PHYSTAT2011 – Workshop on Unfolding

Thursday Jan 20th 2011

Volker Blobel – Universität Hamburg

Unfolding methods for particle physics

1. **Unfolding – direct and inverse process**
2. **Naive unfolding**
3. **Convolution/deconvolution**
4. **Orthogonalization**
5. **Truncation and correlations**
6. **Regularization**
7. **DCT and projection methods**
8. **Iterative unfolding**

\+ **Appendix**

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

---

1. Unfolding – direct and inverse processes

The process of the transition between the **true distribution** $f(t)$ and the **measured distribution** $g(s)$ for linear inverse problems is described by the Fredholm integral equation of the first kind:

$$\int_\Omega K(s,t)\, f(t)\, \mathrm{d}t + b(s) = g(s)$$

($b(s)$ = background contribution). Two types of processes are based on the integral equation:

direct process (MC)	true/MC dist.	$f(t)$	$\Longrightarrow$	$g(s)$ measured dist.
inverse process (unfolding)	measured dist.	$g(s)$	$\Longrightarrow$	$f(t)$ true dist.

Discretization: the integral equation becomes an (usually ill-posed) linear system of equations:

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y} \qquad\qquad y_i = \int_{s_{i-1}}^{s_i} g(s)\, \mathrm{d}s \quad i = 1,\, 2, \ldots,\, m$$

(assuming a case without background contribution) with the representation

true distribution $f(t) \Rightarrow \boldsymbol{x}$	n-vector of unknowns
measured distibution $g(s) \Rightarrow \boldsymbol{y}$	m-vector of measured data
Kernel $K(s,t) \Rightarrow \boldsymbol{A}$	rectangular m-by-n response matrix .

The variables s, t and vectors $\boldsymbol{x}$, $\boldsymbol{y}$ can be multi-dimensional. Elements of the response matrix $\boldsymbol{A}$ are (positive) probabilities, and include efficiency.

Discretization

- The response matrix $\boldsymbol{A}$ has to be determined in particle physics by MC ;
- several different methods can be applied in the discretization: simple binning, quadrature methods (weighted sums), B-splines (allows re-weighting of MC) . . . ;
- unfolding is independent of the assumed $f(t)^{\mathrm{MC}}$, if the Fredholm equation above is correct,
- but for complex physical measurements this assumption may not be true, and the response function is influenced by the assumed $f(t)^{MC}$. In this case the Fredholm integral equation should be rewritten in the form,

$$\int_{\Omega} K(s,t;\, f)\; f(t)\,\mathrm{d}t + b(s) = g(s)\;,$$

which represents a more difficult *nonlinear* inverse problem; $f(t)^{\mathrm{MC}}$ should be close to the expected result.

Problems in Particle Physics differ from problems in other fields:

- input erros are well-known (Poisson data, . . . covariance matrix) $\boldsymbol{V}_y$;
- covariance matrix of result is required, no bias, small correlations;
- dimension parameters are small compared to other fields (with e.g. 10^6 parameters)

Unfolding is more general than "data correction"

Neutral current neutrino interactions (narrow band beam): CHARM collaboration

Measurement of $E_{\rm had}$, $\vartheta_{\rm had}$, $r_{\rm beam}$

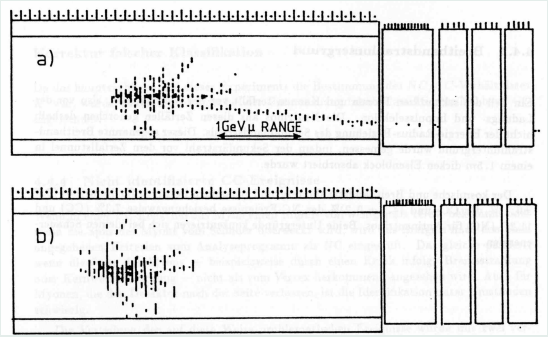
beam flux $\phi(E_\nu, r_{\rm beam})$

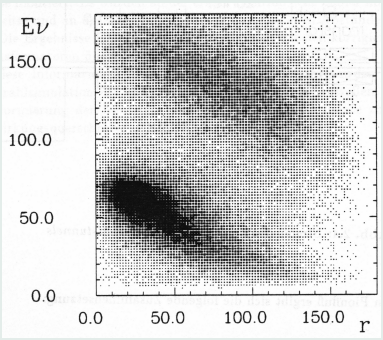
CC: $\nu\ N \to$ [hadrons] μ

$E_{\nu,in} = E_{\rm had} + E_\mu$

NC: $\nu\ N \to$ [hadrons] ν

$E_{\nu,in} = E_{\rm had} +$???





"Indirect" determination of $\mathrm{d}\sigma/\mathrm{d}y$ and $\mathrm{d}\sigma/\mathrm{d}x$:

direct problem (by MC)	$\mathrm{d}\sigma/\mathrm{d}x;\ \mathrm{d}\sigma/\mathrm{d}y \Longrightarrow$	predicted dist. in $E_{\rm had}$, $\vartheta_{\rm had}$, $r_{\rm beam}$
inverse problem (by unfolding)	$\mathrm{d}\sigma/\mathrm{d}x;\ \mathrm{d}\sigma/\mathrm{d}y \Longleftarrow$	measured dist. in $E_{\rm had}$, $\vartheta_{\rm had}$, $r_{\rm beam}$

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

2. Naive unfolding

The data errors are represented by a m-vector $\boldsymbol{e}$, and the actually measured distribution $\boldsymbol{y}$ is given by

measured distribution $\quad \boldsymbol{y} = \boldsymbol{y}_{\text{exact}} + \boldsymbol{e} = \boldsymbol{A}\,\boldsymbol{x}_{\text{exact}} + \boldsymbol{e} \qquad \boldsymbol{e} =$ data errors

$$\text{Unfolding:} \quad \boldsymbol{y};\, \boldsymbol{V}_y \quad \overset{\boldsymbol{A}}{\Longrightarrow} \quad \boldsymbol{x};\, \boldsymbol{V}_x \qquad\qquad \min_{\boldsymbol{x}} \left\{ \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|^2 \right\}$$

The n-by-m pseudoinverse $\boldsymbol{A}^+$ is a generalization of the inverse matrix (also called Moore-Penrose *generalized inverse*); it satisfies the relation $\boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{I}$, and allows the least squares solution by

$$\widehat{\boldsymbol{x}} = \boldsymbol{A}^+\boldsymbol{y} \qquad\qquad \boldsymbol{A}^+ = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}_y^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}_y^{-1}$$

$$\boldsymbol{V}_x = \boldsymbol{A}^+\boldsymbol{V}_y\boldsymbol{A}^{+\mathrm{T}} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}_y^{-1}\boldsymbol{A}\right)^{-1}$$

The response matrix $\boldsymbol{A}$ and the pseudoinverse $\boldsymbol{A}^+$ do not depend on any assumption about $\boldsymbol{x}$.

$$\text{estimate} \quad \widehat{\boldsymbol{x}} = \boldsymbol{A}^+\boldsymbol{y} = \boldsymbol{A}^+\boldsymbol{y}_{\text{exact}} + \boldsymbol{A}^+\boldsymbol{e} = \boldsymbol{A}^+\boldsymbol{A}\,\boldsymbol{x}_{\text{exact}} + \boldsymbol{A}^+\boldsymbol{e}$$
$$= \boldsymbol{x}_{\text{exact}} + \underbrace{\left(\boldsymbol{A}^+\boldsymbol{A} - \boldsymbol{I}\right)\boldsymbol{x}_{\text{exact}}}_{\text{systematic error}} + \underbrace{\boldsymbol{A}^+\boldsymbol{e}}_{\text{statistical error}}$$

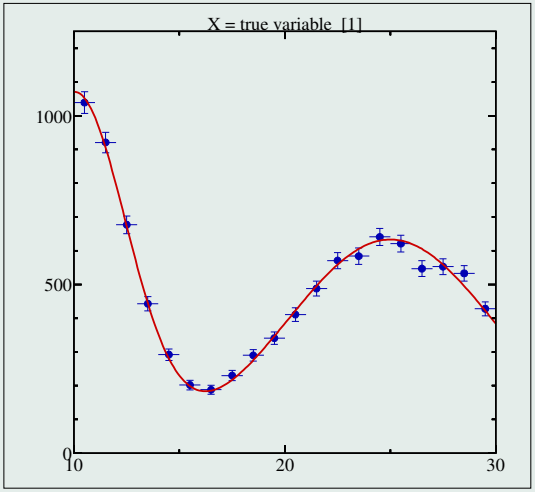
The systematic error (bias) $\left(\boldsymbol{A}^+\boldsymbol{A} - \boldsymbol{I}\right)\boldsymbol{x}_{\text{exact}}$ depends in $\boldsymbol{x}_{\text{exact}}$!

