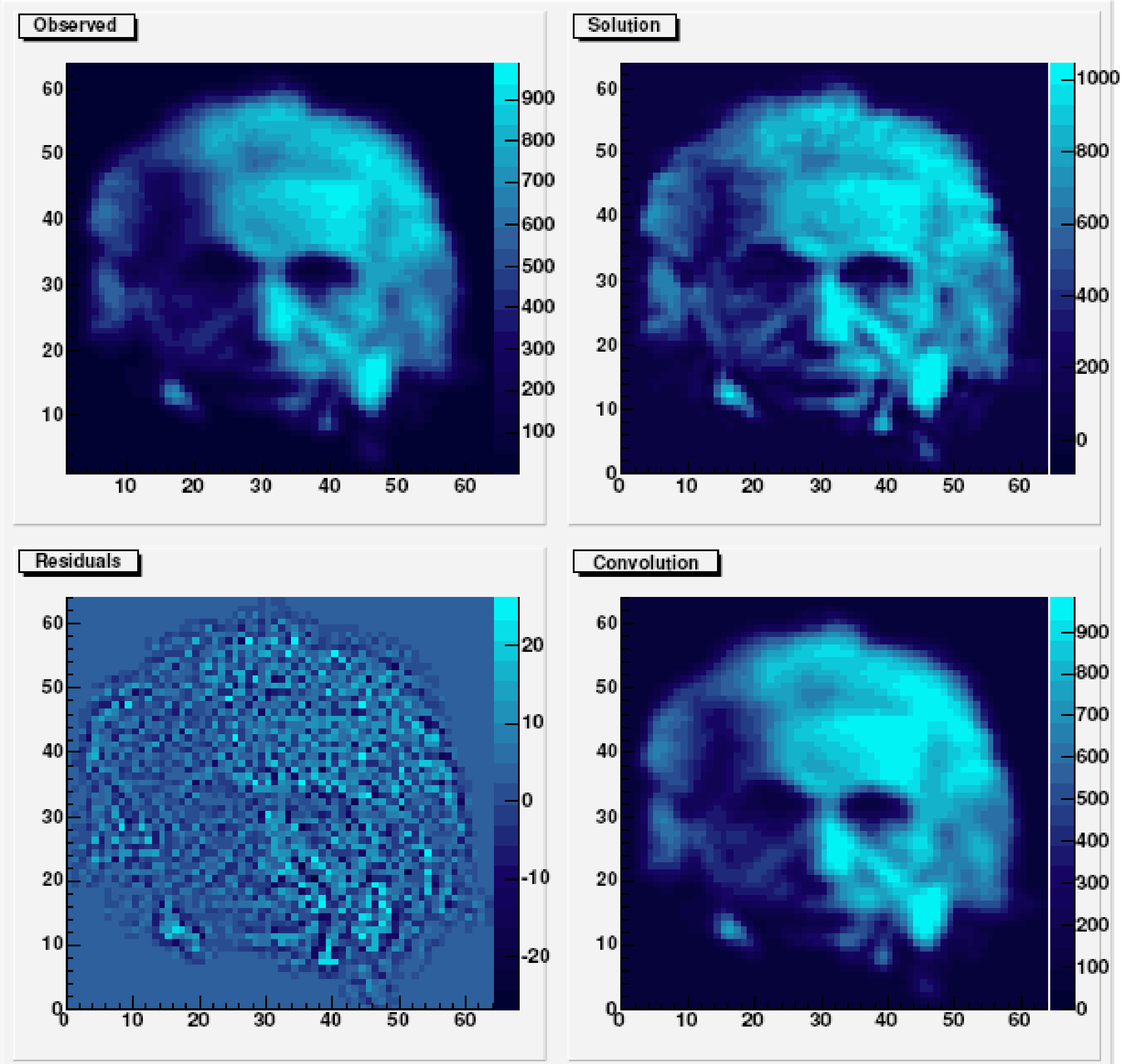
About 40 iterations, regularized with Maximum entropy



