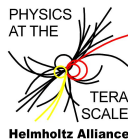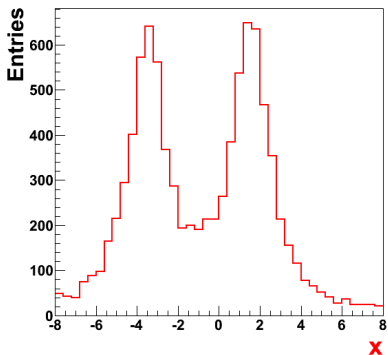# Unfolding
A Statistics Group Project

Gero Flucke
DESY

4th Annual Workshop of the Helmholtz Alliance
"Physics at the Terascale"
December 1-3, 2010
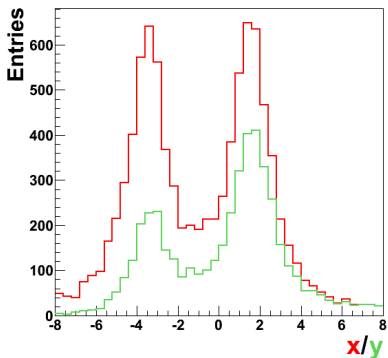Dresden

Outline

- Unfolding: an ill-posed problem
- An Unfolding Framework
- Summary

Measuring Process

We measure a distribution:

- Nature produces $f(x)$.
- Our detector has limited acceptance: $f(x) \Rightarrow g(y)$.
- Probably we suffer from a bias.
- For sure resolution is limited.
- **Background** might be an issue.

We want to measure $f(x)$, but we get $g(y)$...

Measuring Process

We measure a distribution:

- Nature produces $f(x)$.
- Our detector has limited acceptance: $f(x) \Rightarrow g(y)$.
- Probably we suffer from a bias.
- For sure resolution is limited.
- **Background** might be an issue.

We want to measure $f(x)$, but we get $g(y)$...

Measuring Process

We measure a distribution:

- Nature produces $f(x)$.
- Our detector has limited acceptance: $f(x) \Rightarrow g(y)$.
- Probably we suffer from a bias.
- For sure resolution is limited.
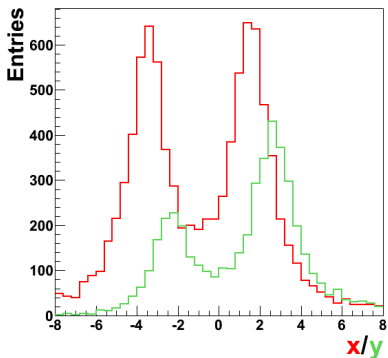- **Background** might be an issue.

We want to measure $f(x)$, but we get $g(y)$...

Measuring Process

We measure a distribution:

- Nature produces $f(x)$.
- Our detector has limited acceptance: $f(x) \Rightarrow g(y)$.
- Probably we suffer from a bias.
- For sure resolution is limited.
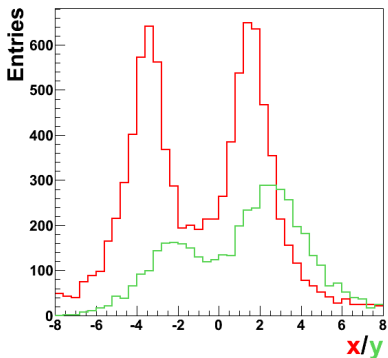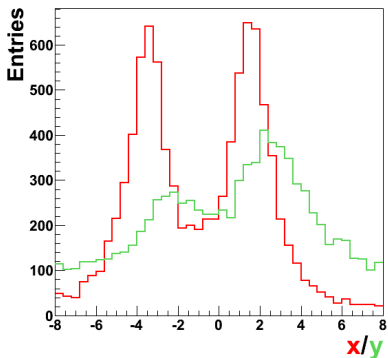- **Background** might be an issue.

We want to measure $f(x)$, but we get $g(y)$...