Least squares $\qquad \boldsymbol{A}^+\boldsymbol{A} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}_y^{-1}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}_y^{-1}\boldsymbol{A} = \boldsymbol{I}$ **no bias!** **Good or bad?**

Naive result with narrow bins

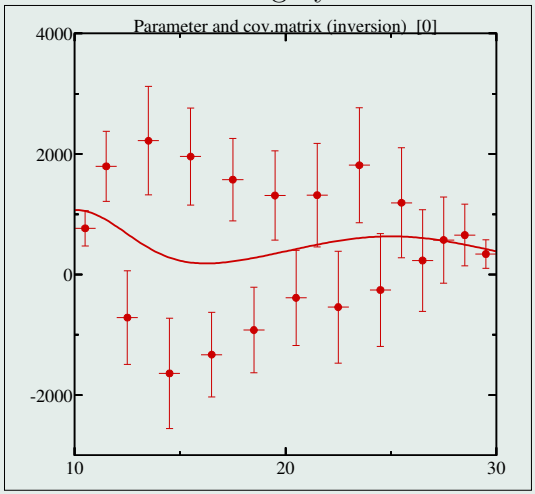
Example of unfolding problem with narrow bins

Perfect resolution (no smearing)



Histogram
for sample with 10 000 entries.

Naive unfolding by inversion



Huge fluctuations, due to large negative correlations: neighbour bin -95% (second $+85\%$).

True curve $f(x)$ is shown in red.

Orthogonalization!

3. Convolution/deconvolution with Gaussian response

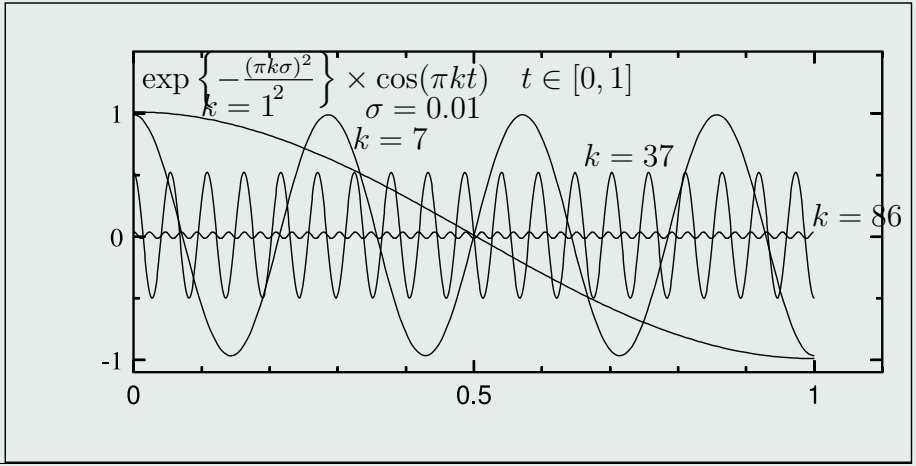
An even function $f(t)$ with period 1 can be approximated by a sum with n terms

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

Convolution of functions $\cos(\pi k t)$ by a kernel function $K(s,t) \equiv K(s-t)$ given by a Gaussian resolution function (standard deviation σ):

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

The amplitude is attenuated by an **exponential factor**, which will become $\ll 1$ for larger values of k.



. . . and deconvolution

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

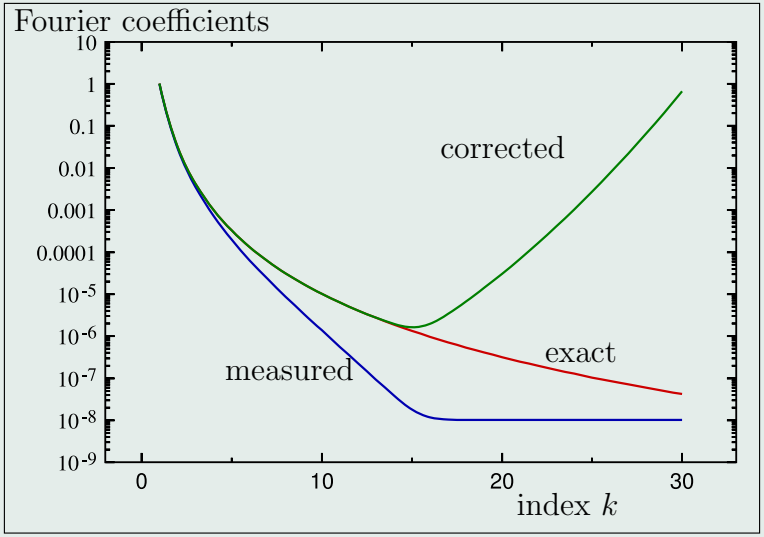
$$\text{Correct coefficients back:} \quad \widehat{a}_k = \exp\left(\frac{\pi^2 k^2 \sigma^2}{2}\right) \times \alpha_k \qquad \text{for small } k$$

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

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, labelled *corrected*, are correct only up to $k = 15$.

Deconvolution with $k \gg 15$ will have result dominated by noise.

For bin width $w = \sigma$ the factor is > 100. A factor below 10 is reached for $w > 1.5\sigma$.



4. Orthogonalization

Singular value decomposition, applied to rectangular m-by-n matrix $\boldsymbol{A}$: $(m \geq n)$

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T = \sum_{i=1}^{n} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^{\mathrm{T}} \qquad \boldsymbol{U}^{\mathrm{T}}\boldsymbol{U} = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{I}$$

$$\begin{pmatrix} & {\scriptstyle (n)} \\ {\scriptstyle (m)} & \boldsymbol{A} \\ & \end{pmatrix} = \begin{pmatrix} & {\scriptstyle (n)} \\ {\scriptstyle (m)} & \boldsymbol{U} \\ & \end{pmatrix} \cdot \begin{pmatrix} & {\scriptstyle (n)} \\ {\scriptstyle (n)} & \boldsymbol{\Sigma} \\ & \end{pmatrix} \begin{pmatrix} & {\scriptstyle (n)} \\ {\scriptstyle (n)} & \boldsymbol{V}^{\mathrm{T}} \\ & \end{pmatrix}$$

Matrix $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$ with ordered singular values $\sigma_1 \geq \sigma_2 \geq \ldots \sigma_n \geq 0$, assuming pre-whitening of eq. $\boldsymbol{A}\boldsymbol{x} \simeq \boldsymbol{y}$, i.e. $\boldsymbol{V}_y = \boldsymbol{I}$.

Diagonalization of symmetric LS matrix (or Hessian of log-Likelihood functions):

$$\boldsymbol{C} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\right)^{\mathrm{T}} \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T = \boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{\mathrm{T}}$$

Eigenvalues λ_j of symmetric matrix $\boldsymbol{C}$ are equal to squared singular values σ_j, and eigenvectors are equal to the singular vectors of matrix $\boldsymbol{V}$.

Least squares solution

SVD . . . "a new way to see into the heart of a matrix" (Gilbert Strang)

Matrix product using SVD: $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{x} = \sum_{j=1}^{n} \sigma_j \left(\boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{x}\right)\boldsymbol{u}_j = \boldsymbol{y}$

Fourier coefficients $\boldsymbol{c} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{y}$ with $\boldsymbol{V}_c = \boldsymbol{I}$ represent measurement $\boldsymbol{y}$.

Least squares solution with SVD:

$$
\begin{aligned}
\boldsymbol{U}^{\mathrm{T}} \cdot | \qquad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\,\widehat{\boldsymbol{x}} &= \boldsymbol{y} \\
\boldsymbol{\Sigma}\boldsymbol{V}^T\,\widehat{\boldsymbol{x}} &= \boldsymbol{U}^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{c} \qquad \text{Fourier coefficients} \quad c_j = \boldsymbol{u}_j^{\mathrm{T}}\boldsymbol{y} \pm 1 \\
\widehat{\boldsymbol{x}} &= \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{c} \\
\boldsymbol{V}_x &= \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}^{-1}\boldsymbol{V}^{\mathrm{T}} \qquad \text{error propagation}
\end{aligned}
$$

$$
\widehat{\boldsymbol{x}} = \sum_{j=1}^{n} \left(\frac{1}{\sigma_j}\right) c_j \boldsymbol{v}_j \qquad\qquad \boldsymbol{V}_x = \sum_{j=1}^{n} \left(\frac{1}{\sigma_j^2}\right) \boldsymbol{v}_j \boldsymbol{v}_j^{\mathrm{T}}
$$

Problem with zero or very small singular values!

