Kernel based unfolding of data obtained from detectors with finite resolution and limited acceptance

N.D. Gagunashvili[*]

University of Akureyri, Borgir, v/Nordurslód, IS-600 Akureyri, Iceland

M. Schmelling

Max-Planck-Institut für Kernphysik, P.O. Box 103980, 69029 Heidelberg, Germany

Abstract

A kernel based procedure for correcting experimental data for distortions due to the finite resolution and limited detector acceptance is presented. The unfolding problem is known to be an ill-posed problem that can not be solved without some a priori information about solution such as, for example, smoothness or positivity. In the approach presented here the true distribution is estimated by a weighted sum of kernels, with the width of the kernels acting as a regularization parameter responsible for the smoothness of the result. Cross-validation is used to determine an optimal value for this parameter. A numerical example with a simulation study of systematical and statistical errors is presented to illustrate the procedure.

1. Introduction

In this paper the 1-dimensional unfolding problem will be addressed. Here the probability density function (PDF) $P(x')$ of an experimentally measured characteristic x' in general differs from the true physical PDF $p(x)$ because of the limited acceptance (probability) $A(x)$ to register an event with true

---

[*]Tel.: +354-4608505; fax: +354-4608998
Email address: nikolai@unak.is (N.D. Gagunashvili)

characteristic x and finite resolution in the response function $R(x'|x)$, the probability to observe x' for a given true value x. Formally the relation between $P(x')$ and $p(x)$ is given by

$$P(x') \propto \int_\Omega p(x)A(x)R(x'|x)\,dx\ .\tag{1}$$

The integration in (1) is carried out over the domain Ω of the variable x. In practical applications the experimental distribution is usually discretized by using a histogram representation, obtained by integrating $P(x')$ over n finite size bins

$$P_j = \int_{c_{j-1}}^{c_j} P(x')dx' \quad j = 1,\ldots,n\tag{2}$$

with c_{j-1}, c_j the bounds of bin j.

If a parametric (theoretical) model $p_T(x, a_1, a_2, \ldots, a_l)$ for the true PDF is known, then the unfolding can be done by determining the parameters in a least squares fit to the binned data [1] or a maximum likelihood fit to the unbinned data. In both cases the a priori information which is needed to correct for the distortions by the experimental setup is the fit model, which allows to describe the true distribution by a finite number of parameter values.

Model independent unfolding to identify a physical distribution, as considered in [2–12], is an underspecified problem and every approach to solving it requires a priori information about the solution. Different methods differ, directly or indirectly, in the use of this a priori information.

The remainder of the paper is organized as follows. In section 2 a new method for solving the unfolding problem will be presented. Properties of the algorithm are discussed in section 3 and illustrated in section 4 by applying it to a numerical example proposed in [3] and also used in Refs. [6, 7]. Conclusions are given in section 5.

2. Description of the unfolding method

To solve the unfolding problem (1) the following ansatz for $p(x)$ will be used

$$p(x) = w_0 + \sum_{i=1}^{s} w_i\,K(x, x_i, \lambda),\tag{3}$$

where the true distribution is written as an offset w_0 plus a weighted sum of s kernel functions (PDFs) $K(x, x_i, \lambda)$, $i = 1, \ldots s$, with non-negative weights w_i, central locations x_i and a scale parameter λ which determines the width of the kernel. Kernels are widely used for the estimation of a PDF [13] as well as in non-parametric regression analysis [14]. Note that Eq.(3) uses only kernels of one type with a common scale parameter. The only difference between different kernels is the location of the center. In this paper we will only consider this simplified case. In principle the approach could be generalized to vary also functional form and scale parameter.

Using Eq.(3) to parametrize the solution $p(x)$ reduces the unfolding problem of finding a solution from the infinitely many dimensional space of all functions to finding a solution in a finite dimensional space. This way a discretization is performed which, in contrast to e.g. a discretization by a histogram, has the advantage to introduce negligible quantization errors for sufficiently smooth distributions.

The following discussion will focus on symmetric kernels, although, depending on the kind of problem one attempts to solve, also asymmetric kernels may be appropriate. The a priori information of $p(x)$ being proportional to a PDF is incorporated by accepting only positive weights. The scale parameter of the kernel functions acts as a regularization parameter which allows to adjust the smoothness of the result. Weights, locations and the number of kernel functions needed to estimate $p(x)$ will be determined by the unfolding procedure described below.

Below examples of smooth symmetric kernels $K(x_i - x) = K(x_i + x)$ are presented. All kernels are PDFs which are normalized to unity when integrating over x. For convenience, in all cases the variable $u = (x - x_i)/\lambda$ is used. With the indicator function

$$I_{\{\cdots\}} = \left\{ \begin{array}{ll} 1 & \text{if } u \text{ satisfies the condition in the brackets} \\ 0 & \text{otherwise} \end{array} \right.$$

a class of polynomial kernels is defined by

$$K(u, \lambda) = \frac{N(a, b)}{\lambda}(1 - |u|^a)^b \, I_{\{|u| \leq 1\}} \; . \tag{4}$$

Often used are the following special cases:

kernel	a	b	$N(a,b)$
Epanechnikov	2	1	4/3
Biweight	2	2	15/16
Triweight	2	3	35/32
Tricube	3	3	70/81

Commonly employed non-polynomial kernels are:

$$\text{Cosine:} \quad K(u,\lambda) = \frac{\pi}{4\lambda} \cos(\frac{\pi u}{2}) \, I_{\{|u| \leq 1\}} \tag{5}$$

$$\text{Cauchy:} \quad K(u,\lambda) = \frac{1}{\lambda \pi}(\frac{1}{1+u^2}) \tag{6}$$

$$\text{Gaussian:} \quad K(u,\lambda) = \frac{1}{\lambda\sqrt{2\pi}} e^{-\frac{u^2}{2}} \tag{7}$$

Also frequently used is the piecewise defined cubic B-spline

$$K(u,\lambda) = \frac{1}{3\lambda} \begin{cases} (2u+2)^3 \, I_{\{-1 \leq u < -0.5\}} \\ (1+3(2u+1)(1-2u(2u+1))) \, I_{\{-0.5 \leq u < 0\}} \\ (1-3(2u-1)(1-2u(2u-1))) \, I_{\{0 \leq u < 0.5\}} \\ (2-2u)^3 \, I_{\{0.5 \leq u < 1\}} \end{cases} . \tag{8}$$