Measuring Process

We measure a distribution:

- Nature produces **$f(x)$**.
- Our detector has limited acceptance: **$f(x)$ ⇒ $g(y)$**.
- Probably we suffer from a bias.
- For sure resolution is limited.
- **Background** might be an issue.

We want to measure **$f(x)$**, but we get **$g(y)$**...

Measuring Process: More Formally

$$\int A(y, x) \, f(x) \, dx + b(y) = g(y)$$

Our Detector
- Limited Acceptance
- Inefficiencies
- Limited Resolution
- ...

True Distribution
- Invariant Mass.
- Transverse Momentum.
- Something multidimensional.
- ...

Measured Distribution
- Another one!
- Including **background $b(y)$!**

Usual goal:
- Find out, whether a theory can describe $f(x)$.
- But the measurement gives $g(y)$!
- Way out: We know our detector response **A** (hopefully pretty well).

$$\int A_{sim}(y, x)\, f_{theo}(x)\, dx + b(y) = g_{theo}(y)$$

Simulate Measurement Process

- Feed theoretical $f_{theo}(x)$ through detector simulation A_{sim}.
- Add background $b(y)$ (simulation / from data).
- Compare $g_{theo}(y)$ to measured $g(y)$.

- Probably the best one can do!
- But how to compare to
    - future theories?
    - another experiment?
- $\Rightarrow$ Need to reconstruct $f(x)$ from $g(y)$,
    - using A_{sim}, $b(y)$.
- $\Rightarrow$ Unfolding!

Here and in the following: Follow V. Blobel's notations.

$$\int A_{sim}(y, x)\, f(x)\, dx + b(y) = g(y)$$

Discretize to use linear algebra (for simplicity: skip background b):

$$A\, x = y$$

- x: n-vector of true variable
  (n limited by resolution)
- y: m-vector of measured variable
- $m \gg n$ recommended
- A: $m \times n$ response matrix,
  independent of theory.
- A_{ij} is probability that event in bin j
  of x is observed in bin i of y
- $\sum_{i=1}^{m} A_{ij} = P_j \leq 1$ (acceptance)
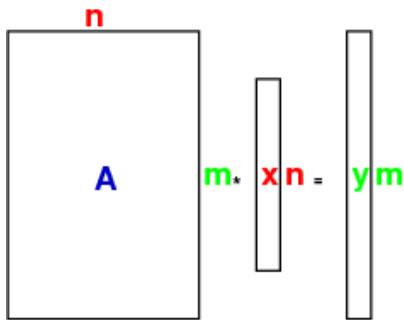
Unfolding is an Inverse Problem

Direct Process

$$A\,x + \epsilon = y$$

ϵ from statistical uncertainties

The Ill-posed Inverse Problem

- Reconstruct input x from a measured output y.
- A washes out fine structures in x, cannot be reconstructed from y.
- $\Rightarrow$ Small perturbation of the data can cause an arbitrary large perturbation of the solution.

Inverse Process

- Find $A^\sharp$ to get estimator $\hat{x}$

$$\hat{x} = A^\sharp y$$

- $A^\sharp$: $n \times m$ matrix, ("generalised inverse").

$$Ax + \epsilon = y$$

- Minimise $(Ax - y)^T V_y^{-1} (Ax - y)$ (LS principle)
- $\Rightarrow A^{\sharp} = (A^T V_y^{-1} A)^{-1} A^T V_y^{-1}$
- $\Rightarrow V_x = A^{\sharp} V_y A^{\sharp T}$ (standard error propagation)

(Small $y_i \Rightarrow$ use Maximum Likelihood with Poisson statistics.)

Result

- Unbiased estimator $\hat{x} = A^{\sharp} y$.
- Smallest variance V_x of all unbiased estimators.
- But you will see
  - large fluctuations
  - and large correlations!
- $\Rightarrow$ next slide.

