

# Information Criteria for Selecting Parton Distribution Function Solutions

Aurore Courtoy<sup>(a,1)</sup>, Arturo Ibsen<sup>(a,2)</sup>

<sup>(a)</sup>*Instituto de Física, Universidad Nacional Autónoma de México,  
Apartado Postal 20-364, 01000 Ciudad de México, México \**

(Dated: November 12, 2025)

In data-driven determination of Parton Distribution Functions (PDFs) in global QCD analyses, uncovering the true underlying distributions is complicated by a highly convoluted inverse problem. The determination of PDFs can be understood as the inference of a function supported on  $[0, 1]$ , a problem that admits multiple acceptable solutions. An ensemble of solutions exists that pass all standard goodness-of-fit criteria.

In this paper, we propose algorithms for the classification, clustering, and selection of solutions to the determination of PDFs, or any functions on  $[0, 1]$ , based on the characterization of their shape. We explore information-theoretic based (Rényi entropy and divergence) and optimal-transport based (Wasserstein distance) criteria. In particular, we advocate for the use of the Rényi entropy as an *absolute* estimator per solution, as opposed to *relative* estimators that compare solutions pairwise. We show that the Rényi entropy can characterize the space of solutions *w.r.t.* the PDF shapes. Paired with the identification of the optimal combination of solutions via Pareto fronts, it provides a plausible and minimalist selection algorithm. Moreover, Rényi entropy proves versatile for use in clustering applications.

## I. INTRODUCTION

Distribution functions play a central role in scattering processes involving hadrons in the deep inelastic regime, where cross sections can be factorized into hard and soft contributions. These non-perturbative mathematical objects, the PDFs, depend on two variables, the momentum fraction  $x$  and the hard scale  $Q^2$ . While the behavior of PDFs with respect to the latter is known from perturbative QCD, their shape as functions of  $x$  is not. This characteristic is inherent to low-energy dynamics and thus largely inaccessible from first principles, although many efforts have been put forth in recent years, *e.g.* by the Schwinger-Dyson or, more importantly, the lattice QCD communities [1, 2].

PDFs are more reliably determined through data-driven analyses in the QCD framework, where they are extracted by solving an inverse problem. The state-of-the-art proton PDFs [3–5] result from analyses that combine experimental, theoretical and statistical tools and frameworks, aiming to tackle the many aspects of the PDF inverse problem. Progress in PDF inference has reached an unprecedented refinement in terms of statistical treatment, in particular, when it comes to quantifying the uncertainties on PDFs, see *e.g.* [6–11].

The inverse problem, formulated through optimization of an objective function, consists in inferring models for PDFs. The convolutions inherent to the theoretical description of observables make the resulting inverse problem ill-posed, *i.e.* it does not admit a unique solution. A model for PDF consists of a mathematical basis and a set of hyperparameters, each leading to a central solution accompanied by an Hessian matrix or, in the Monte Carlo approach, to a distribution of solutions – both reflecting the propagation of the experimental (*aleatoric*) uncertainty. Commonly adopted approaches model PDFs using a single, so-called “fixed,” functional form (a parametric approach) or by neural networks (a non-parametric approach). A recently developed non-parametric method instead generates polynomials by varying hyperparameters [9]. As most non-parametric approaches produce ensembles of solutions, dedicated tools are required to classify them. This is particularly relevant for addressing parametrization bias or model uncertainty. The non-uniqueness of the plausible models manifests itself through the coexistence of multiple solutions, functions of  $x$ , *i.e.* the PDF shapes. How this ensemble of solutions is accounted for contributes to the global analysis’s methodology; for the CTEQ-TEA group, it gives rise to an *epistemic* uncertainty [6].

In this context, a prescription for model combination was designed [8, 12], which raises the following question: How can we select the most diverse shapes that span the space of solutions? Addressing this question requires introducing estimators, such as metrics or information criteria, to characterize the various solutions as shapes, across various flavor combinations. Specifically, we propose using the Rényi entropy, a generalization of the Shannon entropy, to quantify shapes in terms of their *bumpiness* or *flatness*. As we will argue, the Rényi entropy is an *absolute*, as opposed to

---

\*Electronic address: <sup>1</sup>[aurorecourtoy@gmail.com](mailto:aurorecourtoy@gmail.com), <sup>2</sup>[a.i.v.r@ciencias.unam.mx](mailto:a.i.v.r@ciencias.unam.mx)

*relative*, quantity: it quantifies each curve independently, without reference to another solution or curve. It is also combinatorially more efficient than evaluating relative quantities, which will also be discussed and compared to. Once the Rényi space defined, the most representative shapes can be selected. To this end, while there are many algorithms available to cluster and characterize such optimal solutions, we employ the concept of Pareto fronts, developed in the context of multiobjective optimization.

We illustrate how Rényi entropy can be used to select, classify, and cluster families of shapes on  $x \in [0, 1]$ , using the PDFs of the pion generated within the Fantômas framework [7]. For the FantoPDFs, the selection of the most diverse shapes, called *metamorphs*, was originally performed “by hand.” The present manuscript upgrades this feature, which could be extended to families of PDFs with more flavor combinations, effectively using the pion PDF analyses as a sandbox for proton studies.

The manuscript is organized as follows. Section II outlines the motivation for classifying and selecting solutions of PDFs. The core of the paper consists in the selection criteria, distinguishing between *absolute* and *relative* types. We show that the results obtained through the Rényi -entropy based algorithm (Section III A) are qualitatively comparable to using *relative* criteria, such as the Rényi divergence or the Wasserstein distance (Section III B). Comments and caveats are discussed in Section IV, and further in the conclusions (Section V). Mathematical details, specifically of the Pareto fronts, are provided in the Appendices A-B.

## II. PARTON DISTRIBUTION FUNCTIONS

Parton Distribution Functions characterize the behavior of quarks and gluons as a function of the momentum fraction  $x$  of the parent hadron they carry when interacting with a hard probe, at a factorization scale  $\mu$ . PDFs are mathematical objects, formally defined through factorization theorems, with a formulation on the light cone. For example, the quark PDFs  $f_{a/p}$  representing the distribution of quarks of flavor  $a$  in a hadron  $p$  are defined in the  $\overline{\text{MS}}$  factorization scheme as

$$f_{a/p}(x, \mu^2) = \frac{1}{4\pi} \int dz^- e^{-ixP^+z^-} \langle P | \bar{\psi}_a(0, z^-, \mathbf{0}) \gamma^+ W(z^-, 0) \psi_a(0) | P \rangle, \quad (1)$$

where we have used the light-cone coordinates,  $a^\pm = (a^0 \pm a^3)/\sqrt{2}$  and  $a^\mu = (a^+, \vec{a}_\perp, a^-)$ , and

$$W(z^-, 0) = \mathcal{P} \exp \left( -ig \int_0^{z^-} d\bar{z}^- \hat{A}^+(0^+, \bar{z}^-, \vec{0}_T) \right) \quad (2)$$

is the Wilson eikonal line that ensures color gauge invariance. In the light-cone gauge, and to LO, PDFs have been interpreted as Probability Density Functions, which is intuitively the distribution of the momentum fraction  $x$  among all members,  $a \in \{u, \bar{u}, d, \bar{d}, s, \bar{s}, g, \dots\}$  a flavor that contributes to the proton’s structure and  $a \in \{u, \bar{d}, \bar{u} = d = s = \bar{s}, g\}$  a flavor that contributes to the pion’s structure. Beyond those approximations, PDFs do no longer follow a probability density interpretation. Still, their support is  $x \in [0, 1]$  for quarks and gluons ; they are derivable on  $x \in ]0, 1[$  and serve as weights to the  $x$ ’s  $n$  projections, called Mellin moments. These properties lead to the integrability rules

$$\begin{aligned} \int_0^1 dx x^{n>0} f_{\{S,g\}/p}(x, Q^2) & \quad \text{are finite,} \\ \int_0^1 dx x^{n\geq 0} f_{v/p}(x, Q^2) & \quad \text{are finite,} \end{aligned} \quad (3)$$

where  $v$  refers to valence, and  $S, g$  to sea and gluon, respectively. The integrability of PDF’s projections builds on the Operator Product Expansion connection to PDFs and the relation to Mellin moments as defined through local operators. Some of those moments reflect conservation laws through sum rules, *e.g.* the valence sum rule corresponding to  $n = 0$ , and are hierarchically important *first-principle* constraints on PDFs.