The iterative algorithm + best fit + MaxEnt regularization

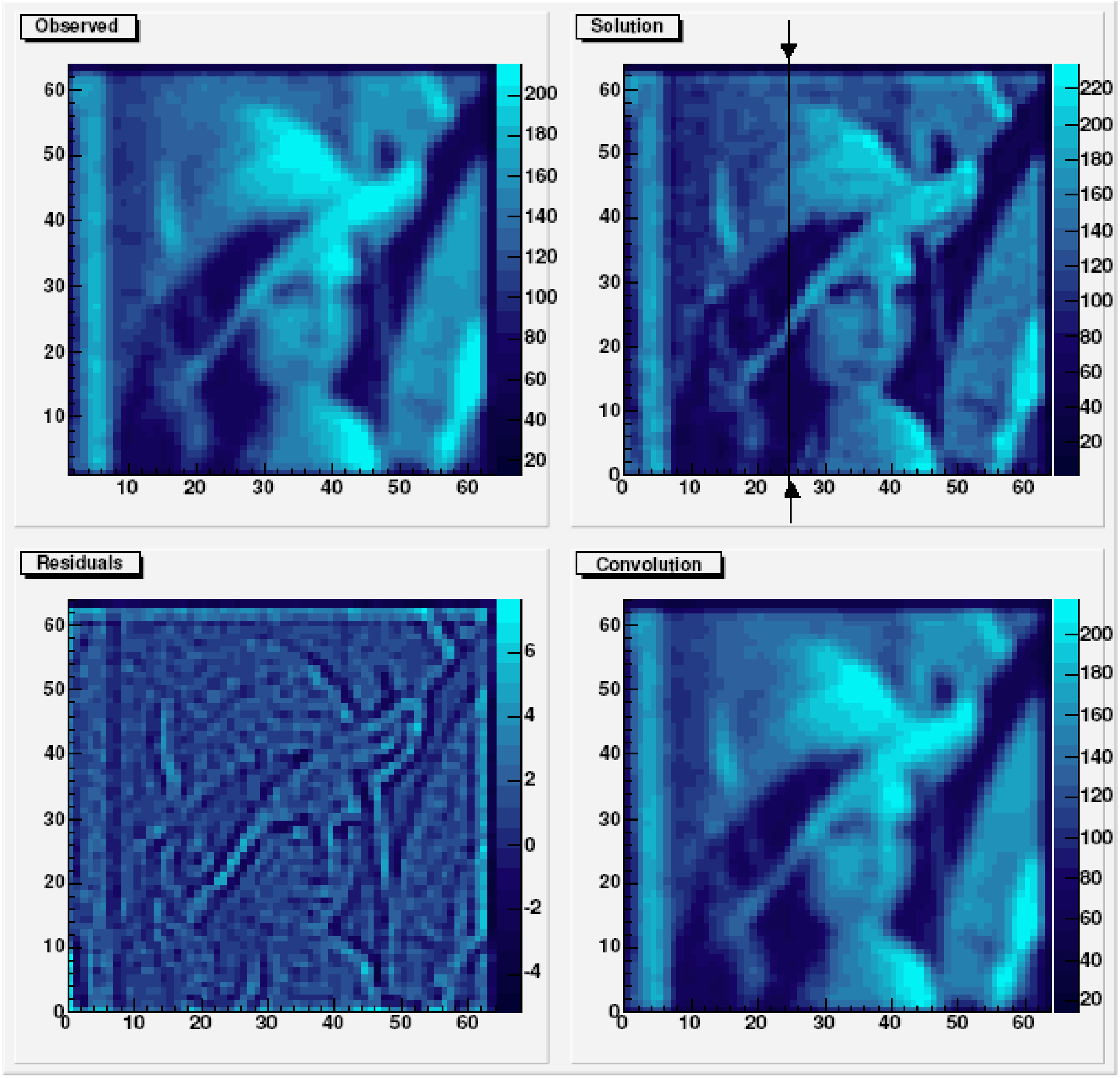
$$\mu_{k+1} = \mu_k + \beta_k [R * n - (R * R + \alpha I) * \mu_k]$$

About 100 iterations, regularized with the sum of squares



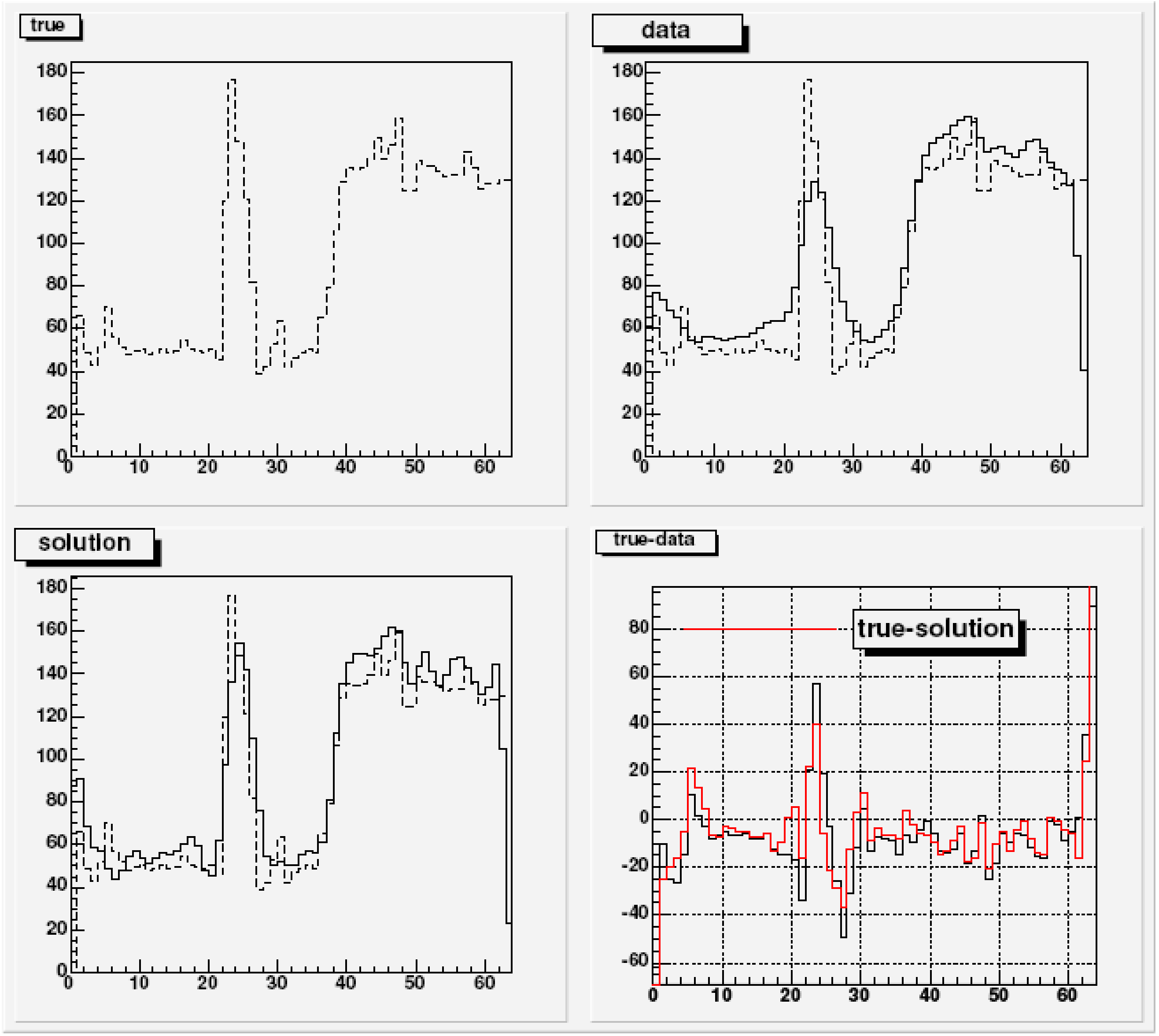
The iterative algorithm + best fit + Tichonov regularization

