Unfolding Methods in Particle Physics[*]

Volker Blobel
University of Hamburg, Hamburg, Germany

Abstract

Measured distributions in particle physics are distorted by the finite resolution and limited acceptance of the detectors. The transformation to the underlying true distribution is called unfolding in particle physics and belongs to the class of linear inverse problems.

1 Inverse problems

1.1 Direct and inverse processes

The distributions $f(t)$ of a physics variable t to be measured in particle physics experiments are often not directly accessible. Because of limited acceptance and finite resolution the distribution $g(s)$ of the measured variable s is related to the distribution $f(t)$ by migration, distortions and transformations. Using Monte Carlo (MC) methods the direct process from an assumption $f(t)^{\text{model}}$ on the true distribution $f(t)$ to the expected measured distribution $g(s)$ can be simulated. The inverse process from the actually measured distribution $g(s)$ to the related *true* distribution $f(t)$ is difficult and ill-posed: small changes in the measured distribution can cause large changes in the reconstructed *true* distribution, if naive methods are used. In particle physics the inverse process is usually called *unfolding*. The direct and the inverse process

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.

are described by the Fredholm integral equation of the first kind

$$\int_\Omega K(s,t)\; f(t)\,\mathrm{d}t = g(s) \tag{1}$$

with a Kernel function $K(s,t)$ describing the physical measurement process (Refs. [1]– [4] and references therein). In particle physics the Kernel function $K(s,t)$ is usually implicitly known from a Monte Carlo sample based on an assumption $f(t)^{\text{model}}$.

1.2 Discretization and linear solution

The inverse problem given by the Fredholm integral equation has to be discretized in order to allow a numerical solution, with the result of the linear equation

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}\ . \tag{2}$$

The relations between the functions/distributions and the matrix and vectors are:

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 .

[*] This paper is the abridged version of a contribution to the forthcoming book O. Behnke et al. (eds.) with the working title: "Contemporary Data Analysis Methods", Wiley-VCH, ISBN 978-3-527-41058-3. The book will also contain more references.

The variables s, t and vectors $\boldsymbol{x}$, $\boldsymbol{y}$ are assumed to be one-dimensional in the following[1]. Several different discretization methods are possible. Real data are collected usually by integrating a signal over a short interval (bin), given by a grid $\{s_0, s_1, \ldots s_m\}$, often with equidistant bin limits in a histogram. The elements y_i correspond to integrals of $g(s)$ from s_{i-1} to s_i for $i = 1, 2, \ldots, m$ and are calculated according to equation (2) by the product $y_i = \boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x}$, where the vector $\boldsymbol{a}_i^{\mathrm{T}}$ is a row vector of matrix $\boldsymbol{A}$ and $y_i = A_{i1}x_1 + A_{i2}x_2 + \ldots + A_{in}x_n$. If the response is determined by a Monte Carlo sample, the same method can be used for the discretization $K(s,t) \Rightarrow \boldsymbol{A}$ and $f(t) \Rightarrow \boldsymbol{x}$; in this case element x_j is the average of $f(t)$ in bin j. Elements of the response matrix $\boldsymbol{A}$ are (positive) probabilities, and include the description of inefficiencies of the measurement detector. Other methods are possible, for example $f(t)$ can be discretized by a superposition of B-splines [3] which avoids discontinuities in the unfolded distribution, or the discretization can be based on numerical quadrature.

Assuming an accurate response matrix $\boldsymbol{A}$ and the relation $\boldsymbol{A}\,\boldsymbol{x}_{\text{exact}} = \boldsymbol{y}_{\text{exact}}$, the measured distribution deviates from the exact one only by statistical data errors. The data errors are represented by an m-vector $\boldsymbol{e}$, and the actually measured distribution $\boldsymbol{y}$ is given by

$$\boldsymbol{y} = \boldsymbol{y}_{\text{exact}} + \boldsymbol{e} = \boldsymbol{A}\,\boldsymbol{x}_{\text{exact}} + \boldsymbol{e}\ .$$

In particle physics the statistical properties of the measurements are usually well known. Often the elements of the vector $\boldsymbol{y}$ are counts, following Poisson statistics. In general the expectation value and variance are

$$\mathrm{E}\left[\boldsymbol{y}\right] = \boldsymbol{y}_{\text{exact}} \qquad\qquad \mathrm{V}\left[\boldsymbol{y}\right] = \mathrm{V}\left[\boldsymbol{e}\right] = \mathrm{E}\left[\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}\right] = \boldsymbol{V}_y\ , \tag{3}$$

i.e., an unbiased measurement $\boldsymbol{y}$ with $\mathrm{E}\left[\boldsymbol{e}\right] = 0$ is assumed, and the covariance matrix $\boldsymbol{V}_y$ of the measurement[2] is known.

In particle physics, unlike other fields, not only the result vector $\boldsymbol{x}$ has to be determined, but also the covariance matrix $\boldsymbol{V}_x$ of the result vector . If the linear Fredholm equation is solved for the estimate $\widehat{\boldsymbol{x}}$ by a linear transformation of the data vector according to $\widehat{\boldsymbol{x}} = \boldsymbol{A}^{\dagger}\boldsymbol{y}$, the propagation of the data uncertainties to the unfolding uncertainties is straightforward: $\boldsymbol{V}_x = \boldsymbol{A}^{\dagger}\boldsymbol{V}_y\boldsymbol{A}^{\dagger\mathrm{T}}$. The case $m = n$ with a quadratic matrix $\boldsymbol{A}$ could be solved by the inverse matrix $\boldsymbol{A}^{\dagger} = \boldsymbol{A}^{-1}$, but often the matrix $\boldsymbol{A}$ has a bad condition or is even singular and $m = n$ should be avoided. In the recommended case $m > n$ the n-by-m matrix $\boldsymbol{A}^{\dagger}$ can be constructed from the m-by-n matrix $\boldsymbol{A}$ and used to determine the estimate $\widehat{\boldsymbol{x}}$:

$$\widehat{\boldsymbol{x}} = \boldsymbol{A}^{\dagger}\boldsymbol{y} = \boldsymbol{A}^{\dagger}\,\boldsymbol{y}_{\text{exact}} + \boldsymbol{A}^{\dagger}\boldsymbol{e} = \boldsymbol{A}^{\dagger}\boldsymbol{A}\,\boldsymbol{x}_{\text{exact}} + \boldsymbol{A}^{\dagger}\boldsymbol{e}\ . \tag{4}$$