Besides the above-mentioned properties, the shape of PDFs is largely unconstrained by the theory of the strong interaction itself, as PDFs appear to be non-perturbative object at the core – they bridge the degrees of freedom of perturbative QCD to bound-states. Hence, the determination of PDFs as functions of  $x$  is truly twofold – uncovering the low-energy dynamics that binds partons into hadrons, and using solely a perturbative framework for data-driven analyses to infer such shapes.

The inference of PDFs results of global QCD analyses of large pools of data with an extended coverage in  $(x, Q^2)$ , where  $Q^2$  is often identified with  $\mu^2$ . Each such optimization problem aims to find the family of functions,  $f_{a/p}$  with

a component of vector, which length and composition depends both on the flavor structure of the hadron and the experimental access to processes that are sensitive to those flavors. For example, while proton PDFs span spaces that can be as large as 9-dimensional, the pion PDFs are experimentally less constrained and a 3-dimensional space is the best that can be explored at this time. Of relevance to the present analysis, we will focus on  $n_{\text{flavor}}$ -dimensional spaces in which the families of functions are defined.

To each direction of the  $n_{\text{flavor}}$ -dimensional space corresponds a function of  $x$ , which, more often than not, is given in terms of polynomials, characterized by its generating function (or basis), its hyperparameters, and its vector of (effective) free parameters. The latter are commonly used to characterize the dimensions of the problem. For a single-family solution, they are best illustrated by the PDFs obtained in the Hessian formalism. For example, for the pion PDFs, the xFitter group used a fixed parametrization with  $4(v) + 3(S) + 2(g) = 9$  parameters [13], corresponding to a 9-dimensional eigenvector space for the Hessian matrix, and the objective function. On the other hand, a baseline solution for the proton PDF spans, at least, a 26-dimensional parameter space [3].

Heavily convoluted inverse problems do not have a unique solution, and, as such, allow for multiple shapes of PDFs. While not all solutions are statistically acceptable, *w.r.t.* goodness-of-fit and analytical properties mentioned above, many constitute very good complementary solutions. Once filtered through  $\chi^2 + \delta\chi^2$  and physical priors (such as soft positivity bounds), those are qualitatively equally good [14]. They are expected to reflect the imperfect kinematical coverage of the  $(x, Q^2)$  plane. As such, the various acceptable solutions should span the extrapolation regions to result in large uncertainties, *i.e.* the uncertainty is large where no data or no theoretical penalties are constraining the shapes.

Accounting for more families of solutions means that diverse models are generated and tried against the data. Models for PDFs are built on a basis  $\mathcal{Q}$

$$x f_{a/h}(x, Q_0^2) = A_a x^{\alpha_a} (1-x)^{\beta_a} \times \mathcal{Q}(g[x]; \underline{c}, \underline{c}^h), \quad (4)$$

where  $\underline{c}^h$  is the vector of hyperparameters inherent to the model, and the optimization determines the best values for the vector of free parameters,  $\{A_a, \alpha_a, \beta_a; \underline{c}\}$ . There exist various proposals to generate ensembles of models to sample from. We will mainly focus on the non-parametric, Fantômas methodology [9]. To generate **metamorphs**,  $\mathcal{Q} = 1 + \mathcal{B}$ , with  $\mathcal{B}$  a Bézier curve, and  $\underline{c}^h = \{N_m, \underline{x}, g[x]\}$ , respectively, the degree of the polynomial, the vector of abscissa for the control points, and the scaling function of the argument. Each set of  $\underline{c}^h$  generates a different model, all of which are not necessarily related one to another.<sup>1</sup> A family of solutions can thus be represented as  $\{x f_{v/\pi}^k(x, Q_0^2), x f_{S/\pi}^k(x, Q_0^2), x f_{g/\pi}^k(x, Q_0^2)\}$  for the pion PDFs, with  $k \in 1, \dots, n_s$  labeling the  $k^{\text{th}}$  of all  $N_s$  acceptable solutions.

The classification of solutions is reduced to the characterization of that  $n_{\text{flavor}}$ -dimensional space. Conceptually, we switch the focus from parameter space to an  $n_{\text{flavor}}$ -dimensional space. In doing so, we acknowledge the difficulty of mapping the free parameters between the different models. In that regard, it is akin to the philosophy underlying the use of neural networks for PDF determination. This peculiarity renders the cross-analysis of the families of solutions rather intricate. To streamline this analysis, we conceptually integrate the free parameters of each flavor into the shape of the PDFs. The latter then need to be characterized, which we do through an elegant strategy, as described below.

### III. SELECTION CRITERIA

The sets of solutions described above must hence be classified or organized to obtain a full span of the space of solutions. In the case of sampling over models, the parametrizations, or models, that led to the solutions do not emerge from an obvious probability distribution. Hence there should be no notion of density in how the space is populated. On the contrary, the entire set of solutions illustrates the tossing of markers into the hyperparameter space, each toss generating a model. The resulting models can cluster in neighboring regions. As such, keeping only those few solutions that best characterize the space is the task that we will be describing here below. Our main results are found using *absolute shape estimators*, Section III A. They are challenged and confirmed using *relative criteria*, such as distances and divergences III B.

---

<sup>1</sup> The Bézier curve formalism allows to relate polynomials of different degrees easily.

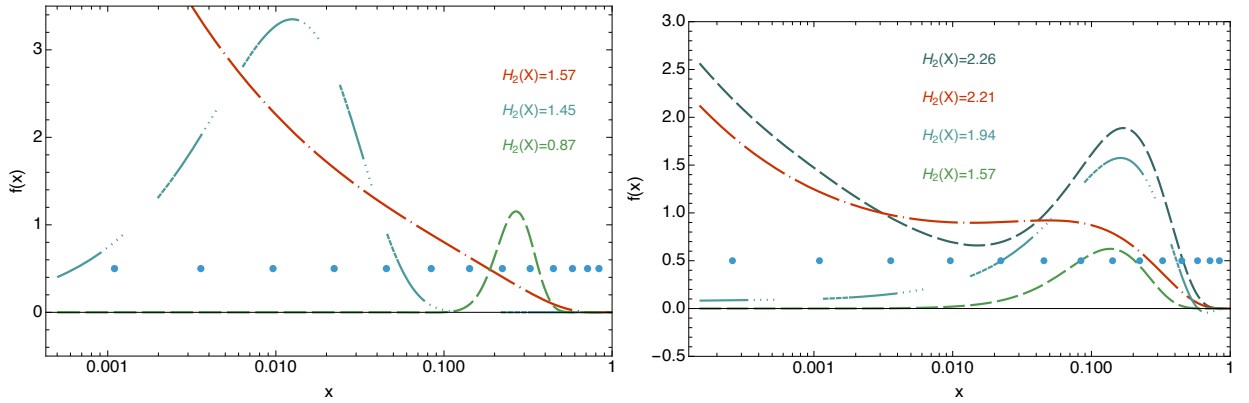


FIG. 1: Illustration of values for  $H_2(X)$  for various shapes. (Left) ideal case of shapes ranging from (in decreasing order) steep (dot-dashed red), spread-out (3-dot-dashed cyan) and highly-concentrated (dashed green), for tweaked Gaussian point (blue dots). (Right) similar color code and characteristics, with the addition of the short-dashed blue-grey curve whose  $H_2(X)$  is similar to the steep curve. Both show a similar pattern in variations, not magnitude, at the evaluation points.