Re-writing Eq.(3) in the form

$$p(x) = \sum_{i=0}^{s} w_i \, K_i(x) \quad \text{with} \quad K_i(x) = \begin{cases} 1 & \text{for } i = 0 \\ K(x,x_i,\lambda) & \text{for } i > 0 \end{cases} \tag{9}$$

and substituting this into the basic equation (1) yields

$$P(x') = \sum_{i=0}^{s} w_i \int_{\Omega} K_i(x) A(x) R(x'|x) \, dx \tag{10}$$

Taking statistical fluctuations into account, the relation between the weights w_i and the histogram of the observed distribution becomes a linear equation

$$\boldsymbol{P} = \mathbf{Q}\boldsymbol{w} + \boldsymbol{\epsilon}, \tag{11}$$

where $\boldsymbol{P}$ is the n-component column vector of the experimentally measured histogram, $\boldsymbol{w} = (w_0, w_1, ..., w_s)'$ is $(s+1)$-component vector of weights and $\mathbf{Q}$ is an $n \times (s+1)$ matrix with elements

$$Q_{ji} = \int_{c_{j-1}}^{c_j} K_i(x) \, A(x) \, R(x'|x) \quad j = 1, \ldots n \quad i = 0, \ldots s. \tag{12}$$

4

The vector $\boldsymbol{\epsilon}$ is an n-component vector of random residuals with expectation value $E[\boldsymbol{\epsilon}] = \mathbf{0}$ and covariance matrix $\boldsymbol{C}$ with diagonal elements $\text{Var}[\boldsymbol{\epsilon}] = \text{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2)$, where σ_i is the statistical error of the measured distribution for the ith bin. Each column of matrix $\mathbf{Q}$ is the response of the system to the true distribution represented by the respective kernel. Numerically the calculation of the column vectors can be done by weighting the events of a Monte Carlo sample such that they follow the distribution the corresponding kernel, see Ref. [1], and taking the histogram of the observed distribution obtained with the weighted entries.

For a given set of kernels the weights $\boldsymbol{w}$ in Eq.(11) can be determined by a linear least squares fit. In order to have an as flexible as possible model, the candidate kernels in principle could have a continuous range of central positions. In practical applications it will usually be sufficient to consider a discrete set with a spacing significantly smaller than the bandwidth λ. The goal then is to find a subset of kernels for the final fit which provides a good description of the data and where all weights are positive and significantly different from zero. This at the same time stabilizes the solution and guarantees positiveness.

To find such an optimal subset, a forward stepwise algorithm [15] is used. It requires a criterion for the quality of the fit which will be taken the test statistic X_l^2,

$$X_l^2 = (\boldsymbol{P} - \boldsymbol{Q}\hat{w})^T \boldsymbol{C}^{-1}(\boldsymbol{P} - \boldsymbol{Q}\hat{w}) \tag{13}$$

where the index l denotes the number of weights in the fit and $\hat{w}$ is determined such that it minimizes X_l^2. The solution $\hat{w}$ and its covariance matrix $\boldsymbol{C}_w$ are given by the well known expressions

$$\hat{w} = (\boldsymbol{Q^T C^{-1} Q})^{-1} (\boldsymbol{Q^T C^{-1}}) \boldsymbol{P} \quad \text{and} \quad \boldsymbol{C}_w = (\boldsymbol{Q^T C^{-1} Q})^{-1} . \tag{14}$$

If the underlying distribution of the measured histogram $\boldsymbol{P}$ can be described by a linear combination of the columns of $\boldsymbol{Q}$, then the X_l^2 statistics follows a χ^2-distribution with $n - l$ degrees of freedom.

Now assume a total of s candidate kernel function $K_i(x), i = 1, \ldots, s$ with centers evenly spaced along the possible values x of the true distribution. In a first step the weight $\hat{w}_0$ is determined by fitting only the constant function K_0 to the data. Then an iterative procedure starts with alternating "Forward" and "Backward" steps described below.

Given a fit model consisting of l kernels, in the next Forward step each of the other $s - l$ kernels is tried for inclusion into the model. From all

combinations that one is selected where all weights are positive and which gives the largest reduction in X_l^2. If no such fit is found then the procedure stops. Otherwise the new kernel is included into the model if

$$\frac{X_l^2 - X_{l+1}^2}{X_{l+1}^2}(n - l - 1) > F_{in} \ , \tag{15}$$

i.e. if the reduction in χ^2 is sufficiently large. Also in case the best fit does not satisfy Eq.(15) the procedure stops. After accepting a new kernel into the model a Backward step is performed. Here in turn each of the previously included kernels is removed from the model and the test-model fitted to the data. From all fits which have only positive weights the one with the smallest increase in X_l^2 is taken. If the increase is below a certain threshold

$$\frac{X_{l-1}^2 - X_l^2}{X_l^2}(n - l) < F_{out} \tag{16}$$

the respective kernel is removed from the model, and the Backward step is iterated with the reduced model. If no kernel is removed then again a Forward Step is tried. The procedure stops if neither a Forward, nor a Backward Step can be done.

For the stepwise method defined above, appropriate thresholds F_{in} and F_{out} must be chosen. Usually one uses $F_{in} = F_{out} = F_0$. There is no common opinion about the best value for this constant. Reference [16] for example used $F_0 = 2.5$, the authors of Ref. [17] used $F_0 = 3.29$ for the same sample of data. To allow the inclusion of as many kernels as possible into the model, very small values F_0 can be used.

