Towards a Unifying View on Deconvolution in Cherenkov Astronomy⋆

Mirko Bunse, Nico Piatkowski, and Katharina Morik

TU Dortmund, AI Group, 44221 Dortmund, Germany
{firstname.lastname}@tu-dortmund.de

Abstract. Obtaining the distribution of a physical quantity is a frequent objective in experimental physics. In cases where the distribution of the relevant quantity cannot be accessed experimentally, it has to be reconstructed from distributions of correlated quantities that are measured, instead. This reconstruction is called deconvolution. While several approaches to deconvolution exist in experimental physics, the problem is rather unknown in data science. Both fields can benefit from each other, but notational differences tend to prevent this. In this work, we outline the unification of existing approaches in order to pave the way for new joint research directions. Cherenkov astronomy is employed as a use case which illustrates the deconvolution problem.

Keywords: deconvolution · unfolding · Cherenkov astronomy.

1   Introduction

Obtaining the distribution of a physical quantity is a frequent objective in experimental physics. An accurate and reliable estimate of the sought-after distribution, e.g. the energy spectrum of an astrophysical particle source, is crucial to understand the underlying physical principles. It is also important for the verification—or exclusion—of certain theoretical models. In cases where the distribution of the relevant quantity cannot be accessed experimentally, it has to be reconstructed from distributions of correlated quantities that are measured, instead. This reconstruction is called *deconvolution* (or *unfolding*). It states a prominent problem faced in experimental physics.

While several approaches to deconvolution exist in physics [2,6,8], the problem is rather unknown in data science. Both fields can benefit from each other, but notational inconsistencies tend to prevent this. In this work, we outline the unification of existing approaches in order to pave the way for new joint research directions. We consider Cherenkov astronomy as a use case of deconvolution which illustrates the problems faced in reconstructing the distribution of a physical quantity from correlated quantities.

Cherenkov astronomy studies the energy distribution of cosmic γ radiation to reason about the characteristics of celestial objects emitting such radiation. One prominent tool in this field are Imaging Air Cherenkov Telescopes (IACTs), which are placed at high altitudes on the surface of our planet. Since it is not viable to detect high-energy γ particles directly there, IACTs record Cherenkov light emitted by a cascade of secondary particles, instead. Such cascades, called air showers, are induced by γ particles interacting with Earth's atmosphere. Since the γ radiation is not directly measured by the telescopes, deconvolution is applied to reconstruct the γ energy distribution from the related Cherenkov light recorded by IACTs.
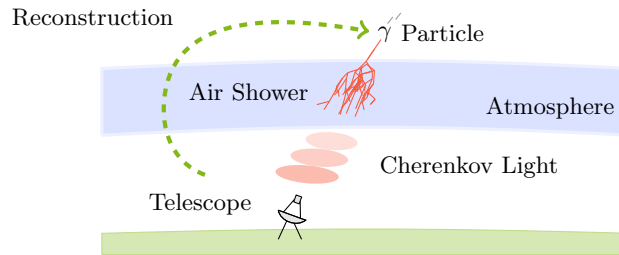
Fig. 1. A γ particle interacting in Earth's atmosphere produces a cascade of secondary particles. This air shower emits Cherenkov light, which is measured by an IACT [3]. The energy distribution of γ particles is reconstructed from IACT measurements.

The Cherenkov light emitted by each shower is recorded as a video sequence, from which features are extracted. These features represent each air shower on a higher level of abstraction, e.g. by its geometrical shape and size. Deconvolution adopts each feature as one observable quantity, from which the energy distribution of observed γ particles is reconstructed.

2   The Deconvolution Problem

Formally, the task of deconvolution is to estimate a density function $f : \mathcal{Y} \to \mathbb{R}$ of a random variable Y which represents a physical quantity, where $\mathcal{Y}$ denotes the state space of Y. In Cherenkov astronomy, the most prominent Y is the energy of γ particles of which the Cherenkov light is observed by an IACT. The task is aggravated by the following deficiencies inherent to the experimental setup [2]:

Transformation When Y cannot be measured directly, one has to measure a related but different quantity X instead. Only with sufficient knowledge about the relation between X and Y, f can be reconstructed from the density g of the measurable quantity X. For IACTs, the features of the observed light cone are measurable quantities which are related to the energy Y.

Finite resolution The measurement may not be exact, resulting in noisy measurements of the quantity X. IACTs may fail to capture the exact shape and size of a light cone if this cone is observed only partially.

Background noise Additional events may be recorded which do not originate from the respective process under study. For example, not every observed air shower is induced by a γ particle emitted by the monitored γ ray source.