Vanishing singular values and truncation

Assumption: p non-zero singular values of total n values with contribution $d_j = c_j/\sigma_j$.

$$\widehat{\boldsymbol{x}} = \underbrace{\sum_{j=1}^{p} d_j \boldsymbol{v}_j}_{\boldsymbol{x}_{\text{range}} \in \mathbb{R}^p} + \underbrace{\sum_{j=p+1}^{n} \widetilde{d}_j \boldsymbol{v}_j}_{\boldsymbol{x}_{\text{null}} \in \mathbb{R}^{n-p}} \qquad \boldsymbol{A}\widehat{\boldsymbol{x}} = \sum_{j=1}^{p} \sigma_j d_j \boldsymbol{v}_j + \underbrace{\sum_{j=p+1}^{n} \sigma_j \widetilde{d}_j \boldsymbol{v}_j}_{=0} = \boldsymbol{y}$$

$(n-p)$ contributions $\widetilde{d}_j$ are arbitrary: $\boldsymbol{x}_{\text{null}} \in \mathbb{R}^{n-p}$ without influence on measured distribution $\boldsymbol{y}$

Alternatives for unfolding solution:

- $\boldsymbol{x}_{\text{null}} =$ plausible contribution; in iterative methods the initial contributions $\boldsymbol{x}_{\text{null}} \in \mathbb{R}^{n-p}$ remain unchanged;
- minimum norm solution: $\boldsymbol{x}_{\text{null}} = 0$ with $n > p$;

$$\widehat{\boldsymbol{x}} = \boldsymbol{x}_{\text{range}} + \boldsymbol{x}_{\text{null}} \qquad \|\boldsymbol{x}_{\text{range}} + \boldsymbol{x}_{\text{null}}\|^2 = \|\boldsymbol{x}_{\text{range}}\|^2 + \|\boldsymbol{x}_{\text{null}}\|^2$$

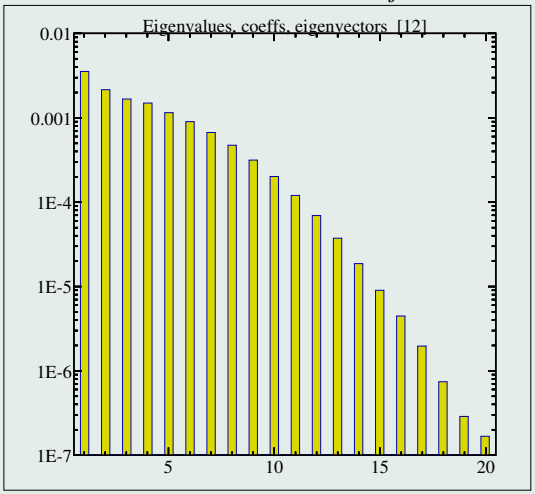
- reduction of dimension: $n' = p$ with full-rank covariance matrix.

What is the "best" approach?

Eigenvalues and Fourier coefficients

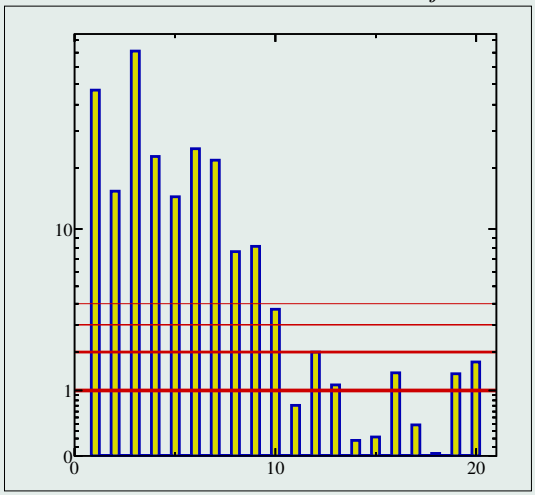
Example of unfolding problem with narrow bins

Eigenvalues $\lambda_j = \sigma_j^2$



Eigenvalues decrease by 4 orders of magnitude.

Normalized coefficients c_j



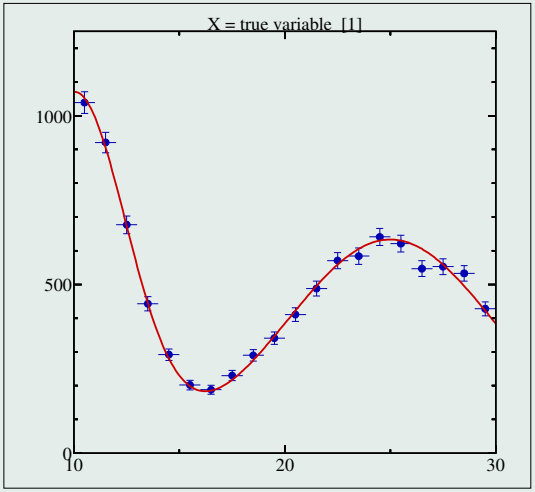
Only 10 of the 20 coefficients are significant.

Red lines are for 1, 2, 3 and 4 standard deviations.
Statistical errors are 1 for all coefficients.

Example with truncation

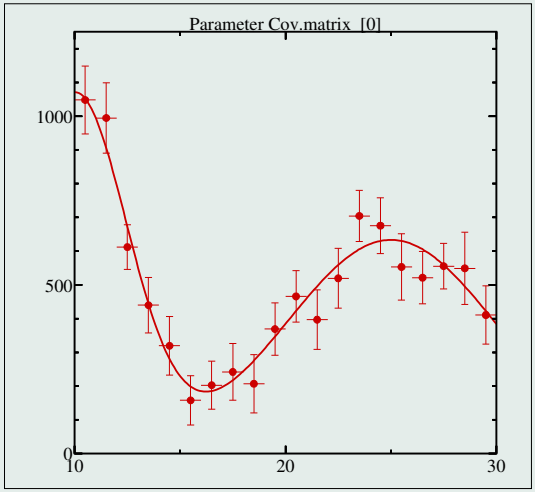
Example of unfolding problem with narrow bins

Perfect resolution (no smearing)



Histogram
for sample with 10 000 entries.

Truncation method (15 terms kept)



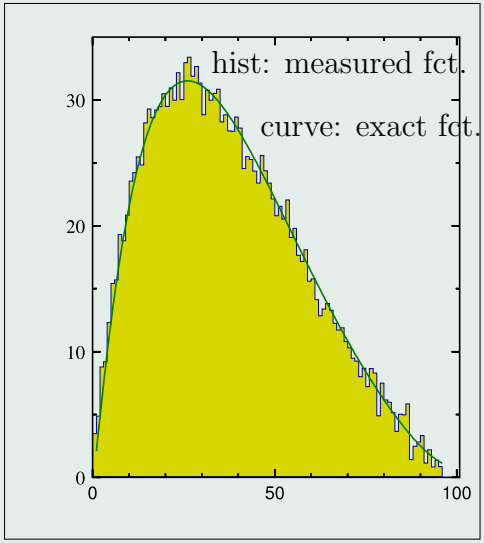
Reduced fluctuations

True curve $f(x)$ is shown in red.

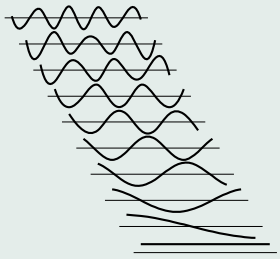
5. Truncation and positive correlations

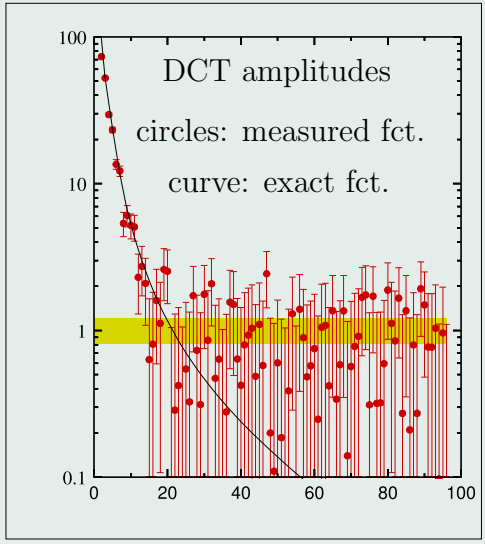
"Can unfolding dist have $\sigma < \sqrt{n}$?"

