The Unfolding Problem in Data Analysis

V. Blobel

II. Inst. für Experimentalphysik, Hamburg

2000-05-26

ZEUS seminar

Web home page: `http://www.desy.de/~/blobel`

I. The unfolding problem

Measurements in High Energy Physics ... ... are distorted by detector effects:

- finite detector resolution: *smearing* of measured quantities
- non-linear response
- limited acceptance (trigger, cuts): loss of events, loss of particles

... and by physical effects:

- radiation of photons (ISR, FSR)
- Fermi motion within the nucleus
- fragmentation (partons $\rightarrow$ hadrons)

Folding $\quad g(y) = \int A(y,x) f(x) \, \mathrm{d}x \qquad$ Fredholm integral equation of first kind

HEP: Measured data have relatively large statistical errors. Response function $A(y,x)$ known only by set of Monte Carlo events (usually limited in statistics).

Possibilities for data analysis:

- either compared folded model ($g(y)$) with uncorrected data distribution ,
- or correct/transform ($f(x)$) the data distribution.

The transformation is called *Unfolding*.

Discretization

Integral equation converted to matrix equation by ansatz $f(x) = \sum_{j=1}^{n} f_j \cdot B_j(x)$

with basis functions $B_j(x)$ = B-splines of order 1 (or 4):

$$g(y) = \int A(y,x) f(x)\,\mathrm{d}x = \int A(y,x) \sum_{j=1}^{n} f_j \cdot B_j(x)\,\mathrm{d}x = \sum_{j=1}^{n} f_j \cdot \int A(y,x) B_j(x)\,\mathrm{d}x$$

$$= \sum_{j=1}^{n} f_j A_j(y)$$

Measured distribution "explained" by superposition of MC histograms $A_j(y)$.

$$g_i = \sum_{j=1}^{n} A_{ij} f_j \qquad \boldsymbol{g} = A\ \boldsymbol{f}\ .$$

Representation of:

- true distribution $f(x)$ by histogram $\boldsymbol{f}$ with n bins
- measured distribution $g(y)$ by histogram $\boldsymbol{g}$ with m bins
- response function $A(y,x)$ by m-by-n matrix $\boldsymbol{A}$

Unfolding by inversion

Interpretation of response matrix in $\boldsymbol{g} = A\ \boldsymbol{f}$:

- A_{ij} = probability, to measure an event from true bin f_j in bin g_i,
- column $A_{\bullet j}$ is detector response to true element f_j,
- all elements $A_{ij} \geq 0$ by definition (probabilities).

Alternative: assumed shape of $f(x)$ included in definition of A; then only correction factors to assumed shape are to be determined.

Standard statistical method: inversion $\boldsymbol{f} = \boldsymbol{A}^{-1}\boldsymbol{g}$ (or ML/LSQ fit)
... has optimal statistical properties:

- unbiased, and
- efficient (smallest possible variance for an estimator with zero bias)

Correlations between bins are unavoidable *negative* and *large*; the result looks unacceptable. The problem are the statistical fluctuations in $\boldsymbol{g}$, which generate oscillating terms in $\boldsymbol{f}$.

Any reduction in variance can only be achieved by introducing a bias.

Art of Unfolding: construct biased estimator such that the bias will be small, if the assumptions (e.g. smoothness of solution) are in fact correct.

[Glen Cowan; Statistical Data Analysis. Oxford Science Publications, Clarendon, Oxford. 1998]

Unfolding by analytical formula

Special case: true distribution $f(x)$ periodic (period $= 2\pi$) with Gaussian x resolution function (standard deviation $= \sigma$)

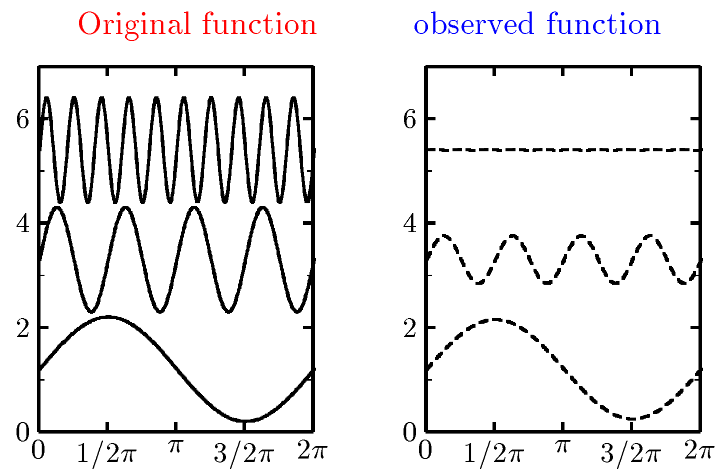
Note: each single Fourier coefficient is modified (damped), independent of all other coefficients!

Unfolding is very simple: the *measured* coefficients of the $\cos \nu x$ and $\sin \nu x$ terms are corrected by multiplication with factors

$$\exp\left(+\frac{\nu^2\sigma^2}{2}\right)$$

in order to get the original coefficients back.

Example for $\sigma = 0.1$



Conclusion: Strongly fluctuating higher order terms have to be ignored!

”A new method has been invented . . . ”

”A new unfolding method has been developed” based on . . .

Bin-by-Bin Correction, Unsmearing, Regularization, a Migration Matrix Technique, Bayes’ Theorem, Wavelet Method, Degree of Belief, Singular Value Decomposition, Intuition, Maximum Entropy, Feeling, . . .

I prefer methods, that

1. start from principle (e.g. Maximum Likelihood, Least Squares),
2. derive estimator from principle ($\rightarrow$ method),
3. derive statistical properties of estimator.

Convergence is well defined, if the method is iterative.

There are several unfolding methods, which consist of a

1. -
2. method of solution, usually iterative,
3. -.

Not derived from a principle. Iteration stopped, if corrections small. Their properties are studied experimentally (. . . the method works . . .). Often the method of solution is explained by *words*, not by a formula.

Note: Properties of a method also depend on details of implementation.

Two cases of unfolding problems

1. measured variable *almost* identical to true variable;
2. measured variable(s) *completely different* from true physics variable, even in dimension.

First case: folding has effect of *smearing*, making distribution smoother. Effect may be negligible or small, if true distribution itself constant or linear.

Traditional method of HEP: bin-by-bin correction

- find MC input distribution that describes (after folding) the measured distribution rather good;
- "correct" measured bin i by ratio

$$\frac{f_i \text{ (MC input)}}{g_i \text{ (MC output)}} ,$$

  even if migration is large;
- error handling: ratio is assumed to have negligible error.

Although the method has no mathematical basis (result must be known before, correlations are neglected), results obtained by this method may be not in contradiction to the true distribution, if MC assumption correct and true distribution smooth.

(but → Gedankenexperiment)

Second case

Measured variable(s) *completely different* from true physics variable, even in dimension.

Example: NC neutrino reactions in target calorimeter

$\nu N \to \nu X$ measure: hadronic energy E_X and r = radius of vertex .

Aim: measurement of distribution

$$\mathrm{d}\sigma/\mathrm{d}y \qquad \text{with} \qquad y = \frac{E_X}{E_\nu}$$

A certain bin in y corrsponds to a well-defined distribution (E_X, r). Unfolding using known Flux $\Phi(E_\nu, r)$ of narrow band beam:

measured variables: E_X, r → kinematical variable y .

II. The case $n = m = 2$ and the problem of correlations

Only $n = 2$ possible true values and $m = 2$ possible observed values.

Example: classification of events into NC (neutral current) and CC (charged current).

$$\begin{pmatrix} NC \\ CC \end{pmatrix}_{\text{obs}} = \begin{pmatrix} NC \to NC & CC \to NC \\ NC \to CC & CC \to CC \end{pmatrix} \cdot \begin{pmatrix} NC \\ CC \end{pmatrix}_{\text{true}}$$

or short $\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f}$ with matrix $\boldsymbol{A}$ containing *positive* probabilities.

Unfolding by matrix inversion (*what else?*):

$$\boxed{\begin{pmatrix} NC \\ CC \end{pmatrix}_{\text{true}} = \boldsymbol{f} \quad = \quad \boldsymbol{A}^{-1}\boldsymbol{g} = \boldsymbol{A}^{-1} \begin{pmatrix} NC \\ CC \end{pmatrix}_{\text{obs}}}$$

i.e. a linear transformation from the observed values to the true values ($\to$ straightforward error propagation).

