Regularization by Control of the Resolution Function

Michael Schmelling – MPI for Nuclear Physics

email: `Michael.Schmelling@mpi-hd.mpg.de`

Outline

- Introduction
- Singular Value Decomposition
- Unfolding
- Conclusions

➜ *The problem:*

$$a(x) = \int dy \; g(x, y) \cdot b(y)$$

Given an estimate for $a(x)$ and known $g(x, y)$ reconstruct an estimate for $b(y)$!

❖ *study Volker Blobel's classical example rescaled to* $[0, 1]$

$$b(y) = \frac{20.334}{100 + (10y - 2)^2} + \frac{2.0334}{1 + (10y - 4)^2} + \frac{4.0668}{4 + (20y - 15)^2}$$

$$g(x, y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(x - \left[y - \beta y^2\right]\right)^2\right) \cdot \left(1 - 4\alpha\left(y - \frac{1}{2}\right)^2\right)$$

- ➜ gaussian resolution function with $\sigma = 0.05$
- ➜ quadratic bias with $\beta = 0.1$
- ➜ parabolic efficiency loss towards phase space limits with $\alpha = 0.5$

The Discrete 1-dim Linear Unfolding Problem

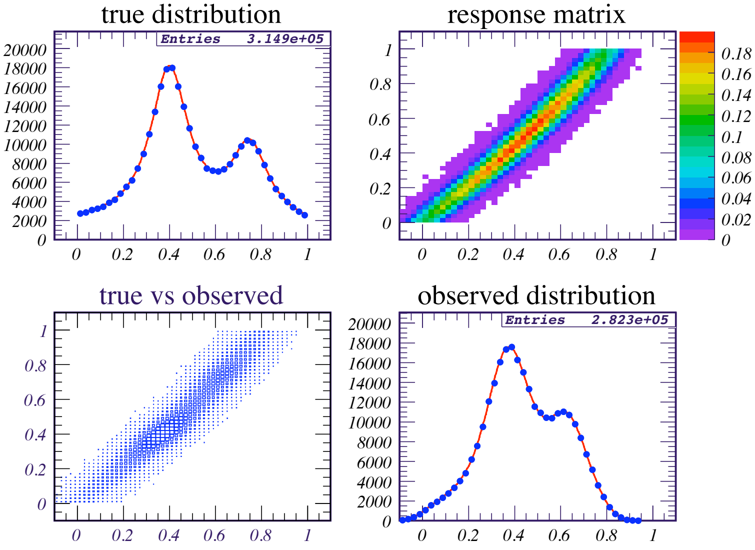
➜ the high energy physics use case

- $a(x)$ is estimated from a counting experiment
  - ➜ represented by a histogram with poisson errors on bin contents
- $b(y)$ is (proportional to) a cross-section, i.e. non-negative
  - ➜ conveniently also represented by a histogram
- $g(x, y)$ becomes a matrix mapping $b \to a$

$$a_k = \sum_{i=1}^{n} G_{ki} \cdot b_i \qquad \text{with} \qquad k = 1, 2, \ldots m$$

- ➜ reduced problem to infer only average densities over finite bin sizes
- ➜ furthermore:
  - ✘ assume the response matrix is exact, i.e. ignore quantization errors
  - ✘ consider “constrained” problems $m = n$

Numerical Simulation



Finite Resolution

➜ *Fourier-analysis of a signal*

schematically:

$$b(y) = \int d\omega \; A(\omega)\cos(\omega y)$$

effect of finite resolution:

$$a(x) = \int dy \; \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-y)^2/2\sigma^2} \cdot b(y) = \int d\omega \; e^{-\omega^2\sigma^2} \cdot A(\omega)\cos(\omega x)$$

- high frequency components in $a(x)$ exponentially suppressed
- accessible only with very large statistics
- complete unfolding even after discretization usually not possible
  - ➜ very unsatisfactory results from $b = G^{-1} \cdot a$
  - ➜ Zhigunov,83: *improvement of the resolution function* rather than *unfolding*
  - ➜ try to make this more quantative. . .