Matrix Inversion

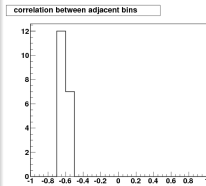**Residuals: Unfolded - Truth**
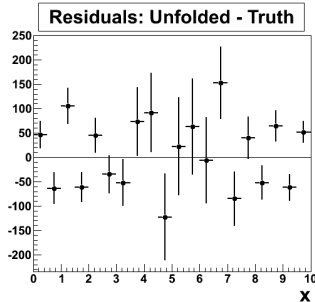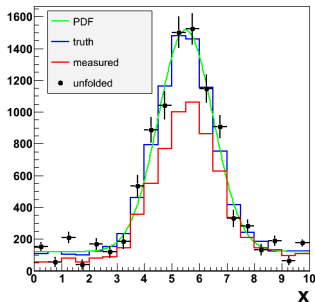
Even Simpler Case

$n = m \Rightarrow A^{\sharp} = A^{-1}$

- **True** distribution (Gauss).
- **Measured** distribution:
  - some acceptance loss,
  - Gaussian resolution: $\sigma_x = 0.5$ (= bin width).
- **Unfolding** result.

- Residuals show large fluctuations.
- Accompanied by large negative correlations of adjacent bins:

correlation between adjacent bins

Using Diagonalisation

- Diagonalise n-matrix $(AV_y^{-1}A^T) = U\Lambda U^T$.
  - $UU^T = U^T U = I$.
- Λ is diagonal matrix of Eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$.
- Now one can write

$$\hat{x} = \sum_{i=1}^{n} \frac{1}{\sqrt{\lambda_i}} c_i u_i, \qquad V_x = \sum_{i=1}^{n} \frac{1}{\lambda_i} u_i u_i^T.$$

- u_i normalised Eigenvectors,
- c_i coefficients with uniform variance: $V_c = I$.

Eigenvalues Spectrum and Coefficients c_i

Eigenvalues λ_i
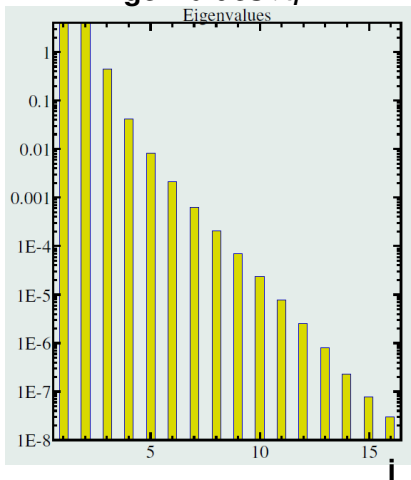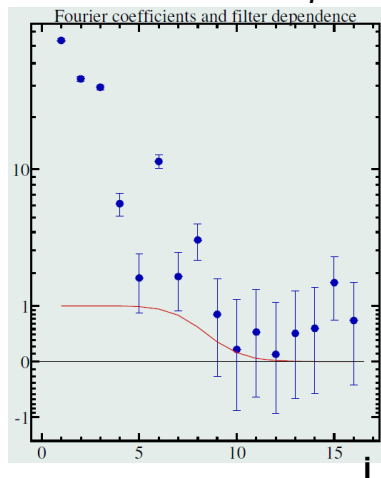
Coefficients c_i

- Eigenvalues decrease by orders of magnitude.
- Coefficients c_i of small λ_i are insignificant (since: $V_c = I$).

(Figures by V. Blobel)

Investigate Solution

$$\hat{x} = \sum_{i=1}^{n} \frac{1}{\sqrt{\lambda_i}} c_i u_i \qquad\qquad V_x = \sum_{i=1}^{n} \frac{1}{\lambda_i} u_i u_i^T$$

$V_c = I$: Insignificant coefficients (noise, follow $N(0,1)$)

- can make large contribution to $\hat{x}$
- if Eigenvalues are small!