The pseudo-inverse $\boldsymbol{A}^{\dagger}$, also called Moore-Penrose generalized inverse, satisfying the relation $\boldsymbol{A}^{\dagger}\boldsymbol{A} = \boldsymbol{I}$, is a generalization of the inverse matrix, and allows the solution in the naive least squares sense, derived from the requirement

$$\min_{\boldsymbol{x}} F(\boldsymbol{x}) \quad \text{with} \quad F(\boldsymbol{x}) = \left(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\right)^{\mathrm{T}}\boldsymbol{V}_y^{-1}\left(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\right)\ , \tag{5}$$

where the inverse of the data covariance matrix $\boldsymbol{V}_y$ is included to take into account the different accuracy of the elements of the data vector. The least squares solution from the normal-equations formalism can be expressed by the pseudo-inverse

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

(with $\boldsymbol{A}^{\dagger}\boldsymbol{A} = \boldsymbol{I}$) with the matrix $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}_y^{-1}\boldsymbol{A} = \boldsymbol{C}$. The covariance matrix $\boldsymbol{V}_x$ is given by $\boldsymbol{V}_x = \boldsymbol{A}^{\dagger}\boldsymbol{V}_y\boldsymbol{A}^{\dagger\mathrm{T}} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{V}_y^{-1}\boldsymbol{A}\right)^{-1} = \boldsymbol{C}^{-1}$. Although the estimate $\widehat{\boldsymbol{x}}$ has the expectation $\boldsymbol{x}_{\text{exact}}$ (see equation

[1] The variables and the vectors can be multi-dimensional in practice, even with different dimensions for the true and the measured distribution.

[2] Covariance matrices are written with a subscript like $\boldsymbol{V}_y$; matrices $\boldsymbol{V}$ without subscripts are orthogonal matrices from a decomposition (Section 2).

(4)) because of $\boldsymbol{A}^\dagger \boldsymbol{A} \equiv \boldsymbol{I}$, this naive solution is often not satisfactory. It can be strongly oscillating with large negative correlation coefficients between neighbouring points and large positive correlation coefficients between next-to-immediate neighbours.

1.3 Parametrized unfolding

Unfolding was considered above to determine a discretized version $\boldsymbol{x}$ of a distribution $f(t)$ without a specific parametrization. A predicted probability density function (pdf) $f(t)$ without unknown parameters can be checked for compatibility with the data by *folding*; however folding does not provide information on the *sensitivity*. If a certain parametrization $f(t) \equiv f(t;\, \boldsymbol{a})$ depending on a vector of parameters $\boldsymbol{a}$ (to be fitted) is assumed, motivated e.g., by the theoretical analysis of the problem, this parametrization can be directly used in unfolding, using the response matrix $\boldsymbol{A}$, without the need to introduce a regularization. The bin content y_i is approximated using the elements of an auxiliary vector $\boldsymbol{x}$:

$$y_i = a_i^{\mathrm{T}} \boldsymbol{x} \quad \text{with} \quad x_j(\boldsymbol{a}) = \int_{t_{j-1}}^{t_j} \mathrm{d}t\, f(t;\, \boldsymbol{a}) \qquad\qquad j = 1,\, 2, \ldots,\, n \tag{6}$$

assuming a grid $\{t_0,\, t_1, \ldots t_n\}$ for the variable t. Unfolding is then the solution of the minimization problem

$$\min_{\boldsymbol{a}} F(\boldsymbol{a}) \quad \text{with} \quad F(\boldsymbol{a}) = (\boldsymbol{A}\boldsymbol{x}(\boldsymbol{a}) - \boldsymbol{y})^{\mathrm{T}}\, \boldsymbol{V}_y^{-1}\, (\boldsymbol{A}\boldsymbol{x}(\boldsymbol{a}) - \boldsymbol{y}) \ . \tag{7}$$

The function value $F(\widehat{\boldsymbol{a}}) = \chi^2_y$ should follow the χ^2-distribution with $m - n_{\text{par}}$ degrees of freedom, if the parametrization has n_{par} parameters. A standard fit program like MINUIT (CERN) with numerical derivatives can determine the parameter vector $\widehat{\boldsymbol{a}}$ and its covariance matrix.

An example of a parametrized unfolding, taken from Ref. [5], is shown in Figure 1. A pdf $f(t) = (1 + a\,t)\,/\,(1 + a/2)$ with t in the interval $[0, 1]$ is measured with a Gaussian resolution with standard deviation of 0.3. Figure 1(a) shows a simulated example for $a = 1$ with 5 000 entries of the measured distribution in the interval $[-0.3, 1.3]$. A 20-by-20 response matrix is determined by a simulation of 50 000 cases, using a uniform distribution (parameter $a = 0$) in $[0, 1]$. The result of the parameter fit according to equation (7) with the result $\widehat{a} = 1.09 \pm 0.18$ is shown in Figure 1(b) together with the simulated true histogram. Figure 1(c) shows the histogram of the fitted slope a from 10^5 simulations, together with a Gaussian curve of standard deviation 0.18; the fitted parameter has on average the correct value $a = 1$ with a slightly asymmetric distribution. These results agree with the results of Ref. [5].

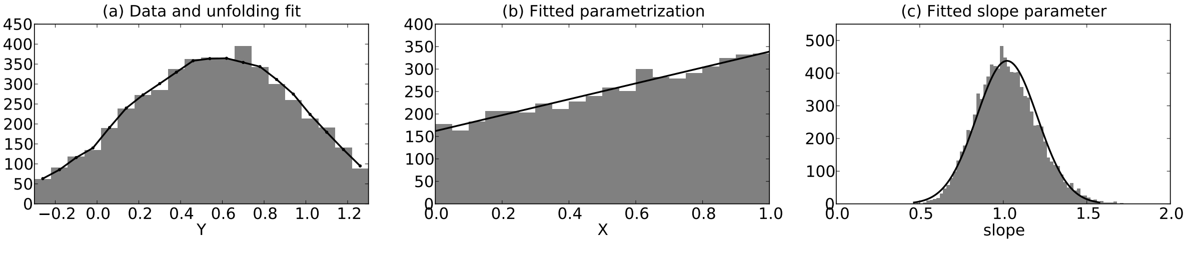

Fig. 1: Example for parametrized unfolding

1.4 Convolution and deconvolution

A function $f(t)$ with period 1 can be approximated by a sum of cosine functions, which is a complete system, periodic in $[0, 1]$ and orthogonal in the interval $0 \le t \le 1$. The approximation is given by $f(t) = a_0 + a_1 \cos(\pi t) + a_2 \cos(2\pi t) + \cdots$. The terms are the basis functions of the discrete cosine transformation, shown in Figure 2. The special case of a Kernel $K(s,\, t) \equiv K(s - t)$ is called a *convolution* and the inverse process is called *deconvolution*. A convolution of the function $f(t)$ by a Gaussian