### A. Absolute criterion

Models for PDFs consist in functional forms of the variable  $x \in [0, 1]$  and a vector of training parameters,  $\underline{c}'$ , while each model is a realization of a basis cast into a functional form through a choice of hyperparameters,  $\underline{c}^h$ . What characterizes the various solutions is the shape of the solution, *e.g.* bumpy, flat, exponentially falling off, etc. An absolute description of each curve can be obtained through the evaluation of its Rényi entropy,  $H_\alpha$ . Defined for probability densities,  $H_\alpha$  reflects how certain (uncertain) a random variable (here, the momentum fraction  $x$ ) is. In other words, a pronounced maximum—a bump—is certain and has a low entropy, while a flat region is uncertain and corresponds to a higher entropy. Even though PDFs are not proper probability densities<sup>2</sup>, using  $H_\alpha$  as an estimator is still a valid shape characterizer or bump-width filter.

The Rényi entropy is defined as

$$H_\alpha(X) = \frac{1}{1-\alpha} \ln \left( \sum_{i=1}^n p(x_i)^\alpha \right), \quad \alpha \neq 1 \quad (5)$$

where  $p(x)$  is the probability density, *i.e.* a discrete random variable  $X$  with possible outcomes  $\{x_1, x_2, \dots, x_n\}$  and corresponding probabilities  $\{p_1, p_2, \dots, p_n\}$ . The value of  $\alpha$  gives emphasis on core or tails, with  $\alpha = 2$  commonly called the collision entropy or the Rényi entropy. Higher values of  $\alpha$  will weigh maxima, or bumps, more and the other way round, with the particular case of recovering the Shannon entropy for  $\alpha = 1$ . The choice  $\alpha = 2$ , which we will adopt from now on, constitutes a good balance between identifying the bumps and accounting for more flat ranges.

For densities normalized according to  $f(x_i)/(\sum_i^n f(x_i))$  with  $f(x)$  a function akin to PDFs,  $H_2(X)$  is bounded by 0, for a delta function, and  $\ln n$ , for a flat distribution. For unimodal functions, *smooth* enough function at  $X$  or function with steep rise (yet integrable) at end-points, the entropy may vary continuously from highly concentrated, to spread-out, to steep, and to flat solutions. This is illustrated, in a linear-log scale, in Fig. 1. The concept of concentration does not reflect the magnitude of the curves, nor the position of the bumps. That is, the most extreme entropies do not necessarily encompass all the ordinates of intermediate solutions.

The evaluation of the Rényi entropy is defined for discrete points,  $x_i$ , and will depend on the  $x$ -grid that we choose, as it is the case for many estimators. In the case of PDFs, the  $x$  spacing can be chosen so to span (or not) the data coverage. We have chosen to use tweaked Gaussian points (see Appendix A), so that we have a direct access to an

<sup>2</sup> Not all PDFs obey  $f_{a/q}(x \rightarrow 0) \rightarrow 0$ .

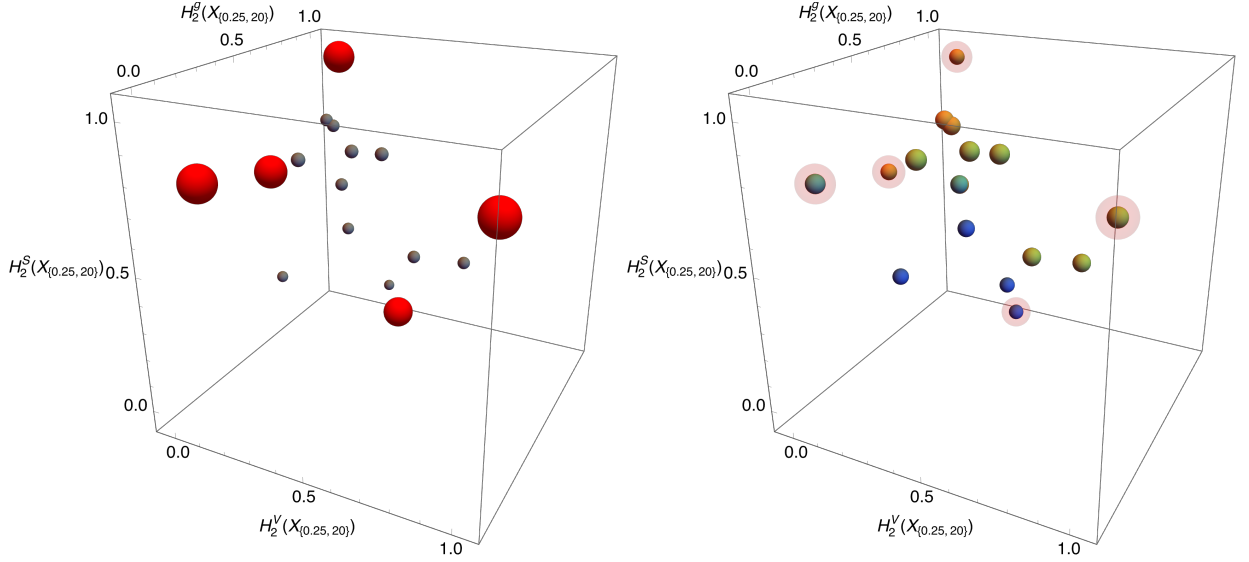


FIG. 2: Rényi 3-dimensional space. Left panel: The light gray points represent all acceptable solutions. Larger red points represent the vertices of the Pareto hypersurfaces, *i.e.* floor and roof. Right panel: The four clusters identified through the UMAP nearest-neighbor algorithm (orange, green, cyan, and dark blue), along with the vertices of the Pareto fronts (low opacity red circles). Both are illustrated for  $X_{0.25, 20}$ .

integral or kernel interpretation of the argument of the logarithm in Eq. (5). In practice, we evaluate

$$H_2^q(X_{\alpha_x, n, \tau}) = - \ln \left( \sum_{i=1}^n \left( \frac{x_i f_{a/h}(x_i)}{\sum_{j=1}^n x_j f_{a/h}(x_j)} \right)^2 \right), \quad (6)$$

with the PDFs evaluated at  $Q_0$  and  $X_{\alpha_x, n, \tau}$  is a truncated ( $\tau$ ) vector of  $n$   $x_i$  defined with a scaling  $\alpha_x$ . We chose to use the  $x f(x)$  product as it highlights features of the solutions that would otherwise be lost ; it is due to the fact that the integrability of the sea and the gluon is ensured starting from the definition of the average momentum fraction, Eq. (3).