Assume distribution $\boldsymbol{y}$ without migration: discrete cosine transformation $\boldsymbol{c} = \boldsymbol{U}_{\mathrm{DCT}}^{\mathrm{T}}\, \boldsymbol{y}$



Histogram with unit cov. matrix





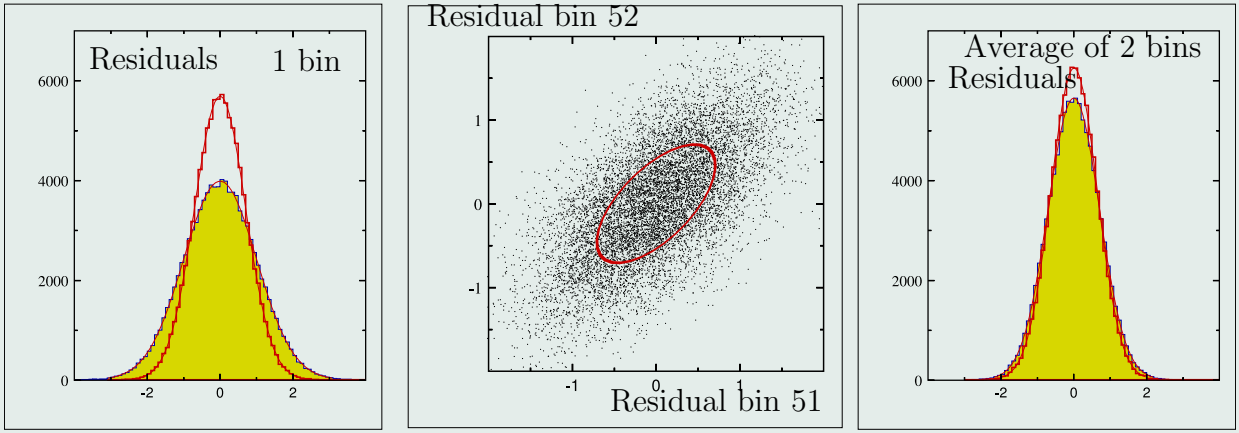
DCT amplitudes with unit cov. matrix

All high-frequency coeffients with index > 20 compatible with zero: truncation of second half of contributions (coeffs. and elements of cov. matrix) and back-transformation $\widehat{\boldsymbol{y}} = \boldsymbol{U}_{\mathrm{DCT}}\, \boldsymbol{c}$ reduces noise.

Note: $\Delta\chi^2 = c_j^2$ for single coefficient c_j!

Increase of accuracy . . .

- Histograms of differences to exact bin content: single bin, bin 51 vs. bin 52, 2-bin average;
- broader distribution from unmodified histogram, narrower from histogram with truncation after half the coefficients;
- truncation results in higher local accuracy (single bins), no bias, but positive correlations;
- perfect agreement of width with calculated errors and correlations.



Truncation of coefficients of high frequency contributions

- no bias introduced; cov.matrix calculation is accurate (but rank defect);
- higher precision for single bins only, due to introduction of positive correlations;
- neither higher total precision by truncation nor distortion of data or errors.

6. Regularization

Norm regularization: $\min_{\boldsymbol{x}} \left\{ \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|^2 + \tau \|\boldsymbol{x}\|^2 \right\} \quad \Longrightarrow \quad \boldsymbol{x} = \boldsymbol{A}^{\#}\boldsymbol{y} = \left[\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \tau \boldsymbol{I} \right)^{-1} \boldsymbol{A}^{\mathrm{T}} \right] \boldsymbol{y}$

Solved for fixed τ by inversion, or (better) ... $\qquad \boldsymbol{A}^{\#}\boldsymbol{A} \neq \boldsymbol{I}$

using SVD $\quad \overline{\boldsymbol{x}} = \boldsymbol{V} \underbrace{\left[\left(\boldsymbol{\Sigma}^2 + \tau \boldsymbol{I} \right)^{-1} \boldsymbol{\Sigma}^2 \right]}_{\text{filter factor } \varphi} \boldsymbol{\Sigma}^{-1} \underbrace{\left(\boldsymbol{U}^{\mathrm{T}} \boldsymbol{y} \right)}_{\text{coeff. } \boldsymbol{c}}$

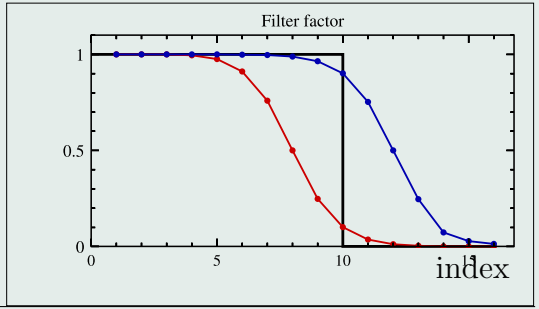
Effect of regularization: introduction of filter factor φ_j:

$$\widehat{\boldsymbol{x}} = \sum_{j=1}^{n} \frac{1}{\sigma_i} \boldsymbol{c}_j \boldsymbol{v}_j \qquad \Longrightarrow \qquad \widehat{\boldsymbol{x}} = \sum_{j=1}^{n} \frac{1}{\sigma_i} \varphi_j \boldsymbol{c}_j \boldsymbol{v}_j$$

$$\boldsymbol{V}_x = \sum_{j=1}^{n} \frac{1}{\sigma_j^2} \boldsymbol{v_j} \boldsymbol{v_j}^{\mathrm{T}} \quad \Longrightarrow \quad \boldsymbol{V}_x = \sum_{j=1}^{n} \frac{1}{\sigma_j^2} \varphi_j^2 \boldsymbol{v_j} \boldsymbol{v_j}^{\mathrm{T}}$$

filterfactor $\varphi_j = \left(\dfrac{\sigma_j^2}{\sigma_j^2 + \tau} \right) = \begin{cases} 1 & \text{for } \sigma_j^2 \gg \tau \\ 1/2 & \text{for } \sigma_j^2 = \tau \\ 0 & \text{for } \sigma_j^2 \ll \tau \end{cases}$

Different dependence possible: $1/\left(1 + (\tau/\sigma_j^2)^{\alpha}\right)$



Regularization with differential operator

$$\boldsymbol{L}\text{ regularization:}\quad \min_{\boldsymbol{x}} \left\{ \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|^2 + \tau \|\boldsymbol{L}\boldsymbol{x}\|^2 \right\} \quad\Longrightarrow\quad \boldsymbol{A}^{\#} = \left[\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \tau \boldsymbol{L}^{\mathrm{T}}\boldsymbol{L} \right)^{-1} \boldsymbol{A}^{\mathrm{T}} \right]$$

Solution by more complicated mathematical operation ("generalized singular value decomposition"), but almost identical unfolding formalism with filter factors φ_j.

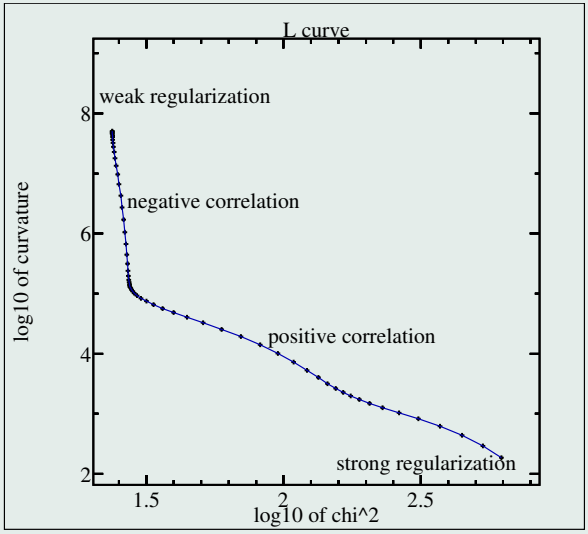
$$\text{most popular: sec.der.}\quad \boldsymbol{L}_2^r = \begin{pmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{pmatrix} = \boldsymbol{U}_{\mathrm{DCT}}\, \boldsymbol{\Lambda}\, \boldsymbol{U}_{\mathrm{DCT}}^{\mathrm{T}}$$