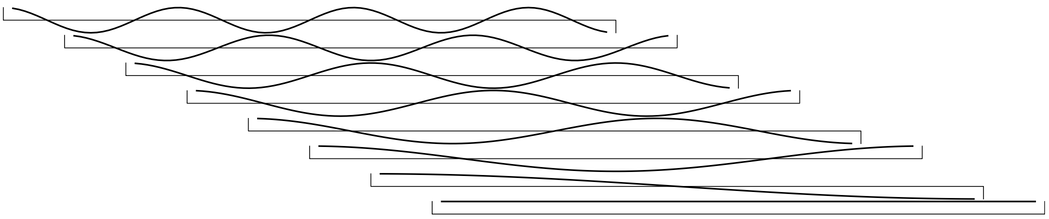

Fig. 2: The first eight basis functions of the discrete cosine transformation over the range $0 \ldots 1$

resolution function with standard deviation σ is considered. For a single term $\cos(k\pi t)$ the form of the term is not changed by the convolution with the Gaussian:

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

but the amplitude is attenuated by an exponential factor, which will become $\ll 1$ for larger indices k. The convoluted function $g(s)$ (see equation (1)) is *smoother* than $f(t)$ and can be approximated again by a cosine sum with coefficients α_k instead of a_k. The coefficients α_k of the convoluted function $g(s)$ will become small and negligible asymptotically much faster than of the original function $f(t)$. Deconvolution is simple in this case: the coefficients α_k, determined from $g(s)$, have to be multiplied by the inverse exponential factor, to reconstruct the coefficients a_k. With increasing index k, the exponential correction factors of the coefficients α_k soon become extremely large, increasing the relative uncertainty of the coefficients by a factor $\gg 1$. Thus the number of terms of the original function $f(t)$ which can be reconstructed is *limited* because of the finite resolution.

2 Solution with orthogonalization

2.1 Singular value decomposition (SVD)

The standard numerical method for the analysis of ill-posed problems $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ is the singular value decomposition (SVD) of the m-by-n matrix matrix $\boldsymbol{A}$, defined for any m and n. Assuming $m \geq n$ (called *thin* SVD) the SVD is of the form[3] with elements

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

where $\boldsymbol{U} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n) \in \mathbb{R}^{m\times n}$ and $\boldsymbol{V}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n) \in \mathbb{R}^{n\times n}$ are matrices with orthonormal columns and the diagonal matrix $\boldsymbol{\Sigma} = \textbf{diag}\{\sigma_1, \ldots, \sigma_n\} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{V}$ has non-negative diagonal elements σ_i, called singular values, in non-increasing order. The *condition* of matrix $\boldsymbol{A}$ is defined as the ratio of the largest to the smallest singular vector: $\mathrm{cond}\left(\boldsymbol{A}\right) = \sigma_1/\sigma_n$. The condition is an upper bound on the magnification factor of the ratio of relative errors of the estimate $\widehat{\boldsymbol{x}}$ to the data $\boldsymbol{y}$. The m-vectors $\boldsymbol{u}_i$ and the n-vectors $\boldsymbol{v}_i$ are called left and right singular vectors of $\boldsymbol{A}$. The SVD matrices with the property $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U} = \boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{I}$ will be used for the least squares solution $\widehat{\boldsymbol{x}}$ of the problem. The singular vectors $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ have an increasing number of sign-changes with increasing index and decreasing singular value, similar to the cosine functions in Figure 2 .

In order to take the uncertainty of the data $\boldsymbol{y}$, given by the covariance matrix $\boldsymbol{V}_y$, into account, a pre-scaling (also called pre-whitening) of the problem is required. For uncorrelated data this is achieved by dividing the rows of the linear system by the standard deviation $\sqrt{(\boldsymbol{V}_y)_{ii}}$ of the data. The fastest

[3] This is a decomposition into *outer* products of two vectors. The outer product $\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}$ of two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, also called dyadic product, is a rank-1 matrix $\boldsymbol{B}$ with elements $B_{jk} = a_j b_k$.

method for correlated data is based on the Cholesky decomposition of the matrix $\boldsymbol{V}_y = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}$ with an upper triangular matrix $\boldsymbol{R}$. In the following it is assumed that a pre-scaling of $\boldsymbol{A}$ and $\boldsymbol{y}$ has already been done (i.e., $\boldsymbol{A} := (\boldsymbol{R}^{-1})^{\mathrm{T}}\boldsymbol{A}$ and $\boldsymbol{y} := (\boldsymbol{R}^{-1})^{\mathrm{T}}\boldsymbol{y}$) with the result $\boldsymbol{V}_y = \boldsymbol{I}$, before the singular value decomposition. If the elements y_i are counts with standard deviation $\sqrt{y_i}$, the magnitude of the singular values is proportional to the number of measured events. For the case of a Gaussian response matrix with standard deviation σ the decrease of the singular values is approximately described by the exponential factor $\exp\left(-ak^2\sigma^2\right)$ (with some constant a) in the convolution example from Section 1.4.

2.2 Symmetric eigenvalue decomposition

The eigenvalue decomposition of a symmetric n-by-n matrix $\boldsymbol{C}$ is the orthogonalization method to be used in maximum likelihood methods based of the Poisson statistics [3] ($\boldsymbol{C} =$ Hessian) and in normal-equations least squares ($\boldsymbol{C} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$, assuming pre-scaling of matrix $\boldsymbol{A}$), and can be achieved by a SVD. In the SVD of this matrix the left and right singular vectors are identical:

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

The diagonal matrix $\boldsymbol{\Lambda} = \textbf{diag}\left\{\lambda_1, \lambda_2, \ldots\right\}$ has non-negative diagonal elements λ_i, called eigenvalues, equal to the square of the singular values σ_i of the matrix $\boldsymbol{A}$. The symmetric eigenvalue decomposition of the matrix $\boldsymbol{C}$ is mathematically equivalent to the singular value decomposition of the matrix $\boldsymbol{A}$.

2.3 Least squares using the SVD

The use of the SVD in least square problems allows some insight into the structure of the matrix $\boldsymbol{A}$ of ill-posed problems $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$. The matrix product $\boldsymbol{A}\boldsymbol{x}$ expressed using the SVD matrices

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

shows that contributions to $\boldsymbol{y}$ with small singular values σ_j, corresponding to higher-frequency contributions, are suppressed. If all singular values are non-zero, the least squares estimate $\widehat{\boldsymbol{x}}$ is given by

$$\widehat{\boldsymbol{x}} = \boldsymbol{A}^{\dagger}\boldsymbol{y} = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{y}\right) = \sum_{j=1}^{n}\frac{1}{\sigma_j}\left(\boldsymbol{u}_j^{\mathrm{T}}\boldsymbol{y}\right)\boldsymbol{v}_j = \sum_{j=1}^{n}\frac{1}{\sigma_j}c_j\boldsymbol{v}_j . \tag{8}$$

