The Structure of Ill-Posed Inverse Problems in Experimental Particle Physics

Sean Gilligan

Abstract

This report provides a brief look at ill-posed inverse problems within the context of measurement related data distortions in experimental particle physics. The methods they use in solving them are in general collectively referred to as unfolding. The specifics of data and data collection methods are generalized. Common features are discussed insofar as they contribute to the necessary understanding of the data and implementation of any covered methods. In order to construct a slightly more holistic picture some additional topics are briefly touched upon if they relate to other common aspects of data analysis in particle physics, but only during parts of relevant discussions where they would otherwise normally appear. One non-iterative unfolding procedure that uses extensions of common linear regression tools is discussed and applied to a simulated example.

1 Introduction

A common problem faced in the quantitative sciences and their associated technologies is the introduction of errors during the data collection process. While the possible sources of these errors are as varied as the possible events which the data might describe, significant work has been done to develop methods the can help would-be analysts reconcile them. The requisite understanding of a scenario's underlying systematic and stochastic processes might not allow researchers to truly reverse entropy or make up for the finite resolution of a detector, but it can approximate them with a quantifiable degree of certainty.

The applied mathematics that this involves falls within the general category of **inverse problems**, and there are a variety of labels used to refer to the procedures in its arsenal. There is the colloquially vague **unsmearing**, but there are also names that reference specific applications and methods. For the sake of simplicity, and satisfying necessary physical constraints, the manner of inverse problems addressed here will only have satisfactory solutions that involve linear operations that map from one Hilbert space[1] to another. Symbolically this can be expressed by the equation

$$Az = u,$$

where A is a linear operator acting on an element $z \in Z$, the sought solution, to produce an element $u \in U$, the observed data. Within the context of the methods described here z and u take the form of continuous or discrete distributions that when integrated or summed over the domain of their arguments result in finite real quantities.

The difficulty of solving for z can be classified into one of two camps. The easiest cases involve conditions that create a **well-posed** problem, which requires that [35]

1. $\forall u \in U$ there exists a solution $z \in Z$,
2. the solution is unique,
3. and the problem is stable on U and Z.

Conditions 1 and 2 work together to imply that the inverse operator A^{-1} exists, and Condition 3 is often worded to describe the inverse as continuous, which implies that small deviations in u should correspond to similar deviations in z. When one or more of these conditions are not met, the problem is said to be **ill-posed**, and some of the consequences of assuming otherwise should hopefully become clear in the coming pages.

Entire books have been written on this subject that do not begin to cover the full scope of the methods developed to deal with ill-posed problems. With that in mind the hope for this short paper is for it to serve as an introduction to ill-posed problems while providing some degree of direction for those who would like to know more. The specific case of data and measurement error being uncorrelated and analyzable as continuous, parametric distributions will serve as a starting point. Being permissible under these conditions the method of deconvolution will be introduced, after which some generalization will occur and nonparametric cases with discrete representations will be become the focus.

1.1 The Deconvolution

One way to characterize a basic example of a situation suitable for being treated as a convolution would be one that should be very familiar to anyone who has ever taken a statistics course. Assume that data collected regarding n statistical events represent the measurement

[1] The definition of a Hilbert space is provided in Appendix B for convenience.

of n independent and identically distributed (i.i.d.) random variables $\boldsymbol{X} = \{X_1, X_2, \ldots, X_n\}$ from a distribution of possible values represented by the probability density function (PDF) $f_X(x)$, such that the probability of a random variable X_i having a value between x_a and x_b is

$$P(\, x_a < X_i < x_b \,) = \int_{x_a}^{x_b} f_X(x)\, dx$$

and

$$\int_{\mathcal{X}} f_X(x)\, dx = 1,$$

where $\mathcal{X}$ represents the domain of x. The error introduced during the measurement process is similarly represented by a set of i.i.d. random variables $\boldsymbol{\varepsilon} = \{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$ with a PDF $f_\varepsilon(\varepsilon)$, where the sets $\boldsymbol{\varepsilon}$ and $\boldsymbol{X}$ are assumed to be independent of each other. This is naive but appropriate for this simple example. The set of measured/reconstructed values $\boldsymbol{Y} = \{Y_1, Y_2, \ldots, Y_n\}$ then are also i.i.d. and can be defined in terms of the preceding sets of variables such that for event $i \in \{1, \ldots, n\}$,

$$\begin{aligned} Y_i &= g(X_i, \varepsilon_i) \\ &= X_i + \varepsilon_i. \end{aligned} \tag{1}$$

In light of this relationship, the corresponding PDF $f_Y(y)$ can be found explicitly through an operation on $f_X(x)$ and $f_\varepsilon(\varepsilon)$ using the mathematics of functional analysis. Stated in more general terms, the empirical density function f_Y is formed from the **convolution** of the true density function f_X and the error density function f_ε, and is defined by [30]

$$f_Y \equiv f_X * f_\varepsilon \tag{2}$$

$$\begin{aligned} f_Y(y) &\equiv \int_{\mathcal{X}} f_X(x) f_\varepsilon(\varepsilon)\, dx \\ &= \int_{\mathcal{X}} f_X(x) f_\varepsilon\Big(g_x^{-1}(y)\Big) \left|J_{g_x^{-1}}(y)\right| dx \\ &= \int_{\mathcal{X}} f_X(x) f_\varepsilon(y - x)\, dx, \end{aligned} \tag{3}$$

where J represents the Jacobian of the transformation involved in performing the change of basis on f_ε from ε to x, which is necessary for the evaluation of the integral for a given y. The magnitude of the Jacobian for transformation of ε to $y - x$ through the manipulation of Equation (1) happens to be 1.

As the collection of measured values $\boldsymbol{Y}$ accumulates an estimate of empirical density $\hat{f}_Y$ can readily be formed. However, a major goal in an analysis of data like this is typically to develop an accurate estimate of the true density $\hat{f}_X$. Using the information contained in

$\hat{f}_Y$ to accomplish this necessarily requires some attempt at finding an inverse process to the convolution, i.e. the **deconvolution**.

For cases in the form of this particular example there are a variety approaches, but they commonly involve the Fourier transform of the density functions $\{f_X, f_\varepsilon, f_Y\}$ into their corresponding characteristic functions $\{\phi_X, \phi_\varepsilon, \phi_Y\}$ [29][30]. Minor aspects of the definition for the Fourier transform can vary slightly between applications, resulting primarily from the use of different scale factors and sign conventions. Here it will be defined for some random variable $U \in \mathbb{R}$ with density function $f_U(u)$ and random variable $T \in \mathbb{R}$ as

$$\phi_T(t) = \int_{-\infty}^{\infty} f_U(u)\, e^{itu}\, du. \tag{4}$$

When conditions permit the inverse Fourier transform can be found via

$$f_U(u) = \int_{-\infty}^{\infty} \phi_T(t)\, e^{-itu}\, dt. \tag{5}$$

The Fourier transform is important in deconvolution methods because when it is applied to the convolution of two density functions the link between their respective characteristic functions becomes purely multiplicative, i.e.

$$f_Y = f_X * f_\varepsilon \implies \phi_Y = \phi_X \phi_\varepsilon.$$

An instructional proof of this result is provided on page 447 of [16]. The steps so far characterize a typical deconvolution scheme, with later steps consisting of various ways to perform density estimation and addressing issues similar to those that will be seen ahead [29].

1.2 Generalizing

The remainder of this paper is dedicated to a more generalized study of these types of problems. With the understanding that even experts can be fairly loose and inconsistent with their vocabulary, this paper will do its best to provide clear definitions. To begin, while most literature on deconvolution methods do use the word "convolution", this operation is also referred to by the German word *faltung* [36]. The latter's English translation, **folding**, is featured prominently in the particle physics community, but refers to a more generalized process than what is described by Equation (3) [22][8][14]. In general, folding and **unfolding** refer to two sets of processes within which the sets of convolution and deconvolution processes form proper respective subsets.

One way to arrive at the desired generalization is with the help of conditional probability.

Thinking of $\{X, Y\}$ as a continuous bivariate random vector with joint PDF $f(x, y)$ and marginal PDFs $f_X(x)$ and $f_Y(y)$, we can define the conditional PDF of Y given that $X = x$ as function of y, $f(y\,|x)$ [19]. The relationship between these PDFs is sufficient to define any one of them in terms of operations involving one or more of the others. As such, for $f_Y(y)$ it can be shown

$$\begin{aligned} f_Y(y) &= \int_{\mathcal{X}} f(x, y)\, dx \\ &= \int_{\mathcal{X}} f(y\,|x) f_X(x)\, dx \\ &= \int_{\mathcal{X}} K(x, y) f_X(x)\, dx. \end{aligned} \tag{6}$$

While integrating over x, $f(y\,|x)$ is implicitly treated as a function of both x and y. Acknowledging this allows for understanding Equation (6) as a Fredholm integral of the first kind with a Kernel function $K(x, y)$ that reflects the physical measurement process [15]. The relationship between x and y in $K(x, y)$ is not defined, but when the kernel is a function of the difference of its arguments, such that $K(x, y) = K(y - x)$, Equation (6) becomes the convolution described in Equation (3).

In particle physics experiments, analysts make use of Monte-Carlo (MC) simulations to estimate detector response to randoms samples from some true distribution $f_X(x)^{\mathrm{MC}}$, which is itself estimated by way of MC simulations using models that typically contain theory being tested by the experiment in question. The resulting measured distribution $f_Y(y)^{\mathrm{MC}}$ grants implicit knowledge of $K(x, y)$ by way of Equation (6) [14]. Finding the inverse of this Kernel is then the goal, as it should in theory allow for the mapping of experimental observations $\boldsymbol{Y}$, as randomly sampled from $f_Y(y)$, back to their true values $\boldsymbol{X}$.

1.3 Discretization

In practice researchers are often dealing with estimates $\hat{f}_X$, $\hat{f}_Y$, $\hat{f}_X^{\mathrm{MC}}$, and $\hat{f}_Y^{\mathrm{MC}}$, and the sets of data that contribute to these estimates are organized by bin into histograms that form unnormalized granular approximations of their true distributions, which to an extent is a natural result of data digitization, and is required for most modern computational methods. Thinking in terms of these histograms allows for the reformulation of Equation (6) into the linear matrix equation:

$$\boldsymbol{\nu} = \boldsymbol{R}\boldsymbol{\mu}. \tag{7}$$

The vectors $\boldsymbol{\nu}$, $\boldsymbol{\mu}$ and matrix $\boldsymbol{R}$ relate to their continuous counterparts by [14]:

$$\begin{aligned}\text{true distribution } f_X(x) &\longrightarrow \boldsymbol{\mu} \in \mathcal{U} \equiv \{\mathbb{R}_+^M \cup \mathbf{0}\} \text{ the unknown true bin counts,}\\ \text{measured distribution } f_Y(y) &\longrightarrow \boldsymbol{\nu} \in \mathcal{V} \equiv \{\mathbb{R}_+^N \cup \mathbf{0}\} \text{ the measured bin counts,}\\ \text{Kernel } K(x,y) &\longrightarrow \boldsymbol{R} \;\text{ the rectangular } N\text{-by-}M \textbf{ response matrix}.\end{aligned}$$

The components of vectors $\boldsymbol{\nu}$ and $\boldsymbol{\mu}$ represent the number of events that have occurred within the regions of x and y that define the components' corresponding bins. For $i = 1, \ldots, N$ and $j = 1, \ldots, M$ the components of matrix $\boldsymbol{R}$ are defined by the conditional probability [21]

$$\begin{aligned} R_{ij} &= P(\text{measured value in bin } i|\text{true value in bin } j)\\ &= \frac{P(\text{measured value in bin } i \text{ and true value in bin } j)}{P(\text{true value in bin } j)}\\ &= \frac{\int_{\text{bin } i}\int_{\text{bin } j} K(x,y) f_X(x) dx\, dy}{\int_{\text{bin } j} dx\, f_X(x)}\\ &\equiv P(\nu_i|\mu_j). \end{aligned} \tag{8}$$

In terms of $P(\nu_i|\mu_j)$ the full response matrix then has the form

$$\boldsymbol{R} = \begin{pmatrix} P(\nu_1|\mu_1) & P(\nu_1|\mu_2) & \ldots & P(\nu_1|\mu_N)\\ P(\nu_2|\mu_1) & P(\nu_2|\mu_2) & \cdots & P(\nu_2|\mu_N)\\ \vdots & \vdots & \ddots & \vdots\\ P(\nu_M|\mu_1) & P(\nu_M|\mu_2) & \ldots & P(\nu_M|\mu_N) \end{pmatrix}. \tag{9}$$

With these definitions Equation (7) tells us that an event produced in bin μ_j has some probability ≥ 0 of being measured in each of the N bins of $\boldsymbol{\nu}$, and that each bin count ν_i receives potential contributions from each of the M bins in $\boldsymbol{\mu}$, i.e.

$$\nu_i = \sum_{j=1}^{M} R_{ij}\mu_j \quad \text{and} \tag{10}$$

$$\frac{\partial \nu_i}{\partial \mu_j} = R_{ij}. \tag{11}$$

The number of bins are typically set such that $M \leq N$, with the convention $N = M + 1$ being common. A higher number of bins in the measured distribution reflects that the measuring process is expected to map some events in $\boldsymbol{X}$ to values of $\boldsymbol{Y}$ that are outside the region of values that define the initial M bins. These one or more extra bins are intended to account for all the possible values that a particular event could be mapped to, such that for a given event starting in bin j one might expect the probabilities of it being measured in each of the N final bins to sum to 1.