Unfolding by inversion

Quantitative example: ε = probability for migration from original class into other class.

$$\begin{pmatrix} g_1 \\ g_2 \end{pmatrix}_{\text{obs}} = \begin{pmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{pmatrix} \cdot \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}_{\text{true}}$$

or short: $\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f}$.

Unfolding with error propagation:

$$\boxed{\begin{aligned} \boldsymbol{f} &= \boldsymbol{A}^{-1}\boldsymbol{g} \\ \boldsymbol{V}_f &= \boldsymbol{A}^{-1}\,\boldsymbol{V}_g\,\boldsymbol{A}^{-1} \end{aligned}}$$

Inverse matrix (with *negative* off-diagonal elements):

$$\boldsymbol{A}^{-1} = \frac{1}{1-2\varepsilon} \begin{pmatrix} 1-\varepsilon & -\varepsilon \\ -\varepsilon & 1-\varepsilon \end{pmatrix}$$

Resulting covariance matrix:

$$\boldsymbol{V}_x = \frac{1}{(1-2\varepsilon)^2} \begin{pmatrix} (1-\varepsilon)^2 g_1 + \varepsilon^2 g_2 & -\varepsilon(1-\varepsilon)(g_1+g_2) \\ -\varepsilon(1-\varepsilon)(g_1+g_2) & (1-\varepsilon)^2 g_2 + \varepsilon^2 g_1 \end{pmatrix}$$

(… becomes singular in case of $\varepsilon = 0.5$)

A numerical example

Measurement and covariance matrix are:

$$g_1 = 100 \pm 10 \qquad g_2 = 81 \pm 9 \qquad V_g = \begin{pmatrix} 100 & 0 \\ 0 & 81 \end{pmatrix}$$

Result of unfolding by inversion:

ε	$\hat{f}_1$	$\hat{f}_2$	ρ
0	100 ± 10	81 ± 9	0
0.1	102.4 ± 11.3	78.6 ± 10.2	-0.22
0.4	138 ± 35	43 ± 33	-0.92
0.5	$\pm\infty$	$\pm\infty$	singular

Statistical errors magnified because of finite resolution (migration probability ε)!

Effect in detail for $\varepsilon = 0.4$:

$$A^{-1} = \frac{1}{1-2\varepsilon} \begin{pmatrix} 1-\varepsilon & -\varepsilon \\ -\varepsilon & 1-\varepsilon \end{pmatrix} = \begin{pmatrix} 3 & -2 \\ -2 & 3 \end{pmatrix}$$

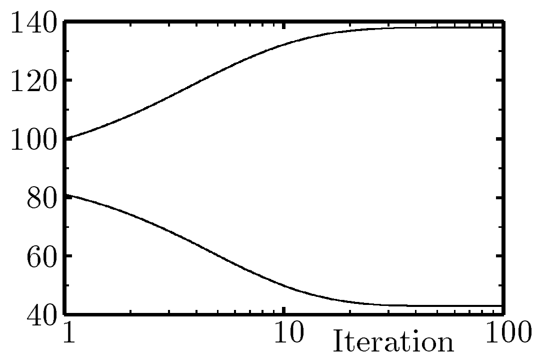
for example $\quad \hat{f}_1 = 3 \cdot g_1 - 2 \cdot g_2 = 300 - 162 = 138$

with $\quad \sigma = \sqrt{3^2 \cdot 100 + 2^2 \cdot 81} = 35 \qquad (\gg 10)$

... the bin-by-bin correction method

What about the bin-by-bin acceptance correction method in this case?

Parameter f_1 and f_2 versus iteration number (for $\varepsilon = 0.4$):



The method converges (of course) to the *same* values as the inversion method, but extremely slow (linear convergence).

$$f_1 = 138 \pm 13.8 \qquad f_2 = 43 \pm 4.8$$

The errors are too small (correlation by finite resolution is ignored).

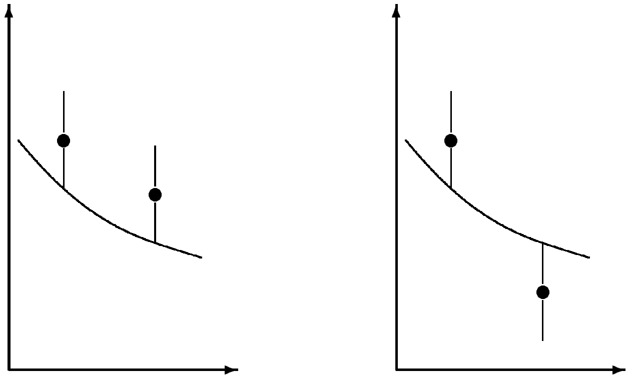
A general comment on *linear* convergent methods: $x_k \to x_{k+1}$

ratio $\quad \dfrac{F(x_{k+1}) - F(x^*)}{F(x_k) - F(x^*)} \approx \left(\dfrac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\max} + \lambda_{\min}}\right)^2 \quad$ often close to 1

Numerical Example for deviations

$\sigma_1 = \sigma_2 = \sigma$ and two deviations $\Delta = \pm 1\,\sigma$

$$\chi^2 = \begin{cases} 2/(1+\rho) & \text{same sign of } \Delta \\ 2/(1-\rho) & \text{opposite sign of } \Delta \end{cases}$$



Probability from χ^2

	same sign	opposite sign
$\rho = 0$	37 %	37 %
$\rho = +0.9$	59 %	0.0005 %
$\rho = -0.9$	0.0005 %	59 %

Conclusion: Difficult interpretation of *common* deviations of > 1 data point in case of *correlated* ($\rho \neq 0$) data points.

Correlations between data points

Case of two *correlated* data points:

$$\text{Covariance matrix} \quad \boldsymbol{V} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

$$\begin{aligned} \sigma_{1,2} &= \text{standard deviations of the variables} \\ \rho &= \text{correlation coefficient} \qquad -1 < \rho < +1 \end{aligned}$$

The χ^2 value for a deviation $\boldsymbol{\Delta}$ from a theroretical prediction is given by:

$$\boxed{\chi^2 = \boldsymbol{\Delta}^T \left(\boldsymbol{V}^{-1}\right) \boldsymbol{\Delta}}$$

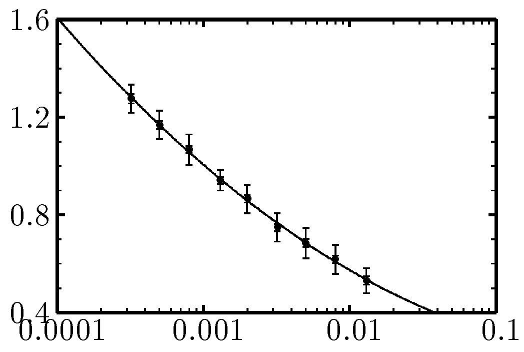
with the inverse of the covariance matrix

$$\boldsymbol{V}^{-1} = \frac{1}{(1-\rho^2)} \begin{pmatrix} 1/\sigma_1^2 & -\rho/\sigma_1\sigma_2 \\ -\rho/\sigma_1\sigma_2 & 1/\sigma_2^2 \end{pmatrix}$$

Example from a publication

What happens, if many "data" points are generated from a low-rank data set?

Nine strongly (+) correlated data points, while effective rank is only two ... three:



Fit of parabola (3 parameters) to the data shows: **either**

- errors to big (by factor > 5), **or**
- fluctuations too small.

- Shape of MC assumption copied to the "data"!

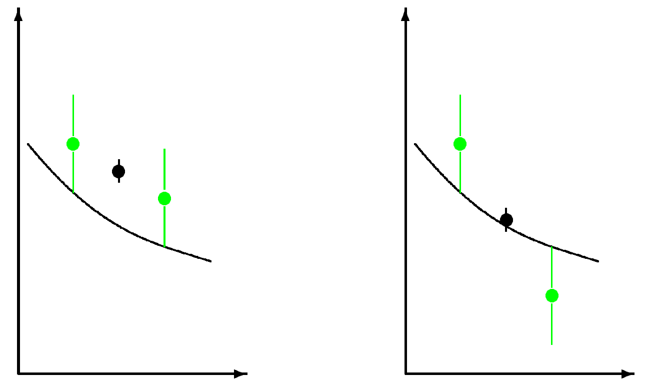
Numerical Example - averaging of two data points

$\sigma_1 = \sigma_2 = \sigma$ and correlation coefficient ρ

Standard deviation of averaged data point:

	$\sqrt{\frac{1+\rho}{2}}\,\sigma$	comment
$\rho = 0$	$0.707\ \sigma$	$1/\sqrt{2}$
$\rho = +0.9$	$0.975\ \sigma$	same precison as single point
$\rho = -0.9$	$0.224\ \sigma$	much more accurate

Averaging in case of negatively correlated data points ($\rho = -0.9$):



Conclusion:

- Averaging of data points with large negative correlations results in very accurate values.
- Averaging of data points with large positive correlation reduces the number of data points without increasing their accuracy.