Limited acceptance The detector may fail to recognize some of the events that occur in the studied process. For example, IACT observations are dropped if the trigger is not sufficiently certain about the recording.

A reliable estimate of the relevant density f is only obtained, if these deficiencies are rectified, i.e. if f is appropriately reconstructed from the measured density g. The name of deconvolution is motivated by g being modeled as a convolution of f with a detector response function $R : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$.

$$g(x) \;=\; \int_{\mathcal{Y}} R(x \,|\, y) \cdot f(y) \; \mathrm{d}y \tag{1}$$

R provides the link between X and Y, incorporating background knowledge about their relation. Specifically, $R(x \,|\, y)$ represents the conditional probability of measuring some $x \in \mathcal{X}$ when the actual value of the relevant quantity is $y \in \mathcal{Y}$. To obtain f, this model of g has to be inverted—it has to be "deconvolved". This is done by fitting f to the model of g, when g and R are given[1].

Classical deconvolution algorithms solve a discrete variant of the continuous deconvolution problem from Equation 1. This discrete variant presented in Equation 2 is a linear system of equations where f, g, and R from the continuous problem are replaced by discretized versions. The relation between the continuous case (Eq. 1) and the discrete case (Eq. 2) is clarified by Equation 3, where we denote the discretized problem for each discretized state of the observed data $\mathbf{g} = (\mathbf{g}_1, \ldots \mathbf{g}_J) \in \mathbb{R}^J$.

$$\mathbf{g} \;=\; \mathbf{R}\,\mathbf{f} \tag{2}$$

$$\Leftrightarrow \quad \mathbf{g}_j \;=\; \sum_{i=1}^{I} \mathbf{R}_{ij}\,\mathbf{f}_i \qquad 1 \leq j \leq J \tag{3}$$

Here, each dimension of the vector $\mathbf{g}$ corresponds to a specific discrete value $\mathrm{disc}_g(x)$, and each dimension of the vector $\mathbf{f}$ corresponds to a specific discrete value $\mathrm{disc}_f(y)$. Since the domain $\mathcal{X}$ of g is usually multi-dimensional, clustering is employed to map each vector-valued observation $\mathbf{x} \in \mathcal{X}$ to the index j of the cluster which the observation belongs to, effectively aggregating the multi-dimensional feature space into a single dimension. The conditional probabilities in $\mathbf{R}$ are estimated from a set of training observations. Two deconvolution methods are considered state-of-the-art, even though they have not yet been bench-marked on IACT data:

---

[1] Readers familiar with deconvolution will notice that our nomenclature is slightly different from the notation that is conventional in particle physics. Specifically, the roles of the letters x and y are exchanged here for compliance with the notation that is common in machine learning: Here, y refers to the value of the target variable and x is the observed value. This notation clarifies the relation between the deconvolution problem and classification—a relation that is important to our unifying view.

Regularized Unfolding performs a maximum likelihood fit of Equation 2, assuming that the absolute number of events for each discretized state of $g(x)$ is Poisson-distributed conditioned on $\mathbf{f}$ [1, 2]. The optimization is carried out via iterative numerical second-order optimization. Since a large variance between consecutive states is not physically plausible, regularization is employed to produce sufficiently stable results. The amount of regularization is controlled by the degrees of freedom in the second-order local model.

Iterative Bayesian Unfolding estimates the target distribution by applying Bayes' theorem to the observed frequencies $\mathbf{g}$ and the conditional probabilities embodied by the detector response matrix $\mathbf{R}$ [5, 6]. Bayes' theorem is applied multiple times, each time updating the prior distribution with the respective latest estimate. Repeating the update reduces the influence of the—potentially inappropriate—initial prior distribution.

3 Deconvolution as a Classification Task

The novel Dortmund spectrum estimation algorithm (DSEA) [8] is of particular interest for the data science community, because it translates the deconvolution problem into a multinomial classification task, instead of solving Equation 2. It thus opens the deconvolution problem to the field of machine learning in a principled way. Here, the discretized target variable Y serves as the label of the classification task. Thus, we may apply any multi-class classification algorithm to predict the range of Y-values which belong to an individual observation. DSEA underlies the intriguing idea that $\mathbf{f}_i = \mathbb{P}(Y \equiv i)$ can be recovered from a classifier's confidence. To see this, consider Equation 4, where the conditional probabilities $\hat{\mathbb{P}}(Y \equiv i | \mathbf{X} = \mathbf{x})$ are estimated by confidence values $c_{\mathcal{M}}(i \,|\, \mathbf{x}_n)$ of the underlying classifier, e.g. a random forest. In addition, a uniform prior is imposed on $\mathbf{x}$ (given that $\mathcal{X}$ is a closed subset of $\mathbb{R}^d$). The outcome is the DSEA estimator shown in Equation 5, which estimates $\mathbf{f}$ from confidence values.