However, in practice there are a variety of constraints on events that can either result in them not being included for analysis or even prevent them from being detected at all. For example, an analyst might cut events observed in regions of a detector that result in insufficient data collection, or maybe some event information carriers miss the detector entirely, resulting in such events going unseen. In either case the effect of these missing events is described using the detector **efficiency**, and represented mathematically by the N-vector $\boldsymbol{\epsilon}$, where component ϵ_j is the efficiency of the jth true bin defined[2] by [21]:

$$\sum_{i=1}^{N} P(\nu_i|\mu_j) = \sum_{i=1}^{N} R_{ij} = \epsilon_j \leq 1. \tag{12}$$

In contrast to this are contributions to measured counts from **background** processes. Just as events produced in a region of interest can be smeared out of it, events produced out of it can be smeared into it. The crossed barrier could correspond to the variable of interest, but it can also include events excluded from analysis due to assigned constraints on other variables that describe the event. Background processes can be studied and dealt with prior to the unfolding procedures described in the paper. It is briefly mentioned here to provide a slightly more holistic picture of particle physics analyses. Mathematically, background would be included by modifying Equation (10) to read

$$\nu_i = \sum_{j=1}^{M} R_{ij}\mu_j + \beta_i, \tag{13}$$

where β_i is the ith component of the N-vector $\boldsymbol{\beta}$, which represents the binned background counts. This leads to equations like $\nu_i^{\text{sig}} = \nu_i - \beta_i$ in order to specify the expected number of measured counts that are from the signal of interest. Going forward background will be assumed to already have been accounted for, and ν_i will refer to the expected signal counts of bin i.

As all these variables so far have been derived from the exact continuous distributions $f_X(x)$ and $f_Y(y)$, they correspond to the expectation values that researchers are estimating during data collection and analysis. As this is a counting process the components of the observed number of signal events $\boldsymbol{n}$, an N-vector, are often related to the components of the expected

[2] In the continuous case it is typically written as $\epsilon(x)$, and understood to be the conditional probability of an event producing any measured value given it has a true value of x. It is typically absorbed into $K(x, y)$ where it goes on to manifest within $\boldsymbol{R}$ in the manner shown in Equation (12) [14].

number of observed counts $\boldsymbol{\nu}$ as a collection of N separate and independent Poisson processes. That is to say the observed counts n_i in bin i are treated as i.i.d. random variables with the probability mass function

$$P(n_i|\nu_i) = \frac{\nu_i^{n_i} e^{-\nu_i}}{n_i!}. \tag{14}$$

As such counts n_i form the estimate $\hat{\nu}_i$ of the expected counts ν_i by

$$\begin{aligned} \nu_i &= \mathrm{E}[\hat{\nu}_i] = \mathrm{E}[n_i] \\ &= \mathrm{Cov}[\hat{\nu}_i] = \mathrm{Cov}[n_i]. \end{aligned}$$

Understanding the probability distribution of $\boldsymbol{n}$ allows for unfolding methods that involve the use of maximum likelihood estimation. Additionally, it will be convenient, and necessary for methods based on least-squares, to estimate the covariance matrix $\hat{\boldsymbol{\Sigma}}_\nu$ (written $\hat{\Sigma}^\nu_{ij}$ when referring to components) of the observations, which for independent Poisson processes has components of the form

$$\begin{aligned} \hat{\Sigma}^\nu_{ij} &= \mathrm{Cov}[\hat{\nu}_i, \hat{\nu}_j] \\ &= \mathrm{Cov}[n_i, n_j] \\ &= \delta_{ij} n_i, \end{aligned} \tag{15}$$

where δ_{ij} is the Kronecker delta[3]. The path to an estimated covariance matrix of the estimated true distribution $\hat{\boldsymbol{\mu}}$, itself a function of $\boldsymbol{n}$ and $\boldsymbol{\nu}$ (or $\hat{\boldsymbol{\nu}}$), can be considered briefly by considering the maximum log-likelihood, where it can be shown

$$\begin{aligned} \log L(\boldsymbol{\mu}) &= \sum_{i=1}^{N} \log \left(\frac{\nu_i^{n_i} e^{-\nu_i}}{n_i!} \right) \\ &= \sum_{i=1}^{N} \left(n_i \log \nu_i - \nu_i - \log n_i! \right) && (16) \\ \frac{\partial \log L}{\partial \mu_k} &= \sum_{i=1}^{N} \frac{\partial \log L}{\partial \nu_i} \frac{\partial \nu_i}{\partial \mu_k} \\ &= \sum_{i=1}^{N} \left(\frac{n_i}{\nu_i} - 1 \right) R_{ik} = 0. && (17) \end{aligned}$$

Some minor algebra here thankfully reproduces the estimate $\hat{\boldsymbol{\nu}} = \boldsymbol{n}$, as expected from an

[3] The Kronecker delta δ_{ij} is a piecewise function of variables i and j defined by $\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$.

earlier observation about n_i. Continuing with an additional derivative shows

$$\begin{aligned}\frac{\partial^2 \log L}{\partial\mu_k \partial\mu_l} &= -\sum_{i=1}^{N}\left(\frac{n_i}{\nu_i^2}\frac{\partial\nu_i}{\partial\mu_l}\right)R_{ik} \\ &= -\sum_{i=1}^{N}\frac{n_i R_{il} R_{ik}}{\nu_i^2}, \qquad (18)\end{aligned}$$

the negative of the expectation value of which is the Fisher information matrix $\boldsymbol{\mathcal{I}}(\boldsymbol{\mu})$. Since the Fisher information's relationship with the Cramér-Rao lower bound can, as its name implies, be used to determine the lower bound on the covariance matrix of an estimator of $\boldsymbol{\mu}$, one can show for one such *unbiased* estimator, say $\boldsymbol{T}(\boldsymbol{n})$, that

$$\begin{aligned}\mathrm{Cov}_{\boldsymbol{\mu}}[\boldsymbol{T}(\boldsymbol{n})] \geq \boldsymbol{\mathcal{I}}(\boldsymbol{\mu})^{-1} &= \left(-E\left[\frac{\partial^2 \log L}{\partial\mu_k \partial\mu_l}\right]\right)^{-1} \\ &= \left(\sum_{i=1}^{N}\frac{E[n_i] R_{il} R_{ik}}{\nu_i^2}\right)^{-1} \\ &= \left(\sum_{i=1}^{N}\frac{R_{il} R_{ik}}{\nu_i}\right)^{-1} \\ &= \left(\boldsymbol{R}^T \boldsymbol{\Sigma}_\nu^{-1} \boldsymbol{R}\right)^{-1} \\ &= \boldsymbol{R}^{-1}\boldsymbol{\Sigma}_\nu(\boldsymbol{R}^{-1})^T. \qquad (19)\end{aligned}$$

Indeed, this matrix must then be the lower bound for the covariance matrix of any unbiased estimator of $\boldsymbol{\mu}$. This is some good insight to have before getting into the weeds of working on actual data.

1.4 A Simulated Example

The following example is not meant to reflect any actual physics. Consider the following three sets of i.i.d. random variables from separate Cauchy distributions.

$$\begin{aligned}X_{1,i} &\sim \mathrm{Cauchy}(x_0^{(1)}, \gamma_1) \\ X_{2,i} &\sim \mathrm{Cauchy}(x_0^{(2)}, \gamma_2) \\ X_{3,i} &\sim \mathrm{Cauchy}(x_0^{(3)}, \gamma_3)\end{aligned}$$

Current models predict that some class of physics events observed in past detectors are solely coming from the first two processes (Model 1), such that for n events the number coming from the first process is an i.i.d. random variable from the binomial distribution $B(n, p)$.

Meanwhile, a new model (Model 2) has been developed that suggests that the third process has been occurring this whole time but has been incorrectly categorized as one or the other of the first two. Napkin math has estimated a contribution rate that is reflective of some probability p_3, such that the binomial distribution is actually a multinomial distribution with probabilities $\boldsymbol{p} = \{p_1, p_2, p_3\}$.

A new experiment is being designed and funded to test this new model, on top of many others, and simulations are being performed to give analyzers plenty of opportunities to develop their collaboration's analysis framework, perform calibrations, and make ready a myriad of studies that hope to shed light on their unanswered questions. The corresponding MC simulations performed during this time include such simulations for the physics events of interest, but also for the detector. For the purposes of this paper I am performing a relatively simple set of simulations, resulting in 200,000 simulated events per model that are meant to represent the MC simulations, and a set of 20,000 simulated events for each model that are meant to represent either hypothetical detector data or MC test sets as need dictates. I will also be writing the code to carry out any of the discussed methods that are applied to this simulated data.

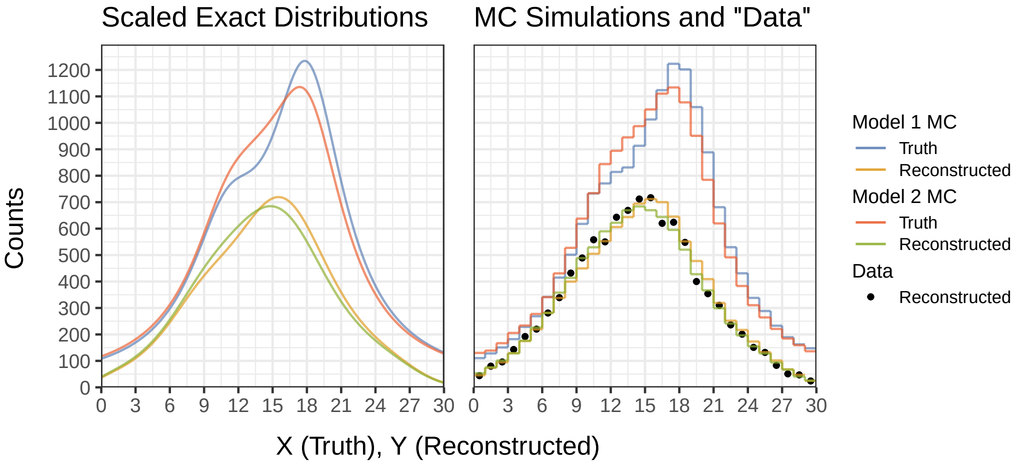

Figure 1: *The distinguishing characteristics between Model 1 and Model 2 become greatly diminished as their respective distributions are smeared and diffused by mechanisms common to both. Ignoring the knowledge granted in this figure's legend, one would be hardpressed to predict which Truth distribution a particular Reconstructed distribution came from.*

Please see Appendix C for more specific information about the performed simulations. Referring to Figure 1, the blue and red colored plots correspond to the true distributions of Model 1 and Model 2 respectively. The colored lines represent the MC simulations. They have been rescaled to represent the same number of possible events as the experimental data. The impact of the third contributing Cauchy process in Model 2 can be seen by way of the notably diminished peak and the partial filling of the dividing indent between the two processes of Model 1. While it is clear for both the continuous and discrete representations that the two larger distributions are distinct, less can be said about them once finite detector resolutions and other inefficiencies have had their effects, as one can see in the similarities between their yellow and green reconstructed distributions.