III. Linear Least Squares Methods

Unfolding means solution of a system of m linear equations with n unknowns.

$$\boldsymbol{A}\boldsymbol{x} \cong \boldsymbol{b} \qquad \text{equivalent to} \qquad \left(\boldsymbol{A}^T\boldsymbol{A}\right)\boldsymbol{x} = \left(\boldsymbol{A}^T\boldsymbol{b}\right)$$

System may be ill-conditioned. Standard techniques for solution do not allow to recognize the problems $\rightarrow$ orthogonalization.

Number of measured events in a bin follows the Poisson distribution. Here a high-statistics measurement is assumed, where the Poisson distribution can be approximated by a Gaussian distribution. Each equation can be divided by the experimental standard deviation.

(1) Least square solution by singular value decomposition (SVD)

$$\text{SVD of } \boldsymbol{A}: \quad \boldsymbol{A}_{m\times n} = \boldsymbol{U}_{m\times n}\,\boldsymbol{S}_{n\times n}\,\boldsymbol{V}^T_{n\times n} \qquad \text{or} \qquad \boldsymbol{S} = \boldsymbol{U}^T\,\boldsymbol{A}\,\boldsymbol{V}$$

where matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal: $\boldsymbol{U}^T\boldsymbol{U} = \mathbf{1}$, $\boldsymbol{V}^T\boldsymbol{V} = \mathbf{1}$ and $\boldsymbol{V}\boldsymbol{V}^T = \mathbf{1}$.

$$\boldsymbol{S} \quad \text{is diagonal matrix of singular values } s_i \; (\geq 0): \quad \boldsymbol{S} = \begin{pmatrix} s_1 & 0 & & 0 \\ 0 & s_2 & & 0 \\ & & \cdots & \\ 0 & 0 & & s_n \end{pmatrix}$$

Singular value decomposition - SVD

If rank k of matrix $\boldsymbol{A}$ is $< n$: some of the singular values s_i vanish.

This means: contributions to $\boldsymbol{x}$ from a $(n-k)$-dimensional subspace (null space) have no effect on r.h.s. of $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. For a given $\boldsymbol{b}$ the solution $\boldsymbol{x}$ is arbitrary: arbitrary contributions to $\boldsymbol{x}$ from the nullspace are possible.

Strategy: force zero contribution from null space.

$$\text{Solution } \boldsymbol{x}: \quad \boldsymbol{x} = \boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{U}^T\boldsymbol{b}$$

Matrix $\boldsymbol{S}^{-1}$ is diagonal matrix with $1/s_i$ on diagonal, or zero, if $s_i = 0$ (i.e. replace ∞ by 0).

Strategy: treat small values of s_i like zero.

(introduces a (small) bias, but is more accurate).

Vector $\boldsymbol{x}$ is linear combination of columns of $\boldsymbol{V}$: ($\boldsymbol{V}_{(i)}$ is column vector of $\boldsymbol{V}$)

$$\boldsymbol{x} = \sum_{i=1}^{m} \left(\frac{\boldsymbol{U}_{(i)}\cdot\boldsymbol{b}}{s_i}\right)\cdot\boldsymbol{V}_{(i)} = \sum_{i=1}^{m} \frac{c_i}{s_i}\cdot\boldsymbol{V}_{(i)}$$

$$\text{Covariance matrix of vector } \boldsymbol{x}: \quad \boldsymbol{V}[\boldsymbol{x}] = \boldsymbol{V}\boldsymbol{S}^{-2}\boldsymbol{V}^T$$

Coefficients c_i are statistically independent: $\boldsymbol{V}[\boldsymbol{c}] = \mathbf{1}$.
This allows definition of significance: c_i follows Gaussian distribution $N(0,1)$ for insignificant coefficients.

A numerical example cont.

$$\text{Matrix } \boldsymbol{A} \quad \begin{pmatrix} 0.6 & 0.2 & 0 & 0 & 0 & 0 & \\ 0.4 & 0.6 & 0.2 & 0 & 0 & 0 & \\ 0 & 0.2 & 0.6 & 0.2 & 0 & 0 & \\ 0 & 0 & 0.2 & 0.6 & 0.2 & 0 & \\ 0 & 0 & 0 & 0.2 & 0.6 & 0.2 & \\ & & & & & \ddots & \end{pmatrix}$$

Measured histogram is sum of the two histograms. Unfolding of upper histograms has exact input distribution as result. Unfolding of lower histogram (errors) gives strongly fluctuating histogram.

Table of singular values and coefficients

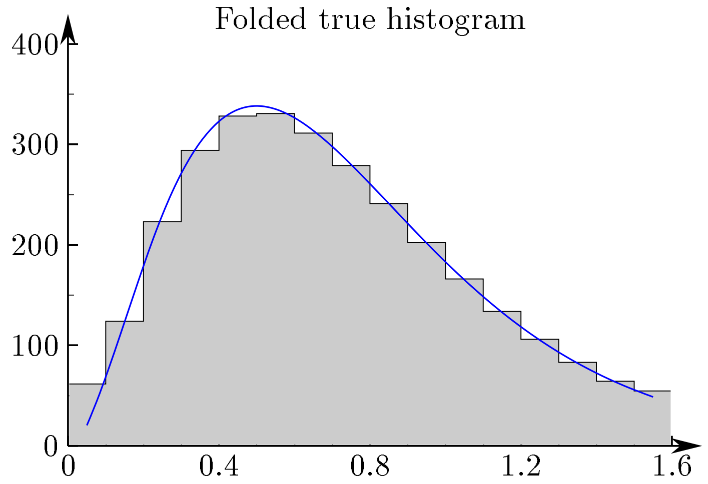
i	s_i	c_i	i	s_i	c_i
1	1.0066	−694.3	9	0.5586	7.6
2	1.0020	194.6	10	0.4757	−6.6
3	0.9733	397.2	11	0.3985	1.9
4	0.9306	−168.7	12	0.3302	−3.5
5	0.8735	93.1	13	0.2739	0.2
6	0.8046	−41.9	14	0.2319	−2.2
7	0.7268	25.0	15	0.1985	1.5
8	0.6436	−14.6	16	0.2059	−0.2

Following figures show unfolding results with different numbers of singular values set to zero.

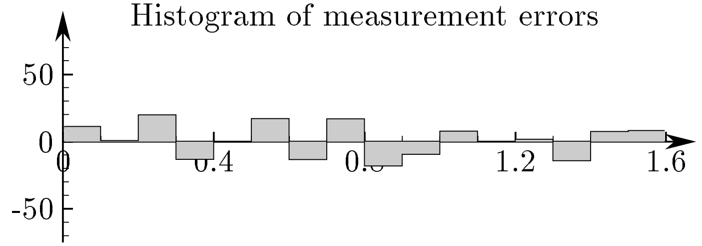
A numerical example

Histogram of 16 bins with square matrix $\boldsymbol{A}$.

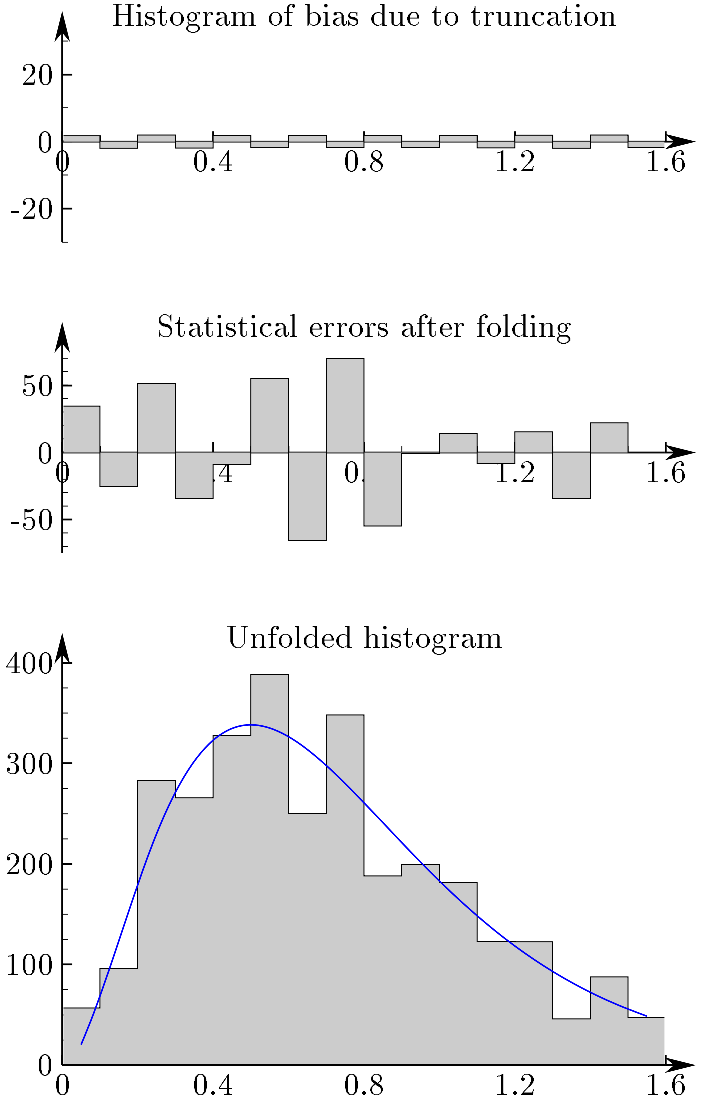
Original distribution (curve) and folded histogram (3000 entries) without statistical errors.