2. Singular Value Decomposition (SVD)

❖ *SVD for any matrix $A[m,n]$ with $m \geq n$*

$$A[m,n] = U[m,n] \cdot W[n,n] \cdot V[n,n]^T \quad \text{with}$$

$$U^T \cdot U = V^T \cdot V = V \cdot V^T = \mathbf{1}_n \quad \text{and positive definite diagonal matrix } W$$

➜ *diagonalization of the unfolding problem*

- transform measurements $x = M\,a$ such that $C(x) = M\,C(a)\,M^T = \mathbf{1}_m$
- SVD of the response matrix $(M\;G)$ of the transformed problem:

$$x = (M\;G)\,b = (U\;W\;V^T)\,b$$

- introduce normalized measurements u

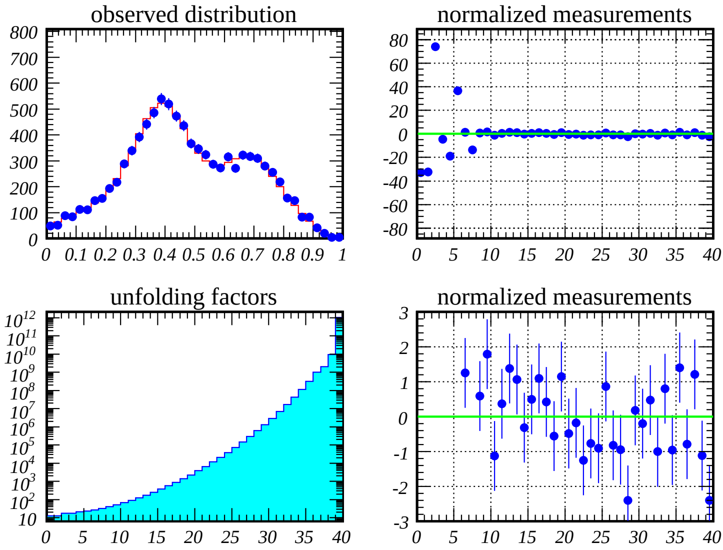
$$u = U^T x \quad \text{with} \quad C(u) = U^T\,C(x)\,U = \mathbf{1}_n$$

- diagonalized problem

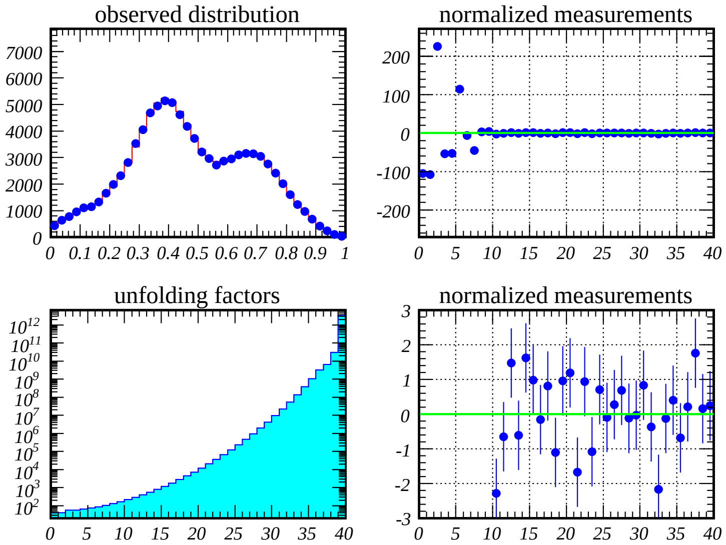
$$W^{-1}u = V^T\,b \quad \text{components:} \quad u_k / W_{kk} = \sum_{i=1}^{n} V^T_{ki}\,b_i$$

➜ expansion of b into orthogonal functions: $b_i = \sum_k (u_k / W_{kk})\,V^T_{ki}$

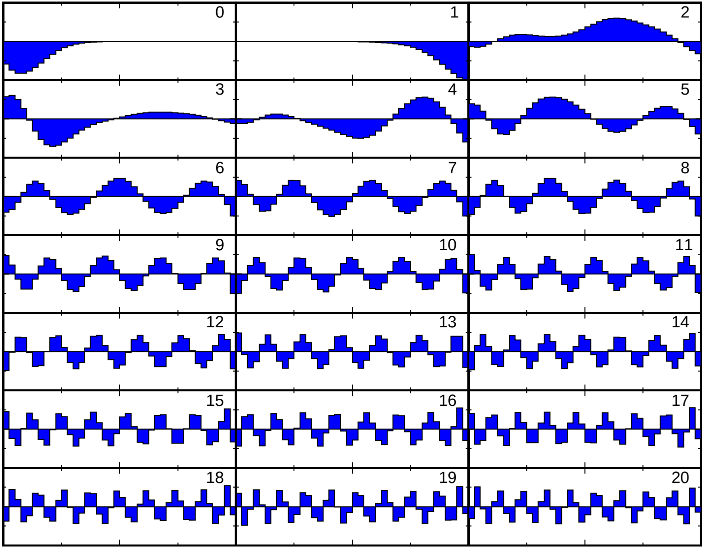
Numerical Example with 10^4 Events



Numerical Example with 10^5 Events



Orthogonal Functions of the Unfolded Distribution



Discussion

- measurements provide only limited information about the true distribution
  - ➜ 10^4 events: 5–10 coefficients
  - ➜ 10^5 events: 10–15 coefficients
- unstable results when using all coefficients, i.e. naive matrix inversion
- regularization: ignore measurements of higher order coefficients
  - ➜ set to zero, i.e. don't use higher order functions
  - ➜ replace by external criteria, using e.g. Maximum Entropy
    - ✘ least informative distribution
    - ✘ non-negative (if exists)
    - ✘ numerically efficient and unique solution
- resolution of unfolding result determined by measured part

➜ quantitative estimates . . .