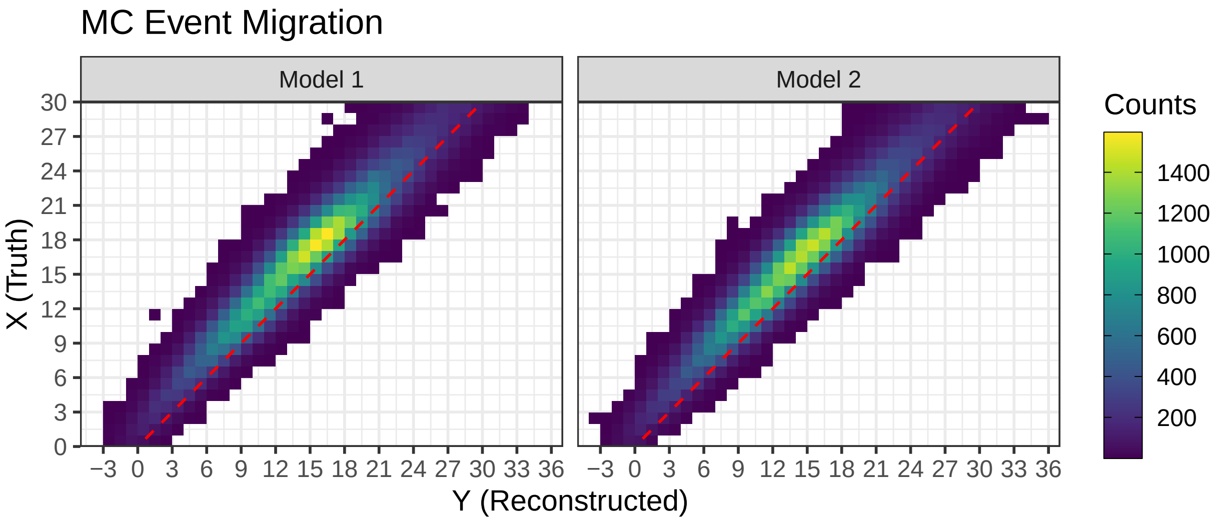

Figure 2: *The added dimensional perspective offered by this migration heat map reveals additional structure and dependencies concerning the smearing process.*

Figure 2 provides an alternative perspective to Figure 1 by providing a visual representation of the two models' migration matrices. The red dashed lines indicate where events would live if there were no skewing or smearing processes. Note that the Reconstructed axis extends past points covered by the Truth axis, providing insight into the behavior with which events produced in the region of interest ($X \in (0, 30)$ here) end up measured or reconstructed outside of it, but still in a detector's **fiducial volume**, which represents the reliable, central region of the detector or other relevant region of the events' **phase space**, which is the space of all possible event states.

2 Unfolding in particle physics

The international particle physics community is highly collaborative. Most of this is likely due to necessity, as advances in the field are requiring larger and more expensive experiments that rely on pooled resources of talent, labor, and capital. This extends to digital resources as well. ROOT [18] is an open-source data analysis framework developed and used primarily by people doing particle physics. Analyses incorporating its resources are primarily written in some combination of Python and C++. While it does contain a strong base level of classes, objects, and methods within its default toolkit, including some for unfolding, extended frameworks are typically built on top of it to meet the needs of specific experiments or to facilitate expanded capabilities that would effectively serve as bloatware for most users. The RooUnfold framework [8] is one such example. Two methods facilitated in whole or in part by the RooUnfold framework will be covered in this section. One is considered naive, as will be shown. The other finds ways past the issues brought up in the first.

2.1 Inverting the Response Matrix

In the event of Equation (7) being well-posed the obvious approach would be to construct the unique inverse of the response matrix $\boldsymbol{R}^{-1}$, as estimated from MC simulations, and map the reconstructed counts back to an estimate of the true counts via

$$\hat{\boldsymbol{\mu}} = \boldsymbol{R}^{-1}\boldsymbol{n}. \tag{20}$$

A statistical justification for this comes from performing generalized least-squares [25] fit to estimate $\boldsymbol{\mu}$, which relies on approximating bin count n_i as normally distributed with mean ν_i and variance $1/\nu_i$. Minimizing the sums of squares yields

$$\begin{aligned}
\min_{\boldsymbol{\mu}} \nabla_{\boldsymbol{\mu}}\boldsymbol{\chi}^2(\boldsymbol{\mu}) &= \nabla_{\boldsymbol{\mu}}(\boldsymbol{R\mu} - \boldsymbol{n})^T\boldsymbol{\Sigma}_\nu^{-1}(\boldsymbol{R\mu} - \boldsymbol{n}) \qquad (21)\\
&= \nabla_{\boldsymbol{\mu}}(\boldsymbol{\mu}^T\boldsymbol{R}^T - \boldsymbol{n}^T)\boldsymbol{\Sigma}_\nu^{-1}(\boldsymbol{R\mu} - \boldsymbol{n})\\
&= \nabla_{\boldsymbol{\mu}}(\boldsymbol{\mu}^T\boldsymbol{R}^T\boldsymbol{\Sigma}_n^{-1}\boldsymbol{R\mu} - \boldsymbol{\mu}^T\boldsymbol{R}^T\boldsymbol{\Sigma}_\nu^{-1}\boldsymbol{n} - \boldsymbol{n}^T\boldsymbol{\Sigma}_\nu^{-1}\boldsymbol{R\mu} + \boldsymbol{n}^T\boldsymbol{\Sigma}_\nu^{-1}\boldsymbol{n})\\
&= \nabla_{\boldsymbol{\mu}}(\boldsymbol{\mu}^T\boldsymbol{R}^T\boldsymbol{\Sigma}_n^{-1}\boldsymbol{R\mu} - 2\boldsymbol{\mu}^T\boldsymbol{R}^T\boldsymbol{\Sigma}_\nu^{-1}\boldsymbol{n} + \boldsymbol{n}^T\boldsymbol{\Sigma}_\nu^{-1}\boldsymbol{n})\\
&= 2\boldsymbol{R}^T\boldsymbol{\Sigma}_\nu^{-1}\boldsymbol{R\mu} - 2\boldsymbol{R}^T\boldsymbol{\Sigma}_\nu^{-1}\boldsymbol{n}\\
&= 0\\
\implies \boldsymbol{R}^T\boldsymbol{\Sigma}_\nu^{-1}\boldsymbol{R\mu} &= \boldsymbol{R}^T\boldsymbol{\Sigma}_\nu^{-1}\boldsymbol{n}\\
\implies \hat{\boldsymbol{\mu}} &= (\boldsymbol{R}^T\boldsymbol{\Sigma}_\nu^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}^T\boldsymbol{\Sigma}_\nu^{-1}\boldsymbol{n} \qquad (22)\\
&= \boldsymbol{R}^+\boldsymbol{n}, \qquad (23)
\end{aligned}$$

where $\boldsymbol{R}^+ = (\boldsymbol{R}^T\boldsymbol{\Sigma}_\nu^{-1}\boldsymbol{R})^{-1}\boldsymbol{R}^T\boldsymbol{\Sigma}_\nu^{-1}$ is the Moore-Penrose generalized inverse (or pseudo-inverse) [14] of $\boldsymbol{R}$. Note that multiplying $\boldsymbol{R}$ to the left side of $\boldsymbol{R}^+$ as calculated here fits the form of the hat matrix $\boldsymbol{H}$ of ordinary regression, such that $\hat{\boldsymbol{\nu}} = \boldsymbol{H\nu}$. This has all so far just been basic regression using generalized least-squares. As such the inverse of $\boldsymbol{R}^T\boldsymbol{\Sigma}_\nu^{-1}\boldsymbol{R}$, using the Moore-Penrose generalized inverse, corresponds to the estimated true counts' covariance matrix:

$$(\boldsymbol{R}^T\boldsymbol{\Sigma}_\nu^{-1}\boldsymbol{R})^{-1} = \boldsymbol{R}^+\boldsymbol{\Sigma}_\nu\boldsymbol{R}^{+^T} = \boldsymbol{R}^+\text{Cov}[\boldsymbol{n}]\boldsymbol{R}^{+^T} = \text{Cov}[\boldsymbol{R}^+\boldsymbol{n}] = \text{Cov}[\hat{\boldsymbol{\mu}}] = \hat{\boldsymbol{\Sigma}}_\mu.$$

This is the same lower bound covariance matrix calculated in Equation (19), which indicates that $\boldsymbol{R}^+$ is indeed an unbiased estimator with the lowest possible variances among unbiased estimators. For the simulations one such inverse matrix was calculated from the response matrix of each MC Model simulation, and were used to form estimates of their respective $\boldsymbol{\Sigma}_\mu$s for each MC-Data combination using each Data Model's estimated covariance matrix.

The resulting true counts estimates, with their ±1 estimated standard deviations, are plotted below in 3. The rescaled true distribution for each MC simulation was provided for comparison. They are in separate plots as their structure is not visible at the scales necessary to show the unfolded Data. The color of the lines indicate the true model behind the simulation and the shaded regions, which represent the calculated uncertainty, are colored to indicate the MC model of the response matrix.

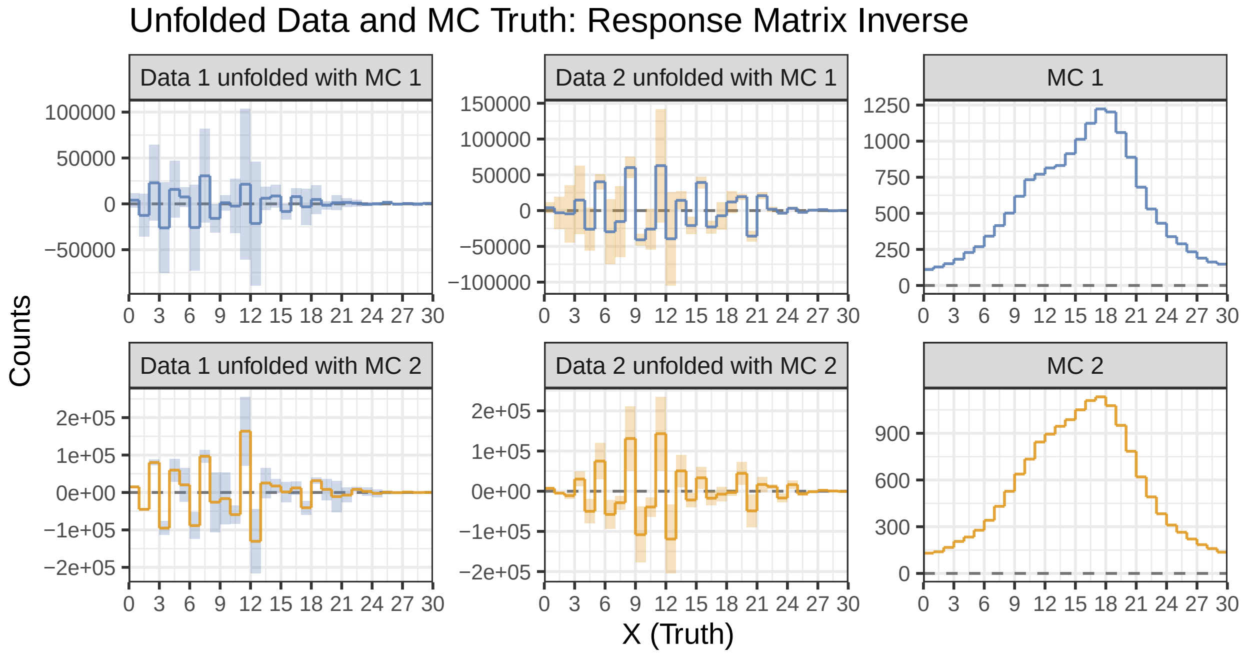

Figure 3: *Plots comparing the unfolded reconstructed data under both model assumptions to the true MC distributions. Note the large uncertainties and the rapidly oscillating positive and negative estimated counts.*

The massive, rapid oscillations and large uncertainties indicate that these are not very good results, regardless of whether or not the Data and the MC came from the same Model. The unfolding process used here cannot account for even small random variations in Reconstructed space because it has been over-fitted by the MC results. In search of an unbiased estimator we have unwittingly created a minimum variance that is simply too large to be in any way useful. Some degree of bias in estimating $\hat{\boldsymbol{\mu}}$ will be needed to probe the neighborhood around the unbiased estimate resulting from minimizing the least-squares of Equation (21) (or equivalent maximum log-likelihood) to find an alternative solution that can actually be used elsewhere. This is done by including an extra term in these equations through a process called **Regularization**.