The data $\boldsymbol{y}$ with unit covariance matrix are transformed by $\boldsymbol{U}^{\mathrm{T}}$ to an n-vector $\boldsymbol{c} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{y}$ with unit covariance matrix $\boldsymbol{V}_c = \boldsymbol{I}$, representing the transformed measurement. The elements $c_j = \boldsymbol{u}_j^{\mathrm{T}}\boldsymbol{y}$ of $\boldsymbol{c}$, called *Fourier coefficients*, tend to decrease rather fast towards small values for larger indices j, if the distribution described by $\boldsymbol{y}$ is smooth. The coefficients c_j are *independent* and, having a *variance of* 1, show the significance of the corresponding contribution to the estimate $\widehat{\boldsymbol{x}}$. The value of a Fourier coefficient c_j will follow a Gaussian $N(0,1)$, if the exact value is small compared to the standard deviation 1. The expression (8) for the estimate $\widehat{\boldsymbol{x}}$ shows that the contribution to the estimate $\widehat{\boldsymbol{x}}$ related to a single Fourier coefficient c_j is multiplied by the *inverse* of the singular value σ_j. Small singular values σ_j will generate large fluctuations in the unfolding result $\widehat{\boldsymbol{x}}$, and can make the result unacceptable. The calculation of the uncertainty of the estimate $\widehat{\boldsymbol{x}}$ is straightforward, because of the linear transformation of the data $\boldsymbol{y}$ in the expression (8); the covariance matrix is

$$\boldsymbol{V}_x = \boldsymbol{A}^{\dagger}\boldsymbol{V}y\boldsymbol{A}^{\dagger^{\mathrm{T}}} = \boldsymbol{V}\boldsymbol{\Sigma}^{-2}\boldsymbol{V}^{\mathrm{T}} = \sum_{j=1}^{n}\left(\frac{1}{\sigma_j^2}\right)\boldsymbol{v}_j\boldsymbol{v}_j^{\mathrm{T}} . \tag{9}$$

In other methods e.g., iterative methods (Section 4) an estimate $\widehat{\boldsymbol{x}}$ is determined without the construction of a transformation matrix like $\boldsymbol{A}^{\dagger}$, which makes the above uncertainty calculation impossible.

2.4 Null space and truncated SVD

The SVD defines by matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ a new basis for the measured data and the unfolding result in a frequency space. The measured data $\boldsymbol{y}$ are transformed to independent Fourier coefficients $c_j = \boldsymbol{u}_j^{\mathrm{T}} \boldsymbol{y}$ with fixed standard deviation 1 (*white* noise). The least-square estimate $\widehat{\boldsymbol{x}}$ can be written in the form $\widehat{\boldsymbol{x}} = \sum_j d_j \boldsymbol{v}_j$ with coefficients $d_j = c_j/\sigma_j$, that are still independent, but have standard deviations $1/\sigma_j$ increasing with index j; this property could be called *blue* noise because the uncertainty is increasing with the frequency. Typically the singular values σ_j of a response matrix $\boldsymbol{A}$ decrease to small values without a clear gap between large and small singular values. Due to rounding and other errors there will be no exactly zero singular values, but taking into account potential uncertainties of the elements of matrix $\boldsymbol{A}$ at least a few singular values may be effectively zero, reducing the *effective rank* of the matrix $\boldsymbol{A}$ to a number p (less than n), which is an upper limit on the number of contributions. Especially if the response matrix is determined by a Monte Carlo simulation there are unavoidable uncertainties in the elements. A tolerance δ can be defined to determine the effective rank p by $\sigma_p > \delta \geq \sigma_{p+1}$ with

$$\delta = \epsilon \times \max_{1 \leq i \leq m} \sum_{j=1}^{n} |A_{ij}| \ , \tag{10}$$

where e.g., $\epsilon = 0.01$, if the elements A_{ij} are correct to about two digits, as is the case of typical Monte Carlo calculations. Small singular values $\sigma_j < \delta$ would give meaningless contributions to the solution. Assuming an effective rank of p (less than n), the estimate $\widehat{\boldsymbol{x}}$ of equation (8) can be written in the form

$$\widehat{\boldsymbol{x}} = \underbrace{\sum_{j=1}^{p} d_j \boldsymbol{v}_j}_{\boldsymbol{x}_{\text{range}} \in \mathbb{R}^p} + \underbrace{\sum_{j=p+1}^{n} \widetilde{d}_j \boldsymbol{v}_j}_{\boldsymbol{x}_{\text{null}} \in \mathbb{R}^{n-p}} \ . \tag{11}$$

The first term $\boldsymbol{x}_{\text{range}}$, with contributions $d_j = c_j/\sigma_j$, is a rather well-defined element of a p-dimensional subspace of the $\mathbb{R}^n$, but the second term $\boldsymbol{x}_{\text{null}}$ has arbitrary contributions $\widetilde{d}_j$, which in the product $\boldsymbol{A}\widehat{\boldsymbol{x}}$, multiplied by the singular value σ_j, have essential no effect on the expected data $\widehat{\boldsymbol{y}}$. Because the two terms in $\widehat{\boldsymbol{x}} = \boldsymbol{x}_{\text{range}} + \boldsymbol{x}_{\text{null}}$ (equation (11)) are orthogonal, the squared norm of $\widehat{\boldsymbol{x}}$ is the sum of the two squared norms $\left\|\boldsymbol{x}_{\text{range}} + \boldsymbol{x}_{\text{null}}\right\|^2 = \left\|\boldsymbol{x}_{\text{range}}\right\|^2 + \left\|\boldsymbol{x}_{\text{null}}\right\|^2$. The solution recommended in textbooks is the minimum-norm solution with $\widehat{\boldsymbol{x}}_{\text{null}} = 0$ and $\|\widehat{\boldsymbol{x}}\| = \left\|\boldsymbol{x}_{\text{range}}\right\|$. In this case the n-by-n covariance matrix $\boldsymbol{V}_x$ has a rank defect of $n - p$ and cannot be inverted. Alternatively the dimension of estimate $\widehat{\boldsymbol{x}}$ can be reduced to p, with a full-rank p-by-p covariance matrix $\boldsymbol{V}_x$ (see Section 3.6). In a simulation a data sample of 5000 events is generated with $n = 20$ and $m = 40$, assuming a Gaussian response. The effective rank of $\boldsymbol{A}$ as estimated from equation (10) with $\epsilon = 0.01$ is 18. Figure 3 shows the Fourier coefficients $|c_j|$ and the contributions $|d_j|$. The coefficients c_j for $j \geq 8$ are insignificant, giving a lower limit of the number of contributions. The (insignificant) contributions increase after $j = 10$ and the last contributions ($j \geq 16$) would dominate the result. Truncation after $j = 10$ gives an acceptable result without bias.