A function is neither uniquely nor unambiguously determined by its entropy. For example, on the *r.h.s* of Fig. 1, two curves which small- $x$  behavior is similar but differs in the large- $x$  region display a similar value for  $H_2(X)$ . This is understood by the trend in the curves's derivatives. Rényi entropy can nonetheless lead to a satisfactory classification of functions on  $[0, 1]$ .

Evaluating the Rényi entropy for all families of solutions is straightforward and fast. Once obtained, the values of the Rényi entropy can be visualized as coordinates in the  $n_{\text{flavor}}$ -dimensional space of Rényi entropy, *e.g.* a 3-dimensional space formed by the values of the entropy for the valence, sea and gluon PDFs of the pion. It is more informative to work with normalized values of  $H_2(X)$ , so all spaces will be unit cubes, as illustrated in Fig. 2. The convex hull of the points can be used to define the volume occupied by the solutions, see *e.g.* [15]. Identifying the most extreme points of that volume would allow to balance the diversity of shapes among all flavors. Instead, we choose to analyze Pareto fronts, which give us the floor (and, conversely, the roof)<sup>3</sup> of enveloping solutions,

$$\begin{aligned} P_{\text{floor}}(Y) &= \{\underline{y} \in Y : \{\underline{y}' \in Y : \underline{y}' \preceq \underline{y}, \underline{y} \neq \underline{y}'\}\} \\ P_{\text{roof}}(Y) &= \{\underline{y} \in Y : \{\underline{y}' \in Y : \underline{y}' \succeq \underline{y}, \underline{y} \neq \underline{y}'\}\} \end{aligned} \quad (7)$$

with  $Y \equiv H_{2, X_{\alpha_x, n, \tau}}^q$  and  $\underline{y}$  correspond to the coordinates, *e.g.*  $\{H_2^V, H_2^S, H_2^g\}$  for the pion PDF. The symbols  $\{\preceq, \succeq\}$  denote the componentwise partial order, and are defined in Appendix B.

<sup>3</sup> In economy, the Pareto floor and roof are referred to as, correspondingly, Pareto front and worse front.

The Pareto fronts represent the optimal balance between all  $n_{\text{flavor}}$  coordinates in either a minimal or maximal fashion. While all PDFs are related via the momentum sum rule, *e.g.* there is already a balance in the momentum fractions, the tradeoff revealed by the Pareto fronts arises in identifying the diversity in entropies across flavors. For a point on the Pareto floor, if one coordinate is large, at least one of the other two must be smaller; similarly, for the Pareto roof, if one coordinate is small, at least one other must be larger. This balance is not guaranteed for points in between the fronts. The obtained Pareto fronts, while discrete, can be seen as approximating continuous hypersurfaces derived through ordering. This allows the approach to be extended to flavor spaces with  $n_{\text{flavor}} > 3$ , as illustrated here, providing a visually intuitive picture.

Since the fronts can still be large ensembles, we study the effect of selecting a subset, corresponding to the extrema of both fronts in all  $n_{\text{flavor}}$  directions,

$$\underline{y}_{\min} \preceq \underline{y}' \preceq \underline{y}_{\max}, \quad (8)$$

with

$$\underline{y}_{\min} / \underline{y}_{\max} = \min / \max_{\underline{y} \in P_{\text{floor}}(Y) \cup P_{\text{roof}}(Y)} y_i \quad \forall i = 1, \dots, n_{\text{flavor}}.$$

The vertices of the Pareto set, taken as the corner of the space, correspond to specific solutions for the PDFs. Given that PDFs are *smooth* enough, defining intervals of extrema of the Rényi entropy, as defined in Eq. (6), is sufficient to encompass most bounding shapes<sup>4</sup>.

In the Fantômas analysis, we tested about 100 solutions, of which we retain only  $N_s \sim 20$  that pass both the  $\chi^2 + \delta\chi^2 = 450$  threshold and a soft positivity constraint<sup>5</sup>. By applying the above-described methodology to the  $N_s$  solutions, we obtained 5 vertices that characterize the extrema of the three-dimensional space formed by the value of Rényi entropy for each flavor. They are shown in red on the *l.h.s.* of Fig. 2 for tweaked Gaussian points  $X_{0.25,20,\{5,18\}}$ . The corresponding *metamorph* solutions are illustrated in Fig. 3. This exercise, for the pion, has the peculiarity of including a zero gluon PDF as an acceptable solution. In the sense of Rényi entropy, it is a flat, uncertainty distribution, but it does not integrate to 1. Hence, it contributes with statistical noise, slightly preventing more general conclusions, see Appendix C for discussion.

The solutions that are selected through the vertices correspond, in methodology, to the Fantômas pion PDF [7], for which we hand-picked the most diverse shapes. The latter were then combined [8, 9] through the *mcgen* technique [12], for which each of the solution is

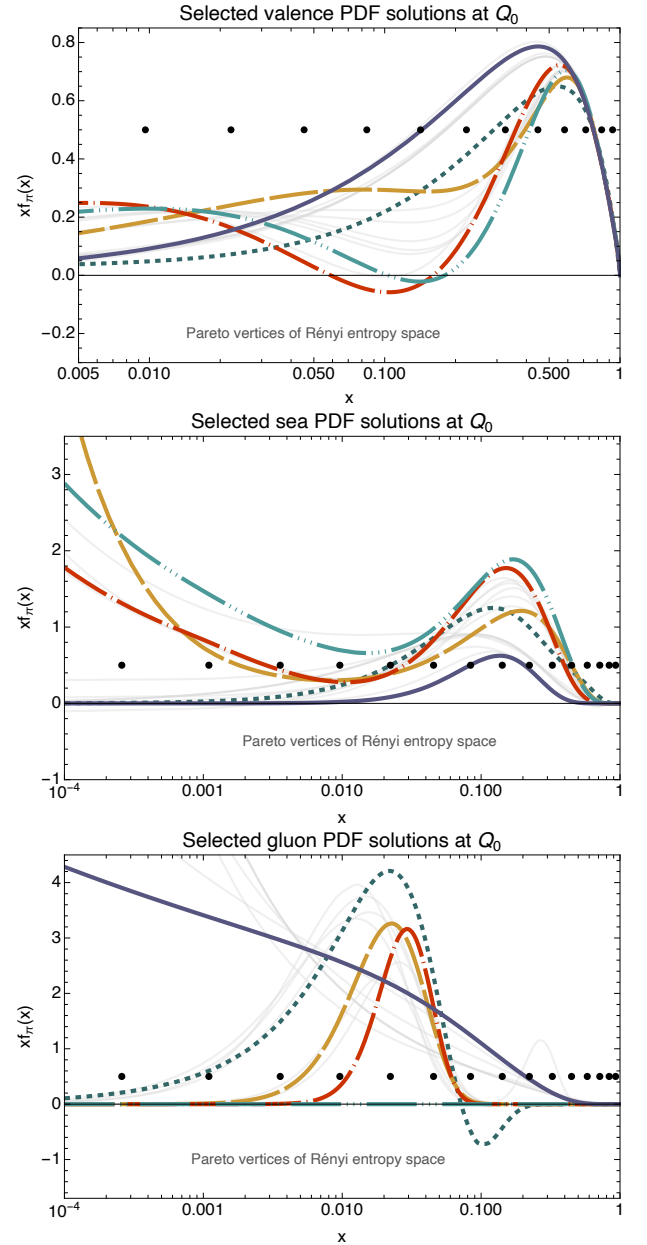


FIG. 3: Selected shapes (color) of the pool of acceptable metamorph solutions generated for the Fantômas pion PDFs (light grey). Rényi entropy-based coordinates are filtered as Pareto vertices. The black dots represent the  $x$  values at which the entropy was evaluated.