2.2 Regularization

This section reviews a method of unfolding that utilizes a form of **Tikhonov regularization**. Tikhonov regularization as discussed here can be described as a form of **regularized ridge regression** with weighted least-squares, and using a ridge penalty known by some as the **ridge fused penalty** [37]. This method is referred to in particle physics as **Singular Value Decomposition** (SVD) [8][14][21], which unfortunately hides much of the underlying concepts. In general, SVD is a matrix factorization method in linear algebra which, to be fair, is featured prominently in this unfolding method. Please refer to Appendix D for a brief description of SVD.

As directed in Höcker and Kartvelishvili's seminal paper on this method [24] one should shift the location of event count information on the right-hand side of equation (7) from the $\boldsymbol{\mu}$ to $\boldsymbol{R}$, such that $\boldsymbol{R}$'s elements are the actual number of corresponding events, much like Figure 2. This new matrix will be referred to by $\boldsymbol{X}$. Next, redefine $\boldsymbol{\mu}$ by its ratio to the MC true counts $\boldsymbol{\mu}^{\mathrm{MC}}$ to produce instead $\boldsymbol{\beta}$ such that

$$\beta_i = \mu_i / \mu_i^{\mathrm{MC}}.$$

This naturally requires an inverse rescaling of any estimated $\hat{\boldsymbol{\beta}}$ to get $\hat{\boldsymbol{\mu}}$. The substitution $\boldsymbol{Y} = \boldsymbol{n}$ will also be used to facilitate broader notation conventions used in discussions of this subject. The weighted sum of squares with these changes comes out to be

$$\chi^2(\boldsymbol{\beta}) = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T \hat{\boldsymbol{\Sigma}}_\nu^{-1} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \tag{24}$$

After rescaling, singular value decomposition is applied to the reconstructed covariance matrix $\hat{\boldsymbol{\Sigma}}_\nu$ which, being symmetric and containing only positive elements, will result in something of the form

$$\hat{\boldsymbol{\Sigma}}_\nu = \boldsymbol{Q}\boldsymbol{T}\boldsymbol{Q}^T, \qquad \text{where} \qquad \hat{\boldsymbol{\Sigma}}_\nu^{-1} = \boldsymbol{Q}\boldsymbol{T}^{-1}\boldsymbol{Q}^T. \tag{25}$$

The elements of the diagonal matrix $\boldsymbol{T}$ are to be written $T_{ij} = t_i^2\delta_{ij}$. Substituting the inverse of this SVD of $\hat{\boldsymbol{\Sigma}}_\nu$ into Equation (24) the weighted sum of squares becomes an unweighted sum of squares, such that

$$\begin{aligned}
\chi^2(\boldsymbol{\beta}) &= (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T \hat{\boldsymbol{\Sigma}}_\nu^{-1} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \\
&= (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T \boldsymbol{Q}\boldsymbol{T}^{-1}\boldsymbol{Q}^T (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \\
&= (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T \boldsymbol{Q}\boldsymbol{T}^{-1/2}\boldsymbol{T}^{-1/2}\boldsymbol{Q}^T (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \\
&= \left(\boldsymbol{T}^{-1/2}\boldsymbol{Q}^T\boldsymbol{Y} - \boldsymbol{T}^{-1/2}\boldsymbol{Q}^T\boldsymbol{X}\boldsymbol{\beta}\right)^T \left(\boldsymbol{T}^{-1/2}\boldsymbol{Q}^T\boldsymbol{Y} - \boldsymbol{T}^{-1/2}\boldsymbol{Q}^T\boldsymbol{X}\boldsymbol{\beta}\right) \\
&= \left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\boldsymbol{\beta}\right)^T \left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\boldsymbol{\beta}\right), \qquad (26)
\end{aligned}$$

where

$$\tilde{\boldsymbol{X}} = \boldsymbol{T}^{-1/2}\boldsymbol{Q}^T\boldsymbol{X} \qquad \text{and} \qquad \tilde{\boldsymbol{Y}} = \boldsymbol{T}^{-1/2}\boldsymbol{Q}^T\boldsymbol{Y}. \qquad (27)$$

The bias introduced by regularizing the sum of squares exists by way of the addition of a weighted functional $\tau\,\boldsymbol{\Omega}(\boldsymbol{\beta})$. As an operator it acts on $\boldsymbol{\beta}$ in a fixed way as determined by the needs of the analyst, and τ (sometimes α or λ) is varied to shift the location of the minimum least-squares [35]. In this circumstance we are most concerned with reducing the oscillations in our final estimate by encouraging smoothness in the estimated distribution. The general practice is to use an estimate of the second derivative between the ordered elements of $\boldsymbol{\beta}$ by way of a matrix operator $\boldsymbol{C}$, with the overall goal of minimizing

$$\begin{aligned}
\chi^2(\boldsymbol{\beta}) &= \left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\boldsymbol{\beta}\right)^T \left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\boldsymbol{\beta}\right) \\
&= \left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\boldsymbol{\beta}\right)^T \left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\boldsymbol{\beta}\right) + \tau(\boldsymbol{C}\boldsymbol{\beta})^T\boldsymbol{C}\boldsymbol{\beta} \\
&= \left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\boldsymbol{\beta}\right)^T \left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\boldsymbol{\beta}\right) + \tau\boldsymbol{\beta}^T\boldsymbol{C}^T\boldsymbol{C}\boldsymbol{\beta}. \qquad (28)
\end{aligned}$$

The matrix $\boldsymbol{C}$ by default assumes a uniform bin width, but if that is not the case the bins can be weighted accordingly by a diagonal matrix $\boldsymbol{B}$ with positive elements $B_{ii} = \Delta x_i$, the width of bin i. Plugged in, this changes the regularizing expression in Equation (28) to

$$\tau\left(\boldsymbol{B}^{1/2}\boldsymbol{C}\boldsymbol{\beta}\right)^T \boldsymbol{B}^{1/2}\boldsymbol{C}\boldsymbol{\beta} \;\; = \;\; \tau\boldsymbol{\beta}^T\boldsymbol{C}^T\boldsymbol{B}\boldsymbol{C}\boldsymbol{\beta},$$

where [21],

$$\boldsymbol{C} = \begin{pmatrix} -1 & 1 & 0 & 0 & \dots & & \\ 1 & -2 & 1 & 0 & \dots & & \\ 0 & 1 & -2 & 1 & \ddots & & \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ & & & & \dots & -2 & 1 \\ & & & & \dots & 1 & -1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{B} = \begin{pmatrix} \Delta x_1 & 0 & \dots & & \\ 0 & \Delta x_2 & \dots & & \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ & & \dots & \Delta x_{M-1} & 0 \\ & & \dots & 0 & \Delta x_M \end{pmatrix}.$$

Uniform bin widths where $\Delta x_i \neq 1$ are equivalent to when $\Delta x_i = 1$, with τ absorbing the appropriate scale factor. Continuing with uniform bin width, Höcker and Kartvelishvili conceptualize an equivalent to Equation (7) that involves stacking $N \times M$ matrix $\sqrt{\tau}\boldsymbol{C}$ below the response matrix and lengthening the Reconstructed vector by N zeros, such that

$$\begin{bmatrix} \tilde{\boldsymbol{X}} \\ \sqrt{\tau}\boldsymbol{C} \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} \tilde{\boldsymbol{Y}} \\ \boldsymbol{0}_N \end{bmatrix},$$

At this point you could perform SVD to calculate a pseudoinverse and be done. However, they instead implement the method of **damped least squares** to express the solution for $\tau > 0$ in terms of the solution for $\tau = 0$. They start with absorbing the matrix $\boldsymbol{C}$ into $\boldsymbol{\beta}$ to get $\begin{bmatrix} \tilde{\boldsymbol{X}}\boldsymbol{C}^{-1} \\ \sqrt{\tau}\,\boldsymbol{I}_{N\times M} \end{bmatrix} \boldsymbol{C}\boldsymbol{\beta}$. The matrix $\boldsymbol{C}$ is notably singular, so $\boldsymbol{C}^{-1}$ does not exist. The writers work around this by adding a negligible value to its diagonal elements and creating the invertable matrix $\tilde{\boldsymbol{C}} = \boldsymbol{C} + \xi\boldsymbol{I}$, where typically $\xi = 10^{-3}$ or 10^{-4}.

Setting $\tau = 0$ and applying singular value decomposition to the remaining matrix on the left-hand side, such that $\tilde{\boldsymbol{X}}\tilde{\boldsymbol{C}}^{-1} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$, will modify the least-squares equation to

$$\begin{aligned} \chi^2 &= \left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\boldsymbol{\beta}\right)^T \left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\boldsymbol{\beta}\right) \\ &= \left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{C}}^{-1}\tilde{\boldsymbol{C}}\boldsymbol{\beta}\right)^T \left(\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{X}}\tilde{\boldsymbol{C}}^{-1}\tilde{\boldsymbol{C}}\boldsymbol{\beta}\right) \\ &= \left(\tilde{\boldsymbol{Y}} - \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T\tilde{\boldsymbol{C}}\boldsymbol{\beta}\right)^T \left(\tilde{\boldsymbol{Y}} - \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T\tilde{\boldsymbol{C}}\boldsymbol{\beta}\right) \\ &= \left(\tilde{\boldsymbol{Y}} - \boldsymbol{U}\boldsymbol{S}\boldsymbol{b}\right)^T \left(\tilde{\boldsymbol{Y}} - \boldsymbol{U}\boldsymbol{S}\boldsymbol{b}\right) && (\boldsymbol{b} = \boldsymbol{V}^T\tilde{\boldsymbol{C}}\boldsymbol{\beta}) \\ &= \left(\tilde{\boldsymbol{Y}}^T - \boldsymbol{b}^T\boldsymbol{S}^T\boldsymbol{U}^T\right) \left(\tilde{\boldsymbol{Y}} - \boldsymbol{U}\boldsymbol{S}\boldsymbol{b}\right) \\ &= \tilde{\boldsymbol{Y}}^T\boldsymbol{U}\boldsymbol{U}^T\tilde{\boldsymbol{Y}} - \boldsymbol{b}^T\boldsymbol{S}^T\boldsymbol{U}^T\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{Y}}^T\boldsymbol{U}\boldsymbol{S}\boldsymbol{b} + \boldsymbol{b}^T\boldsymbol{S}^T\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{S}\boldsymbol{b} && (\boldsymbol{U}\boldsymbol{U}^T = \boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}_{N\times N}) \\ &= \boldsymbol{y}^T\boldsymbol{y} - \boldsymbol{b}^T\boldsymbol{S}^T\boldsymbol{y} - \boldsymbol{y}^T\boldsymbol{S}\boldsymbol{b} + \boldsymbol{b}^T\boldsymbol{S}^T\boldsymbol{S}\boldsymbol{b} && (\boldsymbol{y} = \boldsymbol{U}^T\tilde{\boldsymbol{Y}}) \\ &= (\boldsymbol{y} - \boldsymbol{S}\boldsymbol{b})^T(\boldsymbol{y} - \boldsymbol{S}\boldsymbol{b}). \end{aligned}$$