My opinion about regularization:

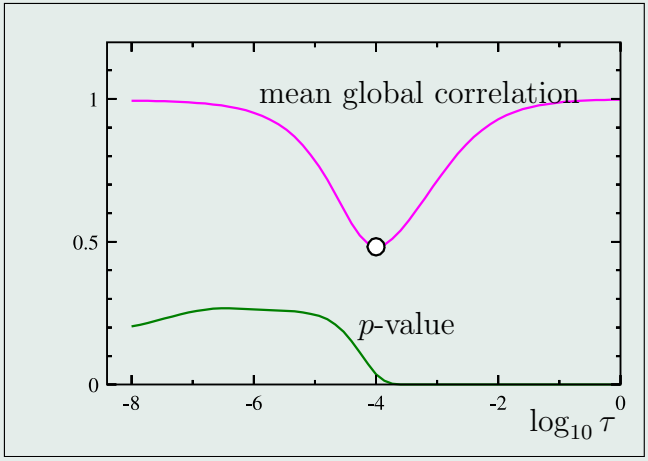
- A result without regularization with many data points represents noise – nothing else;
- removing coefficient c_j (by filtering) means $\Delta\chi^2 = c_j^2$;
- regularization will not introduce unwanted bias, and allows to reduce or suppress insignificant contributions (noise), that would destroy the unfolding result.

Determination of regularization parameter

- Regularization corresponds technically to "weak" a-priori information about $\boldsymbol{x}$ (like "measurement" with standard deviation $\approx \sqrt{1/\tau}$ for norm regularization);
- if $\tau > \sigma_j^2$ of significant Fourier coefficients $c_j \Rightarrow$ **no bias**;
- evaluate several quantities as function of τ for wide range: $\tau \in (\tau_L, \tau_R)$;
- effective number of degrees of freedom: $\sum \varphi_j$



dependence as a function of τ



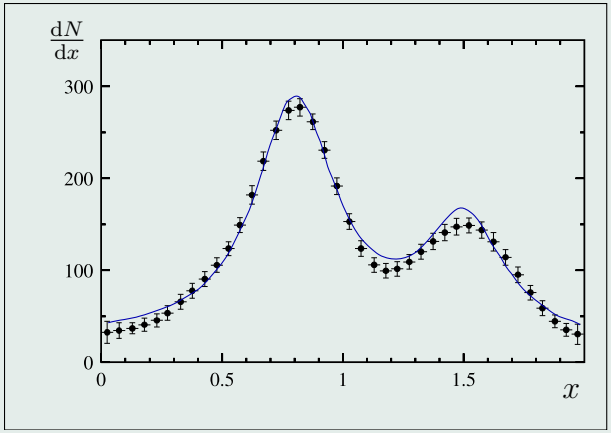
Presentation of regularized result I ... with MC input

(from A. Hoecker and V. Kartvelishvili: NIM A 372)

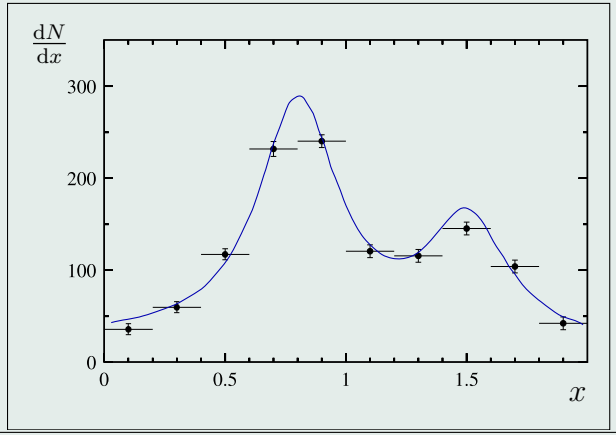
- Result with 40 data points, almost like a "band" representing result
- constructed from 10 significant parameters, with 40-by-40 covariance matrix, singular with rank 10;
  use "effective" weight matrix for fits

$$[\boldsymbol{V}_x]^{-1} = \sum_{j=1}^{n} \sigma_j^2 \, \boldsymbol{v_j} \boldsymbol{v_j}^{\mathrm{T}}$$

- large positive correlations, therefore few sign-changes of the residuals to MC input distribution.

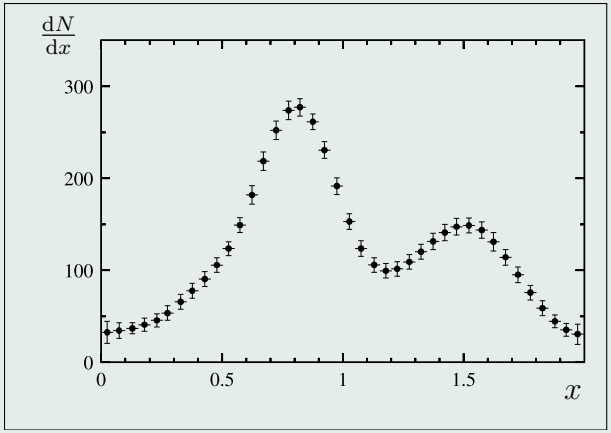


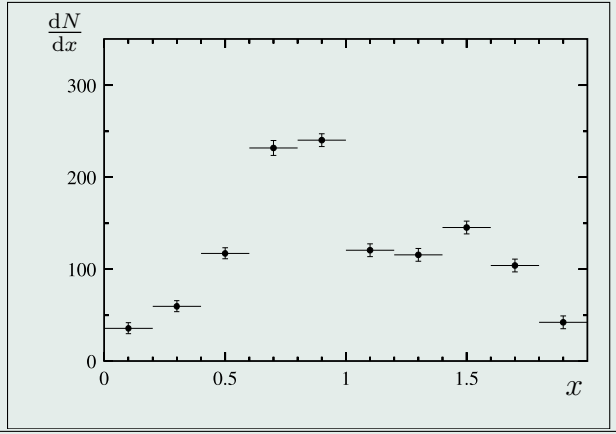
- (Same) result with 10 data points, each point represents a bin average of the result
- 10-by-10 covariance matrix non-singular, inverse is weight matrix.
- small and negligible correlations.



Presentation of regularized result II ... without MC input

- Plots (without the true MC distribution): both data sets represent the same information!
- What is the better data presentation in **Particle Physics?**
- Try to calculate "limits" from the data!
- Note: not obtained by different methods, but (only) different presentation of identical information.





Example: two unknowns, three measured values

The measured vector $\boldsymbol{y}$ and the response matrix $\boldsymbol{A}$ are given by

$$\boldsymbol{y} = \boldsymbol{A}\underbrace{\begin{pmatrix} 1 \\ 1 \end{pmatrix}}_{\text{exact}} + \underbrace{\begin{pmatrix} 0.01 \\ -0.03 \\ 0.02 \end{pmatrix}}_{\text{error}} = \underbrace{\begin{pmatrix} 0.27 \\ 0.25 \\ 3.33 \end{pmatrix}}_{\text{measured}} \qquad \boldsymbol{A} = \begin{pmatrix} 0.16 & 0.10 \\ 0.17 & 0.11 \\ 2.02 & 1.29 \end{pmatrix}$$

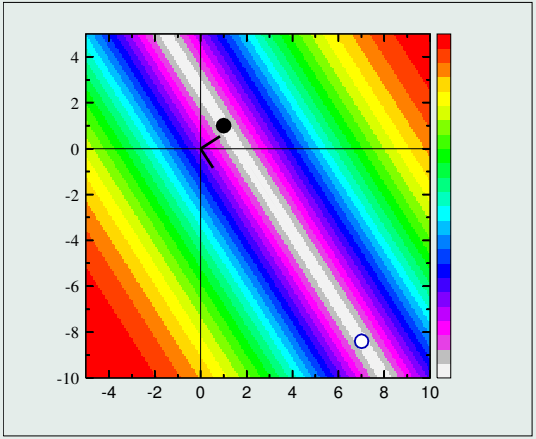
If analysed by the singular value decomposition the singular values are

$$\sigma_1 = 2.4127 \qquad \sigma_2 = 0.0022$$

The least-squares solution $\widehat{\boldsymbol{x}}_{\text{LS}}$, assuming a standard deviations of 0.02 for the three elements of $\boldsymbol{y}$

$$\widehat{\boldsymbol{x}}_{\text{exact}} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad \widehat{\boldsymbol{x}}_{\text{LS}} = \begin{pmatrix} 7.01 \pm 4.90 \\ -8.40 \pm 7.67 \end{pmatrix}$$

correlation coeff. $\rho = -0.999998$



Example from Per Christian Hansen: **Discrete Inverse Problems – Insight and Algorithms**