When the method stops an estimate $\hat{p}(x)$ has been found, defined by the locations $x_i, i = 1 \ldots, k$ of a set of kernel functions which are summed with weights $w_i, i = 0, \ldots, k$ to yield

$$\hat{p}(x) = \sum_{i=0}^{k} \hat{w}_i K_i(x) \ . \tag{17}$$

The error band around $\hat{p}(x)$ is given by $\sqrt{\mathrm{var}[\hat{p}(x)]}$, obtained by setting $x = y$ in the expression for the covariance between any two points x and y

$$\mathrm{cov}[\hat{p}(x), \hat{p}(y)] = \sum_{i,j=0}^{k} K_i(x) \, K_j(y) \, (\boldsymbol{C}_w)_{ij} \ . \tag{18}$$

A histogram representation for the unfolded distribution $\hat{p}(x)$ with m bins integrating over the x-intervals $[b_{i-1}, b_i]$, $i = 1, \ldots, m$ is obtained by

$$\hat{\boldsymbol{p}} = \mathbf{K}\,\hat{\boldsymbol{w}}, \tag{19}$$

where $\mathbf{K}$ is an $m \times (k+1)$ matrix with elements

$$\mathrm{K}_{ij} = \int_{b_{i-1}}^{b_i} K_j(x)\,dx \;. \tag{20}$$

The covariance matrix of $\hat{\boldsymbol{p}}$ is given by

$$\boldsymbol{C}_p = \mathbf{K}^T \boldsymbol{C}_w \mathbf{K} \;. \tag{21}$$

Note that this matrix is singular when the number of weights is smaller than the number of bins in the histogram of the unfolded distribution.

3. Discussion

The unfolding algorithm described above defines a generic approach to represent measured information about a true physical distribution in a compact way. The fact that the model is specified with proper statistical errors allows a quantitative comparison between an independent theoretical model and the unfolding result when working on the subspace spanned by the model $\hat{p}(x)$. To test the hypothesis that the underlying distribution of the unfolding result has the shape $p_T(x)$, one can use the histogram representation of $\boldsymbol{p}_T$ with the same binning as for $\hat{\boldsymbol{p}}$. In case of a non-singular covariance matrix $\boldsymbol{C}_p$ a χ^2-test can be applied directly on the binned distributions. If the number of bins for the unfolded distribution is larger than the number of weights, the comparison can still be done in the space spanned by the weights. In this case the weight vector $\boldsymbol{w}_T$ for expanding $\boldsymbol{p}_T$ into the kernels $\mathbf{K}$ is given by

$$\boldsymbol{w}_T = (\mathbf{K}^T\mathbf{K})^{-1}\,\mathbf{K}\,\boldsymbol{p}_T \tag{22}$$

which, in analogy to Eq.(14) is simply the unweighted fit of the kernel functions used to describe the model to the theoretical prediction $\boldsymbol{p}_T$. If $p_T(x)$ is indeed the underlying distribution of the unfolding result, then the test statistic

$$X^2 = (\hat{\boldsymbol{w}} - \boldsymbol{w}_T)^T \boldsymbol{C}_w^{-1} (\hat{\boldsymbol{w}} - \boldsymbol{w}_T) \tag{23}$$

has a χ^2-distribution with $k+1$ degrees of freedom, the rank of the matrix $\boldsymbol{C}_w$. It has to be emphasized that the above test constitutes only a necessary condition for a theoretical prediction to describe the data. It is not a sufficient one, as examples can be constructed where additional kernels would be needed to properly model the prediction, which may be known to be absent in the data and thus are ignored in the test. In practical applications one therefore also should make sure that $\mathbf{K}\,\boldsymbol{w}_T$ provides a good model for $p_T(x)$.

In principle any smooth kernels can be used and in practice results do not vary significantly when switching between the functions discussed before. The choice of the optimal type of kernel function and the value of the scale parameter λ for a given problem is driven by the quality of the fit. Common tools to asses the fit quality in regression analysis [15] are:

1. p-value of fit
2. analysis of the normalized residuals of the data
   (a) as a function of the estimated value $\hat{\boldsymbol{P}}$
   (b) as a function of the observed value x'
3. Q-Q plot: quantile of normalized residuals versus the theoretical quantile expected from a standard normal $\mathcal{N}(0,1)$ distribution

The positions of the kernels considered in the algoritm should cover the entire allowed range of x with a spacing significantly smaller than the width given by the scale parameter λ. In order to avoid loss of information due to binning the number of bins for the measured histogram $\boldsymbol{P}$ should be as large as possible although, in order to have meaningful error estimates for the least squares fits that determine $\hat{\boldsymbol{w}}$, the number of entries in a single bin should not be less than ~ 25.

An issue left open in the definition of the unfolding algorithm is the determination of the scale parameter λ of the kernel functions. Evidently, larger values will in general result in a more smooth estimate for the true distribution but may lead to bad fits of the observed distribution when narrow features cannot be accommodated. Too small values of λ, on the other hand, will favor overfitting of statistical fluctuations in the data. In general one will therefore try a range of values for λ and, in order to find some optimal balance between smoothness of the result and overfitting of the data, select a parameter in the region just below the largest value which provides a satisfactory fit to the data. In the literature [21–24] the use of cross-validation or bootstrap methods is suggested to find the optimal solution. Here we will

use a simple leave-one-out cross-validation approach [24] to determine the best value for λ.

Finally it should be noted that the unfolding method described above does not take into account uncertainties in the matrix $\boldsymbol{Q}$ which relates the weight vector $\hat{\boldsymbol{w}}$ to the measurements $\boldsymbol{P}$. Therefore, when $\boldsymbol{Q}$ is determined by means of a Monte Carlo simulation the Monte Carlo sample should be significantly larger than the data sample.

4. A numerical example

The method described above is now illustrated with an example proposed by Blobel [3] and for illustration also used elsewhere [6, 7]. The true distribution, defined on the range $x \in [0, 2]$ is described by a sum of three Breit-Wigner functions

$$p(x) \propto \frac{4}{(x - 0.4)^2 + 4} + \frac{0.4}{(x - 0.8)^2 + 0.04} + \frac{0.2}{(x - 1.5)^2 + 0.04} \qquad (24)$$

from which the experimentally measured distribution is obtained by

$$P(x') \propto \int_0^2 p(x) A(x) R(x'|x) dx, \qquad (25)$$

with an acceptance function $A(x)$

$$A(x) = 1 - \frac{(x - 1)^2}{2} \qquad (26)$$

and a resolution function describing a biased measurement with gaussian smearing

$$R(x'|x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x' - x + 0.05x^2)^2}{2\sigma^2}\right) \quad \text{with} \quad \sigma = 0.1 \ . \qquad (27)$$

The acceptance and resolution functions are shown in Fig. 1. Also shown is an example for the measured distribution obtained by simulating a sample of $N = 5000$ events.