Since $\boldsymbol{S}$ is a diagonal matrix with elements $S_{ii} = s_i > 0$ the solution can be found easily enough to be

$$\hat{b}_i^{(0)} = \frac{y_i}{s_i} \quad \text{followed by} \quad \hat{\boldsymbol{\beta}}^{(0)} = \tilde{\boldsymbol{C}}^{-1}\boldsymbol{V}\hat{\boldsymbol{b}}^{(0)}. \tag{29}$$

The estimated Truth counts would then come from reweighting $\hat{\boldsymbol{\beta}}$ such that $\hat{\mu}_i = \hat{\beta}_i \mu_i^{\mathrm{MC}}$. Now since $\tau = 0$ in this case there is yet no regularization. Höcker and Kartvelishvili provide a source regarding this part of damped least-squares that explains that introducing $\tau > 0$ is accomplished just by modifying the non-regularized result to

$$\begin{aligned} y_i^{(\tau)} &= y_i^{(0)} \frac{s_i^2}{s_i^2 + \tau} \\ \Longrightarrow \hat{b}_i^{(\tau)} &= \frac{y_i^{(0)} s_i}{s_i^2 + \tau}, \end{aligned}$$

and so forth for $\hat{\boldsymbol{\mu}}$. The description of this last equation as a low-pass filter is fairly apt, as the extreme influence of small singular values, represented by diagonals $\{s_i\}$ in the diagonal matrix from the SVD of $\tilde{\boldsymbol{X}}\tilde{\boldsymbol{C}}^{-1}$, are muffled by the presence of a nonzero τ.

The analysts work out the resulting covariance matrices for these estimates, which are dependent on the value of τ such that

$$\begin{aligned} \hat{\Sigma}_{ij}^{b} &= \frac{s_i^2}{(s_i^2 + \tau)^2}\delta_{ij}, \\ \hat{\boldsymbol{\Sigma}}_{\beta} &= \tilde{\boldsymbol{C}}^{-1}\boldsymbol{V}\hat{\boldsymbol{\Sigma}}_b\boldsymbol{V}^T(\tilde{\boldsymbol{C}}^{-1})^T, \quad \text{and} \\ \hat{\Sigma}_{ij}^{\mu} &= \mu_i^{\mathrm{MC}}\hat{\Sigma}_{ij}^{\beta}\mu_j^{\mathrm{MC}}. \end{aligned}$$

As far as selecting a value of τ, Höcker and Kartvelishvili recommend first plotting $|y_i|$ with respect to i. They point out that $\boldsymbol{y}$ makes up the coefficients of the decomposition of the reconstructed counts $\boldsymbol{n}$ ($\boldsymbol{Y}$ above), with an orthogonal basis given by the columns of $\boldsymbol{U}$. It is expected that $|y_i|$ will decay exponentially with i until some critical value $i = k$ in which all following coefficients will be indistinguishable from random samples from an $N(0,1)$ distribution. Once that value of k has been identified the idea is to set $\tau = s_k^2$. For this paper's simulations these identifications are graphically identified in Figure 4, which shows the plot of $|y_i^{(0)}|$ and the resulting $|y_i^{(\tau)}|$. The same value of k was chosen for all four combinations for consistency.

This manner of choosing the regularization parameter is just one suggestion, and while its

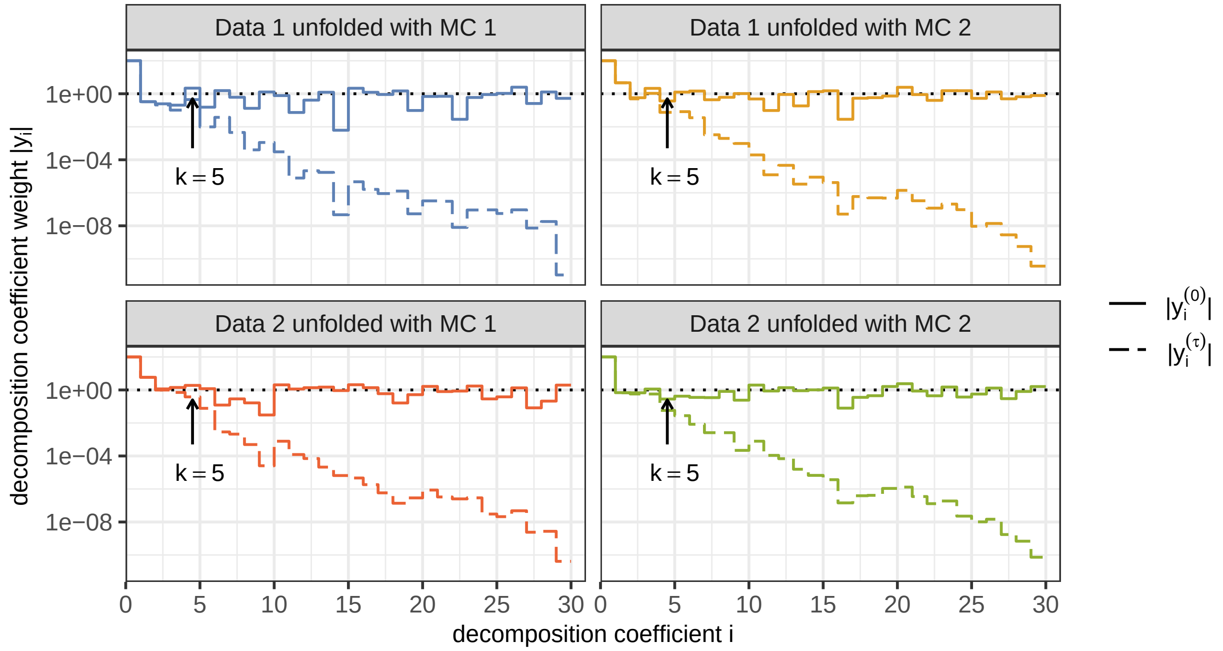

Figure 4: *For each of the four cases the point at which $|y_i|$ drops consistently below 1 is approximately identified. This approach helps determine effective rank. This plot also shows how adding τ to get $|y_i^{(\tau)}|$ suppresses spurious influences due to statistical fluctuation.*

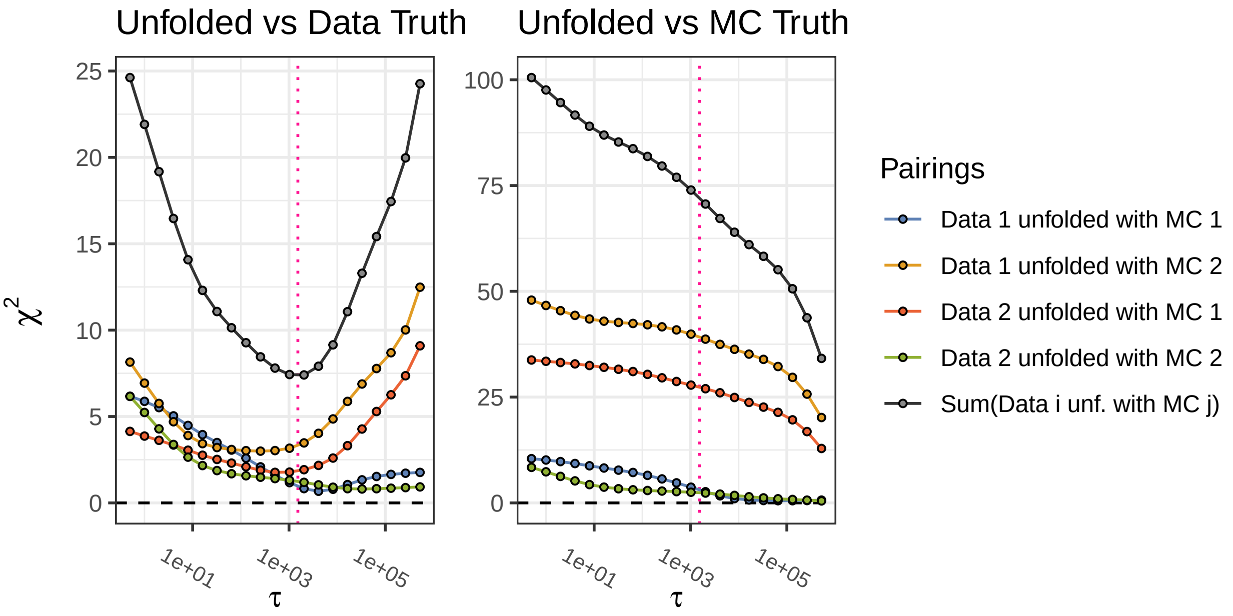

Figure 5: *Above are plots of the individual χ^2s and their summation $\sum \chi^2$ with respect to selected values of τ. On the left the unfolded counts were compared to their actual true counts. On the right the unfolded counts were compared to the true counts of the MC model used to unfold them. The dotted vertical pink line indicates the value of τ that minimizes $\sum \chi^2$ for the plot on the left.*

reasoning is fairly clear there is another way that considers values of τ not restricted to squares of the singular values. Figure 5 features two plots concerning the sums of squares for different values of τ, which were calculated by comparing the unfolded distributions to their true distributions (LEFT) and to the true distributions for the MC model used in the unfolding process (RIGHT). The Data, being also simulated, are playing the role of training sets in this capacity. The plot on the left reveals the existance of a critical value of τ above which an unfolded distribution is overfitted to the associated MC distribution, which will hide features of the original distribution that are a departure from the model being tested. The plot on the right contains no such indicators.

As the purpose of unfolding is to undo effects related to the measurement process, it would be ideal for a method to be model agnostic. As a MC simulation is used to generate the response matrix this is unfortunately impossible. However, if an unknown true model is expected to produce similar looking distributions to an assumed model, we can rely on them behaving similarly under an unfolding procedure using the same regularization parameter, as the behaviors shown in Figure 5 depict.

MC	Data	$\tau = s_5^2$	Data χ^2	MC χ^2	$\min_\tau \sum \chi^2$	Data χ^2	MC χ^2
1	1	3766.77	0.68	1.77	1527.14	0.96	3.09
1	2	3766.77	3.94	37.61	1527.14	3.32	39.21
2	1	3766.77	2.13	26.15	1527.14	1.85	27.35
2	2	3766.77	1.06	2.11	1527.14	1.24	2.41
All		$\sum \chi^2$	7.81	67.63	$\sum \chi^2$	7.36	72.06

Table 1: *The χ^2s and their summations $\sum \chi^2$ for the listed values of τ as chosen by their respective selection methods. Data χ^2 columns are a result of comparing unfolded Data to the Data's true counts. Unfolded Data from Model 1* ***is never compared*** *to Model 2's Data true counts, and vice versa. For the MC χ^2 columns the unfolded Data is compared to the true counts of the MC Model under which it was unfolded. Unfolded Data from Model 1 and Model 2* ***are compared*** *to Model 1's MC true counts when they are unfolded using Model 1 MC and they* ***are compared*** *to Model 2's MC true counts when they are unfolded using Model 2 MC.*

In particular, when both sets of Data are paired with the wrong MC model during unfolding, they both consistently behave similarly across the observed regions of τ values, and reach a minimum at a value of τ that aligns well with confirmatory χ^2 values for the same Data when paired with the correct MC models. As a contrast to finding k as depicted in Figure 4, these observations inform a method for choosing the value of τ which minimizes the combined sum of squares of the residuals between the unfolded counts of test Data sets and their true counts.