$$\mu_{k+1} = \mu_k + \beta_k[R * n - (R * R + \alpha I) * \mu_k]$$



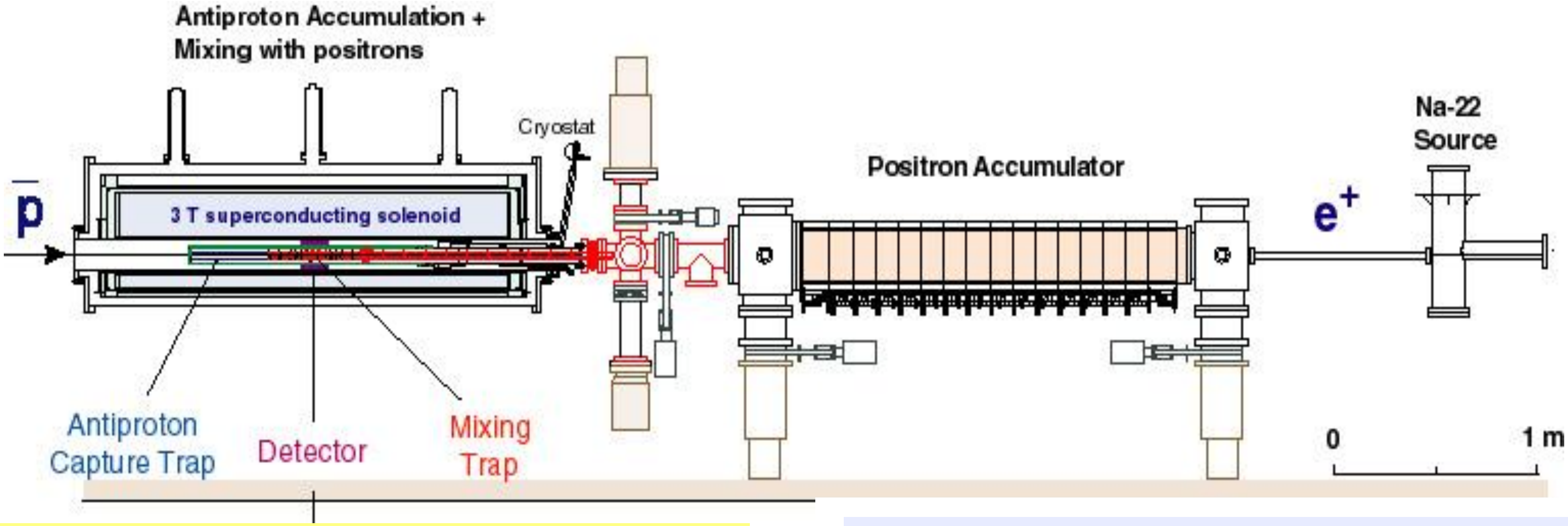
The iterative algorithm + best fit + Tichonov regularization

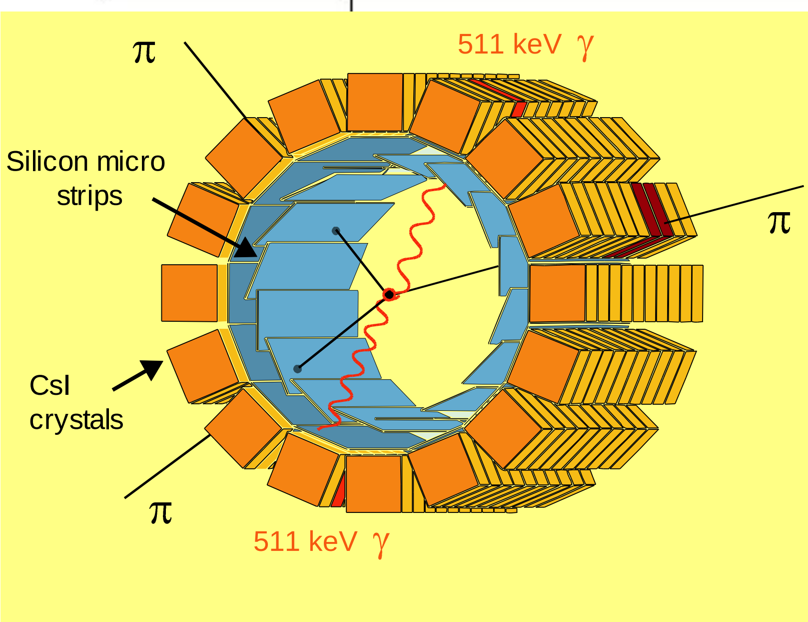
$$\mu_{k+1} = \mu_k + \beta_k[R * n - (R * R + \alpha I) * \mu_k]$$