normalised Eigen-vectors

$\Longleftarrow$

Eigen-vectors scaled by $\frac{1}{\sqrt{\lambda_i}}$ $\Longrightarrow$

increa-sing signifi-cance of c_i

decrea-sing λ_i
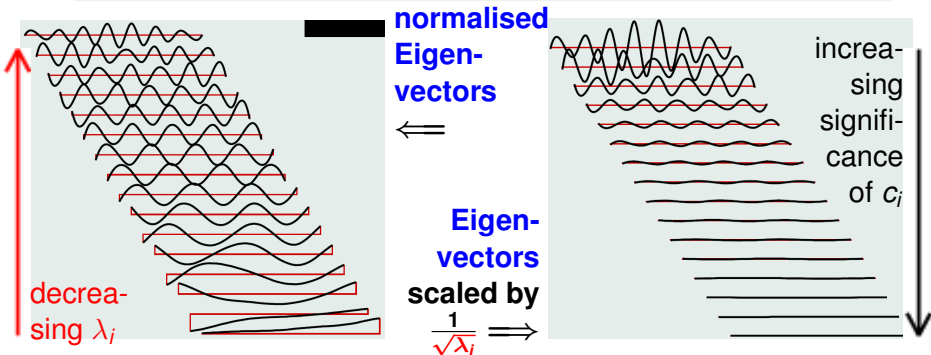
(Figures by V. Blobel)

How to deal with Insignificant High Frequency Eigenvectors?

Simple Approach

- Cut-off, i.e. ignore insignificant coefficients c_i with $k < i \leq n$
- $\hat{x} = \sum_{i=1}^{k} \frac{1}{\sqrt{\lambda_i}} c_i u_i$ $\qquad V_x = \sum_{i=1}^{k} \frac{1}{\lambda_i} u_i u_i{}^T$
- Without any bias.
- Problems:
  - V_x singular (rank k < size n), i.e. V_x^{-1} does not exist.
  - Exact k not unambiguous:
    Prefer smooth cut-off.

Art of Unfolding: Regularisation

- Add penalty term to the minimised expression to make a smooth cut-off.
- Stay (almost) bias free.

Regularisation

New Expression to minimise

- Minimise $(Ax - y)^T V_y^{-1}(Ax - y) + \tau(Lx)^T Lx$
- L: Matrix, depending on regularisation approach.
- τ: Regularisation parameter.

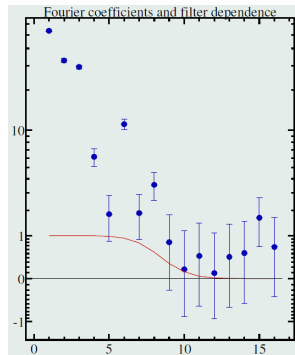- With $L = I$ (identity):
  $$\Rightarrow \hat{x} = \sum_{i=1}^{k} \frac{\phi_i}{\sqrt{\lambda_i}} c_i u_i,$$
  $$\phi_i = \frac{\lambda_i}{\lambda_i + \tau} = 1 \ldots 0.$$
- $\Rightarrow$ Damping of (insignificant) contributions of small Eigenvalues.
- But rank deficit of covariance V_x remains:
  - Fitting cumbersome without V_x^{-1}!
  - Collapse bins?

Fourier coefficients and filter dependence

- Minimise $(Ax - y)^T V_y^{-1}(Ax - y) + \tau(Lx)^T Lx$

Choice of Regularisation Scheme

- $L = I$: Norm of solution.
    - Easily usable for multi-dimensional distributions.
- L based on second derivatives expression.
    - Seems natural for physics: Assumption of smooth result.
    - Extension for more dimensions not straight forward.
- Entropy method adds a sum like $\tau \times \sum_i x_i \log x_i$.
    - Easily usable for multi-dimensional distributions.
- . . .

Choice of Regularisation Parameter

- Depending on curvature vs. χ^2 ("L-curve").
- Choose point of smallest global correlations.
- Match $\sum \phi_i = \sum \frac{\lambda_i}{\lambda_i + \tau}$ to number of significant coefficients c_i.
- . . .

Message to take away

Main inherent problems of unfolding:

- Smearing detector $A(y,x)$ washes out fine structure of truth $f(x)$ in measurement $g(y)$.
- Straight forward unfolding introduces high frequency terms from fluctuations in $g(y)$ (and large correlations).
- (Smooth) cut-off needed: regularisation.
- Often leads to covariance V_x being singular.