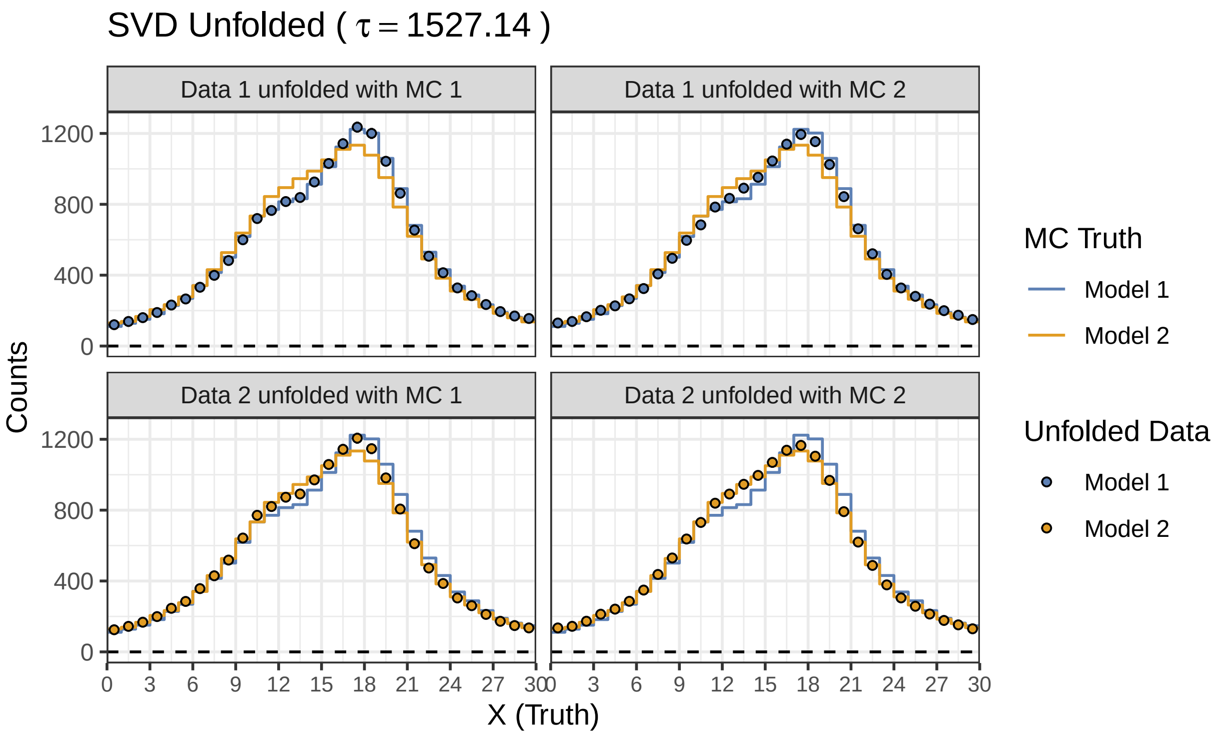

Figure 6: *Unfolded Data superimposed on both MC Truth histograms. Data from both Model 1 and 2 had a difficult time unfolding using the opposite Model's MC response matrix.*

A numerical comparison of both methods' sums of squares is given in Table 1 for unfolded Data against its true counts and for unfolded data against the MC true counts from the Model under which the Data was unfolded. From the individual χ^2s it is evident that the method of minimizing the total sums of squares $\sum \chi^2$ does better at minimizing Model bias while unfolding Data to estimate their true counts, and the $k = 5$ method does a better job fitting unfolded Data to the MC true counts regardless of whether or not the correct Model was used, indicating a higher degree of bias. However, by a quite noticeable margin the MC χ^2 columns for both methods clearly and correctly indicate which Model is more likely to have produced the Data being unfolded.

Complementing this observation is Figure 6, which depicts each of the unfolded results plotted against both MC simulations' true counts. It is clear from the plots that Data unfolded under the correct Model better approximates true counts of the MC. Which Model the unfolded Data belongs to in the mismatched unfolding cases is visually more ambiguous. Numerically, for Data from Model 1 unfolded using Model 2 MC, the χ^2 when comparing to Model 1 MC true counts is 6.49 (vs. 39.21 for Model 2), and for Data from Model 2 unfolded using Model 1 MC, the χ^2 when comparing to Model 2 MC true counts is 4.67 (vs. 27.35 for Model 1). As such, SVD unfolding continues to clearly support the correct Model.

3 Conclusion

The SVD Unfolding method described here has had a storied career in particle physics experiments since its introduction [13][12][6][10][33][9][5][2][3][17][4][7][1][26][11]. There are however many other unfolding methods that are currently heavily implemented. One is an iterative unfolding method based on Bayes' theorem [22][23]. Popular among experiments at the Large Hadron Collider is the unfolding method SPlot [31], which focuses on unfolding the individual contributions to a signal of interest, such as those making up Model 1 and Model 2 here. Also of note is another unfolding method based on least-squares and Tikhonov regularization called TUnfold [34], as well as a self-described iterative, dynamically stabilized (IDS) method of unfolding [27][28]. In contrast to these methods is an ongoing discussion about whether unfolded histograms need to be, or should be, used to test hypotheses. Some work has been done towards exploring this question and developing a "bottom-line-test" that directly compares theoretical predictions that have been smeared and distorted by a simulated detector response to detector data that has not been unfolded [20]. With that said, unfolding continues to be an extremely important part of particle physics analyses. It is not uncommon to consult multiple methods of unfolding for an analysis, and any such method of hypothesis testing that does not involve unfolding would at this time likely exist as another supplementary method in this way.

4 Acknowledgements

I am grateful to my physics advisor, Heidi Schellman, for making it possible for me to study statistics these last few years in concurrence with my physics PhD research. Through her I have been funded in part by the National Science Foundation. It is my hope that the material I've learned in pursuing a statistics degree will makeup for any misgivings she may have had along the way. I would also like to thank my statistics advisor, James Molyneux, for taking me in when my previous advisor left Oregon State University for a position elsewhere. His active engagement and endless optimism have been a guiding light towards my completion of this less traditional statistics MS project. He really unfolded away some of the negative biases and uncertainties in my own perception of my work. I would also like to thank Sarah Emerson and Sharmodeep Bhattacharyya for being my committee members as well as being excellent disseminators of statistics knowledge and wisdom. I could not have asked for better instructors for the foundational courses I took under them. Finally, I need to thank my partner, Katarina. She constantly inspires me to be a better person in ways that positively impact my relationship with others, with my work, and with myself. I love you, Kat.

A Extra Plots

Below are some supplementary plots relevant to Section 2.2.

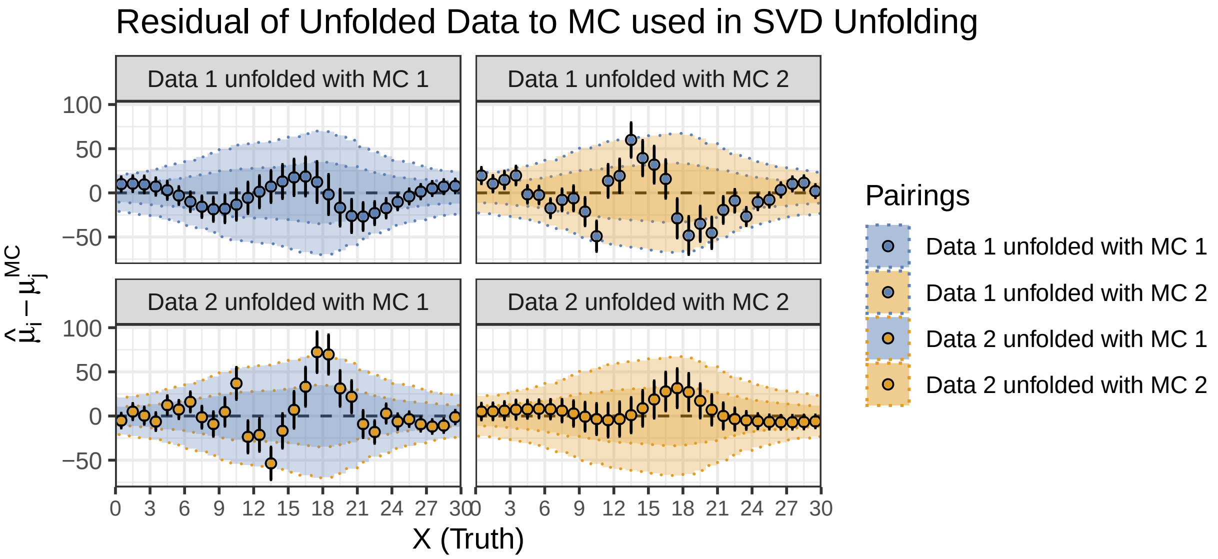

Figure 7: *The above is a plot of the residuals of the SVD unfolded Data with respect to the MC model used in the unfolding. The shaded regions represent 1 and 2 SD based on a per-bin Poisson model assumption for the MC's Truth counts.*

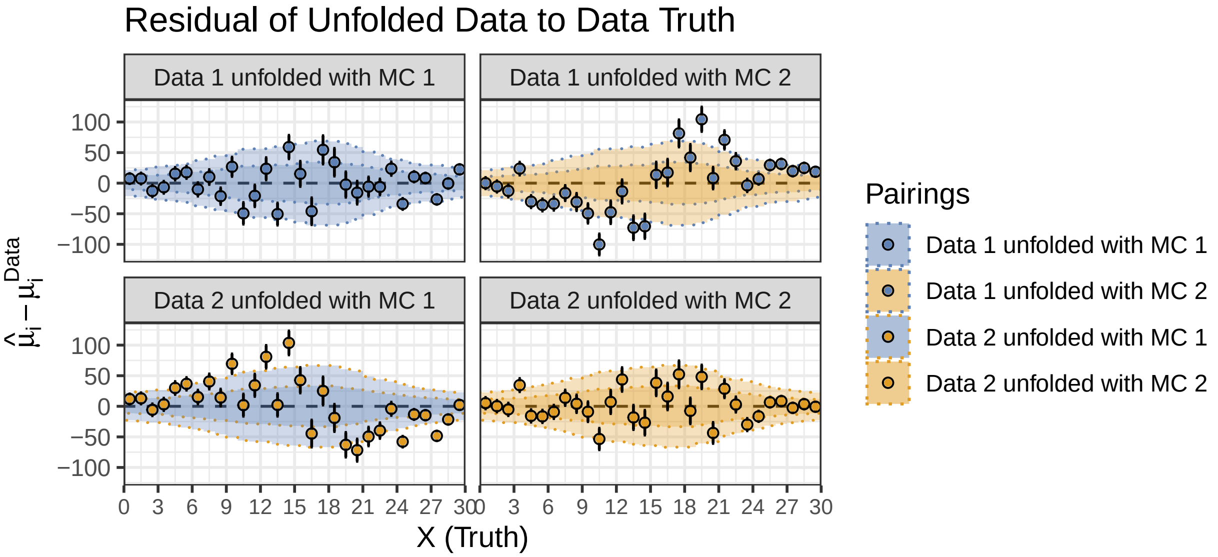

Figure 8: *The above is a plot of the residuals of the SVD unfolded Data with respect to the same Data's Truth counts. SVD unnfolded Data from Model 1 is only compared to Model 1 Data's Truth counts, and similarly for Model 2. The shaded regions represent 1 and 2 SD based on a per-bin Poisson model assumption for the Data's Truth counts.*