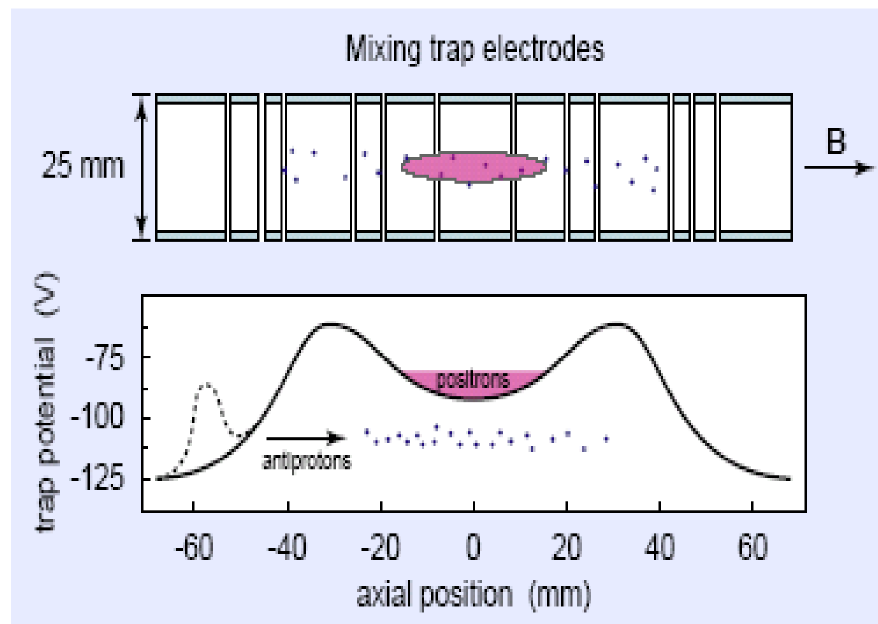


The iterative algorithm + best fit + Tichonov regularization

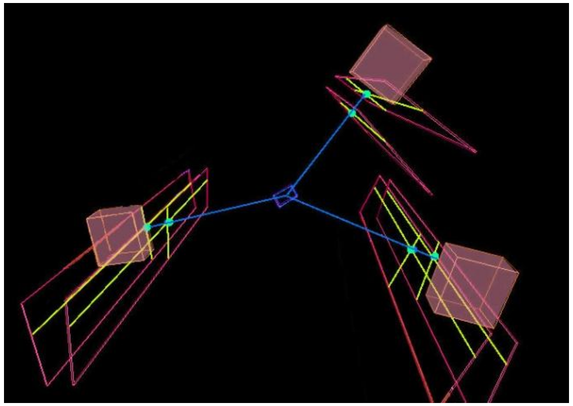
ATHENA apparatus

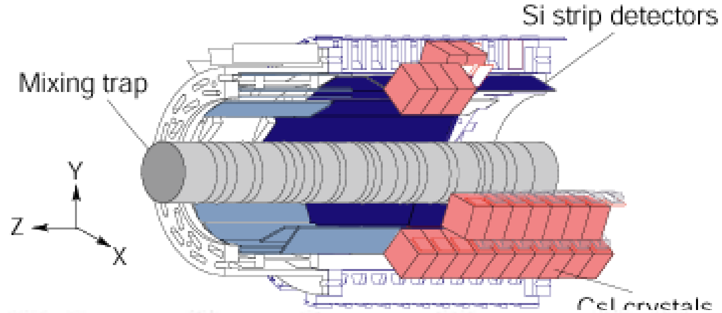






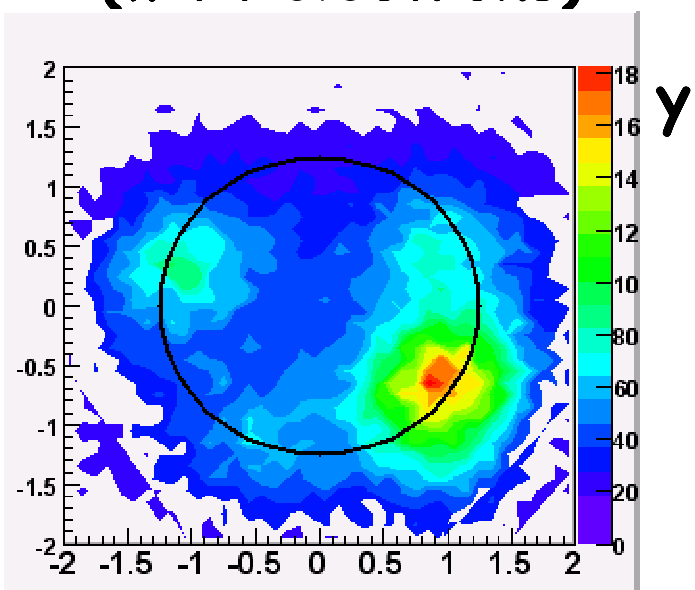
From the ATHENA detector





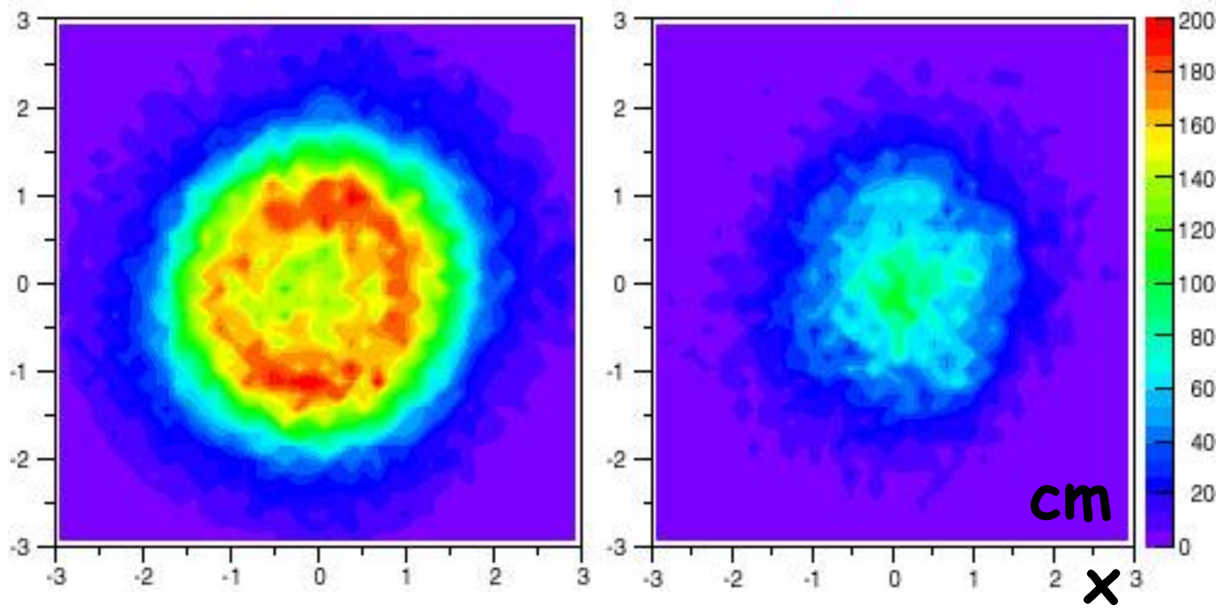
Distribution of annihilation vertices when antiprotons are mixed with ...