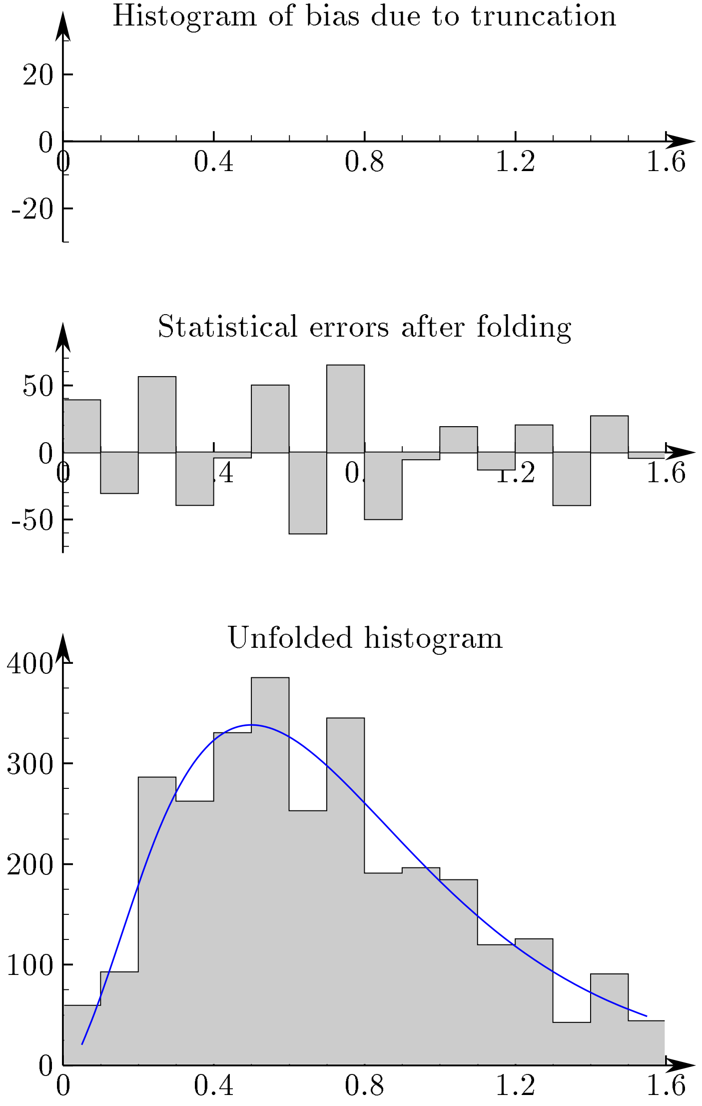
Statistical errors, from a sample of a Poisson distribution:



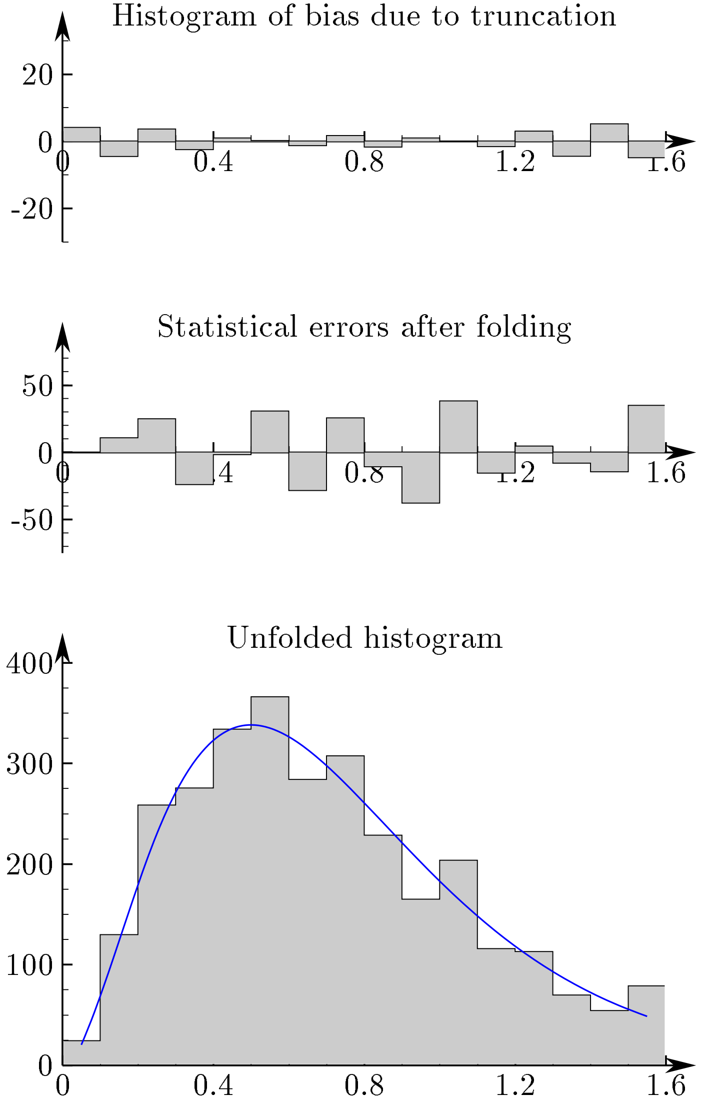
One singular value set to zero



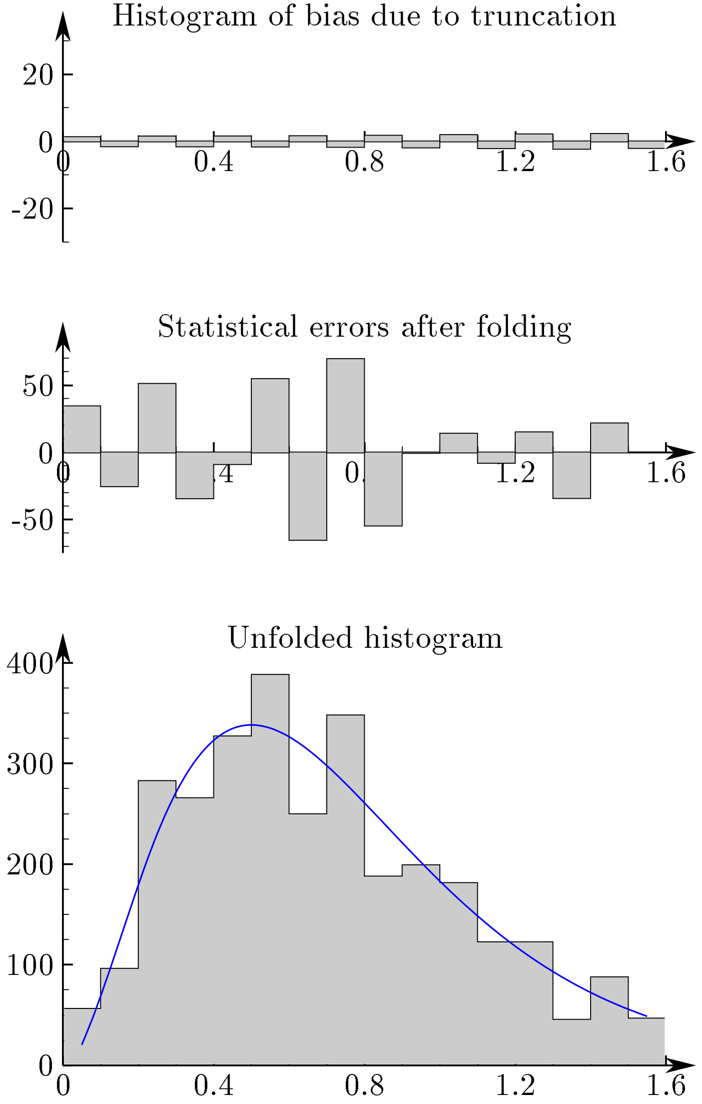
No singular value set to zero



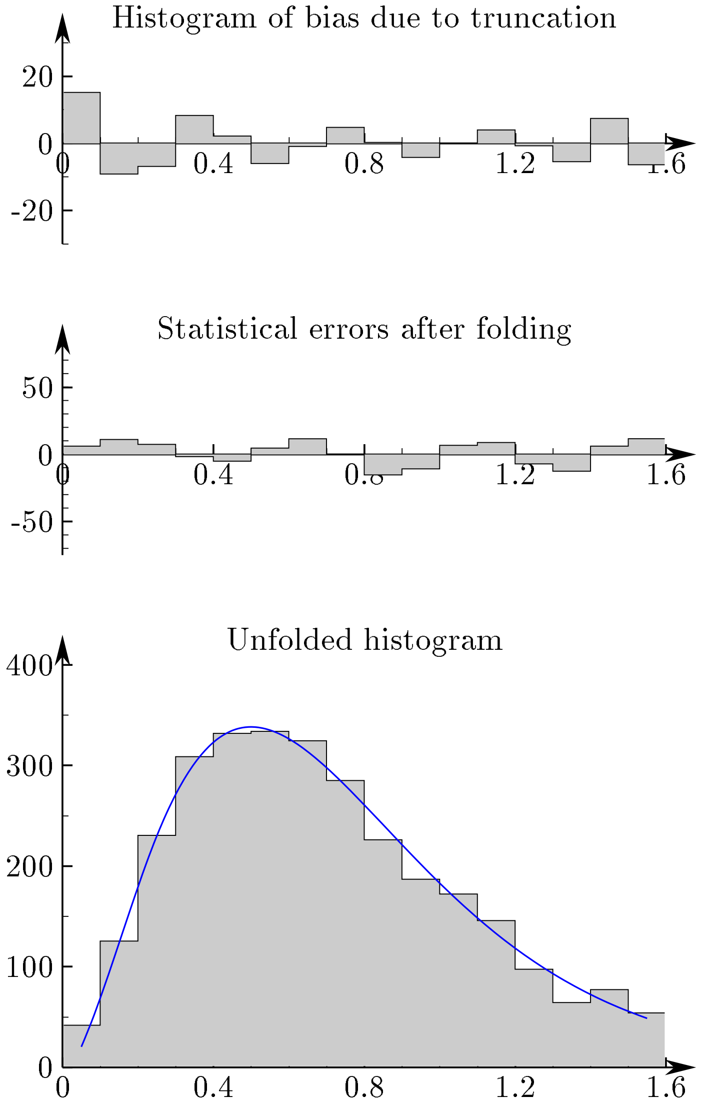
Three singular values set to zero



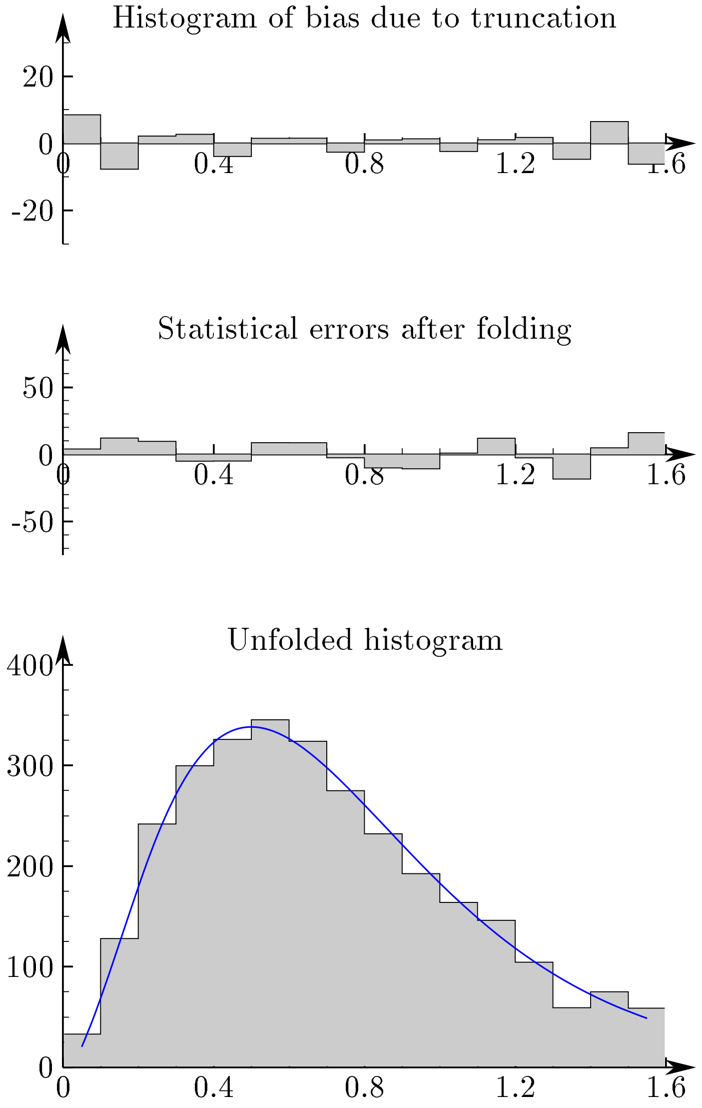
Two singular values set to zero



8 singular values set to zero



6 singular values set to zero



Rank of covariance matrix

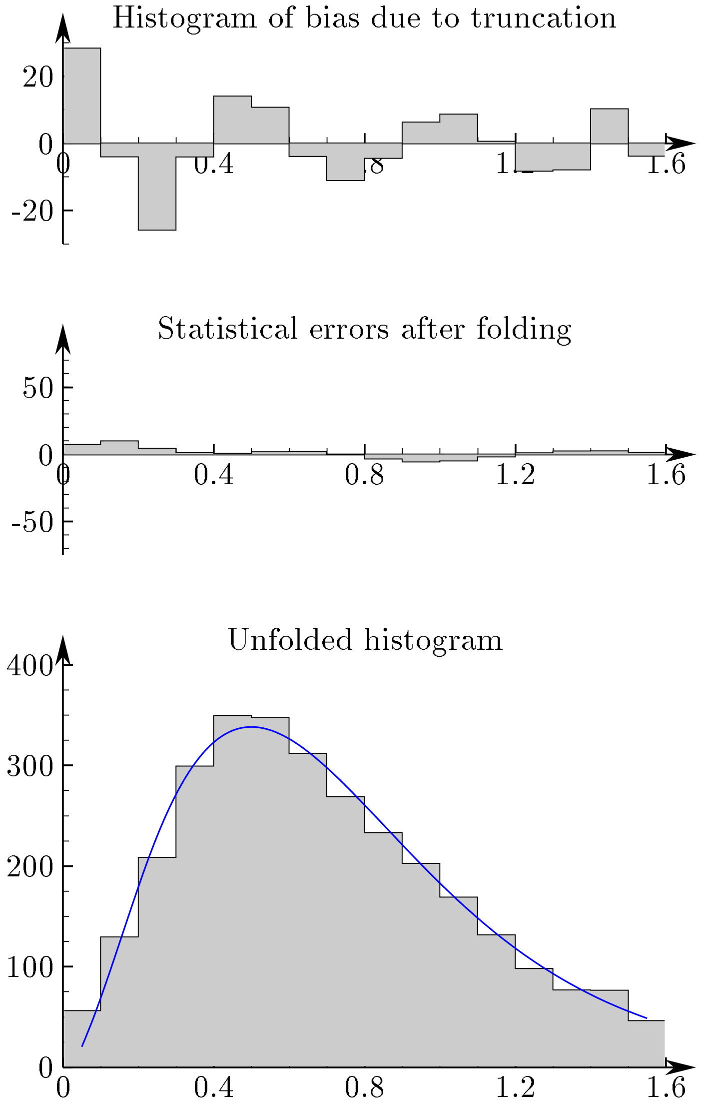
The rank k of the problem (and of the covariance matrix) is reduced, if singular values are set to zero. This introduces positive correlations, which fake an accurate result.