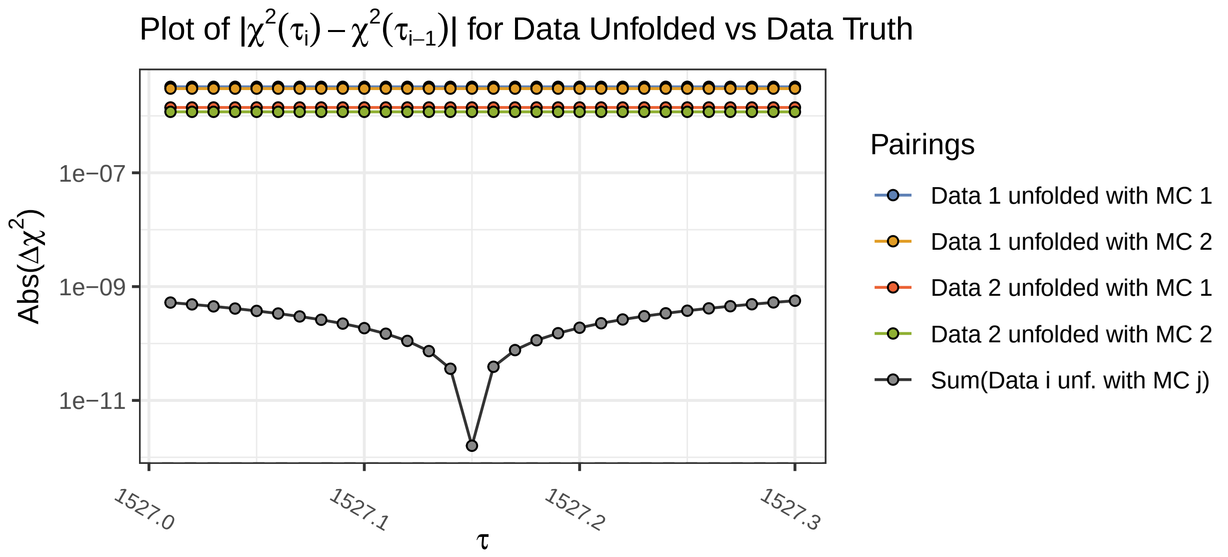

Figure 9: *For adjacent tested values of τ, the above plot shows the absolute value of the change in the weighted sums of squares of the residuals of the unfolded Data with respect to its true counts. The minimum value for the sum of the pairings' χ^2's indicates where the $\sum \chi^2$ shifts from decreasing with τ to increasing with τ.*

B Hilbert Spaces

Hilbert spaces are a prominent feature in the field of functional analysis. They see significant application in partial differential equations, quantum mechanics, and signal processing, where they are commonly implemented in the performance of Fourier analysis. Mathematically they represent an extension beyond the real and complex geometric-like vector spaces developed by earlier generalizations of Euclidean spaces in the 19th century. Developments in real analysis at the beginning of the 20th century lead to spaces of functions and sequences being conceptualized as linear spaces in their own right.

As extensions of previously understood spaces they necessarily exist at the intersection of several other important spaces that aught to be understood beforehand. With that said, the following definitions come from Rudin in [32]. To start, a **vector space**, as defined here, consists of a set X of vectors for which addition and scalar multiplication are defined such that for all $x, y, z \in X$ and any complex number $\alpha \in \mathbb{C}$

1. there exists a vector in X such that
   (a) addition is commutative: $x + y = y + x$,
   (b) addition is associative: $x + (y + z) = (x + y) + z$,

2. αx exists in X such that $1x = x$, $0x = 0$ (the zero vector), and multiplication is distributive:

   (a) $\alpha(\beta x) = (\alpha\beta x)$,

   (b) $\alpha(x + y) = \alpha x + \alpha y$, and

   (c) $(\alpha + \beta)x = \alpha x + \beta x$.

The range of α above describes a complex vector space. If α is restricted to the reals $\mathbb{R}$, then X is considered a real vector space. Note that vector spaces include more than just traditional coordinate-style vectors, but also include function spaces such as the vector space of all polynomials with degree of at most n, which has the basis $\{1, x, x^2, \ldots, x^{n-1}, x^n\}$.

Typically associated in applications, metric spaces form a another relevant set of spaces that has some significant overlap with the vector spaces. A space X is said to be a **metric space** if for all $x, y \in X$ there exists an operator $d(x, y)$ that maps them to a nonnegative real number that defines their distance from each other within X. The properties of this operator are

1. $0 \leq d(x, y) < \infty$ for all x and $y \in X$,
2. $d(x, y) = 0$ iff $x = y$,
3. $d(x, y) = d(y, x)$ for all x and $y \in X$,
4. $d(x, z) \leq d(x, y) + d(y, z)$ for all x, y, $z \in X$.

For a metric space X, the distance operator d is referred to as the metric on X. The intersection of the vector and metric spaces form the set of normed spaces. As an extension of the conditions thus far, a space X is a **normed space** if $\forall x \in X$ there exists a nonnegative real number $||x||$, called the **norm** of x such that

1. $||x + y|| \leq ||x|| + ||y|| \; \forall x, y \in X$,
2. $||\alpha x|| = |\alpha| \, ||x||$ if $x \in X$ and α is a scalar,
3. $||x|| > 0$ if $x \neq 0$.

Such a set is said to be **complete** if every **Cauchy sequence** in X converges to a point in X. A Cauchy sequence in a metric space X is any sequence $\{x_n\}$ that $\forall \varepsilon > 0$ there exists an integer N such that $d(x_m, x_n) < \varepsilon$ when $m > N$ and $n > N$. A quick example of this is the sequence defined by $x_n = \sqrt{n}$. For some starting x_m and x_n where $m - n = \delta$, we have

$$\begin{aligned}
d(x_m, x_n) &= \sqrt{m} - \sqrt{n} \\
&= \sqrt{n+\delta} - \sqrt{n} \\
&= (\sqrt{n+\delta} - \sqrt{n}) \frac{\sqrt{n+\delta} + \sqrt{n}}{\sqrt{n+\delta} + \sqrt{n}} \\
&= \frac{n + \delta - n}{\sqrt{n+\delta} + \sqrt{n}} \\
&= \frac{\delta}{\sqrt{n}(\sqrt{1+\delta/n} + \sqrt{1})} \\
&< \frac{1}{\sqrt{n}} \left(\frac{\delta}{2} \right) < \varepsilon \\
\implies n &> \left(\frac{\delta}{2\varepsilon} \right)^2 .
\end{aligned}$$

Noting that for constant δ the limit of $\frac{1}{\sqrt{n}}\left(\frac{\delta}{2}\right)$ as $n \longrightarrow \infty$ is the zero vector (the point of convergence) would also be sufficient to show that $x_n = \sqrt{n}$ is a Cauchy sequence.

Incidentally, a normed vector space that is complete as defined here meets the definition of a **Banach space**. An additional subset of the normed vector spaces consists of those spaces in which for all $x, y \in X$ there exists a real or complex number $\langle x, y \rangle$ defined by an operator called the **inner product**. For all $x, y, z \in X$ this operation must satisfy

1. $\langle x, y \rangle = \langle y, x \rangle^*$ (where the * represents the complex conjugate),
2. $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$,
3. $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$ (for $\alpha \in \mathbb{C}$),
4. $\langle x, x \rangle \geq 0$, and
5. $\langle x, x \rangle = 0$ iff $x = 0$.

A space that satisfies these requirements forms an **inner product space**, and the inner product defined in such a space relates to the form of its norm, such that $||x|| = \langle x, x \rangle^{1/2}$. Finally, at the intersection of Banach spaces and inner product spaces are the Hilbert spaces. I.e. a **Hilbert space** is a complete vector space with an inner product defined by its norm.

A commonly presented example of a Hilbert space is the L^2 function space, which consists of functions that are square integrable, i.e. if $f(x) \in L^2 \implies ||f(x)||^2 = \int_\chi |f(x)|^2 dx < \infty$, where χ is the domain of x. The subset $L^2[-\pi, \pi]$, where $\chi = [-\pi, \pi]$, has the well known

Fourier series as a basis, which is commonly written such that for $f(x) \in L^2[-\pi, \pi]$

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos(nx) + b_n \sin(nx)\right],$$

where

$$a_n = \tfrac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(nx) dx$$

$$\text{and}$$

$$b_n = \tfrac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(nx) dx.$$

Verification that this basis meets all the requirements laid out here for the basis of a Hilbert Space is beyond the scope of this paper.

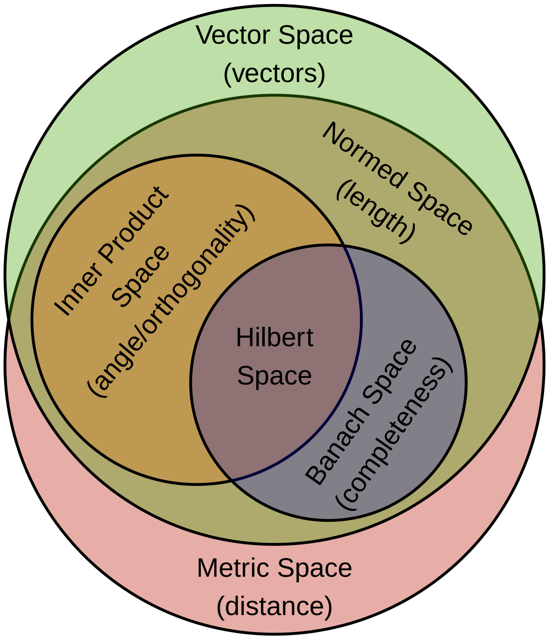

Figure 10: *A Venn diagram representing the intersection and nesting of the spaces described in Appendix B.*

C Description of Simulations

The Cauchy processes specifically are of the form

$$X_{1,i} \sim \text{Cauchy}(11, 4)$$

$$X_{2,i} \sim \text{Cauchy}(18, 4)$$

$$X_{3,i} \sim \text{Cauchy}(14, 5)$$

The probabilities under Model 1 (in which only the first two processes take place) is $\boldsymbol{p} = \{0.3, 0.7\}$. The probabilities governing Model 2 are generated from $\boldsymbol{p}$ by

$$\boldsymbol{p'} = \boldsymbol{\Lambda p} = \begin{pmatrix} 0.75 & 0 \\ 0 & 0.75 \\ 0.6 & 0.1 \end{pmatrix} \begin{pmatrix} 0.3 \\ 0.7 \end{pmatrix} = \begin{pmatrix} 0.225 \\ 0.525 \\ 0.25 \end{pmatrix}.$$

The individual weighted distributions contributing to each model's physical processes are shown below in Figure 11.

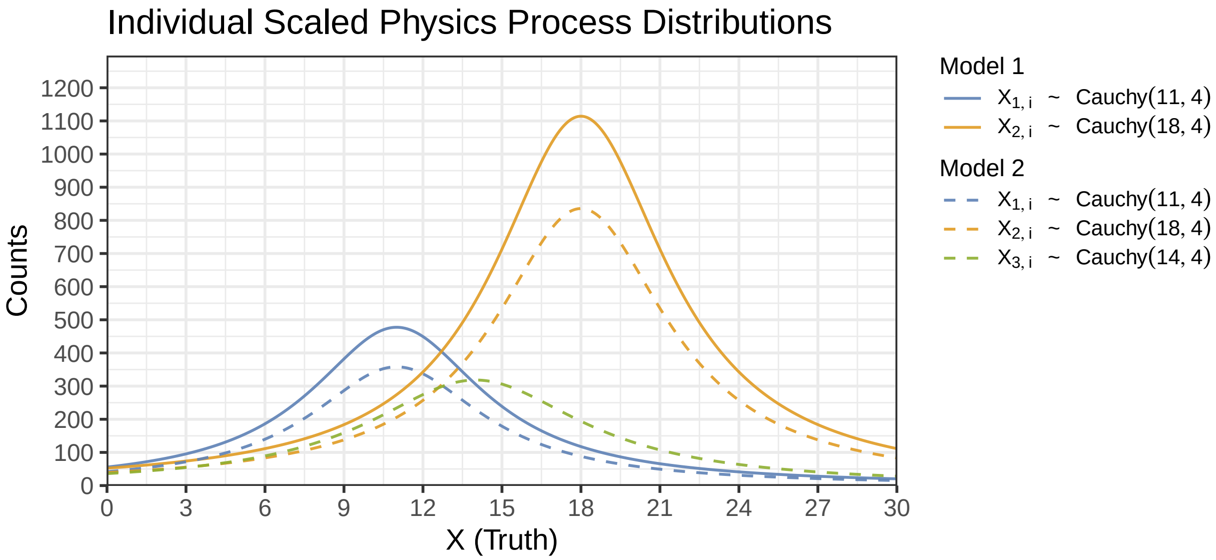

Figure 11: *The figure contains the individual underlying distributions guiding each contributing physics process per model. They are weighted to reflect their relative contributions.*

The effects of detector smearing is represented by i.i.d random variables generated by the conditional Gaussian process

$$\varepsilon_i \sim N\left(\mu(X_i), \sigma(X_i)^2\right),$$

the mean and variance of which are functions defined by

$$\mu(X_i = x) = -x^{1/4} \quad \text{and}$$
$$\sigma(X_i = x) = \log\left(\frac{x + 10}{4}\right).$$

The efficiency is similarly conditional on X_i, and is modeled here as a Bernoulli process with i.i.d random variables $\epsilon_i \sim \text{Bernoulli}\big(p(X_i)\big)$, where the average detection rate (when $\epsilon_i = 1$)

is a function of the form

$$p(X_i = x) = 1 - e^{-\sqrt{x}/4}.$$

D Singular Value Decomposition

The below definitions were provided by the Wikipedia page for Singular Value Decomposition [38] or by a page it directly links to.