Example with regularization

- Regularization term $\tau \|\boldsymbol{x}\|^2$ is equivalent to assumption of measurement information

$$(\boldsymbol{x})_j = 0 \pm \sqrt{\frac{1}{\tau}}$$

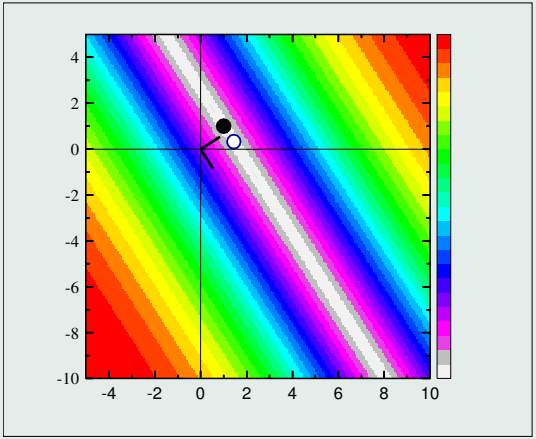
- Example: assume solution to have norm $\|\boldsymbol{x}\|$ of the order of 1.
- Regularization parameter value $\tau = 1/4$ corresponds to

$$(\boldsymbol{x})_j = 0 \pm 2 \qquad j = 1,\ 2$$

Regularized solution of example:

$$\widehat{\boldsymbol{x}}_{exact} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad \widehat{\boldsymbol{x}}_{\text{reg LS}} = \begin{pmatrix} 1.44 & \pm & 1.05 \\ 0.33 & \pm & 1.65 \end{pmatrix}$$

correlation coeff. $\rho = -0.99996$



Calculation using normal fit program `aplcon`.

7. DCT and projection methods

Discrete cosine transformation DCT = orthonormal transformation by matrix $\boldsymbol{U}_{\mathrm{DCT}}$ with $\boldsymbol{U}_{\mathrm{DCT}}\,\boldsymbol{U}_{\mathrm{DCT}}^{\mathrm{T}} = \boldsymbol{I}$,

transformations: $\boldsymbol{X} = \boldsymbol{U}_{\mathrm{DCT}}^{\mathrm{T}}\boldsymbol{x}$ $\qquad$ $\boldsymbol{x} = \boldsymbol{U}_{\mathrm{DCT}}\boldsymbol{X}$

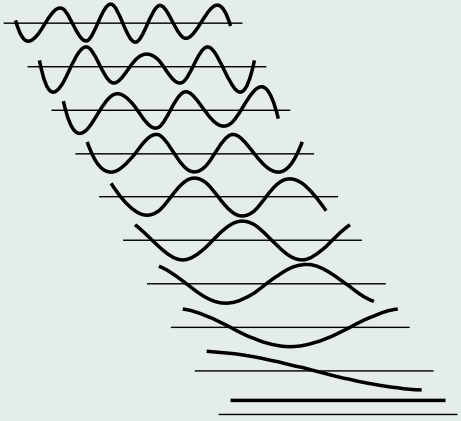
$$U_{jk} = \begin{cases} \sqrt{1/n} & k = 0 \\ \sqrt{2/n}\cos\left[\pi k(j+1/2)/n\right] & k = 1,\, 2 \ldots n-1 \end{cases}$$

Relation to derivative regularization matrix $\boldsymbol{L} = \boldsymbol{L}_2^r$:

$$\boldsymbol{L} = \boldsymbol{U}_{\mathrm{DCT}}\,\boldsymbol{\Lambda}\,\boldsymbol{U}_{\mathrm{DCT}}^{\mathrm{T}} \qquad \boldsymbol{L}^{\mathrm{T}}\boldsymbol{L} = \boldsymbol{U}_{\mathrm{DCT}}\,\boldsymbol{\Lambda}^2\,\boldsymbol{U}_{\mathrm{DCT}}^{\mathrm{T}}$$

with eigenvalues: $\lambda_k = 4\sin^2\left(\dfrac{k\pi}{2n}\right) \qquad k = 0,\, 1, \ldots$

- DCT is purely real, and concentrates "energy" into lower order coefficients better than discrete Fourier transformation;
- separability – perform DCT in any of the directions first and then apply to second direction: coefficients wills not change: 2-D DCT is 2× 1-D DCT;
- used in modern coding standards like JPEG, MPEG.



$$u_k(t) = \cos(\pi k t)$$
$$k = 0,\, 1, \ldots n-1$$
$$t \in (0, 1)$$

Projection methods

So far: approximation of solution $\boldsymbol{x}_{\text{exact}} \in \mathbb{R}^n$ in low-dimensional subspace of low-frequency components:

- by truncation, keeping the first p Fourier coefficients of SVD method (TSVD), or
- by regularization using e.g. second-derivative matrix $\boldsymbol{L}_2^r$ with regularization parameter τ.

Alternative: use of fixed projection n-by-p matrix $\boldsymbol{W}_{(p)} = \{\boldsymbol{v}_1, \boldsymbol{v}_2 \ldots \boldsymbol{v}_p\}$ with $\overline{\boldsymbol{x}} \in \mathbb{R}^p$:

$$\boldsymbol{x} = \boldsymbol{W}_{(p)}\, \overline{\boldsymbol{x}} \qquad\qquad \overline{\boldsymbol{A}} = \boldsymbol{A}\boldsymbol{W}_{(p)}$$

$$\min_{\boldsymbol{x}} Q(\boldsymbol{x}) \quad \text{with} \quad Q(\boldsymbol{x}) = \|\boldsymbol{A}\,\boldsymbol{x} - \boldsymbol{y}\|^2 = \left\|\boldsymbol{A}\boldsymbol{W}_{(p)}\, \overline{\boldsymbol{x}} - \boldsymbol{y}\right\|^2 = \left\|\overline{\boldsymbol{A}}\ \overline{\boldsymbol{x}} - \boldsymbol{y}\right\|^2$$

- Basis vectors $\boldsymbol{v}_j$ = e.g. **DCT eigenvectors**, or SVD singular vectors;
- solutions $\overline{\boldsymbol{x}}$ and $\boldsymbol{x} = \boldsymbol{W}_{(p)}\, \overline{\boldsymbol{x}}$ dominated by low-frequency components;
- truncation or filtering with filter functions φ_j possible;
- fixed one- or two-dimensional transformation, to be used with different MC data sets to study MC differences.

8. Iterative unfolding

Why iterative methods? What is the advantage compared to a direct solution?

- Unfolding with extremely large dimensions (e.g. 10^6 parameters in picture deblurring) requires iterative methods;
- dimension parameters in Particle Physics problems are small, no cpu-time problems for direct analytic methods, like SVD;
- certain iterative methods for inverse problems are popular in Particle Physics: they have semi-convergence, with implicit regularization.

Gauß had used iterative methods at least since the year 1823:

> You will in future hardly eliminate directly, at least not when you have more than two unknowns. The indirect procedure can be done while one is half asleep, or is thinking about other things. [Carl Friedrich Gauß, Werke IX, p.278]

Landweber iteration

Formula for one Landweber[+] iteration step $k = 0,\ 1,\ 2, \ldots$:

$$\boldsymbol{x}^{[k+1]} := \boldsymbol{x}^{[k]} + \omega \boldsymbol{A}^{\mathrm{T}} \left(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}^{[k])} \right) \qquad 0 < \omega < 2/\sigma_1^2$$

Semi-convergence corresponds to regularization ($j = 1,\ 2, \ldots,\ n$): iteration number k

$$\boldsymbol{x}^{[k]} = \sum_{j=1}^{n} \frac{\varphi_j^{[k]}}{\sigma_j} \left(\boldsymbol{u}_j^{\mathrm{T}} \boldsymbol{y} \right) \boldsymbol{v}_j \qquad \varphi_j^{[k]} \approx 1 - \left(1 - \omega\sigma_j^2 \right)^k \approx \begin{cases} 1 & \text{for large } \sigma_j^2 \\ k \left(\omega\sigma_j^2 \right) & \text{for } \sigma_j^2 \ll 1/\omega \\ \to 0 & \text{for } \sigma_j^2 \to 0 \end{cases}$$

Iteration number k plays the role of a regularization parameter:

- convergence fast for components with a large singular values σ_j;
- very slow for components with small singular value: components unchanged after few iterations;
- after a large number of iterations the unique (oscillating) solution of linear system;
- no prescription for covariance matrix calculation.