For the determination of the matrix $\boldsymbol{Q}$ a sample of $500\,000$ Monte Carlo events was simulated. The true distribution was taken uniform and the kernel responses were calculated by weighting the Monte Carlo events with weights proportional value of kernel function [1]. A set of 100 gaussian kernels was
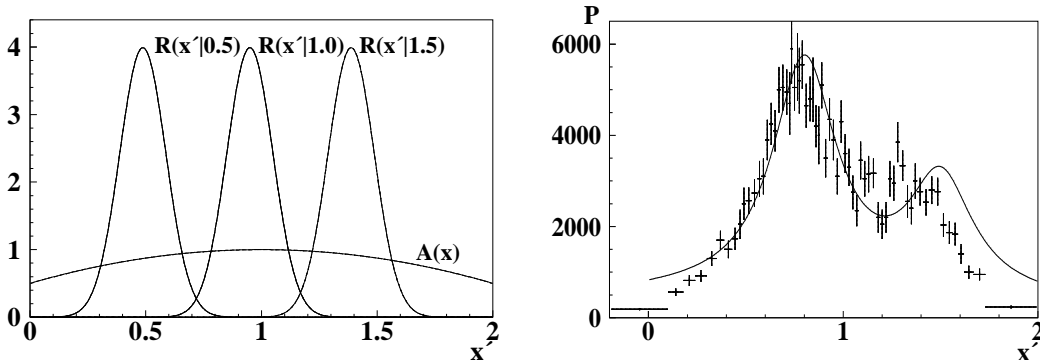
Figure 1: The acceptance function $A(x)$ and resolution function $R(x'|x)$ for $x = 0.5, 1.0$ and 1.5 (left) and histogram of the measured distribution $\boldsymbol{P}$ based on a sample of 5000 events generated for the true distribution (right). The true distribution $p(x)$ is shown by the curve.

used with positions uniformly distributed over the interval $[0, 2]$. For the nominal analysis a scale parameter $\lambda = 0.175$ and a threshold value $F_0 = 10^{-4}$ in the stepwise algorithm was chosen.

The estimate for the true distribution obtained by the unfolding method described above is represented by a constant plus a weighted sum of seven kernels. The positions and weights of the kernels determined by the stepwise algorithm together with the errors and correlation matrix of the weights is listed in Tab. 1. The quality of the unfolding result is illustrated by Fig. 2. It shows the superposition of the folded kernels approximates the measured distribution together with the analysis of the residual and the quantile-quantile plot. No structure in either of the control plots is observed. The p-value from the test for the comparison of the histogram of the measured distribution $\boldsymbol{P}$ and the fitted histogram $\hat{\boldsymbol{P}}$, Fig. 1(a), is $p = 0.23$.

Table 1 gives the results for a scale parameter $\lambda = 0.175$ of the gaussian kernels. To illustrate the effect of this parameter, Fig. 3 shows how the unfolding results varies with λ. The components of the unfolding results are shown together with the estimate $\hat{p}(x)$. Also shown are the error bands $\pm 2\sqrt{\mathrm{var}[\hat{p}(x)]}$ compared to the true distribution $p(x)$. Figure 4 illustrates for an even larger range of λ how the fit quality varies with the scale paramater. One clearly sees that large values $\lambda > 0.2$ lead to a bad fit with a p-value $p < 0.05$. Here also significant structures in the residuals and in the quantile-quantile plots are observed. The smallest value shows some indication of overfitting. The best parameters for this example evidently are in the range
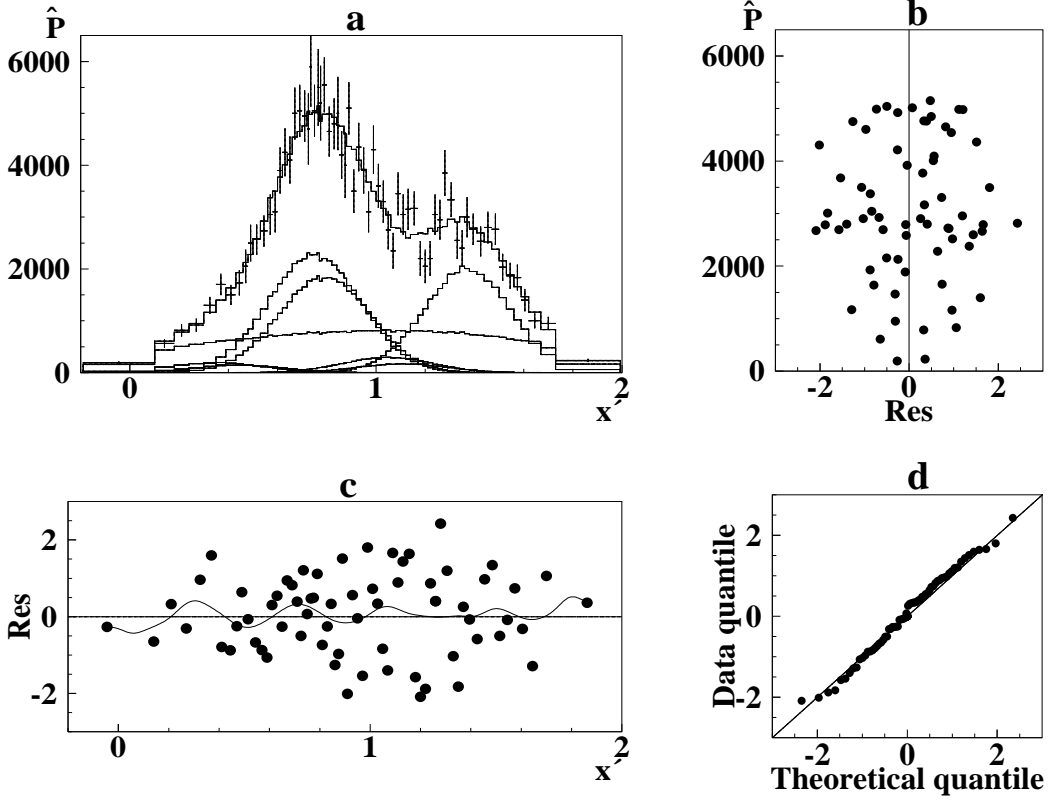
Figure 2: Illustration of the quality of the unfolding result. (a) folded kernels of the estimate of the true distribution compared to the measured distribution; (b) normalized residuals of the fit as a function of $\hat{\boldsymbol{P}}$; (c) normalized residuals as a function of x'; (d) quantile-quantile-plot for the normalized residuals.

between $0.15 < \lambda < 0.20$.