3 Regularization methods

3.1 Regularization

The standard method for the solution of ill-posed problems is the *regularization* method [1, 2]. The expression to be minimized w.r.t. the unfolding result includes the least squares (equation (5)) or (negative) log-likelihood expression, which ensure a good description of the measured distribution. A second term $\Omega(\boldsymbol{x})$, often of the form $\Omega(\boldsymbol{x}) = \|\boldsymbol{L}\boldsymbol{x}\|^2$ with a certain matrix $\boldsymbol{L}$, requires certain properties like smoothness of the unfolding result and contributes with a weight, given by a regularization parameter $\tau > 0$:

$$\min_{\boldsymbol{x}} F(\boldsymbol{x}) + \tau \|\boldsymbol{L}\boldsymbol{x}\|^2 \ .$$

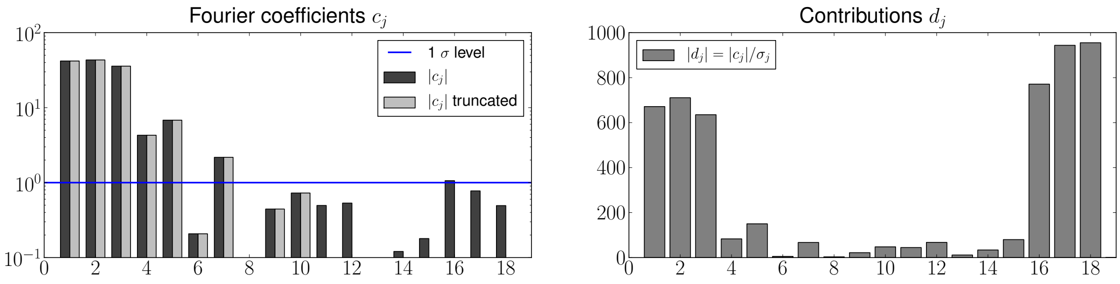

Fig. 3: Truncation

In the regularized solution of the least squares case (equation (5)) the matrix $\boldsymbol{A}^\dagger$ is replaced by the regularized matrix $\boldsymbol{A}^{\#}$:

$$\widehat{\boldsymbol{x}} = \boldsymbol{A}^{\#}\boldsymbol{y} = \left[\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \tau\boldsymbol{L}^{\mathrm{T}}\boldsymbol{L}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\right]\boldsymbol{y} \; . \tag{12}$$

The regularization term $\tau\boldsymbol{L}^{\mathrm{T}}\boldsymbol{L}$ is added to the matrix $\boldsymbol{C} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ of the normal equations, and inserting $\boldsymbol{y} = \boldsymbol{A}\,\boldsymbol{x}_{\text{exact}} + \boldsymbol{e}$ one obtains

$$\widehat{\boldsymbol{x}} = \boldsymbol{A}^{\#}\boldsymbol{A}\,\boldsymbol{x}_{\text{exact}} + \boldsymbol{A}^{\#}\boldsymbol{e} = \boldsymbol{x}_{\text{exact}} + \underbrace{\left(\boldsymbol{A}^{\#}\boldsymbol{A} - \boldsymbol{I}\right)\boldsymbol{x}_{\text{exact}}}_{\text{systematic error}} + \underbrace{\left(\boldsymbol{A}^{\#}\boldsymbol{e}\right)}_{\text{statistical error}} \quad . \tag{13}$$

The product $\boldsymbol{\Xi} \equiv \boldsymbol{A}^{\#}\boldsymbol{A}$ is called the *resolution matrix*. For the regularization scheme the resolution matrix is not equal to the unit matrix, and thus the method has a systematic bias $(\boldsymbol{\Xi} - \boldsymbol{I})\,\boldsymbol{x}_{\text{exact}}$. The fact that the regularized solution has a potential bias of the estimate, which depends on the details of the exact distribution $\boldsymbol{x}_{\text{exact}}$, is connected with the attempt to reduce the unnatural oscillations, which are unmeasurable. The *smoothing* effect of the resolution matrix gives no or small systematic errors for smooth exact distributions, and large systematic deviation for unphysical oscillating distributions. The measured distribution $\boldsymbol{y}$ has to be compared with the distribution $\widehat{\boldsymbol{y}}$ corresponding to the estimated unfolded distribution $\widehat{\boldsymbol{x}}$ and given by $\widehat{\boldsymbol{y}} = \boldsymbol{A}\widehat{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{A}^{\#}\boldsymbol{y}$, where the m-by-m product matrix $\boldsymbol{A}\boldsymbol{A}^{\#}$ is often called the *influence matrix*. The agreement between the measured data $\boldsymbol{y}$ and the vector $\widehat{\boldsymbol{y}}$ predicted by the influence matrix and checked with $\chi^2_y = (\widehat{\boldsymbol{y}} - \boldsymbol{y})^{\mathrm{T}}(\widehat{\boldsymbol{y}} - \boldsymbol{y})$ has to be acceptable.

The deviation of the resolution matrix $\boldsymbol{\Xi} = \boldsymbol{A}^{\#}\boldsymbol{A}$ from the unit matrix $\boldsymbol{I}$, which corresponds to a potential bias, should avoid the unnatural properties of naive unregularized solutions. Sometimes this term is called a penalty function, which seems to express a certain impact on the solution with a bias of the regularized result. For applications in particle physics a non-negligible bias is not acceptable. Below it is shown that the regularization ansatz can be used to separate the significant from the insignificant contributions of the result *without the introduction of a disturbing bias.*

3.2 Norm regularization

The simplest case is the *norm regularization* with $\boldsymbol{L} = \boldsymbol{I}$. For a given value of τ the estimate $\widehat{\boldsymbol{x}}$ can be determined by standard methods of linear algebra (matrix inversion), because of the good condition of the combined matrix. However the solution by the SVD is simple in this case and has several advantages, especially as it allows a clear understanding of the effects of regularization. Using the SVD the solution can be written in the form

$$\widehat{\boldsymbol{x}} = \boldsymbol{V}\;\underbrace{\left[\left(\boldsymbol{\Sigma}^2 + \tau\boldsymbol{I}\right)^{-1}\boldsymbol{\Sigma}^2\right]}_{\text{filter factor matrix } \boldsymbol{F}}\;\boldsymbol{\Sigma}^{-1}\underbrace{\left(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{y}\right)}_{\text{coeff.}\boldsymbol{c}} = \left(\boldsymbol{V}\;\boldsymbol{F}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathrm{T}}\right)\boldsymbol{y} \; ,$$

where the matrix $\boldsymbol{F}$ is diagonal with elements φ_j. Comparison with the unregularized solution (8) shows the additional filter factors φ_j for each term with a strength which depends on the regularization parameter τ, while the Fourier coefficients $\boldsymbol{c}_j$ are defined as before. The estimate $\widehat{\boldsymbol{x}}$ and its covariance matrix (compare (9)) can be expressed by sums:

$$\widehat{\boldsymbol{x}} = \sum_{j=1}^{n} \frac{1}{\sigma_j} \varphi_j \, c_j \boldsymbol{v}_j \qquad \boldsymbol{V}_x = \sum_{j=1}^{n} \left(\frac{1}{\sigma_j^2} \right) \varphi_j^2 \, \boldsymbol{v_j} \boldsymbol{v_j}^{\mathrm{T}} \qquad \text{with} \quad \varphi_j = \frac{\sigma_j^2}{\sigma_j^2 + \tau} \tag{14}$$

(the squared singular values σ_j^2 are replaced by eigenvalues in case of diagonalization). The filter factors φ_j represent a smooth cut-off (with $\varphi_k = 0.5$ if $\tau = \sigma_k^2$), which can avoid a certain oscillating behaviour (Gibbs phenomenon) in the truncation case. No bias will be introduced if the selected regularization parameter τ is small enough to reduce only the insignificant Fourier coefficients.