[+] Other names associated with the algorithm are Richardson, Fridman, Picard and Cimmino.

Iterative methods in particle physics

In general there is the attempt, by an iterative tuning process, to use the "correct" distribution in the MC simulation, i.e. that distribution that should be extracted from the measured data $\boldsymbol{y} \simeq \boldsymbol{A}\boldsymbol{x}$.
Iterative improvement of a matrix $\boldsymbol{M}_x^{[k]}$, which depends on the solution $\boldsymbol{x}$ and which should allow to extract the solution

$$\boldsymbol{x}^{[k+1]} = \boldsymbol{M}_x^{[k]}\,\boldsymbol{y} \quad \Longrightarrow \quad \boldsymbol{x}^{[k+1]} = \left(\boldsymbol{M}_x^{[k]}\,\boldsymbol{A}\right)\boldsymbol{x} \qquad \text{but} \quad \left(\boldsymbol{M}_x^{[k]}\,\boldsymbol{A}\right) \neq \boldsymbol{I}$$

"equation" valid only for certain $\boldsymbol{x}$

bin-by-bin correction factor:	$\boldsymbol{M}_x =$ diagonal
other iterative methods:	$\boldsymbol{M}_x =$ matrix with non-negative elements

Advantage: popular and accepted by collaborations; simple mathematics: no "complicated" operations like SVD;

Disadvantage: no direct error propagation with matrix $\boldsymbol{M}_x$ possible; unknown regularization strength; correlations unknown and ignored in bin-by-bin correction method; questionable statistically and mathematically; applicable only to "correct" for migration effects, no general unfolding.

Summary

- Folding is a direct process – it is robust and simple, but a folded model prediction does not show the sensitivity of the measurement (and is insensitive to contribution $\boldsymbol{x}_{\text{null}}$).
- Unfolding is the inverse process – it is a discrete ill-posed problem, mathematically complex, with a response matrix with large condition number, but it allows to study the sensitivity of the measurement:
  - test of black-box algorithms on a few selected examples are not sufficient;
  - it is essential to understand the statistical and mathematical properties of the algorithm, and to understand the detector;
  - orthogonalization and regularization are generally accepted concepts in other fields;
  - there are still several open questions to apply successfully unfolding for particle physics experiment.

Appendix

- Types of unfolding problems
- Open questions
- Is the unfolding result allowed to dependend on the MC input dependence?
- Comparison of codes
- Unfolding program $\mathcal{RUN}$
- Literature
- Correlations
- Averaging correlated data
- Smoothing by truncation of DCT amplitudes
- Low-pass regularization
- Example with low-pass filter

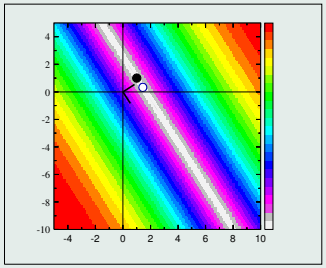
Types of unfolding problems

Parametrized unfolding

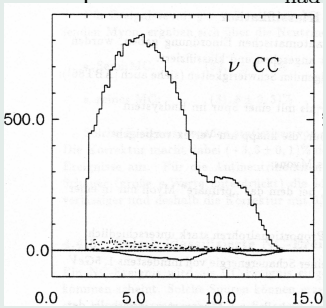
parameterized expression with few parameters no regularization necesssary – using normal fit program.

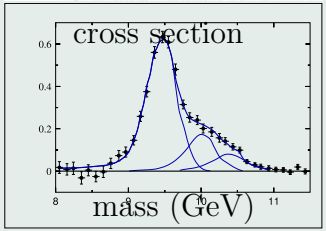
Parameter-free unfolding bins or classes, no parametrization

- Classification problems, small number of bins: no reduction of number of bins possible;
- Correction of distributions with small migration effects;
- Unfolding with transformation, evtl. > 1 measured distribution;
- Structureless distributions (small number of Fourier coefficients):
  - Regularization with possibility to reduce number of bins;
  - Steeply falling: try transformation to remove steepness, e.g. $\sqrt{E_{\mathrm{had}}}$, evtl. use parametrization
- Distributions with narrow structures, peaks: try to use parametrization e.g. Gaussian peaks)



Square root of E_{had}





Open questions

- discretization: continuous B-splines of higher order (necessary for re-weighting of MC – RUN) instead of simple discontinuous bins
- input of n-tuples instead of vectors/matrices
- strategies for steeply falling distributions and for class data (e.g. 2- ,3- ...jet events)
- alternative low-pass filter strategies
- data presentation with small or large number of data points
- projection methods (DCT) in one- and two-dimensions
- case of limited acceptance regions (e.g. low p_{T} – using constraints (Lagrange multiplier – RUN)) in one- and two-dimensions
- using general fit program (like aplcon) for unfolding? (flexible: e.g. fit may include uncertainty of $\boldsymbol{A}$ or background fit)
- uncertainties of response matrix elements
- algorithm for automatic variable bin definition (RUN)
- statistical comparison of information content for different number of degrees of freedom
- use of detailed estimates of the unfolding result (multiplicative, additive)

Is the unfolding result allowed to dependend on the MC input dependence?

My opinion: **NO**.
Do 'blind analysis", whenever possible. Avoid any bias w.r.t. an expected result.

From a paper on the CFM: *'The correction of 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 migration 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."*

In an iterative method: *' . . . it gives the best results (in terms of its ability to reproduce the true distribution) if one makes a realistic guess about the distribution that the true values follow . . . .*

What happens in the case of a completely insensitive detector?

Regularization methods will not be able to get any result!

In the iterative method: *"One finds then that the final probabilities are equal to the initial ones . . . ."*

Comparison of codes

Method	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	✔			

GURU (Fortran, SVD, by Andreas Hoecker and Vato Kartvelishveli) and TUNFOLD (by Stefan Schmitt) in RooUnfold (ROOT Unfolding Framework, by Tim Adye et al.)

RUN (Fortran), converted to C++ by Natalie Milke (Uni Dortmund)

Unfolding program $\mathcal{RUN}$

Development started 1979/1980 for neutral current neutrino experiment CHARM:

- First problem was reconstruction of cross section $\mathrm{d}\sigma/\mathrm{d}y$ from the measured values E_{had} and radius $r_{\mathrm{interaction}}$ (only these 2 quantities were measureable).
- Input are n-tuples, and 1-dim. or 2-dim. or 3-dim. measured histograms.
- Instead of Least Squares the ML method with Poisson statistic was used (sometimes only few entries/bin for $>$ 1-dim. histograms), with diagonalization of Hessian.
- For the intermediate result cubic B-splines were used to avoid discontinuities.
- A special option allows to check the consistency of MC simulation.
- Used in other experiments:
  Neutrino physics, 2-photon-physics at e^+e^- colliders, in astrophysics, and still used in 2010 (LHC, D0 at FNAL).
- Conversion to C++ by Natalie Milke (Uni Dortmund).

Literature

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.

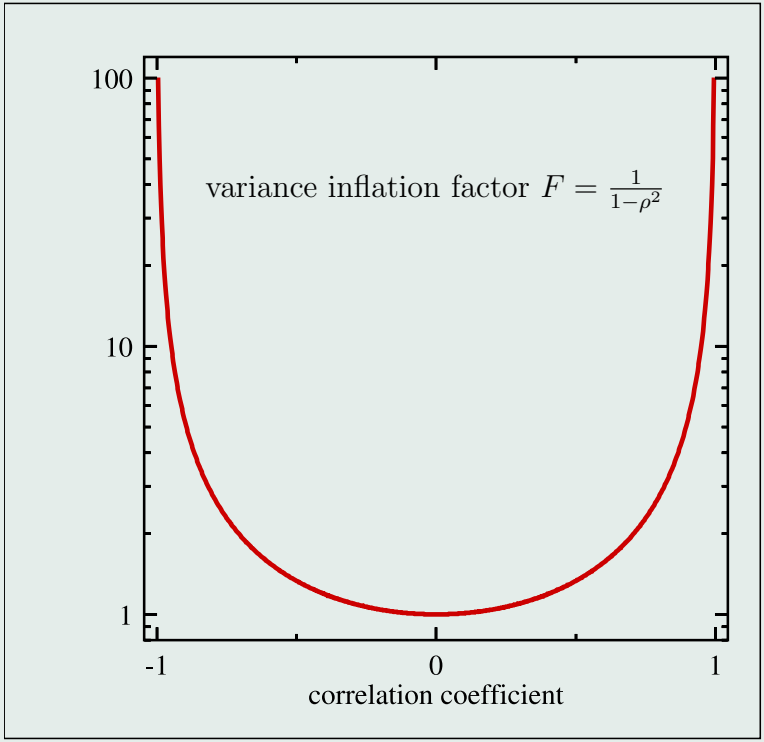
Curtis R. Vogel.
Computational Methods for Inverse Problems,
Volume 160 of *SIAM Frontiers in Applied mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, 2002.