This is confirmed when doing a most simple leave-one-out cross validation, removing in turn each bin of the measured distribution and calculating the predicted residual sum of squares [24] as a function of λ

$$X_{pr}^2 = \sum_{i=1}^{n} \frac{(P_i - \hat{P}_{(i)})^2}{\sigma_i^2} \; . \tag{28}$$

Here $\hat{P}_{(i)}$ is the estimator for the content of the ith bin of the observed distribution $\boldsymbol{P}$, calculated by excluding this bin from the unfolding procedure or from the determination of the weights for the kernels selected by the unfolding procedure. The results of the calculation of X_{pr}^2/n for different

11

Table 1: Positions of kernels x_i, weights $\hat{w}_i$, errors of weights δ_i^w and correlation matrix for the weights determined by the unfolding algorithm.

i	x_i	$\hat{w}_i$	δ_i^w	0	1	2	3	4	5	6
0	—	1456.3	268.4							
1	0.33	122.8	282.6	-0.70						
2	0.77	1111.4	1691.1	-0.51	0.81					
3	1.18	79.1	919.4	0.54	-0.65	-0.86				
4	1.11	137.3	1139.7	-0.55	0.68	0.90	-0.99			
5	0.82	891.3	1816.6	0.49	-0.78	-0.99	0.88	-0.92		
6	0.43	85.1	350.5	0.53	-0.96	-0.88	0.66	-0.70	0.85	
7	1.50	1029.1	164.4	-0.82	0.68	0.66	-0.82	0.80	-0.66	-0.58

scale parameters λ is given in Tab. 2. The minimal value of $X_{pr}^2/n = 1.29$ is achieved for $\lambda = 0.2$. The choice of $\lambda = 0.175$ with $X_{pr}^2/n = 1.32$ gives a solution with $\sim 20\,\%$ larger statistical errors than for $\lambda = 0.2$ but, as will be discussed in more detail below, has a lower bias. The solutions with $\lambda < 0.15$ can be considered as overfitting the data while $\lambda \geq 0.20$ underfits them.

Table 2: The p-values and average predicted residual sum of squares X_{pr}^2/n for different values of the scale parameter λ.

λ	0.1	0.125	0.15	0.175	0.2	0.225	0.25	0.275
p-value	0.21	0.22	0.20	0.23	0.21	0.05	0.01	0.00
X_{pr}^2/n	1.77	1.42	1.46	1.32	1.29	1.58	1.84	2.02

To investigate the statistical properties of the unfolding procedure, $M = 1000$ simulation runs were performed producing statistically independent measured histograms, each based on $N = 5000$ events for the same true distribution (24). The unfolded distribution was calculated for each measured distribution. For the comparison between the unfolding results and the true distribution a histogram representation is used with $m = 12$ and alternatively $m = 40$ bins. Bin contents are normalized to the bin width in order to make the bin contents independent of the binning. The following quantities are considered for each bin i of the unfolded distribution.

- p_i: exact value of bin i of the true distribution

$$p_i = \frac{N}{x_i - x_{i-1}} \int_{x_{i-1}}^{x_i} p(x)\, dx$$
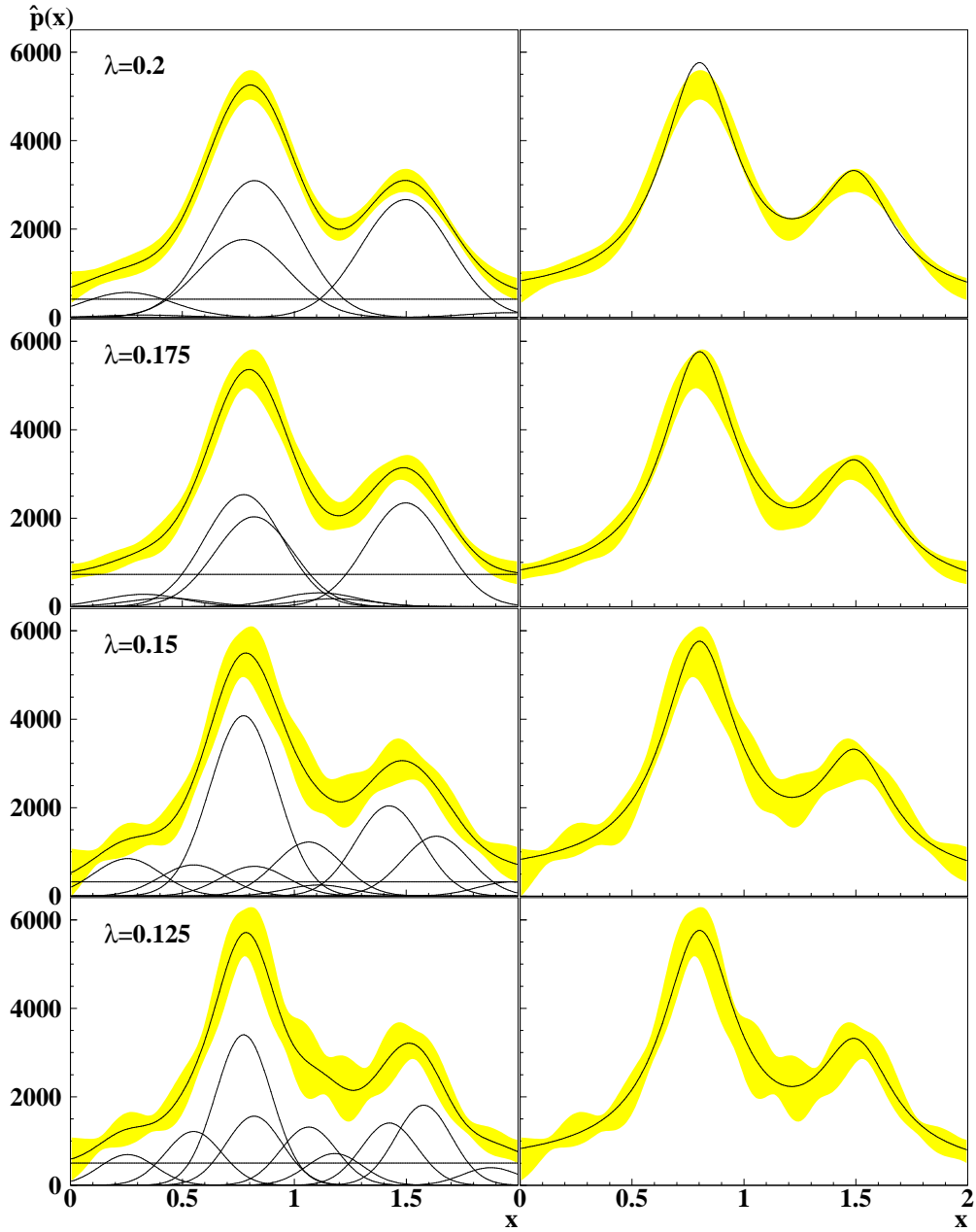
12