Resolution of Truncated Expansions

- in principle given by Nyquist-Shannon theorem, however,
- want gaussian resolution to compare with initial response function. . .

❖ *study response to delta-functions*

input: $b_i = \delta_{iI}$

expansion: $u_k = \sum_{i=1}^{n} v_{ki} b_i = v_{kI}$

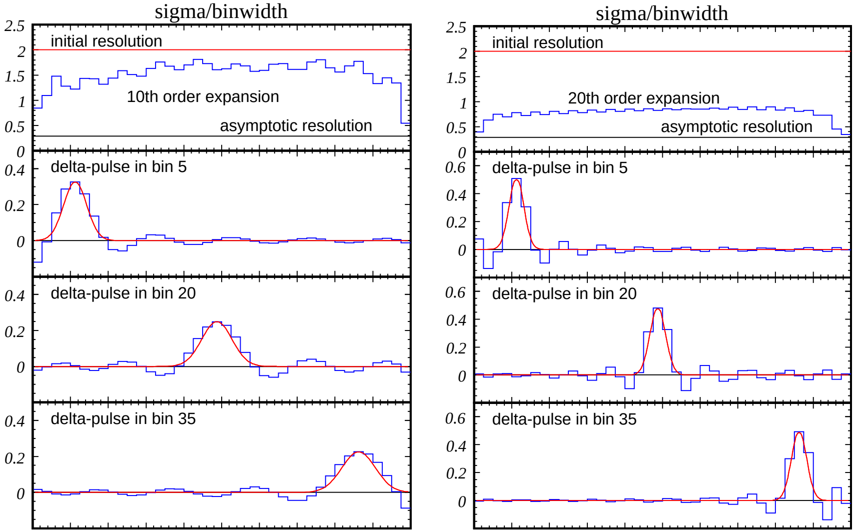
re-sythesis: $b_I = \sum_{k=1}^{m} u_k v_{kI} = \sum_{k=1}^{m} v_{kI}^2$ note: $b_I = 1$ if $m = n$

➜ content of central bin b_I fixes width of (normalized) gaussian

$$\frac{\sigma}{w} = \frac{1}{\sqrt{8}\, erf^{-1}(b_I)} \oplus \frac{1}{\sqrt{12}}$$ with $w =$ bin width