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.

Correlations

Correlation coefficients $|\rho| > 0.5$ are dangerous:

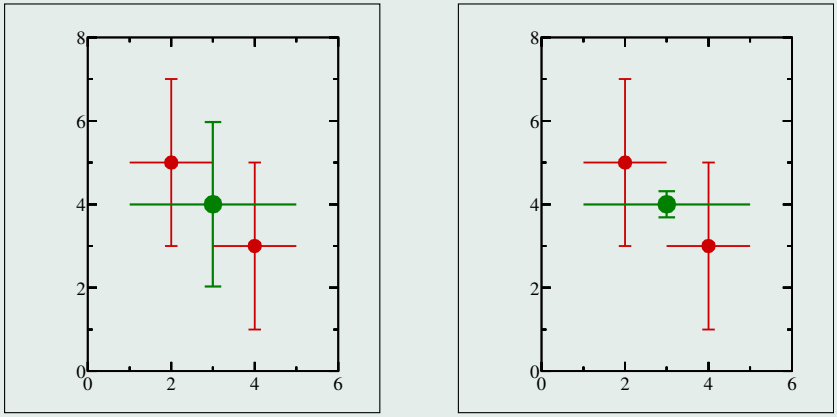


Averaging correlated data

The Figures show two adjacent data points d_1 and d_2 with a large positive (left) and negative (right) correlation coefficient, assuming the same standard deviations: $\sigma = \sigma_1 = \sigma_2$.

$$\boldsymbol{V} = \begin{pmatrix} \sigma^2 & \rho_{12}\sigma^2 \\ \rho_{12}\sigma^2 & \sigma^2 \end{pmatrix} \qquad \text{with} \quad \rho_{12} = \pm 0.95$$

Average: the average value $\overline{\boldsymbol{d}} = \frac{1}{2}(d_1 + d_2)$ has a variance of $\boldsymbol{V}_{\overline{d}} = \frac{1}{2}(1 + \rho_{12})\,\sigma^2$ (see middle point in figures).

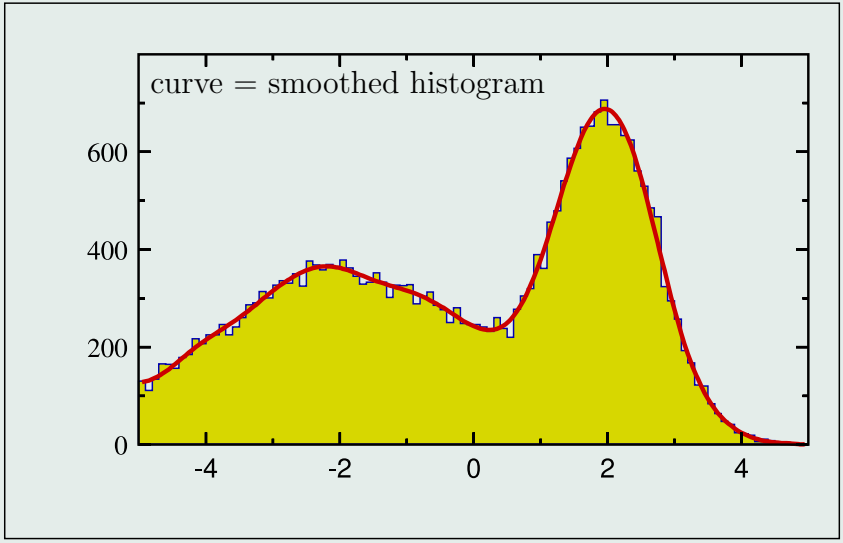


For highly correlated data the properties for averaging and χ^2-comparison with predicted values are not intuitive, but have to be performed with an explicit calculation, based on the inverse covariance matrix $\boldsymbol{V}_x^{-1} = \boldsymbol{W}_x$ (weight matrix).

Smoothing by truncation of DCT amplitudes

Example for smoothing of histogram by DCT and truncation:

- Transformation ($\sqrt{n_j}$) of bin content (Poisson) to stable variance;
- discrete cosine transformation and truncation of insignificant coefficients;
- back transformations.



Low-pass regularization

Without regularization there are large bin-to-bin fluctuations due to negative correlations between neighbour bins. These fluctuations can be suppressed in a low-pass filter by averaging 3-to-1 bins:

$$\overline{x}_j = \tfrac{1}{4}x_{j-1} + \tfrac{1}{2}x_j + \tfrac{1}{4}x_{j+1} \qquad \text{or general} \qquad \overline{x}_j = a_j x_{j-1} + (1-2a_j)\, x_j + a_j x_{j+1}$$

The factor a_j can be chosen to minimize[1)] the variance of $\overline{x}_j$, using the known matrix $\boldsymbol{V}_x$.

Pro: Fluctuations are really suppressed and the true dependence is clearer visible.

No bias, if number of bins large and no strong structure.

Con: In regions of larger second-derivatives a bias is introduced, because the above filter assumes an almost linear dependence over 3-point regions.
First and last bins disappear.

The general averaging algorithm for this 'local"-regularization method:

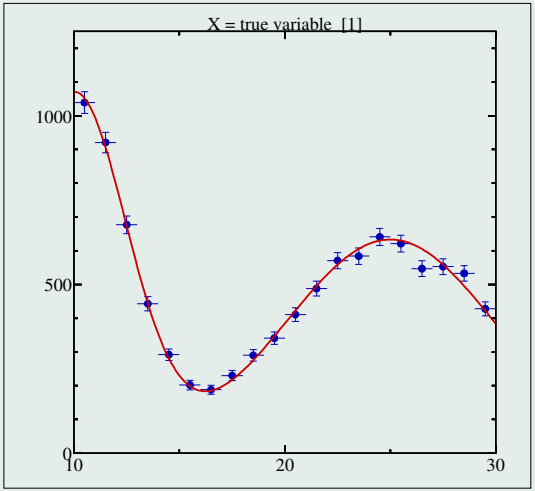
$$\overline{\boldsymbol{x}} = \boldsymbol{T}\boldsymbol{x} \qquad \boldsymbol{V}_{\overline{x}} = \boldsymbol{T}\boldsymbol{V}_x\boldsymbol{T}^{\mathrm{T}} \qquad \text{with} \quad \boldsymbol{T} = \frac{1}{4}\begin{pmatrix} 1 & 2 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 2 & 1 \end{pmatrix}$$

1) O. Helene et al., NIM A 523 (2004) 186; NIM A 580 (2007) 1466 - 1473

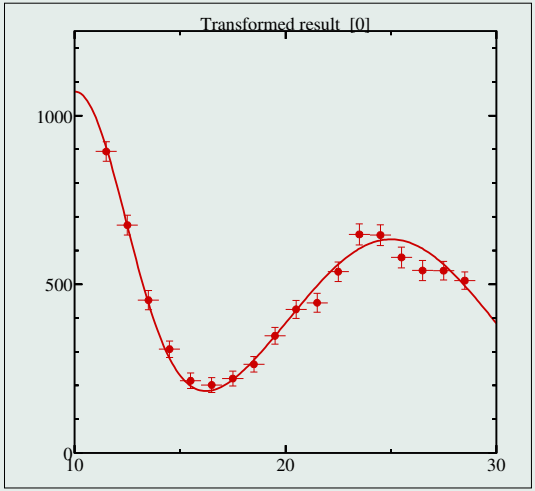
Example with low-pass filter

Example of unfolding problem with narrow bins

Perfect resolution (no smearing)



regularization $\tau = 10^{-5}$ + low-pass filter



Histogram
for sample with 10 000 entries.

Reduced correlations: neighbour bin -10%
and second neighbour -30%.

True curve $f(x)$ is shown in red.

Contents