The Analysis Centre and Unfolding

Alliance Workshop in May 2010

- At DESY:
  https://indico.desy.de/conferenceDisplay.py?confId=3009.
- Presentation of experts.
- ⇒ Revealed controversy among experts.
- Presentation of experience with available implementations.

Some Algorithms

RUN

- by Volker Blobel (Uni Hamburg)
- regularised unfolding (*L* based on second derivatives)
- data internally represented by cubic B-splines instead of bins
- > 20 years old: FORTRAN

TRUEE

- C++ version of RUN
- by Natalie Milke (Uni Dortmund)
- extension for time dependence foreseen

GURU

- by Höcker and Kartvelishvili
- singular value decomposition (SVD - similar to diagonalisation) with regularisation

Some more Algorithms

TUnfold

- by S. Schmitt (DESY)
- least squares fit with free choice of regularisation scheme
- already shipped with ROOT (correct version since ROOT 5.27/04)

Iterative "Bayesian"

- by G. D'Agostini (University and INFN Roma1)
- Bayesian "knowledge update", starting with MC input
- various implementations (and also development)

Bin-by-Bin Corrections

- implemented in various analyses: $\hat{x}_i = N_i^{data} \left(\frac{N_i^{gen}}{N_i^{rec}} \right)_{MC}$
- not recommended: correct only for MC = truth

A Framework

RooUnfold

- by Tim Adye (RAL) et al.
- Framework containing
    - iterative "Bayesian" algorithm,
    - singular value decomposition (SVD),
    - bin-by-bin method (for comparison - not recommended).
- Since summer extended by
    - interface to TUnfold,
    - unregularised matrix inversion (for comparison).

Analysis Centre: Advertisement

Unfolding Framework

Getting people together,

- within and beyond Alliance.

to develop a **Framework for Unfolding Algorithms**:

- common code accessibility,
- unified data handling,
- test framework,
- documentation.

Will **not** work as a black box:

- Cannot relieve the user from the need to understand mechanisms, input,. . . .

Unfolding Framework

- Decided to base work on RooUnfold
    - add response object $A(y, x)$ with unbinned information (T. Adye),
    - integrate TRUEE (G.F., N. Milke),
    - work on improved version of "Iterative Bayesian" method
      (K. Bierwagen [Göttingen], J. Therhaag [Bonn]),
    - probably add simple Likelihood fit for comparison,
    - standalone, configurable main program (P. Cipriano [DESY]),
    - more test features.
- Discussing a systematic test of algorithms
    - "blind test" proposal (N. Gagunashvili [Iceland]),
    - exactly known detector response $A(y, x)$
    - comparing residuals, covariance, χ^2, . . .
        - (best "figure(s) of merit" yet unclear)
    - still in starting phase.

Interested to learn more?

- More details on Alliance Wiki:
  https://www.wiki.terascale.de/index.php/Unfolding_Framework_Project
- Alliance Workshop in May 2010 at DESY (proceedings to come):
  https://indico.desy.de/conferenceDisplay.py?confId=3009
- To follow the activities, sign-up the mailing list
  https://lists.desy.de/sympa/review/hep-unfolding
- Last day of PHYSTAT 2011 (17.-20.1.2011 at CERN) dedicated to unfolding:
  http://indico.cern.ch/conferenceDisplay.py?confId=107747

You want to contribute?

- Contact me: gero.flucke@desy.de
- Sign up mailing list
  https://lists.desy.de/sympa/review/hep-unfolding-dev
- Monthly EVO meetings.

Summary

- Unfolding needed to compare
  - experiments with each other
  - and with (future) theories.
- Ill-posed problem due to smearing effect of response matrix *A*:
  - high frequency terms need to be regularised.
- Experts debate about correctness of each others method.
- Analysis Centre is working on a
  Framework for unfolding algorithms, including tests of algorithms.
  - Your experience (and expertise) welcome.
  - Will pay back in future!

For the Future:

Stay tuned!