<sup>4</sup> See Appendix C for discussion of exceptions.

<sup>5</sup> A similar analysis without the soft constraint on positivity, retaining about  $\sim 40$  solutions, leads to similar results.

associated to a distribution through its corresponding uncertainty in the Hessian formalism. Then, appealing to the de Finetti's theorem, we argued that exchangeable, and independent and identically distributed distributions must find a common origin from an underlying truth, yet unknown, distribution. The selection criteria described here completes the full methodology of model combination.

Yet another way to use Rényi entropy is as an estimator for clustering the solutions. Indeed, the coordinate vector can be used either in an in-house clustering prescription or in more established techniques requiring marker vectors. We pursued this exercise using the online embedding projector tool<sup>6</sup> with the UMAP (Uniform Manifold Approximation and Projection) dimensionality reduction algorithm. While UMAP is not a clustering algorithm *per se*, it leverages distances among nearest neighbors to define the relevant axes in the space of effective dimensions. For the pion case, dimensionality reduction is not strictly necessary (flavor space is three-dimensional:  $v, g, S$ ), but examining the closest neighbors in the evaluated estimator (*vector*) provides useful information for classifying solution IDs (*metadata*). By using the Rényi entropy for all three flavors as coordinates of the (*vector*) estimator and the solution labels as (*metadata*), we identified four distinct regions, or clusters, with UMAP. These clusters are then mapped back into the actual  $n_{\text{flavor}}$ -dimensional space, as illustrated in the right panel of Fig. 3, along with the vertices of the Pareto hypersurfaces. In this case, each cluster contains at least one Pareto vertex, *i.e.*, it confirms that the Pareto vertices are representative of all clusters. Fig. 4 illustrates average estimator patterns in Rényi space for each identified cluster, confirming the difference between clusters, as well as a balance between smaller and larger entropies, on average, across flavors.

UMAP for estimator Rényi  $\alpha=2$ , rescaled

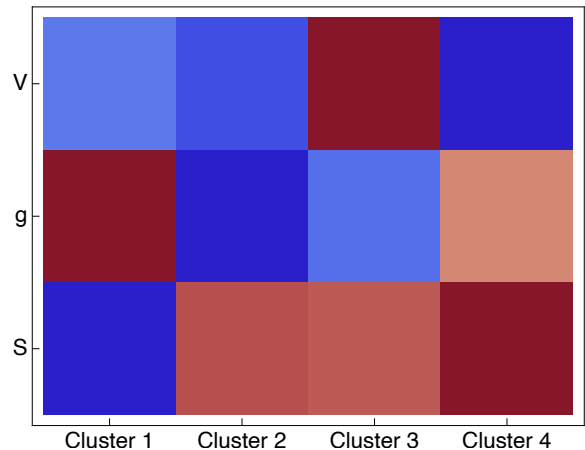


FIG. 4: Color matrix based on the value of the estimator  $H_2^{\{v,g,S\}}(X)$  for UMAP-identified clusters displayed in the right panel of Fig. 3. The values of the average Rényi entropy are rescaled between  $[0, 1]$  (from dark blue to dark red) for improved visualization. To each cluster corresponds a different pattern in projected entropies.

## B. Relative criteria

A common approach is to use divergences, which have an information-theoretic basis, and metrics, which are topological in nature, to classify and quantify distances between a pool of solutions and a reference baseline. On the information-theoretic side, the KullbackLeibler (KL) divergence has been widely used both to quantify solutions [16] and to profile PDFs [17].

Since the problem that we tackle in this paper is not necessarily characterized by a central value, evaluating relative measures requires comparing each solution with the others. It can become combinatorially expensive for a large number of trialed solutions. In the present exercise, only about  $N_s = 20$  solutions pass the pre-processing filters, rendering this cross-analysis manageable. We take advantage of this to show that the Rényi entropy produces satisfactory results, thereby reducing the need for longer computations. This is crucial when applying selection criteria to problems with larger solution sets, such as the case of the proton PDF [3, 18].

In this Section, we consider both an information-theoretic measure, the Rényi divergence, and a metric measure, the Wasserstein distance, as a way to cross-check the results obtained with an absolute criterion.

<sup>6</sup> <https://projector.tensorflow.org/>

### 1. Information character – Rényi divergence

A natural extension of our framework for comparing relative shapes involves the Rényi divergence, defined as,

$$D_\alpha(p||q) = \frac{1}{\alpha - 1} \ln \left( \sum_{i=1}^n \frac{p(x_i)^\alpha}{q(x_i)^{\alpha-1}} \right), \quad \alpha \neq 1 \quad (9)$$

with  $p$  the probability distribution from which  $p(X)$  is drawn, and  $Q$  for  $q(X)$ . For  $\alpha = 1$ ,  $D_\alpha(p||q)$  is nothing else than the KL divergence, where  $q$  plays the role of the reference density and  $p$  comes from the pool of trials.

Here, nor  $p$  or  $q$  represent a baseline. In practice, we will take  $\{p, q\}$  from the poll of solutions  $\underline{f} = \{xf_k, k = 1, \dots, N_s\}$ , and  $\alpha = 2$ ,

$$D_2(f_j||f_k) = \ln \left( \sum_{i=1}^n \frac{\left( x_i f_{a/h}^j(x_i) / \sum_{l=1}^n x_l f_{a/h}^j(x_l) \right)^2}{\left( x_i f_{a/h}^k(x_i) / \sum_{l=1}^n x_l f_{a/h}^k(x_l) \right)} \right), \quad j \neq k \quad (10)$$

By definition, the Rényi divergence is asymmetric, *i.e.*,  $D_2(f_j||f_k) \neq D_2(f_k||f_j)$ . The evaluation of all possible divergences involves  $N_s(N_s - 1)$  terms per flavor. For the example we are considering, the time involved for  $N_s \sim 20$  is nearly 30 times longer than for the evaluation of the set of Rényi entropies.

All divergences are normalized to  $D_2 \in [0, 1]$ , and then, again, considered as coordinates in the  $n_{\text{flavor}}$ -dimensional space. Since we would like to select the most diverging pairs, we identify the  $n_{\text{sol}} = 5$  points on the Pareto fronts with the highest norm as our best pairs – they are all located on the roof. We take the set of the **metamorphs** appearing in the vertices pairs as the desired set. This result is illustrated by the (long-dashed) mustard curves in Fig. 5. For all flavors but the gluon<sup>7</sup>, the outer limits of the selection based on the Rényi divergence are close to those obtained through taking the vertices of the Pareto fronts of the Rényi entropy space.

### 2. Metric character – Wasserstein distance

The Wasserstein metric exploits the topology of the problem, and measures the distance between two probability distributions  $P$  and  $Q$  on a metric space, characterized by a set  $\Omega$  and a metric  $d$ . It evaluates the minimum cost of transforming one distribution into another, where the cost is a function of the distance between the two distributions – it is akin to the work per unit mass for moving the masses. Hence,  $W_p$  represents a “horizontal distance” as opposed to the “vertical distances,” like  $L^p$  [19]. As such, it captures the displacements of the curve along the horizontal axis, which, for functions with support on  $[0, 1]$  translates into shifts in the shape. It is a key tool in the context of optimal transport theory [19, 20].