Figure 3: Components of the unfolded distribution and the unfolded distribution $\hat{p}(x)$ given by the sum of the components with $\pm 2\delta(x)$ interval (left) and the error band overlaid with the true distribution $p(x)$ (right) for different values of the scale parameter λ.
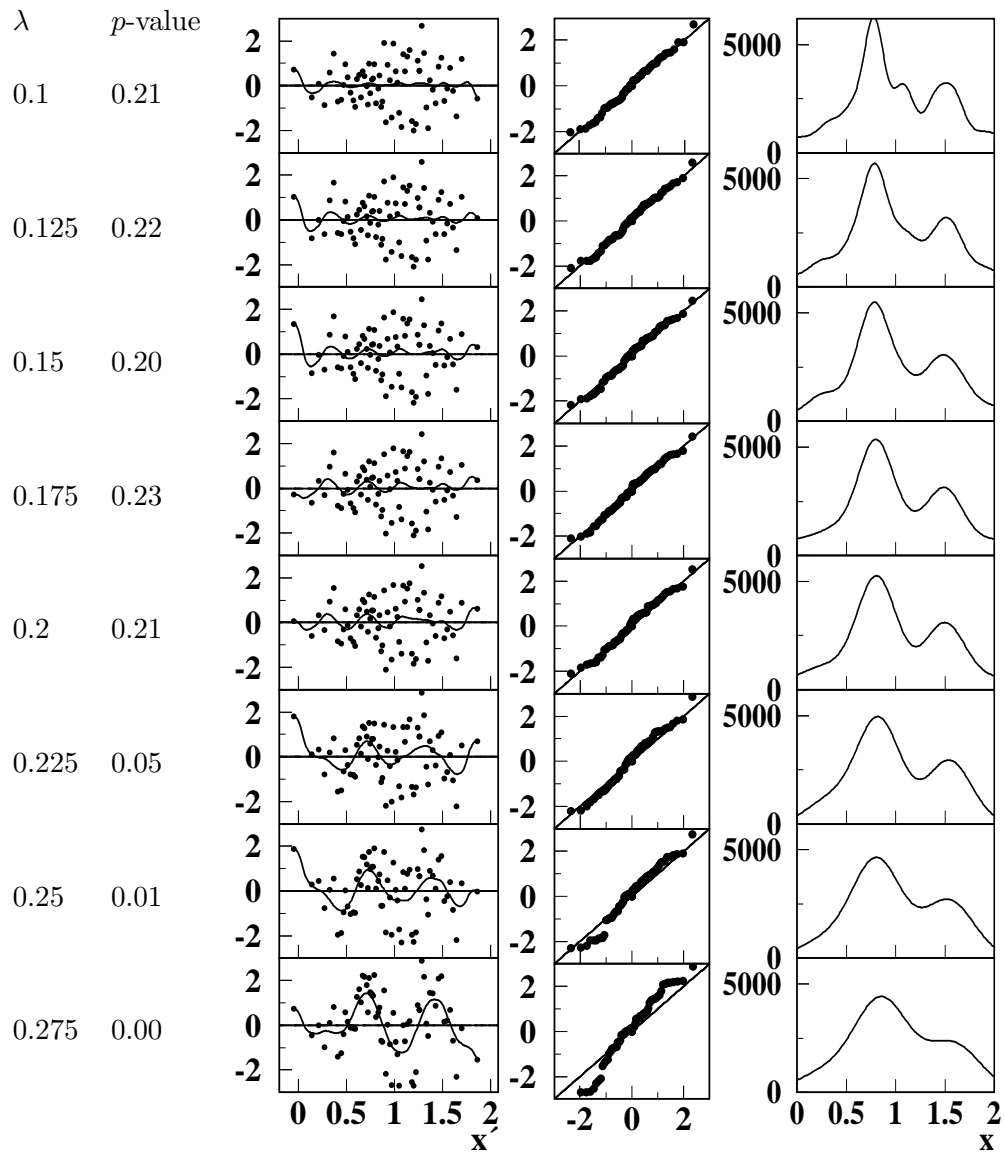
13

λ	p-value
0.1	0.21
0.125	0.22
0.15	0.20
0.175	0.23
0.2	0.21
0.225	0.05
0.25	0.01
0.275	0.00

Figure 4: The p-values, residuals, Q-Q plots and unfolded distributions for different values of the scale parameter λ.

14

- $\bar{\hat{p}}_i$: run-averaged value of bin i of the unfolded distribution

$$\bar{\hat{p}}_i = \frac{1}{M \cdot (x_i - x_{i-1})} \sum_{j=1}^{M} \hat{p}_i(j)$$

- $B[\hat{p}_i]$: bias in bin i of the unfolded distribution

$$B[\hat{p}_i] = \bar{\hat{p}}_i - p_i$$

- s_i: run-averaged standard deviation for bin i

$$s_i^2 = \frac{1}{M-1} \sum_{j=1}^{M} (\hat{p}_i(j) - \bar{\hat{p}}_i)^2$$

- δ_i: run-averaged error estimate for bin i

$$\bar{\delta}_i = \frac{1}{M \cdot (x_i - x_{i-1})} \sum_{j=1}^{M} \delta_i(j)$$

- $B[\delta_i]$: bias on the error of bin i

$$B[\delta_i] = \bar{\delta}_i - s_i$$

- $RMSE_i$: run-averaged Root Mean Square Error for bin i

$$RMSE_i^2 = \frac{1}{M} \sum_{j=1}^{M} (\hat{p}_i(j) - p_i)^2 = s_i^2 + B[p_i]^2$$

In addition to the bin-dependent quantities some global measures for the quality of the unfolding result are defined by summing over all m bins of the unfolded distribution.