Singular value decomposition involves the factorization of an $N \times M$ matrix $\boldsymbol{A}$ into

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T, \tag{30}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are respectively $N \times N$ and $M \times M$ unitary matrices and $\boldsymbol{S}$ is an $N \times M$ rectangular diagonal matrix consisting of non-negative real numbers.

A **unitary** matrix is any matrix in which its conjugate transpose is also its inverse, i.e. $\boldsymbol{U}^\dagger = \boldsymbol{U}^{-1}$. When the contents of a unitary matrix are all real it is also referred to as an **orthogonal** or **orthonormal** matrix. A **rectangular diagonal matrix** is simply a rectangular (not necessarily square) in which all of its off diagonal components are 0, i.e. $S_{ij} = 0$ if $i \neq j$. The matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ can be thought of as two rotation matrices that sandwich a rescaling matrix $\boldsymbol{S}$.

References

[1] Alexander Aab et al. “Depth of Maximum of Air-Shower Profiles at the Pierre Auger Observatory: Measurements at Energies above $10^{17.8}$ eV”. In: *Phys. Rev. D* 90.12 (2014), p. 122005. DOI: 10.1103/PhysRevD.90.122005. arXiv: 1409.4809 [astro-ph.HE].

[2] Georges Aad et al. “Measurement of inclusive jet and dijet production in *pp* collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector”. In: *Phys. Rev. D* 86 (2012), p. 014022. DOI: 10.1103/PhysRevD.86.014022. arXiv: 1112.6297 [hep-ex].

[3] Georges Aad et al. “Measurement of the jet radius and transverse momentum dependence of inclusive jet suppression in lead-lead collisions at $\sqrt{s_{NN}}$= 2.76 TeV with the ATLAS detector”. In: *Phys. Lett. B* 719 (2013), pp. 220–241. DOI: 10.1016/j.physletb.2013.01.024. arXiv: 1208.1967 [hep-ex].

[4] T. Aaltonen et al. "Measurement of the top quark forward-backward production asymmetry and its dependence on event kinematic properties". In: *Phys. Rev. D* 87.9 (2013), p. 092002. DOI: 10.1103/PhysRevD.87.092002. arXiv: 1211.1003 [hep-ex].

[5] R. Abbasi et al. "Measurement of the atmospheric neutrino energy spectrum from 100 GeV to 400 TeV with IceCube". In: *Phys. Rev. D* 83 (2011), p. 012001. DOI: 10.1103/PhysRevD.83.012001. arXiv: 1010.3980 [astro-ph.HE].

[6] G. Abbiendi et al. "Inclusive analysis of the b quark fragmentation function in Z decays at LEP". In: *Eur. Phys. J. C* 29 (2003), pp. 463–478. DOI: 10.1140/epjc/s2003-01229-x. arXiv: hep-ex/0210031.

[7] B. Abelev et al. "Measurement of charged jet suppression in Pb-Pb collisions at $\sqrt{s_{NN}}$ = 2.76 TeV". In: *JHEP* 03 (2014), p. 013. DOI: 10.1007/JHEP03(2014)013. arXiv: 1311.0633 [nucl-ex].

[8] Tim Adye. "Unfolding algorithms and tests using RooUnfold". In: *PHYSTAT 2011*. Geneva: CERN, 2011, pp. 313–318. DOI: 10.5170/CERN-2011-006.313. eprint: 1105.1160.

[9] Alexis A. Aguilar-Arevalo et al. "Measurement of ν_μ and $\bar{\nu}_\mu$ induced neutral current single π^0 production cross sections on mineral oil at $E_\nu \sim \mathcal{O}(1\mathrm{GeV})$". In: *Phys. Rev. D* 81 (2010), p. 013005. DOI: 10.1103/PhysRevD.81.013005. arXiv: 0911.2063 [hep-ex].

[10] A. Aloisio et al. "Measurement of $\sigma(e^+e^- \to \pi^+\pi^-\gamma)$ and extraction of $\sigma(e^+e^- \to \pi^+\pi^-)$ below 1-GeV with the KLOE detector". In: *Phys. Lett. B* 606 (2005), pp. 12–24. DOI: 10.1016/j.physletb.2004.11.068. arXiv: hep-ex/0407048.

[11] Feng Peng An et al. "Measurement of the Reactor Antineutrino Flux and Spectrum at Daya Bay". In: *Phys. Rev. Lett.* 116.6 (2016). [Erratum: Phys.Rev.Lett. 118, 099902 (2017)], p. 061801. DOI: 10.1103/PhysRevLett.116.061801. arXiv: 1508.04233 [hep-ex].

[12] S. Anderson et al. "Hadronic structure in the decay tau- —> pi- pi0 neutrino(tau)". In: *Phys. Rev. D* 61 (2000), p. 112002. DOI: 10.1103/PhysRevD.61.112002. arXiv: hep-ex/9910046.

[13] R. Barate et al. "Measurement of the spectral functions of axial - vector hadronic tau decays and determination of alpha(S)(M**2(tau))". In: *Eur. Phys. J. C* 4 (1998), pp. 409–431. DOI: 10.1007/s100520050217.

[14] Volker Blobel. "Unfolding". In: *Data analysis in high energy physics: A practical guide to statistical methods*. Ed. by Olaf Behnke et al. Weinheim, Germany: Wiley-VCH, 2013. Chap. 6, pp. 187–226.

[15] Volker Blobel. "Unfolding Methods in Particle Physics". In: *PHYSTAT 2011*. Geneva: CERN, 2011, pp. 240–251. DOI: 10.5170/CERN-2011-006.252. eprint: 1105.1160.

[16] Mary L. Boas. *Mathematical Methods in the Physical Sciences*. Third. Wiley, 2005. ISBN: 9780471198260.

[17] Jolanta Brodzicka et al. "Physics Achievements from the Belle Experiment". In: *PTEP* 2012 (2012), p. 04D001. DOI: 10.1093/ptep/pts072. arXiv: 1212.5342 [hep-ex].

[18] R. Brun and F. Rademakers. "ROOT: An object oriented data analysis framework". In: *Nucl. Instrum. Meth. A* 389 (1997). Ed. by M. Werlen and D. Perret-Gallix, pp. 81–86. DOI: 10.1016/S0168-9002(97)00048-X.

[19] G. Casella and R.L. Berger. *Statistical Inference*. Second. Cengage Learning, 2001. ISBN: 9780534243128.

[20] Robert D. Cousins, Samuel J. May, and Yipeng Sun. "Should unfolded histograms be used to test hypotheses?" In: (July 2016). arXiv: 1607.07038 [physics.data-an].

[21] G. Cowan. *Statistical Data Analysis*. Oxford University Press, USA, 1998. ISBN: 9780198501558.

[22] G. D'Agostini. "A Multidimensional unfolding method based on Bayes' theorem". In: *Nucl. Instrum. Meth. A* 362 (1995), pp. 487–498. DOI: 10.1016/0168-9002(95)00274-X.

[23] G. D'Agostini. "Improved iterative Bayesian unfolding". In: *Alliance Workshop on Unfolding and Data Correction*. Oct. 2010. arXiv: 1010.0632 [physics.data-an].

[24] Andreas Hocker and Vakhtang Kartvelishvili. "SVD approach to data unfolding". In: *Nucl. Instrum. Meth. A* 372 (1996), pp. 469–481. DOI: 10.1016/0168-9002(95)01478-0. arXiv: hep-ph/9509307.

[25] R. Johnson and D.W. Wichern. *Applied Multivariate Statistical Analysis*. Sixth. Pearson, 2007. ISBN: 9780131877153.

[26] Vardan Khachatryan et al. "Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8\,\mathrm{TeV}$". In: *Eur. Phys. J. C* 75.11 (2015), p. 542. DOI: 10.1140/epjc/s10052-015-3709-x. arXiv: 1505.04480 [hep-ex].

[27] Bogdan Malaescu. "An Iterative, dynamically stabilized method of data unfolding". In: (July 2009). arXiv: 0907.3791 [physics.data-an].

[28] Bogdan Malaescu. "An Iterative, Dynamically Stabilized(IDS) Method of Data Unfolding". In: *PHYSTAT 2011*. 2011, pp. 271–275. DOI: 10.5170/CERN-2011-006.271. arXiv: 1106.3107 [physics.data-an].

[29] Alexander Meister. *Deconvolution Problems in Nonparametric Statistics*. Vol. Lecture Notes in Statistics. Springer, 2009. ISBN: 9783540875567.

[30] Victor M. Panaretos. "A Statistician's View on Deconvolution and Unfolding". In: *PHYSTAT 2011*. Geneva: CERN, 2011, pp. 252–259. DOI: 10.5170/CERN-2011-006.252. eprint: 1105.1160.

[31] Muriel Pivk and Francois R. Le Diberder. “SPlot: A Statistical tool to unfold data distributions”. In: *Nucl. Instrum. Meth. A* 555 (2005), pp. 356–369. DOI: 10.1016/j.nima.2005.08.106. arXiv: physics/0402083.

[32] W. Rudin. *Functional Analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, 1991. ISBN: 9780070542365.

[33] S. Schael et al. “Branching ratios and spectral functions of tau decays: Final ALEPH measurements and physics implications”. In: *Phys. Rept.* 421 (2005), pp. 191–284. DOI: 10.1016/j.physrep.2005.06.007. arXiv: hep-ex/0506072.

[34] Stefan Schmitt. “TUnfold: an algorithm for correcting migration effects in high energy physics”. In: *JINST* 7 (2012), T10003. DOI: 10.1088/1748-0221/7/10/T10003. eprint: 1205.6201.

[35] Andrey N. Tikhonov and Vasiliy Y. Arsenin. *Solutions of ill-posed problems*. Scripta Series in Mathematic. V. H. Winston & Sons, 1977. ISBN: 9780470991244.

[36] Eric W. Weisstein. *Convolution. From MathWorld—A Wolfram Web Resource*. Last visited on 15/2/2022. URL: https://mathworld.wolfram.com/Convolution.html.

[37] Wessel N. van Wieringen. *Lecture notes on ridge regression*. 2021. arXiv: 1509.09169 [stat.ME].

[38] Wikipedia. *Singular value decomposition — Wikipedia, The Free Encyclopedia*. http://en.wikipedia.org/w/index.php?title=Singular%20value%20decomposition&oldid=1087254983. [Online; accessed 22-May-2022]. 2022.