My opinion: combine bins to get same number of data points and non-zero singular values (then the covariance matrix has full rank).

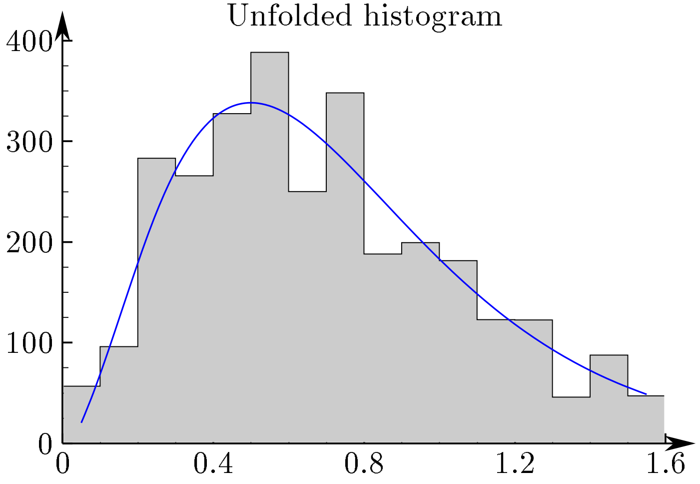
The next figures show the unfolded histograms with 16 bins and, by combining two adjacent bins, with 8 bins.

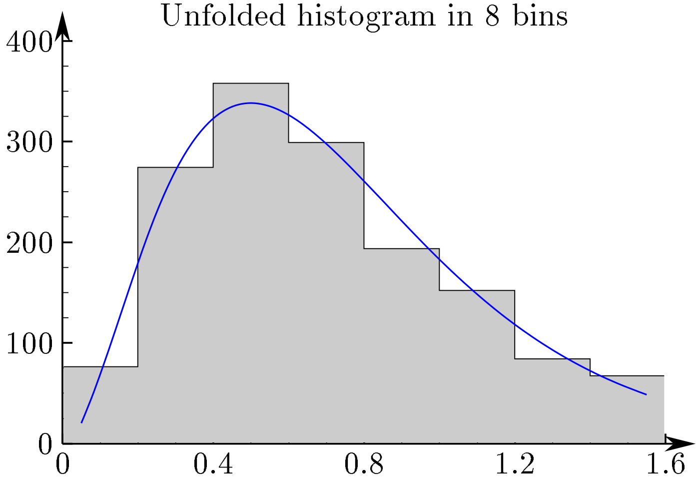
Optimal are $16 - 6 = 10$ non-zero singular values and 10 data points.