The norm regularization corresponds to the original regularization proposal by Tikhonov and by Philipps. The regularization parameter τ can be interpreted as the introduction of the a-priori measurement error $s_{\text{reg}} = 1/\sqrt{\tau}$ for each component of the vector $\boldsymbol{x}$. Individual values of $s_{j,\text{reg}}$ for the components could be introduced, corresponding to a regularization term $\Omega(\boldsymbol{x}) = \sum_j x_j^2/s_j^2$. Norm regularization can be used for unfolding problems with rather smooth solutions $\overline{\boldsymbol{x}}$, requiring only a small number of Fourier coefficients; in other cases some modifications are advisable. One possibility is to change the regularization term $\Omega(\boldsymbol{x}) = \|\boldsymbol{L}\boldsymbol{x}\|^2$ to a term $\Omega(\boldsymbol{x}) = \|\boldsymbol{L}(\boldsymbol{x} - \boldsymbol{x}_0)\|^2$ with some a-priori assumption $\boldsymbol{x}_0$ on the resulting vector $\boldsymbol{x}$; this will reduce the number of significant terms. Another possibility is to make the Monte Carlo simulation with an a-priori assumption $f(t)^{\text{model}}$ about the function $f(t)$, and to include the function $f(t)^{\text{model}}$ already in the definition of the response matrix,

$$\int_{\Omega} \left[K(s,t) \; f(t)^{\text{model}} \right] f^{\ddagger}(t) \, \mathrm{d}t = g(s) \ ;$$

only an almost constant correction function $f^{\ddagger}(t)$ has to be determined with $f(t) = f(t)^{\text{model}} f^{\ddagger}(t)$. This option is available in unfolding methods of particle physics [3, 4]. The elements A_{ij} of the matrix $\boldsymbol{A}$, which includes $f(t)^{\text{model}}$, are now integers, the number of Monte Carlo events from bin j of $\boldsymbol{x}$, measured in bin i of $\boldsymbol{y}$.

3.3 Regularization based on derivatives

Another regularization scheme is based on *derivatives*; most popular are the second derivates, and this scheme often has advantages over the norm regularization. The matrix $\boldsymbol{L}$ is rather simple if equidistant bins are used. For example the second derivative in bin j is proportional to $(-x_{j-1} + 2x_j - x_{j+1})$ and corresponds to a row $\ldots \; -1 \quad 2 \; -1 \ldots$ of the matrix $\boldsymbol{L} \in \mathbb{R}^{(n-2)\times n}$. The solution (12) can again be obtained, for a given regularization factor τ, by matrix inversion. However, a solution using orthogonalization provides an understanding of the details and the separation of significant from insignificant contributions.

Orthogonalization is more complicated than for the norm regularization, because the term $\tau \boldsymbol{L}^{\mathrm{T}} \boldsymbol{L}$ is not diagonal. The *generalized singular value decomposition* can to be used for the corresponding orthogonalization. The orthogonal solution is formally equivalent to the solution (12), with a different definition of the singular or eigenvalues. Compared to the norm regularization the Fourier coefficients refer to a rotated system according to $\boldsymbol{L}^{\mathrm{T}} \boldsymbol{L}$. If, alternatively, the eigenvalue decomposition is used, two rotations (and a scaling) are required to diagonalize simultaneously [3] the two symmetrical matrices $\boldsymbol{C} = \boldsymbol{A}^{\mathrm{T}} \boldsymbol{A}$ and $\boldsymbol{L}^{\mathrm{T}} \boldsymbol{L}$ for the solution of the normal equation $\left(\boldsymbol{C} + \tau \boldsymbol{L}^{\mathrm{T}} \boldsymbol{L}\right) \boldsymbol{x} = \boldsymbol{b}$ with $\boldsymbol{b} = \boldsymbol{A}^{\mathrm{T}} \boldsymbol{y}$. The first diagonalization $\boldsymbol{C} = \boldsymbol{U}_1 \boldsymbol{\Lambda} \boldsymbol{U}_1^{\mathrm{T}}$ is used to rewrite the equation in the form

$$\boldsymbol{U}_1 \boldsymbol{\Lambda}^{1/2} \left(\boldsymbol{I} + \tau \boldsymbol{M} \right) \boldsymbol{\Lambda}^{1/2} \boldsymbol{U}_1^{\mathrm{T}} \boldsymbol{x} = \boldsymbol{b}$$

with the transformed regularization matrix $\boldsymbol{M} = \boldsymbol{\Lambda}^{-1/2} \boldsymbol{U}_1^{\mathrm{T}} \left(\boldsymbol{L}^{\mathrm{T}} \boldsymbol{L} \right) \boldsymbol{U}_1 \boldsymbol{\Lambda}^{-1/2}$. The second diagonalization $\boldsymbol{M} = \boldsymbol{U}_2 \boldsymbol{S} \boldsymbol{U}_2^{\mathrm{T}}$ is used to rewrite the equation in the form

$$\boldsymbol{R} \left(\boldsymbol{I} + \tau \boldsymbol{S} \right) \boldsymbol{R}^{\mathrm{T}} \boldsymbol{x} = \boldsymbol{b}$$