For probability distributions  $P$  and  $Q$  on  $(\Omega, d)$ , the  $p$ -th Wasserstein distance  $W_p(P, Q)$  is defined as:

$$W_p(P, Q) = \left( \inf_{\gamma \in \Gamma(P, Q)} \int_{\Omega \times \Omega} d(x, y)^p d\gamma(x, y) \right)^{\frac{1}{p}}, \quad (11)$$

where  $\Gamma(P, Q)$  is the set of all joint distributions  $\gamma$  on  $\Omega \times \Omega$  with marginals  $P$  and  $Q$ .  $d\gamma(x, y)$  represents the hypersurface where the mass that is moved from  $x$  to  $y$  lies, and

$$F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y d\gamma(u, v) du dv, \quad (12)$$

is its joint cumulative distribution function. The measures of masses, from  $x$  to  $y$ , are hence mapped by  $d\gamma(x, y)$ .  $d(x, y)$  is the distance between those mass distributions, and the cost is expressed as  $d(x, y)^p$ . The infimum,  $\inf$ , is the greatest lower bound, here, of  $\gamma \in \Gamma(P, Q)$ .

---

<sup>7</sup> See Appendix C for discussion on the gluon case.

In one dimension and for the specific case of a translation  $\gamma(x, y) = Q(x)\delta(y - T(x))$ , the Wasserstein distance can be expressed as

$$W_p(P, Q) = \left( \int_{\Omega} |x - T(x)|^p dQ(x) \right)^{\frac{1}{p}}. \quad (13)$$

In practice, we will use projected PDFs,  $xf(x)$ . We choose  $p = 1$ , known as the Earth Mover's Distance (EMD), in which case the flow of “mass” can be recast in such a way as to write  $W_1$  directly in terms of the cumulative distributions in  $x$ . As we will consider the discrete case, our final expression reads

$$W_1(f_j, f_k) = \sum_{i=2}^n \left| \left( \sum_{l=2}^i x_l f_{a/h}^j(x_l) \right) - \left( \sum_{l=2}^i x_l f_{a/h}^k(x_l) \right) \right| \times |x_i - x_{i-1}|, \quad (14)$$

where the discretization, in the Riemann sense, is accounted for through  $|x_i - x_{i-1}|$ . The Wasserstein distance is symmetric, hence it involves the combinatorics of  $(N_s, 2)$ , reducing the CPU time *w.r.t.* the Rényi divergence case by a factor 2. Yet, it is still 15 times longer than evaluating Rényi entropy, for the  $N_s \sim 20$  case. The selection of the most different shapes proceeds as for the Rényi divergence. The  $W_1$  distances act as coordinates in the  $n_{\text{flavor}}$ -dimensional space, the Pareto fronts of the scattered points are evaluated, and the  $n_{\text{sol}}$  points with largest norm are picked.

The resulting **metamorphs** are shown in (3-dots dashed) cyan on Fig. 5. The selection highlights complementary aspects *w.r.t.* those obtained through information criteria, due to  $W_1$  focusing on horizontal changes in the shapes. Yet, the results for the valence and sea sectors are qualitatively consistent with those obtained through evaluating directly the Rényi entropies as an absolute shape estimator. The gluon case shows differences between the selections, a peculiarity that comes from the presence of atypical shapes for PDFs. Yet, the results are acceptable, considering that the curves presented here will be accompanied by their respective error band, to be combined [8, 12]. This is further discussed in Appendix C.

While the case of collinear PDF seems too simple to resort to such topological metrics, the advantage lies in the generalization to higher dimensional problems. We note that the Wasserstein distance has been used in jet algorithms as a shape selection metric, in particular EMD [21–23], due to its topological connection to the measure of mass distribution.

#### IV. DISCUSSION

The objective of the present analysis is to provide an algorithm for shape selection through classification via estimators. The studied shapes correspond to the models for the pion PDFs, defined on  $x \in [0, 1]$ . As shape estimator, Rényi entropy provides a reliable *absolute* characterization of each curve separately. The CPU time that its evaluation requires is advantageous. The other two estimators that we have explored both characterize shapes in

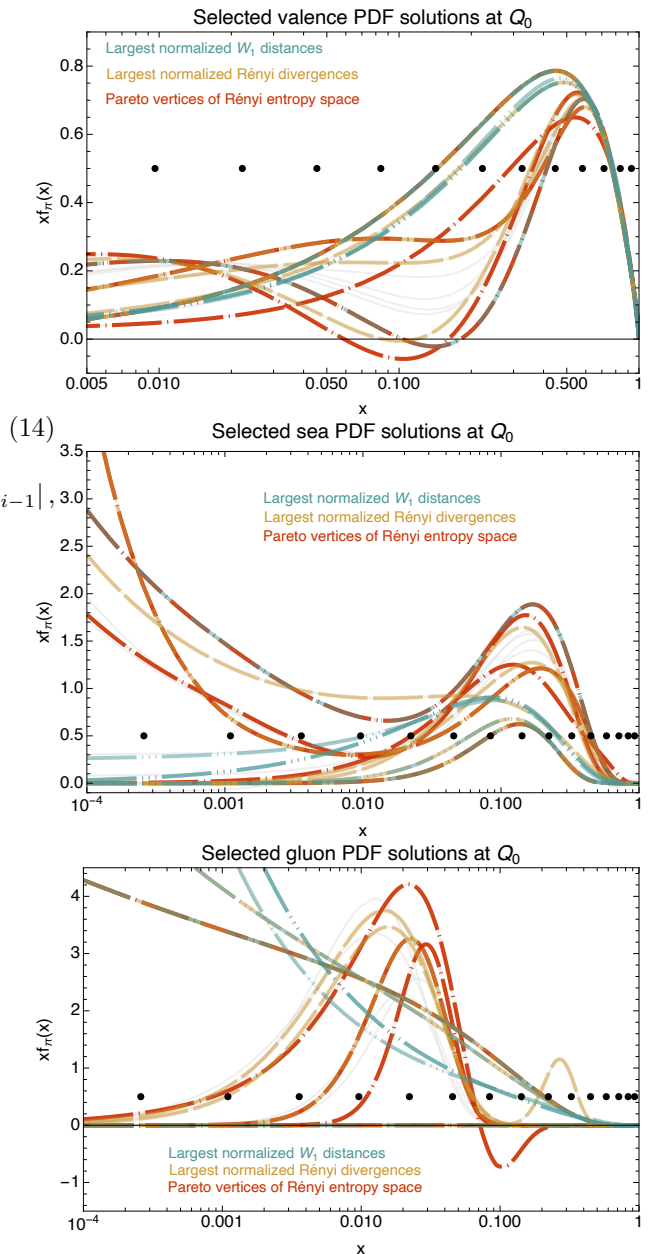


FIG. 5: Comparison of the selection obtained by the absolute criterion (dotted-dashed red), and two relative criteria – Rényi divergence (long dashed mustard) and Wasserstein distance (3 dots-dashed cyan).

pairs, *i.e.* they constitute *relative* estimators. The Rényi divergence, with its information-theoretic character, and the Wasserstein distance, with its metric character, turn out to be complementary measures.