10 singular values set to zero

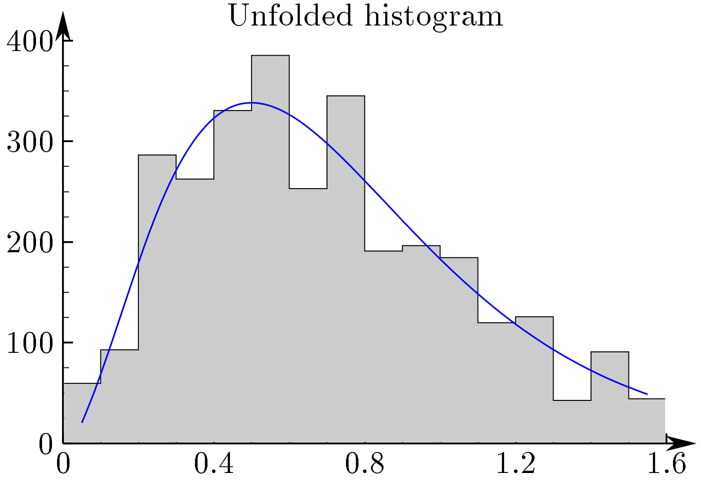


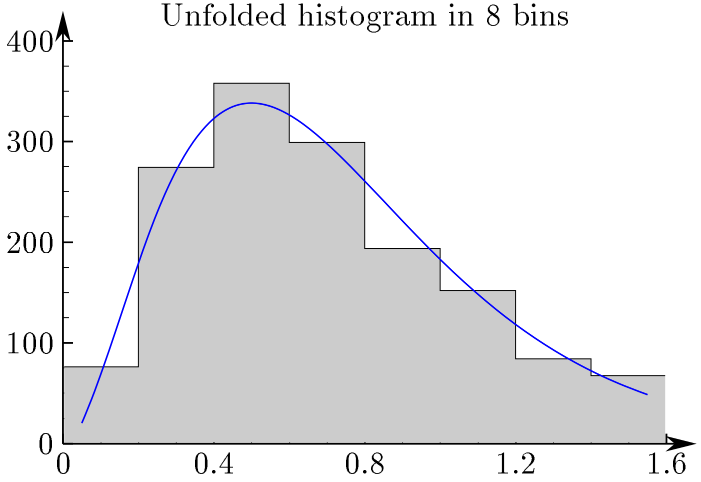
One singular value set to zero



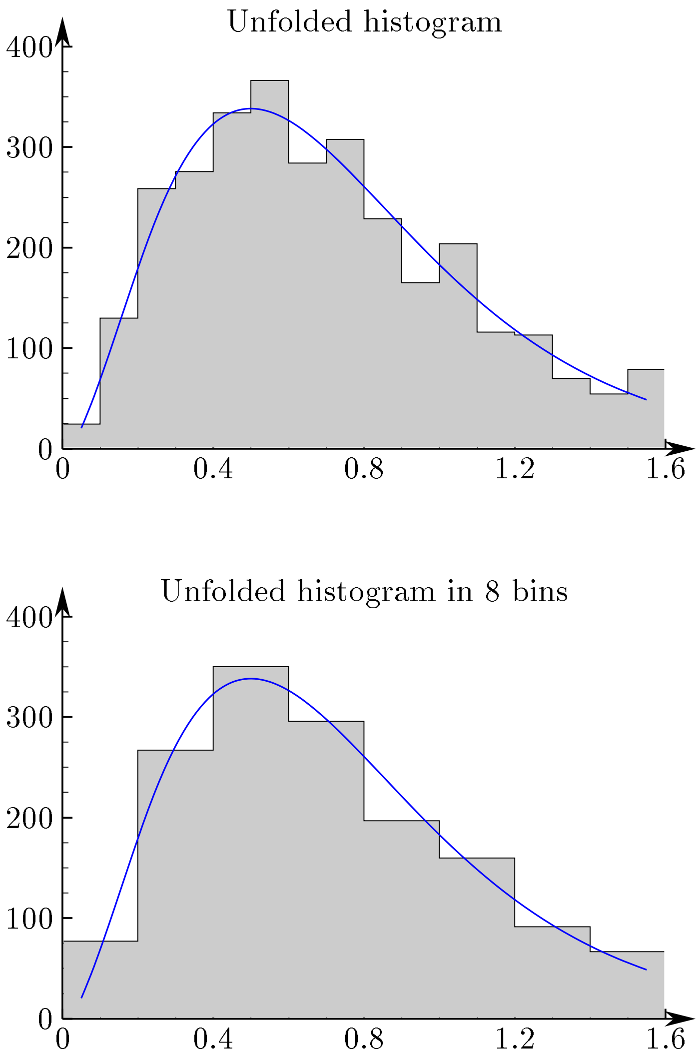


No singular value set to zero

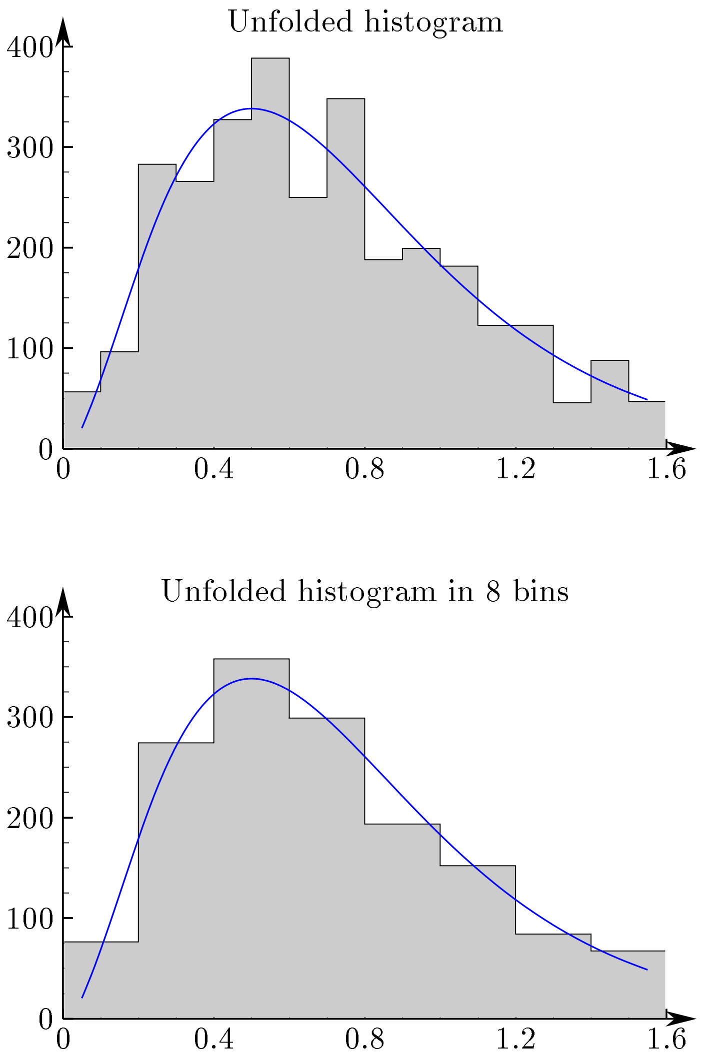




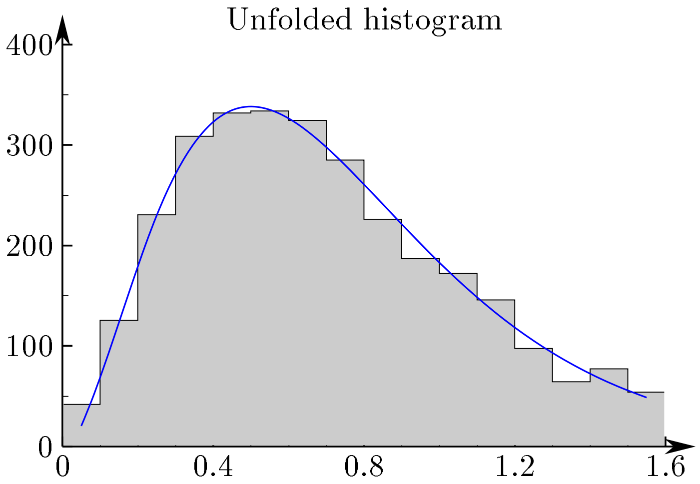
Three singular value set to zero

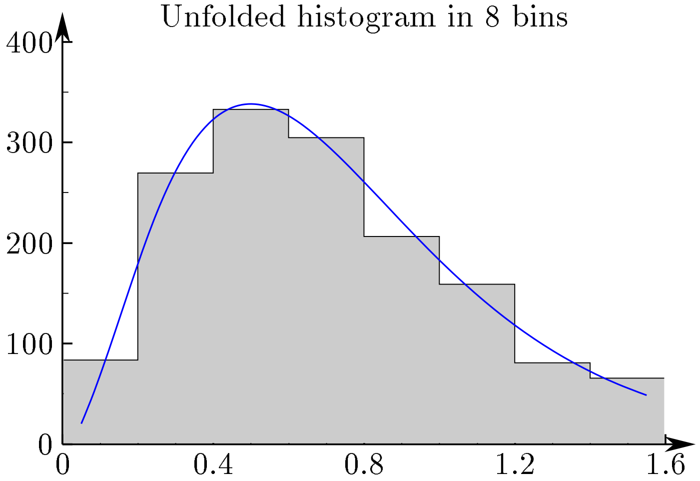


Two singular values set to zero

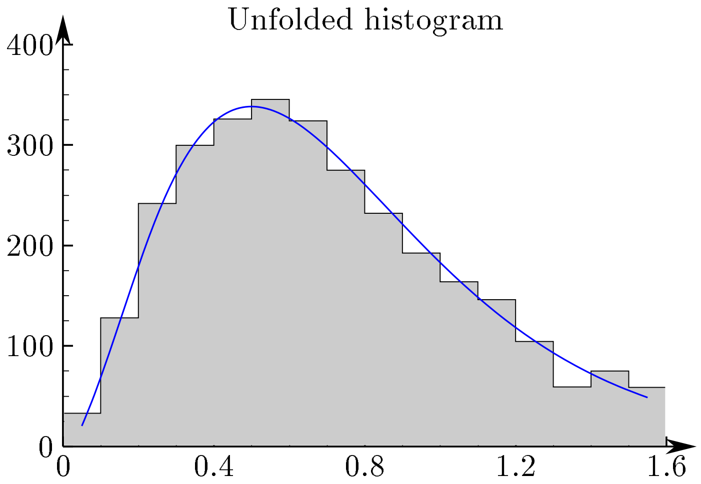


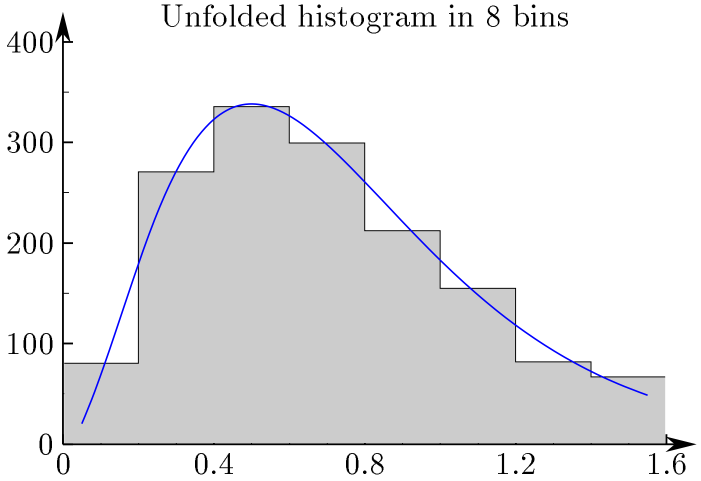
8 singular values set to zero



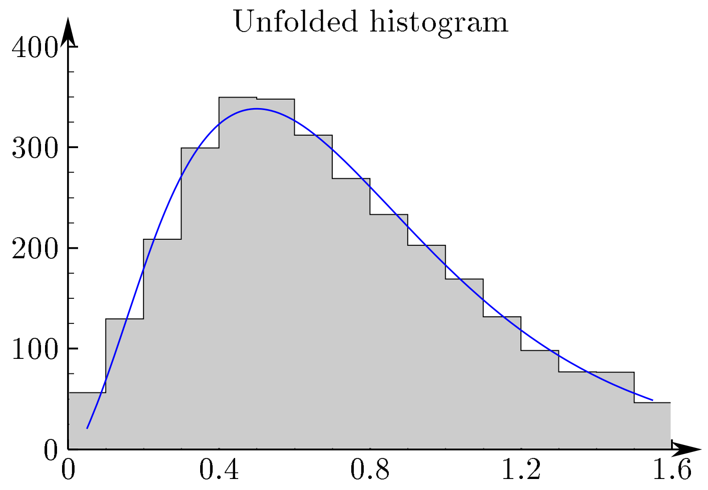


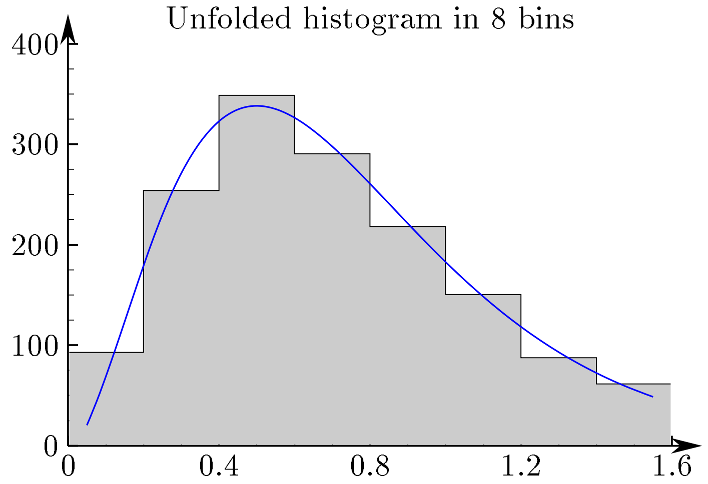
6 singular values set to zero





10 singular values set to zero





Normal equations

(2) Least square solution by normal equations (with diagonalization)

minimize $\|\boldsymbol{Ax}-\boldsymbol{b}\|^2$

$= F(\boldsymbol{x}) = (\boldsymbol{Ax}-\boldsymbol{b})^T(\boldsymbol{Ax}-\boldsymbol{b}) = \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} - 2\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{b} + \boldsymbol{b}^T\boldsymbol{b}$ quadratic in $\boldsymbol{x}$