$$\hat{\mathbb{P}}(Y \equiv i) \;=\; \sum_{\mathbf{x} \in \mathcal{X}} \hat{\mathbb{P}}(Y \equiv i | \mathbf{X} = \mathbf{x}) \cdot \hat{\mathbb{P}}(\mathbf{X} = \mathbf{x}) \tag{4}$$

$$\hat{\mathbf{f}}_i \;=\; \frac{1}{N} \sum_{n=1}^{N} c_{\mathcal{M}}(i \,|\, \mathbf{x}_n) \tag{5}$$

The deconvolution result is then improved by iterating the reconstruction, updating the distribution of the training set by weighting the examples according to the latest estimate of the target distribution. We suggest to scale the update step between iterations to ensure convergence of the algorithm. In fact, the original DSEA diverges from the true $\mathbf{f}$ after having found a suitable estimate in some iteration. Our suggestion is inspired by a common algorithmic pattern in numerical optimization, where the basic algorithm selects a search direction and a sub-routine called *line search* selects an appropriate step size for this direction.
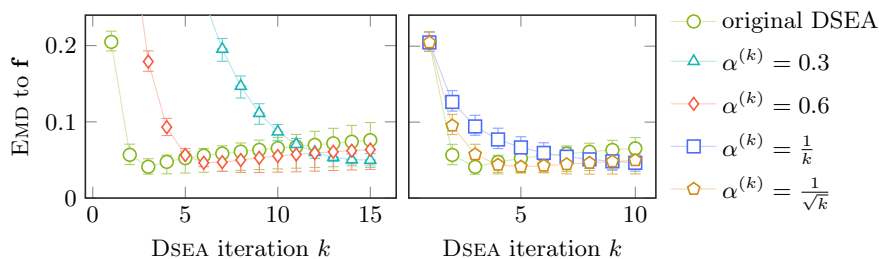
Fig. 2. The convergence of DSEA is controlled by scaling each step with $\alpha^{(k)}$, as indicated by the Earth Mover's Distance [7] between the estimate $\hat{\mathbf{f}}^{(k)}$ in iteration k and the true density $\mathbf{f}$. The original DSEA uses $\alpha^{(k)} = 1$. **Left:** Constant step sizes approach and leave the optimal solution slowly, facilitating the choice of a convergence threshold. **Right:** Decaying step sizes enforce convergence in proximity of the optimum.

4   Conclusion and Future Work

We have illustrated the problem of deconvolution, employing Cherenkov astronomy as a use case. Our presentation provides a basis for unifying the view on current state-of-the-art algorithms from particle physics in the context of machine learning. Against the background of many possible synergies arising from collaborative research on deconvolution, we have given one example of an algorithmic improvement inspired by numerical optimization.

Even though deconvolution methods have been surveyed elsewhere [4], a comparative study evaluating these methods on benchmark data is still missing. We plan to deliver such a study on data taken with IACTs.

References

1. Blobel, V.: Unfolding methods in high-energy physics experiments. Tech. rep., CERN (1985)
2. Blobel, V.: An unfolding method for high energy physics experiments. In: Adv. Stat. Techniques in Part. Phys. pp. 258–267 (2002)
3. Bockermann, C., Brügge, K., Buss, J., Egorov, A., Morik, K., Rhode, W., Ruhe, T.: Online analysis of high-volume data streams in astroparticle physics. In: Proc. of the ECML-PKDD 2015. pp. 100–115. Springer (2015)
4. Cowan, G.: A survey of unfolding methods for particle physics. In: Adv. Stat. Techniques in Part. Phys. (2002)
5. D'Agostini, G.: A multidimensional unfolding method based on Bayes' theorem. Nucl. Instrum. Methods Phys. Res. A **362**(2-3), 487–498 (1995)
6. D'Agostini, G.: Improved iterative Bayesian unfolding. arXiv:1010.0632 (2010)
7. Rubner, Y., Tomasi, C., Guibas, L.J.: A metric for distributions with applications to image databases. In: Proc. of the 6th ICCV. pp. 59–66. IEEE (1998)
8. Ruhe, T., Börner, M., Wornowizki, M., et al.: Mining for spectra – the Dortmund spectrum estimation algorithm. In: Proc. of the ADASS XXVI (2016)