- TRMSB: Total Root Mean Square Bias

$$TRMSB = \sqrt{\frac{1}{m} \sum_{i=1}^{m} B[p_i]^2}$$

15

- TRMSV: Total Root Mean Square Variance

$$\text{TRMSV} = \sqrt{\frac{1}{m}\sum_{i=1}^{m} s_i^2}$$

- TRMSE: Total Root Mean Square Error

$$\text{TRMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\text{RMSE}_i^2} = \sqrt{\text{TRMSB}^2 + \text{TRMSV}^2}$$

Numerical calculations of the characteristics of the unfolding procedure for 12 bins and gaussian kernels with $\lambda = 0.175$ are presented in Tab. 3. One sees that the bias is small compared to the statistical errors of the unfolding result and that the error estimates agree well with the actual scatter of the results. A visual representation of these findings for different values λ is given in Fig. 5 for $m = 12$ and $m = 40$ bins of the unfolded distribution. At the resolution of 12 bins the unfolding result is consistent with the true distribution, at 40 bins and $\lambda = 0.2$ one observes some systematic effects in the bias distributions. The bias gets smaller with decreasing values λ, but the errors become larger, which illustrates the well known "bias-noise complementary law" [2] that the noise grows when the regularization parameter tends to zero.

The behavior of the global characteristics using 12 or 40 bins for the unfolded distribution is shown in Fig. 6. The behavior in both cases is very similar. The plots show how with increasing scale parameter λ, i.e. stronger regularization, statistical errors decrease while the bias increases. Adding both contributions in quadature, the Total Root Mean Square Error shows a minimum around $\lambda = 0.175$, i.e. in the region also favored by the cross-validation approach for the determination of λ.

5. Conclusions

A new method for unfolding the true distribution from experimental data is presented. The unfolding problem is known as an ill-posed problem which can not be solved without some a priori information about solution. Smoothness and positiveness are examples for this type of information. In the proposed algorithm the unknown true distribution is represented as a weighted
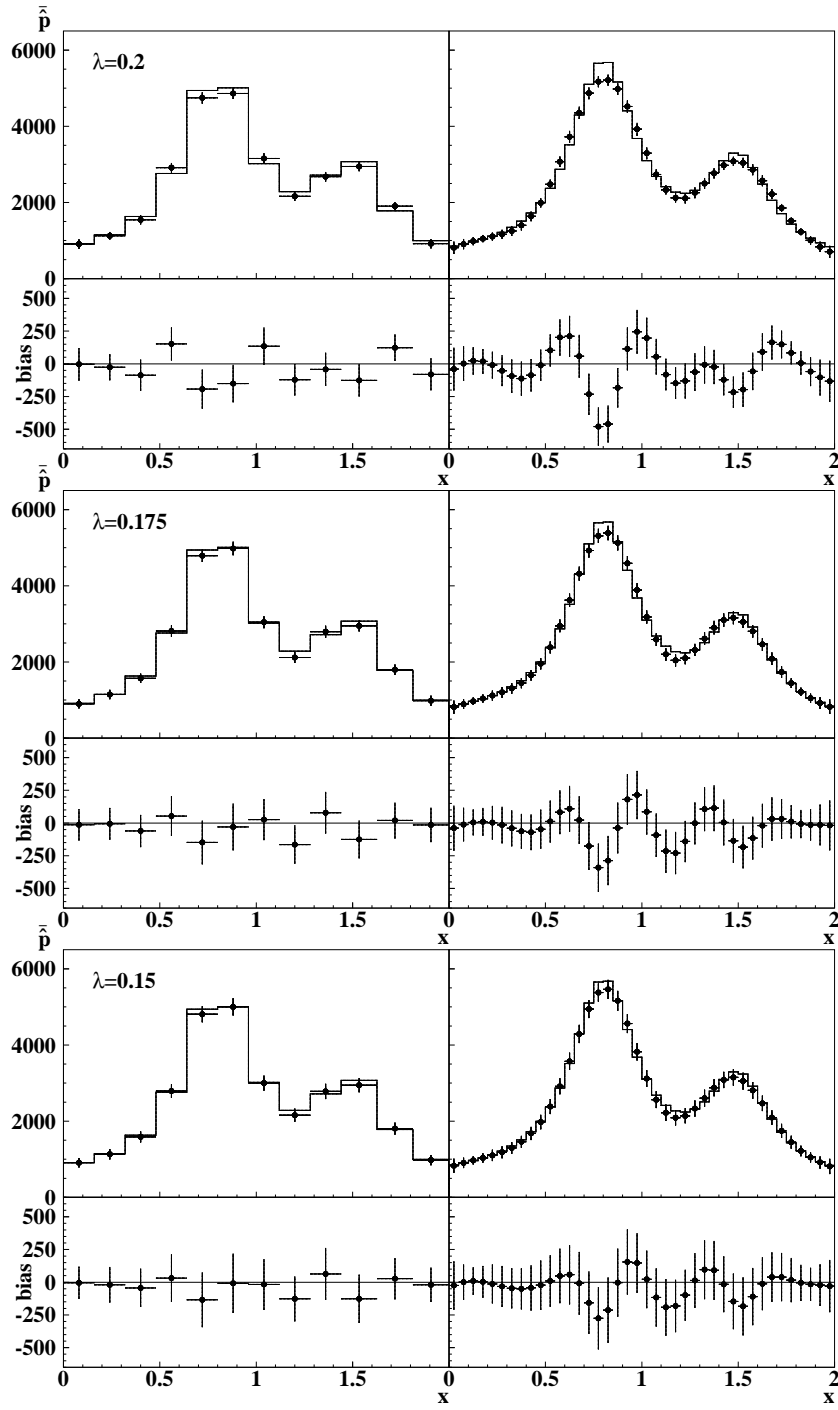
Figure 5: Average unfolding results $\bar{\hat{p}}$ and bias $B[\hat{p}_i]$ using $\lambda = 0.2$, $\lambda = 0.175$, and $\lambda = 0.15$ for $m = 12$ (left) and $m = 40$ (right) bins. The vertical error bars denote the standard deviations s_i. The histograms show the true bin contents p_i.