Once the estimator evaluated, it fills the  $n_{\text{flavor}}$ -dimensional space with the coordinates of each PDF model. Now, at the selection step, the optimal distribution of the diverse shapes is nothing else than the Pareto fronts (Pareto front and worse front). In the case of the absolute estimator, we pick the vertices of both fronts as the compromised selection of most representative PDF models. This is justified by the expected forms of PDFs. As for the relative estimator, the  $n_{\text{sol}}$  models on the Pareto fronts with largest norm are chosen. All three estimators lead to consistent and coherent results for sets of continuously changing shapes. The evaluation of Rényi entropies is less computationally demanding, hence more advantageous.

Ideally, one would like to provide a full clustering algorithm. Instead, we show that clustering solutions for PDFs may not be necessary if the goal is the identify representative solutions. In Rényi entropy space, the Pareto fronts supply the role played by a two-step selection, that is the clustering followed by the identification of the best solution in each cluster. This is true if the solutions have been filtered by their  $\chi^2$  values prior to the selection procedure. However, Rényi entropy makes a particularly suitable estimator for clustering. While we have not fully pursued this path, we have shown its combined usage to UMAP dimensionality-reduction method.

A few comments and caveats shall still be discussed before we draw our final conclusions.

As Rényi entropy estimates the amount of information, it differs from the measure of *wiggleness*, the kinetic energy, proposed in Ref. [24] to tackle signs of overfitting, and that effectively accounts for the length of the curve. In this work, we did not attempt to characterize overfitting using information-theoretic criteria.

It may appear at a first glance that both Wasserstein and Pareto’s approaches consider hypersurfaces, or surfaces whose continuous limit are hypersurfaces in the case of the Pareto fronts defined earlier, as the result of optimization. Yet, their nature and usage are completely different. Wasserstein distance maps hypersurfaces, it is of topological nature and characterizes relative distances. On the other hand, Pareto orders discrete points to form an optimal front, which is used to define borders in some discrete parametric space.

In this analysis, we use Pareto fronts, not as an efficiency measure, but as a divider of coordinates in Rényi space. If the entropies were indeed a measure of optimization, *e.g.* [25, 26], then the Pareto fronts would lead to an efficiency measure. However, we can interpret the Pareto floor as the optimally best known distributions – a balance between the smallest combination of  $n_{\text{flavor}}$  values, which correspond to more concentrated, less entropic, shapes – and the Pareto roof as the optimally worse known distributions – conversely, a balance between the largest combinations of  $n_{\text{flavor}}$  values, which correspond to flat, more entropic, shapes.

The selection of the most diverse shapes based on the relative estimators presents an extra layer of complexity *w.r.t.* that based on the absolute estimator. While it is easy to identify the largest differing pairs from the Pareto roof of the  $n_{\text{flavor}}$ -dimensional space, some solutions may pair among each other to give small distances. Such pairs lie on the Pareto floor and they can easily be identified, and removed from the selection.

Another important aspect of the classification algorithms that we have proposed is the dependence on the choice for the vector  $X$ . In the above analysis, we have reported results for a specific choice for the evaluation of the various entropies, divergences and distances. We have explored the results with different choices of either  $\alpha_x$  parameter,  $n$  or even different spacing choices. While the results mildly vary with the choice of the  $X$  vector, the qualitative output is not substantially affected. As a matter of fact, the choice of the  $x$  span is part of the strategy to emphasize regions of expected spreads in parametrizations. The model for  $X$  is captured through the “hyperparameters”  $(\alpha_x, n, \tau)$ .

In the same line of thought, the freedom to choose  $X$  also allows to define partial Rényi entropies, for example at large vs. small- $x$  values, a strategy that we have explored, yet not pursued, for one or various of the  $n_{\text{flavor}}$  dimensions. We retain that this choice is an integral part of the definition of the shape estimator in the classification process.

The Rényi entropy-based algorithm could be extended to the study of solutions for the proton PDF. In recent analyses of the CTEQ-TEA collaboration, the role of parametrization bias has been accounted for through the concept of tolerance, consisting in expanding the  $\Delta\chi^2$  cut, which encompasses most solutions [3]. On-going analyses aim to quantify the parametrization bias, an effort to which the present work sums up. For the proton, spaces are 6-dimensional, at least. A preliminary usage of the Rényi entropy as an estimator (*vector*) in the UMAP technique displays promising results for pattern classification [27]. On the other hand, given the properties of the Pareto fronts (see Appendix B), the extension of the algorithm for the Rényi entropy selection could straightforwardly be extended to  $n_{\text{flavor}} = 6$ -dimensional spaces.

## V. CONCLUSIONS

Parton Distribution Functions embody the probability of finding a parton with a momentum fraction  $x$  of its hadron parent, at a given scale. Being highly non-perturbative objects, PDFs, as unknown functions of  $x \in [0, 1]$ , are best determined through data-driven analyses within the theoretical framework of QCD. Many models can be designed that will lead to optimized solutions, or multiple distinct solutions to a single model can be found, in the Monte Carlo approach, that show different patterns. The question arises as to how these solutions should be interpreted.

On the one hand, sampling over models, and hence generating more solutions, contributes to a specific case of epistemic uncertainty, namely model uncertainty [6–9]. Its contribution to the total uncertainty of PDFs is crucial for obtaining robust and accurate predictions. It is achieved via a combination method [12, 28], which produces a single set of PDFs. At present, the distributional nature of functional forms is not fully understood, although hints point toward the existence of higher parent distributions [8]. The pragmatic view-point on understanding the impact of the various possible solutions, that we adopt here, consists in identifying the most diverse solutions. Those solutions are regarded as  $n_{\text{flavor}}$  shapes, which we map onto estimators, so to classify and select the most diverse among them.

On the other hand, identifying patterns in PDF shapes in the Monte Carlo approach may also prove valuable, *e.g.* for assessing the role played by priors in the original bootstrapped distributions.

In this paper, we have proposed algorithms to address the problem of selecting shapes of PDFs. We have used both information-theoretic and optimal-transport based estimators, which differ significantly in how they measure diversity. The former quantifies how much a distribution is structured (a measure of *bumpiness*), while the other reflects the cost of shifting the shapes. Furthermore, we classify the estimators as *absolute*, referring to single-shape measures, as opposed to *relative*, which involve pairwise comparisons.

We advocate for the use of the Rényi entropy as a shape estimator for PDF solutions. It characterizes curves by a single number, hence we referred to Rényi entropy as an *absolute* estimator. While the Rényi entropy neither uniquely nor unambiguously determines a function, it appears to perform well in characterizing the space of solutions – each of them corresponds to  $n_{\text{flavor}}$  coordinates in Rényi space. The crucial step following that characterization consists in selecting the most diverse shapes, which encompass a large fraction of all possible solutions. This is achieved identifying the Pareto floor and roof of the ensemble of coordinates in Rényi space. The fronts lie inside the convex hull formed by the coordinates of all solutions, and the range of values spanned is determined by the vertices of the union of these fronts. Tested on the Fantômas pion PDFs, the results thus obtained are qualitatively similar to those presented in the original FantoPDFs [7]. More interestingly, Rényi entropy also proves to be a promising estimator for use in clustering algorithms. Both directions should be further explored in the context of proton PDF selection.

## Acknowledgments

AC thanks her CTEQ-TEA colleagues as well as the Fantômas team for fruitful discussions on model uncertainty and selection criteria. AC and AI are grateful to the PICO initiative Summer 2024 at UNAM during which this project started. AC is supported by the UNAM Grant No. DGAPA-PAPIIT IN102225. This work was performed in part at the Aspen Center for Physics, which is supported by a grant from the Simons Foundation (1161654, Troyer).