Pbar-only
(with electrons)



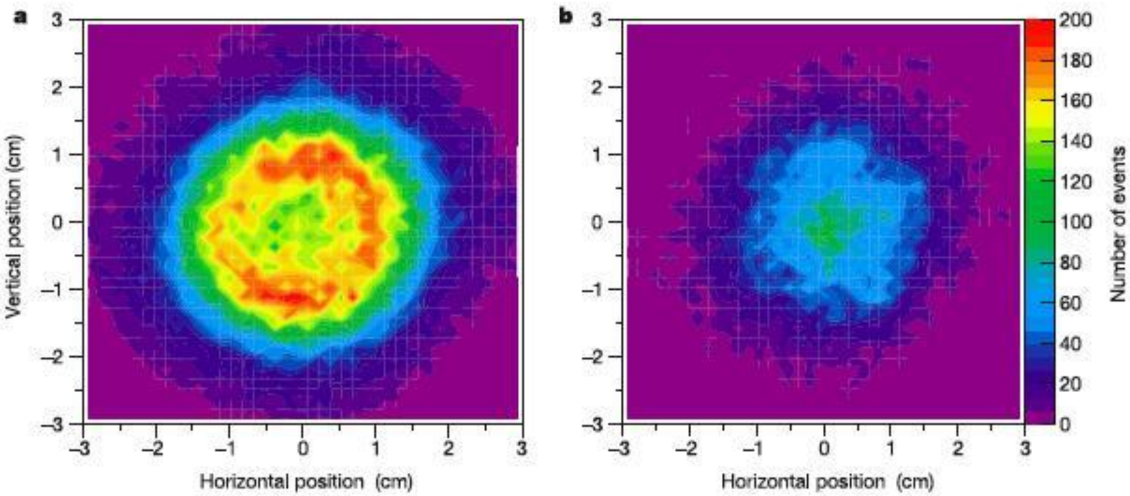
cold positrons

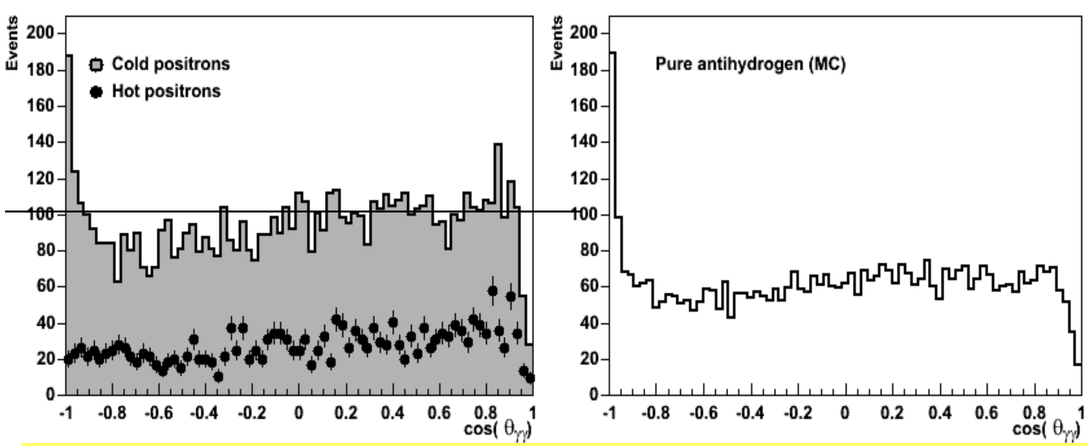
hot positrons



antihydrogen !!!!!!!!!!

FIRST COLD ANTIHYDROGEN PRODUCTION & DETECTION (2002)
M. Amoretti et al., Nature 419 (2002) 456
M. Amoretti et al., Phys. Lett. B 578 (2004) 23





$$\frac{80}{\sqrt{190+110}} = 4.7; \quad \frac{80}{\sqrt{190}} = 6.5; \quad \frac{80}{\sqrt{110}} = 8$$

SIGNAL ANALYSIS:

opening angle
xy vertex distribution
radial vertex distribution

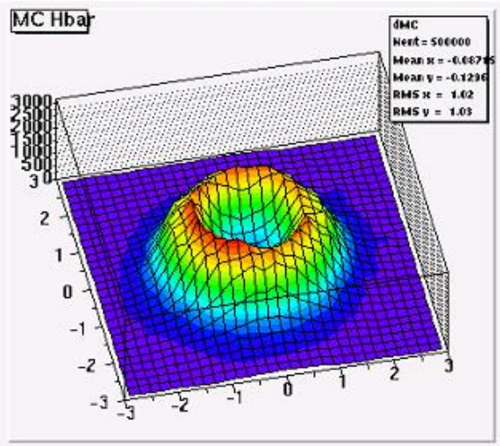
65 % +/- 10% of annihilations are due to antihydrogen

between 2002 & 2004
more than 2 millions antihydrogen atoms have been produced

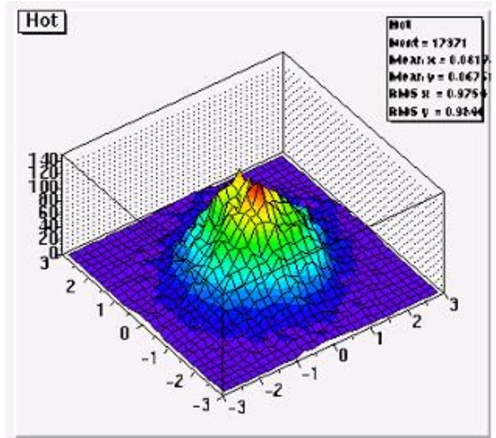
that's about 2 x 10^{-15} mg
.. or .. 1000 Giga years for a gram