Table 3: Exact values of the bins of the true distribution p_i, average values $\bar{\hat{p}}_i$ from the unfolding procedure, bias $\mathrm{B}[\hat{p}_i]$, standard deviation s_i, mean error $\bar{\delta}_i$, bias of the calculated errors $\mathrm{B}[\delta_i]$ and Root Mean Square Errors $RMSE_i$ for $\lambda = 0.0175$.

i	p_i	$\bar{\hat{p}}_i$	$\mathrm{B}[\hat{p}_i]$	s_i	$\bar{\delta}_i$	$\mathrm{B}[\delta_i]$	RMSE_i
1	913.	900.	-13.	119.	123.	4.	120.
2	1152.	1146.	-5.	121.	123.	2.	121.
3	1631.	1570.	-61.	123.	129.	6.	137.
4	2760.	2813.	53.	152.	156.	5.	161.
5	4941.	4793.	-149.	167.	184.	17.	223.
6	5011.	4981.	-30.	177.	190.	13.	180.
7	3018.	3044.	26.	157.	164.	7.	159.
8	2284.	2119.	-165.	146.	157.	11.	220.
9	2718.	2797.	79.	159.	167.	8.	177.
10	3073.	2948.	-125.	144.	165.	21.	191.
11	1779.	1798.	19.	136.	150.	14.	138.
12	997.	983.	-14.	131.	127.	-4.	132.

sum of smooth kernels. The scale parameter of the kernels acts as a regularization parameter allowing to adjust the smoothness of the result. A cross-validation approach is proposed to determine an optimal value of this parameter. The method avoids discretization of the integral equation which is often done by unfolding methods and is an additional source of bias for the solution of unfolding problem. Various criteria were discussed to gauge the quality of the unfolding result. The methods provides a solution for the unfolding problem with a non-singular error matrix which can be used to test the consistency of a theoretical prediction with the experimental data. A numerical example including extensive simulation studies of the statistical properties of the method was presented to illustrate and to validate the procedure. For the example typical execution times per unfolding were found to be around 0.1 s on a 2 GHz CPU. The method can be extended to deal with steeply falling spectra or multidimensional distributions and to handle properly the case of limited statistics in the determination of the response function.
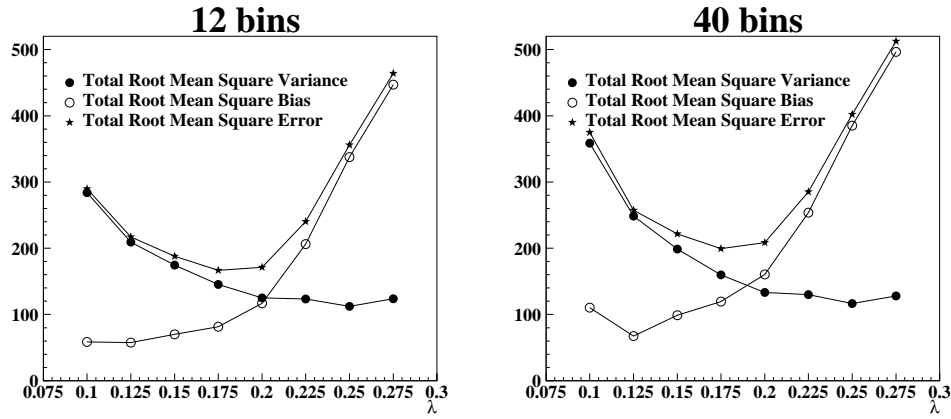
Figure 6: Global characteristics of the unfolding result for 12 bins (left) and 40 bins (right) as a function of the scale parameter λ.

References

[1] N. D. Gagunashvili, Nucl. Instr. Meth. A 120 (2010) 183–190.

[2] V. P. Zhigunov, Nucl. Instr. Meth. 216 (1983) 183–190.

[3] V. Blobel, CERN 85-02, 1985.

[4] V. B. Anikeev, A. A. Spiridonov, V. P. Zhigunov, Nucl. Instr. Meth. A 322 (1992) 280–285.

[5] N. Gagunashvili, Nucl. Instr. Meth. A 343 (1993) 606–609.

[6] M. Schmelling, Nucl. Instr. Meth. A 340 (1994) 400–412.

[7] A. Höcker, V. Kartvelishvili, Nucl. Instr. Meth. A 372 (1996) 469–481.

[8] L. Lindemann, G. Zech, Nucl. Instr. Meth. A 354 (1995) 516–521.

[9] G. D'Agostini, Nucl. Instr. Meth. A 362 (1995) 487–498.

19

[10] V. Blobel, An unfolding method for high-energy physics experiments, in: Proceedings of the Conference on Statistical Problems in Particle Physics, Durham, England, 18–22 March 2002, 258–267.

[11] N. Gagunashvili, Unfolding with system identification, in: Proceedings of the Conference on Statistical Problems in Particle Physics, Astrophysics and Cosmology, 12–15 September, 2005, Oxford, Imperial College Press, London, 2006, pp. 267–210.

[12] J. Albert et al., Nucl. Instr. Meth. A 583 (2007) 494–506.

[13] B. W. Silvermen, Chapman and Hall, 1986.

[14] W. Hardle, Applied nonparametric regression, Cambridge University Press, 1992.

[15] G. A. F. Seber and A. J. Lee, Linear Regression Analysis, John Wiley & Sons, 2003.

[16] M. A. Efroymson, Multiple regression analyses, In A. Ralston and H. S. Wilf (Eds.), Mathematical Methods for Digital Computers, Vol. 1, (1960) 191–203.

[17] N. Draper, H. Smith, Applied Regression Analysis, John Wiley & Sons, 1966.

[18] N. D. Gagunashvili, Nucl. Instr. Meth. A 614 (2010) 287–296.

[19] N. D. Gagunashvili, Comp. Phys. Comm. 183 (2012) 193–196.

[20] M. C. Jones, J. S. Marron, S. J. Sheather, J. Amer. Stat. Assoc. 91 (1996) 401–407.

[21] R. R. Picard and R. D. Cook, J. Amer. Stat. Assoc. 79 (1984) 575–583.

[22] N. J. Delaney and S. Chatterjee, J. Business & Econ. Stat. 4 (1986) 255–262.

[23] G. H. Golub, M. Heath and G. Wahba, Technometrics 21 (1979) 215–223.

[24] D. M. Allen, Technometrics 16 (1974) 125–127.