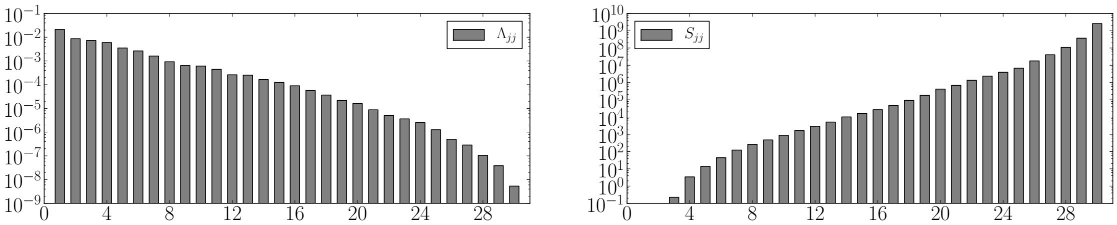

Fig. 4: Eigenvalues Λ_{jj} and S_{jj}

$$\widehat{\boldsymbol{x}} = \left(\boldsymbol{R}^{\mathrm{T}}\right)^{-1} \underbrace{\left(\boldsymbol{I} + \tau \boldsymbol{S}\right)^{-1}}_{\text{filter factor matrix } \boldsymbol{F}} \underbrace{\left(\boldsymbol{R}^{-1}\boldsymbol{b}\right)}_{\text{coeff.}\boldsymbol{c}}$$

using matrix $\boldsymbol{R} = \boldsymbol{U}_1 \boldsymbol{\Lambda}^{1/2} \boldsymbol{U}_2$ and the inverse $\boldsymbol{R}^{-1} = \boldsymbol{U}_2^{\mathrm{T}} \boldsymbol{\Lambda}^{-1/2} \boldsymbol{U}_1^{\mathrm{T}}$. The filter factor is now given by $\varphi_j = 1/(1 + \tau S_{jj})$ with the element S_{jj} of the diagonal matrix $\boldsymbol{S}$. Figure 4 shows the eigenvalues Λ_{jj} and S_{jj} for the example of Section 3.5, both with increasing frequency from the left to the right; the stronger separation of low- and high frequency contributions by the curvature is visible. Note that the definition of the elements S_{jj} is inverse to the definiton of the elements Λ_{jj}, and the first two eigenvalues S_{11} and S_{22}, corresponding to a constant and to a linear contribution (without a curvature), are zero.

3.4 Determination/Selection of the regularization parameter

There is no generally accepted and unique method to determine the regularization parameter τ, applicable for all cases. An often used method is the L-curve method [1,2]. A *lower limit* of $\tau \approx \sigma_p^2$ is given by the size of the singular value σ_p for an effective rank of p (Section 2.4). An *upper limit* of the regularization parameter τ is determined by the overall χ_y^2 of the agreement of the observed distribution. Each Fourier coefficient, removed in the truncation method, will increase the χ_y^2 value by c_j^2 and n_{df} by one. As long as the coefficients c_j of variance one are compatible with mean zero, the p-value will not change significantly. The p-value will decrease towards zero, if significant Fourier coefficients are removed; this defines the upper limit of τ. It is recommended to study the dependence of several statistical quantities on the value of the parameter τ in repeated solutions over the acceptable range of τ-values.

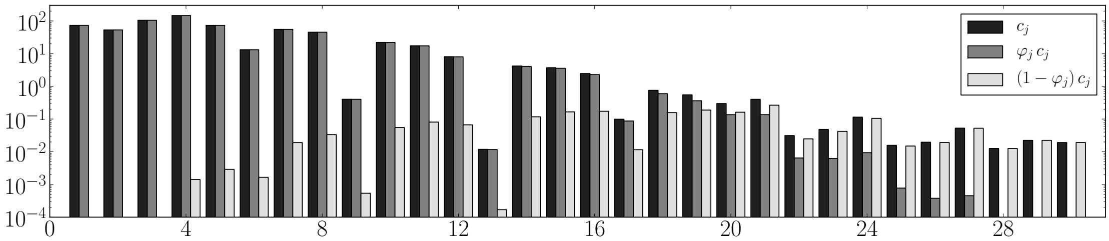

Fig. 5: Fourier coefficients c_j, with filter factors φ_j

3.5 Example: a steeply falling distributions

An example for a difficult unfolding problem is the measurement of the inclusive jet production cross section as a function of the transverse momentum p_{T} in collisions at very high energy, e.g., Reference [6]. The distribution is steeply falling. The transverse momentum p_{T}, as measured in the calorimeter, is systematically underestimated; the bias and the accuracy of the measurement can be determined in a MC simulation. In the publication [6] bin-by-bin correction (see Section 4) is applied, which is essentially

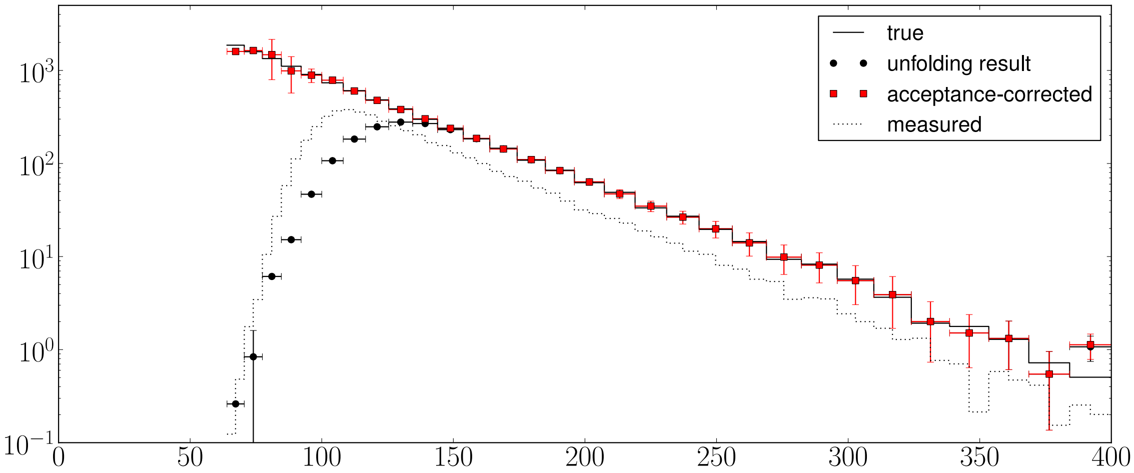

Fig. 6: Unfolding of a steeply falling distribution

only an acceptance correction, unable to correct for a bias; a bias correction is done in a separate step before.

In a simple MC simulation a problem with similar properties is solved by true unfolding according to the method of Section 3.3. Experimental conditions are assumed in analogy to the publication [6]. A pure exponential distribution is assumed with a systematic bias of the measured p_T-value to smaller values up to 10 %, and a Gaussian smearing with a relative standard deviation of $\sigma(p_\mathrm{T})/p_\mathrm{T} = 100\%/\sqrt{p_\mathrm{T} \text{ in GeV}/c}$. In addition a trigger acceptance with a rapid decrease below 100 GeV/c is assumed. Because the p_T-distribution at low values of p_T is unmeasurable, the measured and unfolded p_T-range is restricted to 64 to 400 GeV/c, assuming a realistic model function, with a separate acceptance correction after unfolding. The unfolding is performed in the transformed variable $q_\mathrm{T} = \sqrt{p_\mathrm{T}}$, which has a constant standard deviation $\sigma(q_\mathrm{T}) = 0.5$, with a back-transformation to p_T after unfolding, resulting in a bin-width increasing with p_T. The Fourier coefficients without and with filter factor are shown in Figure 5. The change of the coefficients is always less than the statistical error 1 – thus essentially no bias is introduced. The true, measured and unfolded distribution is shown in Figure 6; below 75 GeV/c the errors are larger than the cross section value.