Annihilation vertex in the trap x-y plane

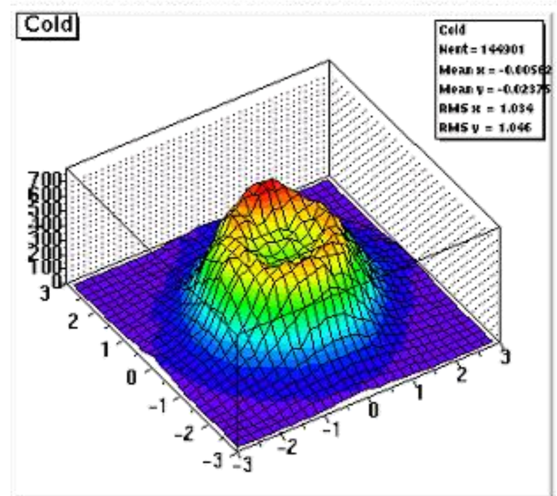
Hbar (MC)



BCKG (HotMixData)



Cold Mix data



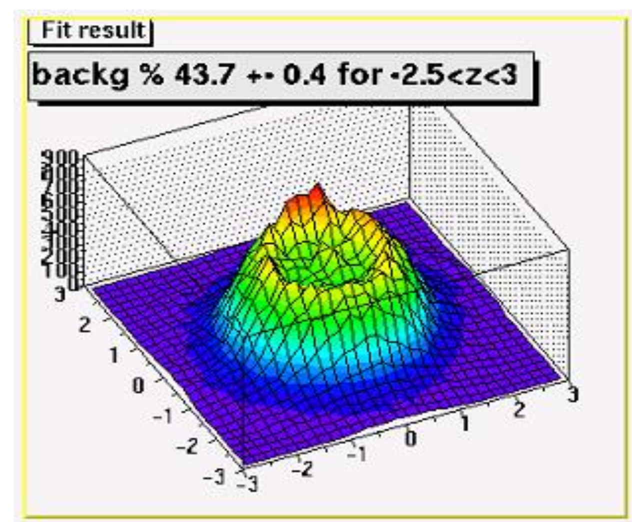
Pbar vertex XY projection (cm)

x Hbar + (1-**x**) BCKG =

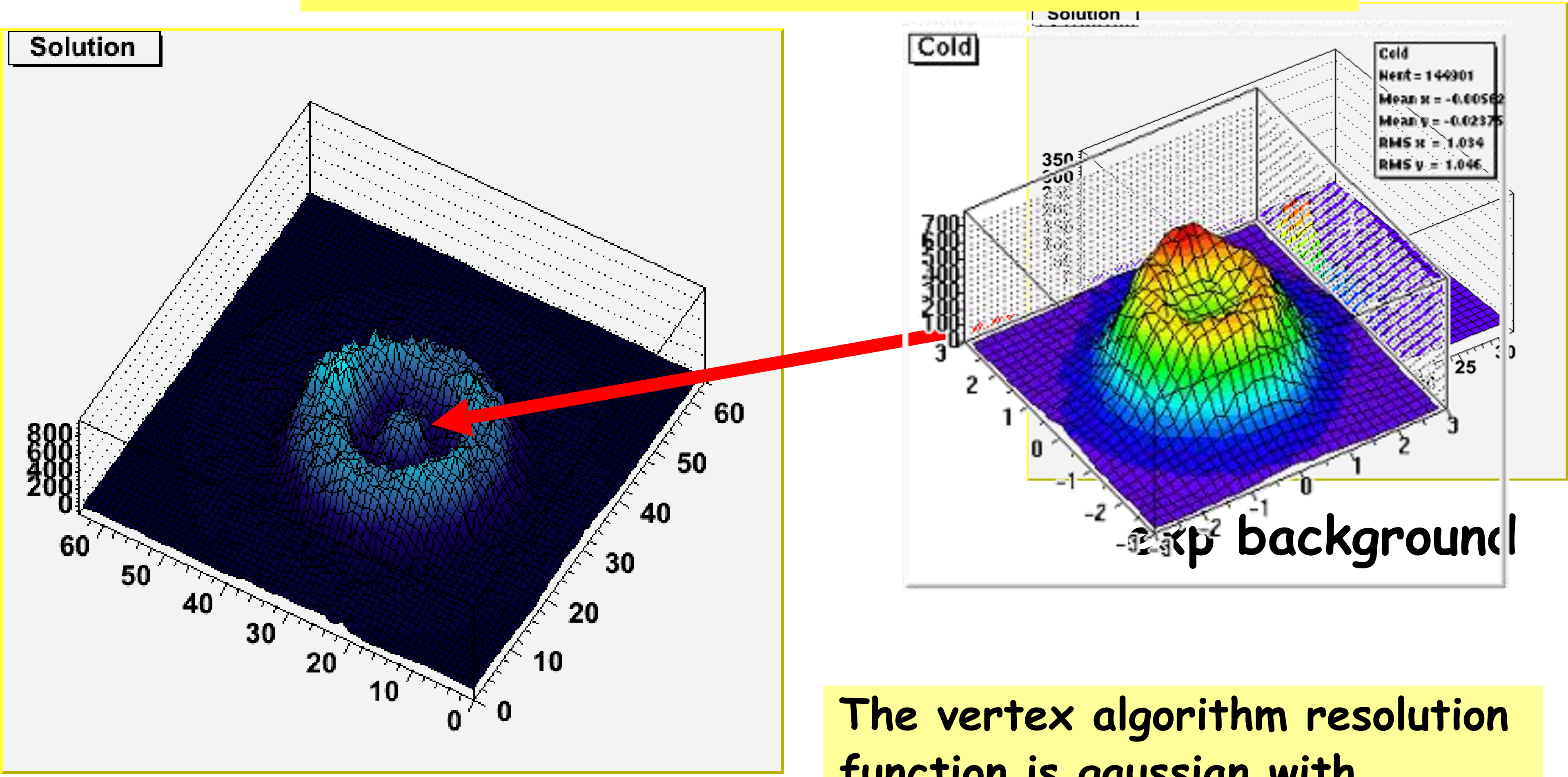
ML Fit Result

Hbar percentage

x = 0.65 ± 0.05



Iteratve best fit method



The vertex algorithm resolution function is gaussian with $\sigma \cong 3$ mm

Cold Mix

The 2D deconvolution reveals two different annihilation modes

The iterative algorithms + best fit + regularization

- iterative algorithms are used in unfolding (ill posed) problems

- they need a Bayesian regularization term

- when there are degrees of freedom, one can use a best fit of a signal+background function to the data