➜ illustration

Behaviour of the Numerical Example



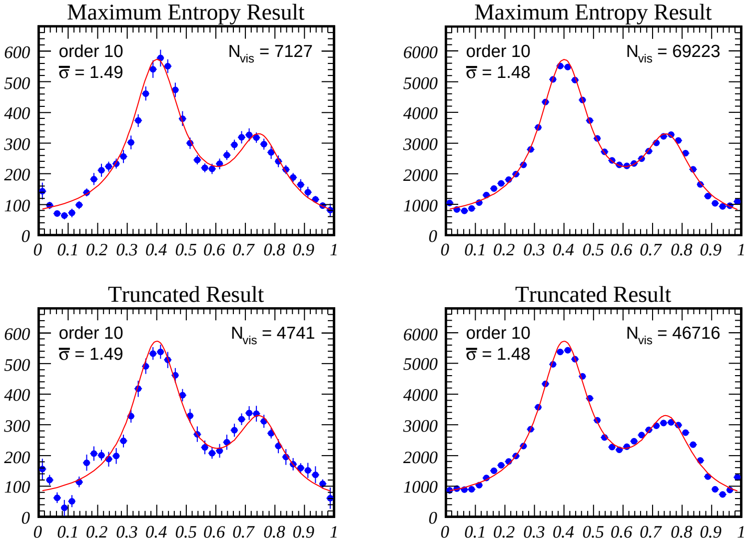
3. Unfolding

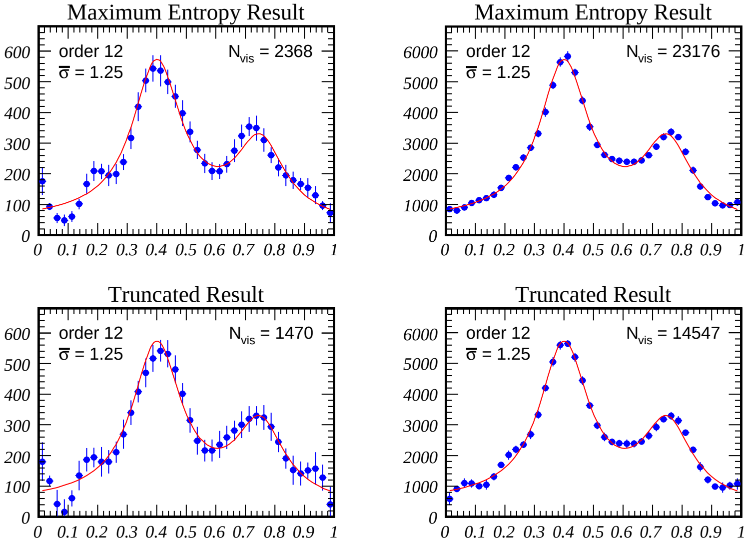
➜ *build unfolding result from measured leading order coefficients*

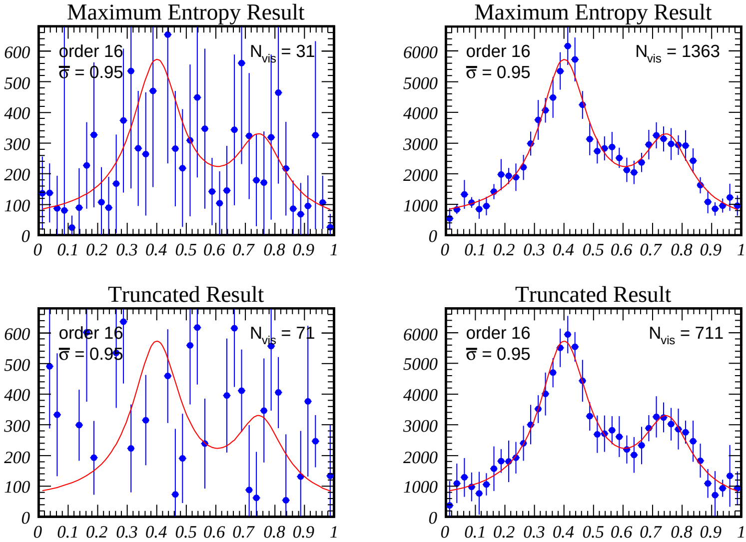
- regularization by truncation, i.e. limiting unfolded resolution
- compare two methods to supply higher order coefficients
  - ➜ case a: fixing to zero
  - ➜ case b: Maximum Entropy
- error propagation
  - ➜ case a: analytic linear error propagation
  - ➜ case b: error Monte Carlo to catch non-linearities
- observable figures-of-merit to look at
  - ➜ average resolution: $\bar{\sigma}$ (in units of bin width)
  - ➜ visible statistics: N_{vis}

$$N_{vis} = \frac{B^2}{\mathrm{Tr}\, C(b)} \qquad \text{with} \qquad B = \sum_{i=1}^{n} b_i$$

➜ study 10^4 and 10^5 observed events


Maximum Entropy Result
order 10
$\bar{\sigma}$ = 1.49
N_{vis} = 7127
Maximum Entropy Result
order 10
$\bar{\sigma}$ = 1.48
N_{vis} = 69223
Truncated Result
order 10
$\bar{\sigma}$ = 1.49
N_{vis} = 4741
Truncated Result
order 10
$\bar{\sigma}$ = 1.48
N_{vis} = 46716
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
0 100 200 300 400 500 600
0 1000 2000 3000 4000 5000 6000


Maximum Entropy Result
order 12
σ̄ = 1.25
N_vis = 2368
Maximum Entropy Result
order 12
σ̄ = 1.25
N_vis = 23176
Truncated Result
order 12
σ̄ = 1.25
N_vis = 1470
Truncated Result
order 12
σ̄ = 1.25
N_vis = 14547


Maximum Entropy Result
order 16
$\bar{\sigma}$ = 0.95
N_{vis} = 31
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
0 100 200 300 400 500 600
Maximum Entropy Result
order 16
$\bar{\sigma}$ = 0.95
N_{vis} = 1363
0 1000 2000 3000 4000 5000 6000
Truncated Result
order 16
$\bar{\sigma}$ = 0.95
N_{vis} = 71
Truncated Result
order 16
$\bar{\sigma}$ = 0.95
N_{vis} = 711

4. CONCLUSIONS

❖ *ansatz: determine regularization by looking at resolution*

- resolution controlled by the order of the expansion
- errors scale with initial statistics and resolution
- Maximum Entropy approach more stable than simple truncation
- within resolution, some artefacts may survive
  - ➜ combine bins to match FWHM, i.e. $2.35\bar{\sigma}$ (initially use 60 bins)
  - ➜ extend regularization to lower order coefficients

Work in progress – more to come. . .