Minimum determined by

$$\text{grad}\, F(\boldsymbol{x}) = 0 \quad \rightarrow \quad \underbrace{\left(\boldsymbol{A}^T\boldsymbol{A}\right)}_{\boldsymbol{C}}\boldsymbol{x} = \boldsymbol{A}^T\boldsymbol{b} \quad \text{linear in } \boldsymbol{x}$$

Diagonalization of the symmetric matrix $\boldsymbol{C}$: $\quad \boldsymbol{C} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^T \quad$ or $\quad \boldsymbol{D} = \boldsymbol{U}^T\boldsymbol{C}\boldsymbol{U}$

$\boldsymbol{D}$ is diagonal matrix of eigenvalues $\lambda_i\ (\geq 0)$: $\quad \boldsymbol{D} = \begin{pmatrix} \lambda_1 & 0 & & 0 \\ 0 & \lambda_2 & & 0 \\ & & \ddots & \\ 0 & 0 & & \lambda_n \end{pmatrix}$

Normal equations cont.

Relation to SVD: $\lambda_i = s_i^2$
SVD numerically more stable, but smaller matrices with normal equations. Same strategy possible for zero or small eigenvalues λ_i.

$$\text{Solution } \boldsymbol{x}: \quad \boldsymbol{x} = \boldsymbol{U}\boldsymbol{D}^{-1}\boldsymbol{U}^T\left(\boldsymbol{A}^T\boldsymbol{b}\right)$$

Vector $\boldsymbol{x}$ is linear combination of columns of $\boldsymbol{U}$: ($\boldsymbol{U}_{(i)}$ is column vector of $\boldsymbol{U}$)

$$\boldsymbol{x} = \sum_{i=1}^{n}\left(\frac{\boldsymbol{U}_{(i)}\cdot\left(\boldsymbol{A}^T\boldsymbol{b}\right)}{\lambda_i}\right)\cdot\boldsymbol{U}_{(i)} = \sum_{i=1}^{n}\frac{c_i}{\sqrt{\lambda_i}}\cdot\boldsymbol{U}_{(i)}$$

$$\text{Covariance matrix of vector } \boldsymbol{x}: \quad \boldsymbol{V}[\boldsymbol{x}] = \boldsymbol{C}^{-1} = \boldsymbol{U}\boldsymbol{D}^{-1}\boldsymbol{U}^T$$

Coefficients c_i are statistically independent: $\boldsymbol{V}[\boldsymbol{c}] = \mathbf{1}$

Solution by normal equations (square matrices) is equivalent to SVD.

IV. Regularisation methods

Usually number m of bins of measured histogram much larger than number n of final unfolded data points: $m \gg n$.

Statistical limitation of events (and often of MC simulation) requires iterative Maximum Likelihood fit, based on Poisson statistic.

$\rightarrow$ SVD can not be used directly, but diagonalization of Hesse matrix in iterations possible.

If MC simulation statistics limited, method of R. Barlow can be used to take into account Poisson statistic of MC data (this introduces many additional parameters, but is still feasible).

A general method

Regularization methods introduce a smooth cut-off instead of a sharp cut-off in the number of coefficients.

Use $m > n$ and start from standard solution: in first iteration χ^2 minimization with

$$(\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f})^T\ \boldsymbol{W}\ (\boldsymbol{g} - \boldsymbol{A}\boldsymbol{f}) = \text{minimum}$$

and in later iterations binned maximum likelihood (Poisson statistic in each bin)

$$-2\ln\mathcal{L}(\boldsymbol{f}) = \text{minimum}$$

in order to allow case of few (even zero) events in bins.

Add regularisation term $\tau \cdot R(\boldsymbol{f})$, e.g. curvature

$$R(\boldsymbol{f}) = \int |f''(x)|^2 \, \mathrm{d}x$$

with regularisation parameter τ of unknown magnitude:

- τ small: no effect of regularisation
- τ large: resulting function will be **linear** → bias.

A different regularisation term is based on the "Maximum Entropy" principle.

Practical problems in unfolding

Knowledge of resolution function $A(y, x)$ essential.

The most important practical problem in unfolding is **control** of the resolution function! There are potential ...

- problems in the event generation,
- problems in the detector simulation,
- problems in the measurement (data taking, reconstruction),

and usually MC statistics is too small.

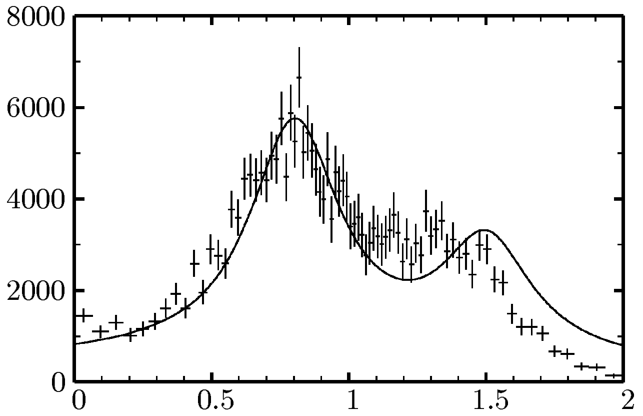
→ Unfolding procedure has to include back-projection of the result (included in RUN).

Any discrepancy has to be understood. Understanding has to be used to improve calculation.

Numerical example

Curve = test function (Sum $\sum_i \mathrm{BW}_i$). Histogram of 10000 generated events:

- Gaussian resolution function with $\sigma = 0.1$
- nichtlinear detector response
- reduced acceptance at both ends



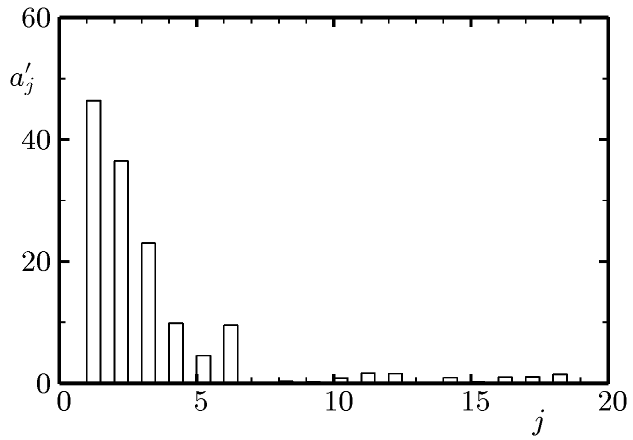
(Width of Histogram bins smaller in the center)

Minimum between the two peaks is filled up!

Amplitudes

$$f(x) = \sum_j a_j\, B_j(x) = \sum_j a'_j P_j(x)$$

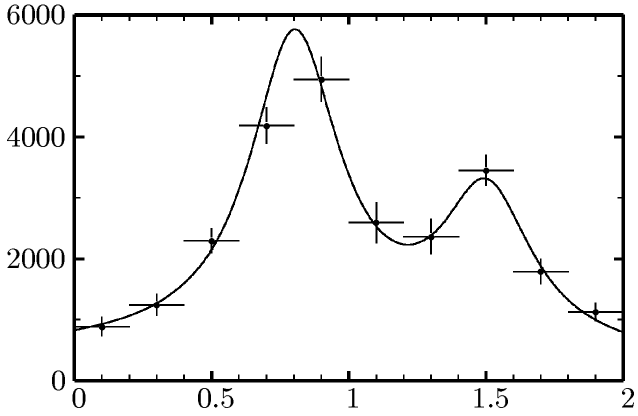
$$\begin{aligned} B_j(x) &= \text{B-Spline, } \neq 0 \text{ only in small area} \\ P_j(x) &= \text{orthonormal linear combinations} \end{aligned}$$



- Coefficients a'_j have variances of 1 and are independent!

Result of unfolding

- 10 data points by integration over interval (only 10 data points)
- comparison with input curve



- Minimum between two peaks *visible* again

V. Conclusions

- Unfolding is necessary for many measurements in HEP experiments
- Unfolding is a difficult statistical and mathematical problem
- It requires a detailed understanding of the detector response
- Orthogonalization provides insight into the potential sensitivity of unfolding results
- Limited resolution necessarily reduces the number of data points, which can be reconstructed
- data points should have only small correlations – otherwise comparison to theoretical expectation difficult