- in this case there are no Bayesian terms (pure frequentist approach) end

Conclusions

- don't be dogmatic

- use Bayes to parametrize the a priori knowledge if any, not the ignorance

- in the case of poor a priori knowledge, use the frequentist methods

The objective function to be minimized is

$$-F(\boldsymbol{\mu}) = -2\ln L(\boldsymbol{n}|\boldsymbol{\mu}) - \alpha\, C(\boldsymbol{\mu}) + \lambda(n_T - \sum_i \nu_i) \quad (24)$$

Some choices of $\alpha > 0$ are:

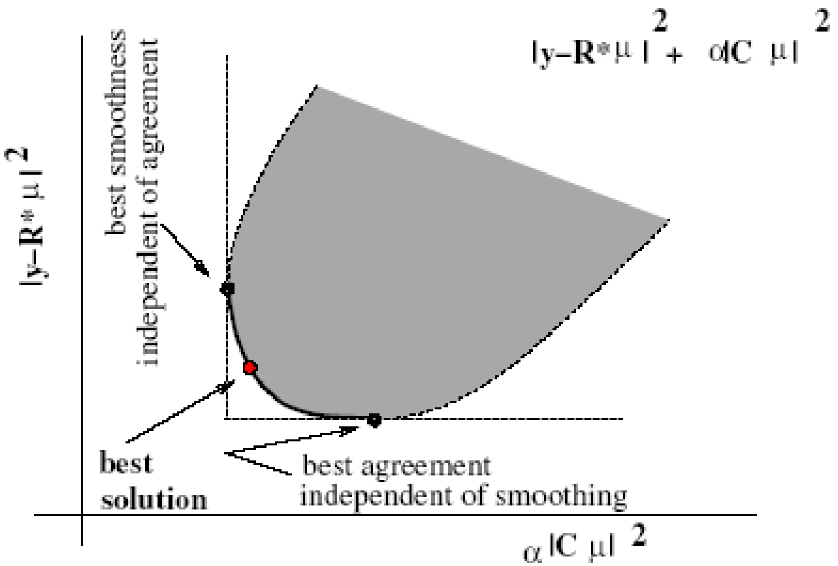
- Bayesian

$$\alpha = \frac{1}{\mu_T}$$

  usually too much smoothing

- $A = \chi^2 + \alpha||C\mu||^2 > 0$
  we can regularize the solution by choosing $\chi^2 \simeq DoF \equiv$ number of pixel N, with the condition:

$$\alpha = \frac{A - N}{||C\mu||^2}$$



Regularization parameter

The problem with fluctuations

From $\nu \to \mu$ deterministic methods can be used

$$\boldsymbol{\nu} = R\,\boldsymbol{\mu} \tag{97}$$

From $n \to \nu \to \mu$ statistical methods must be used

$$\boldsymbol{n} = \boldsymbol{\nu} + \boldsymbol{\rho} = R\,\boldsymbol{\mu} + \boldsymbol{\rho}$$

In the poissonian or binomial case we have to minimize:

$$-\ln L(\boldsymbol{\mu}) = -\sum_i \ln P(n_i, \nu_i)$$

In the gaussian case we must minimize

$$\chi^2(\boldsymbol{\mu}) = \sum_{ij} (\nu_i - n_i)(V^{-1})_{ij}(\nu_j - n_j)$$

These estimators are unbiased:

$$E[\hat{\mu}_j] = \sum_i (R^{-1})_{ji}\, E[n_i - \rho_i] = \sum_i (R^{-1})_{ji}\, \nu_i = \mu_j$$

In 2-D, when $N_{ij}(\mathrm{exp})$ contains fluctuations, we have to minimize:

$$\chi^2 = \sum_{ij} \frac{[N_{ij}(\mathrm{exp}) - N P_{ij}(\mathrm{obs})]^2}{N P_{ij}(\mathrm{obs})} \tag{98}$$

where

$$P(\mathrm{obs}) = P(\nu|\mu)\, P(\mu)$$

Fourier techniques

$$f(x) = \int F(t)\, e^{2\pi i\, xt}\, dt$$

Convolution:

$$f(x) = \int g(y)\delta(x-y)\, dy$$

$$\int F(t)\, e^{2\pi i xt}\, dt = \int G(t)\, e^{2\pi i\, yt}\, \Delta(t)\, e^{2\pi i\,(x-y)t}\, dt$$

$$\int F(t)\, e^{2\pi i\, xt}\, dt = \int G(t)\, \Delta(t)\, e^{2\pi i\, xt}\, dt \rightarrow F(t) = G(t)\, \Delta(t)$$

Correlation

$$\mathrm{Corr}(g,\delta) \equiv \int g(x+y)\, \delta(y)\, dy \rightarrow G(t)\Delta^*(t)$$

if the functions are real

$$G(t) = G(-t)^* \ , \quad \mathrm{Corr}(g,\delta) \rightarrow G(t)\Delta(-t)$$

Autocorrelation (Wiener theorem)

$$\mathrm{Corr}(g,g) \rightarrow |G(t)|^2$$

Total Power:

$$P(f) \equiv \int |f(x)|^2\, dx = \int |F(t)|^2\, dt$$

Power Spectral Density (in the Fourier space):

$$PSD(f) \equiv |F(t)|^2 + |F(-t)|^2 \quad \overset{\text{f(x) real}}{\rightarrow} 2|F(t)|^2 \quad 0 \leq t \leq \infty$$

Conclusions

- best fit minimization methods are crucial in physics. They are mainly frequentist
- they are based on the ML and LS algorithms (they are implemented in the ROOT-MINUIT framework)
- to judge the quality of the result, frequentists use the χ^2 test bayesians use the hypothesis probability
- Bayesian a priori hypotheses should be used with informative priors!!!