3.6 Presentation of the regularization result

The result of regularized unfolding is an n-vector $\boldsymbol{x}$, representing the "true" function $f(t)$, together with a covariance matrix $\boldsymbol{V}_x$. In general the covariance matrix is singular with rank $k < n$ (k = number of non-zero eigenvalues after diagonalization of $\boldsymbol{V}_x$). The n bin contents originate from a small number k of effective parameters, and there will be large *positive bin-to-bin correlations*, which give the plot of the unfolded data a very *smooth* appearance, as illustrated in Figure 7 on the left (taken from Reference [4]). The plot with error bars given by the diagonal elements of $\boldsymbol{V}_x$ may be difficult to interprete and to compare with predictions; in principle the (inverse) covariance matrix has to be used for a χ^2 calculation, but this is not possible because of the rank defect. A fit of a parametrization is of course possible with the original data, as described in Section 1.3. The effective number k of degrees of freedom can be estimated by the sum $n_\mathrm{df} \approx \sum_j \varphi_j$ of the filter factors [3]. A method to avoid the singular-matrix problem is to present the unfolding result with only n_df data points. Combining four positively correlated data points to one will reduce the error by less than a factor $1/2$. This is illustrated in Figure 7 on the right, where the 40 data points are reduced to *almost uncorrelated* 10 bins, because of $n_\mathrm{df} \approx 10$, showing the *true information content* of the unfolded data.

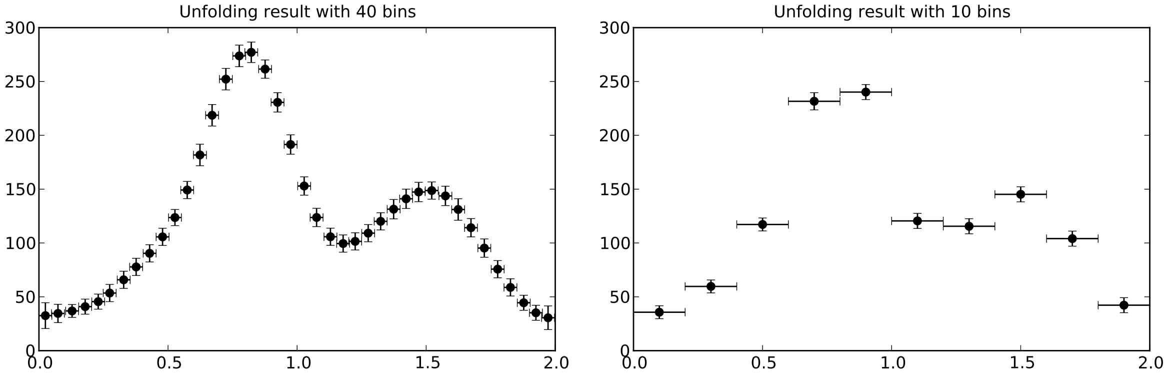

Fig. 7: Unfolding result (data from Ref. [4]) with 40 and with 10 data points

4 Iterative unfolding

Direct matrix methods (i.e. non-iterative methods) like the SVD cannot be used for problems with very large dimension parameters m and n. In those cases iterative methods for unfolding are used where the often sparse response matrix $\boldsymbol{A}$ appears only in products, for example Landweber iteration. These iterative methods are characterized by an implicit regularization, where contributions corresponding to small singular values will have very slow convergence or no convergence at all. Starting from some initial assumption $\boldsymbol{x}^{[0]}$ the first iterations show substantial improvement, but the convergence becomes then rather slow and after a very large number of iterations often a solution with large noise components similar to the naive least squares solution is obtained. This behaviour is called *semi-convergence*. In practice the iteration is stopped early; the number of iterations is the regularization parameter [1,2]. An objective criterion for stopping the iteration is not known.

Iterative methods are rather popular in particle physics although the number of parameters is rather small and there will be neither cpu-time nor memory-space problems for direct matrix methods. If iterative methods are used, usually an attempt is made, by iterative tuning with reweighting, to perform the MC simulation already with the *correct* input distribution $f(t)^{\text{model}}$, i.e., that distribution that *on average* gives a reasonable description of the observed distributions $\boldsymbol{y}$. In these methods an x-dependent unfolding matrix $\boldsymbol{M}_x$ is iteratively improved and applied to the data $\boldsymbol{y}$ to give an improved estimate $\boldsymbol{x}^{[k+1]} = \boldsymbol{M}_x^{[k]}\ \boldsymbol{y}$. Usually the *unfolding* matrix $\boldsymbol{M}_x$ in iterative methods has only positive elements (and $\boldsymbol{\Xi} = \boldsymbol{M}_x\boldsymbol{A} \neq \boldsymbol{I}$). In the *bin-by-bin correction factor method* the matrix is diagonal with elements $(\boldsymbol{M}_x)_{ii} = x_i^{\text{mc}}/y_i^{\text{mc}}$, determined from a tuned MC simulation. The methods provide a reasonable solution, often however with large but unknown positive correlations between the data points, which is equivalent to a strong smoothing. Because no matrix like the effective regularized inverse $\boldsymbol{A}^{\#}$ is available, no prescription for a direct covariance matrix calculation exists. Estimates of the covariance matrix require e.g., Monte Carlo methods.

Acknowledgement

I would like to thank the organizers of the unfolding workshop within the PHYSTAT2011 meeting, especially Louis Lyons. For a careful reading of the manuscript, valuable comments and detailed suggestions I thank Louis Lyons and Gero Flucke.

References

[1] Per Christian Hansen, *Rank-deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*, SIAM monographs on mathematical modeling and computation, Philadelphia, 1997.

[2] Per Christian Hansen, *Discrete Inverse Problems – Insight and Algorithms*, SIAM Fundamentals of algorithm series, Philadelphia, 2010.

[3] Volker Blobel, *Unfolding methods in high energy physics experiments*, Report DESY 84-118, 1984 (also in Proceedings of the 1984 CERN School of Computing, CERN 85-09, pp. 88-127; see also `http://www.desy.de/~blobel/`).

[4] Andreas Höcker and Vakhtang Kartvelishvili, *SVD approach to data unfolding*, Nucl. Instrum. Methods Phys. Res. A 372, 469 – 481, 1996.

[5] N.D. Gagunashvili, *Parametric fitting of data obtained from detectors with finite resolution and limited acceptance*, Nucl. Instrum. Methods Phys. Res. A 635, 86 – 91, 2011.

[6] A. Abulencia et al., *Measurement of the inclusive jet cross section using the k_T algorithm in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV with the CDF II detector*, Phys. Rev. D 75, 2007.