Image restoration

$$\mu_i \stackrel{PSF}{\rightarrow} \nu_i \stackrel{random}{\rightarrow} n_i$$

$$L(\boldsymbol{n}|\boldsymbol{\nu}, \boldsymbol{\mu}) \; P(\boldsymbol{\mu})$$

Bayesians say: posterior = likelihood × prior One maximizes $P(\text{true}_{ij}|\text{obs}) \equiv F(\boldsymbol{\mu})$ (or minimize $-F(\boldsymbol{\mu})$):

$$F(\boldsymbol{\mu}) = \ln L(\boldsymbol{n}|\boldsymbol{\nu}, \boldsymbol{\mu}) + \ln P(\boldsymbol{\mu}) \tag{21}$$

following the Maximum Likelihood (ML) principle. The practical (no Bayesian) experimentalist introduces an empirical regularization parameter α and considers the prior $P(\boldsymbol{\mu})$ as a regularization function $C(\boldsymbol{\mu})$:

$$F(\boldsymbol{\mu}) = \ln L(\boldsymbol{n}|\boldsymbol{\nu}, \boldsymbol{\mu}) + \alpha \, C(\boldsymbol{\mu}) \tag{22}$$

The present status of the Bayesian-Frequentist dispute:

- The The Bayesian: always choose o noninformative prior; reject the concept of the ensemble of identical experiments and give a probability to all the hypotheses.
- The (frequentists) physicists: usually I avoid to give any probability to my hypotheses: they are rejected by the experience (falsification). I use only informative priors into the analysis when certain *a priori information* is available.

2D Unfolding

A picture in a $x-y$ plane is the result of a double dimensional folding, where the true points are smeared out by detector effects.

$$N = \sum_{ij=1}^{n_c} N_{ij}(\mathrm{exp}) \ , \tag{93}$$

N is the total number of events and $N_{ij}(\mathrm{exp})$ is the recorded number of event in the pixel placed at the ith-row and jth-column.

The observed $N_{ij}(\mathrm{exp})$ events have to be compared with the expected values $N_{ij}(\mathrm{th})$ predicted by a model.

$$N_{ij}(\mathrm{th}) = N P_{ij}(\mathrm{obs}) = N \sum_{i'j'} P_{i'j'}(\mathrm{true})\, P_v(\mathrm{obs}_{ij}|\mathrm{true}_{i'j'}) \ , \tag{94}$$

that is, the number of events observed in the ijth-cell is due to the presence into the $i'j'$th-cell, times the probability P_v that the PSF shifts the point from the $i'j'$ to the ij-cell. One has to sum on all the cells near the $ij-$one.
In eq. (94) the normalization is understood. In practice, from (93, 94), one has

$$P(\mathrm{true}_{ij}|\mathrm{obs}_{i'j'}) = \frac{P_{i'j'}(\mathrm{true})\, P_v(\mathrm{obs}_{ij}|\mathrm{true}_{i'j'})}{\sum_{ij} \left[\sum_{i',j'} P_{i'j'}(\mathrm{true})\, P_v(\mathrm{obs}_{ij}|\mathrm{true}_{i'j'})\right]} \tag{95}$$

In the case of a two dimensional Gaussian point spread function PSF:

$$P_v(\mathrm{obs}_{ij}|\mathrm{true}_{i'j'}) = \frac{1}{2\pi\, \sigma_x \sigma_y} \exp\left[-\frac{(x_{ij} - x_{i'j'})^2}{2\,\sigma_x^2} - \frac{(y_{ij} - y_{i'j'})^2}{2\,\sigma_y^2} \right] , \tag{96}$$

In summary, we use

$$\mu_{k+1} = \mu_k + \beta[n - R * \mu_k] \quad (28)$$

The iterative principle

with the initial condition

$$\mu_0 = n \ .$$

The convergence is assured if

$$|I - \beta R| < 1$$

Since $|1 - \beta x| < 1$ implies $0 < \beta < 2/x$, in the case of the operator R, which can be transformed in a square matrix

$$R' = (R * \mu)\, \mu^{-1}$$

we obtain the condition:

$$0 < \beta < \frac{2}{\text{max eigenvalue of R}'}$$

Note that we works always with square matrices $R * \mu$, μ, μ^{-1} and R'.

However, this step must be repeated at each iteration

This method sometimes gives spectacular result!
However, often it gives irregular solutions.

Stopping Rules

- χ^2 variation

$$\chi_k^2 = ||y - R * \mu_k||^2 = \sum_i^{M \times N} \frac{(y_i - \sum_{ij} R_{ij}\mu_{j(k)})^2}{\sum_{ij} R_{ij}\mu_{j(k)}}$$

$$\frac{\chi_k^2 - \chi_{k-1}^2}{\chi_{k-1}^2} < 10^{-6}$$

If the regularization is good, one has $\chi_k^2 \simeq$ DoF.

- Signal to noise ratio (usually measured in decibel)

$$SNR = 10 \log_{10} \left[\frac{\sum_i (\mu_i - y_i)^2}{\sum_i (\mu_i - \mu_{true\ i})^2} \right]$$

where μ_{true} is the true image. This quantity is used in the MC simulations during when the true image is known.

- convergence of the solution

$$\frac{||\mu_k - \mu_{k-1}||^2}{||\mu_{k-1}||^2} < 10^{-6}$$

Unfolding techniques