## Appendix A: Tweaked Gaussian points

The  $i$ -th Gauss node/point,  $x_i$ , is the  $i$ -th root of  $P_n$ , the Legendre polynomial of degree  $n$ , with respective weights  $w_i$  given by the formula

$$w_i = \frac{2}{(1 - x_i^2) [P'_n(x_i)]^2}$$

Using the Gauss-Legendre quadrature for functions with support in  $x \in [0, 1]$  as well as for the usual stretched/tweaked  $x$  power,  $x^{\alpha_x}$ , the Gaussian points and weights are such that

$$\int_0^1 f(x) dx \simeq \sum_{i=1}^n w'_i f(x'_i) \quad \text{with} \quad P_n(x'_i{}^{\alpha_x}) = 0 \quad . \quad (\text{A1})$$

We will refer to  $X_{\alpha_x, n, \tau}$  for the vector of truncated tweaked Gaussian points obtained from  $P_n(x^{\alpha_x})$ . The truncation will be specified by  $\tau$ .

## Appendix B: Pareto fronts as boundaries

The properties of the Pareto fronts have been largely studied. The most relevant ones for our purposes can be found in Emmerich [29] and Ehrgott [30].

Pareto fronts arise in optimization problems where more than one objective function/estimator must be considered simultaneously. The Pareto dominance in such problems defines a strict *partial order* for the objective space  $\mathbb{R}^n$ , *i.e.* it is defined by componentwise order (inequalities) among solutions in the  $n$ -dimensional space. Following the definition of the Pareto floor, *i.e.* Pareto front, and roof, *i.e.* worse front, we summarize a few key properties.

A simple geometric property is that the union of floor and roof points lies inside the convex hull of all points,

$$P_{\text{floor}}(\mathcal{Y}) \cup P_{\text{roof}}(\mathcal{Y}) \subset \text{conv}(\mathcal{Y}) \quad , \quad (\text{B1})$$

with  $\mathcal{Y}$  the set of points, and  $\text{conv}(\mathcal{Y})$  is the convex hull formed by all points in the objective space. Following [30], in optimization problems, the argument is strengthened, *i.e.* the boundary points of the convex hull correspond to Pareto optimal points,

$$P_{\text{floor}}(\mathcal{X}) \cup P_{\text{roof}}(\mathcal{X}) \subseteq \text{conv}(\mathcal{Y}) \quad , \quad (\text{B2})$$

where  $\mathcal{X}$  is the set of all possible solutions to the optimization problem taking  $\mathcal{Y} = f(\mathcal{X})$ . For example, given  $x_1, x_2 \in \mathcal{X}$ , possible solutions to the optimization problem, we say that  $x_1$  Pareto dominates  $x_2$  if and only if  $f(x_2) \succ f(x_1)$ , denoted  $x_2 \succ_f x_1$ . The symbol  $\preceq$  is used to denote a partial order in  $\mathbb{R}^k$ , defined componentwise,

$$\underline{y}' \preceq \underline{y} \iff y'_i \leq y_i, \quad \forall i = 1, \dots, k \quad (\text{B3})$$

which is the definition of the Pareto front given in Eq. (7).

Of relevance for the extension of our algorithm to higher dimensions is the following: the property that all solutions are bounded by both fronts, *i.e.*

$\mathcal{X}$  is *bounded in the Pareto sense* by  $P_{\text{floor}}(\mathcal{X}) \cup P_{\text{roof}}(\mathcal{X})$  if there exist  $y^{\min}, y^{\max} \in P_{\text{floor}}(\mathcal{X}) \cup P_{\text{roof}}(\mathcal{X})$  such that  $y^{\min} \preceq y \preceq y^{\max}$  for all  $y \in \mathcal{X}$  [30].

That is, every element of  $\mathcal{X}$  lies between its Pareto floor and roof limits with respect to the Pareto order,  $\preceq$ .

## Appendix C: The case of the gluon PDF of the pion

The Fantômas PDF for the pion admits solutions for the gluon that include cases with a nearly zero contribution. Motivated by DrellYan data [7], such solution yields phenomenologically interesting gluon momentum fractions, that are somewhat in tension with lattice QCD predictions.

Because its integral vanishes, this solution cannot be interpreted as a probability density. Its Rényi entropy therefore arises purely from fluctuations in the gluon contribution. However, since Rényi entropy, like many information criteria, characterizes only the shape of the distribution, not its magnitude, we expect the zero-gluon solution to be close to that of a flat distribution. Following Eq. (6), we obtain  $H_2^{x f(x)=1}(X) = \ln n$ . The zero-gluon distribution gives  $H_2^{x g(x)=0}(X) \lesssim \ln n$ , which is sufficient to justify keeping this solution for our analysis.

On the other hand, the pion's gluon PDF is mainly determined by low- $x$  data from tagged deep-inelastic scattering data from a virtual pion target and, through the DGLAP equations, by the better-constrained valence sector at very large- $x$ . This leaves an extrapolation region around  $x \sim 0.1$ , where more atypical forms can

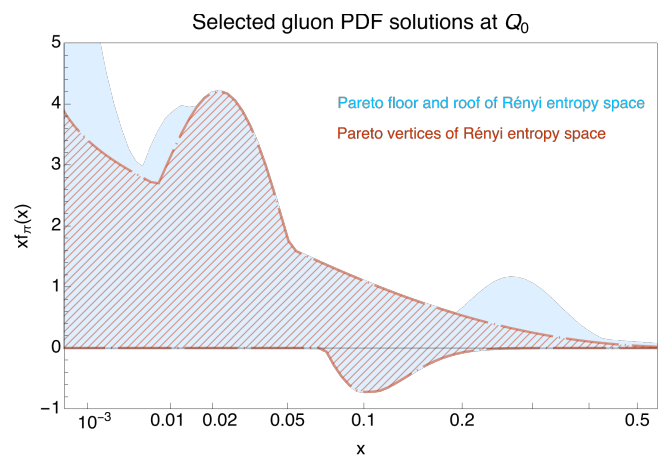


FIG. 6: Comparison of spanned solutions for the Pareto roof+floor vs. only vertices solutions for the gluon PDF of the pion.

emerge. The bumpiest of the admitted solutions appear in this extrapolation region, although some are shifted in  $x$ . Because Rényi entropy is insensitive to the  $x$ -position of the bump, the entropy of the function does not vary with the sliding of its maximum. Consequently, the peculiar shapes of the gluon PDF of the pion do not fully satisfy the simpler rule that the Pareto vertices should encompass all shapes selected from both Pareto fronts. This implies that, in this particular case, taking the vertices of the Pareto fronts is not sufficient, *i.e.* some of the steep solutions will be missed, as illustrated in Fig. 6. This issue is not expected to arise in the proton case, for which the extrapolation regions lie toward the end points.

- 
- [1] H.-W. Lin et al., Prog. Part. Nucl. Phys. **100**, 107 (2018), 1711.07916.
  - [2] M. Constantinou et al., Prog. Part. Nucl. Phys. **121**, 103908 (2021), 2006.08636.
  - [3] T.-J. Hou et al., Phys. Rev. D **103**, 014013 (2021), 1912.10053.
  - [4] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne, Eur. Phys. J. C **81**, 341 (2021), 2012.04684.
  - [5] R. D. Ball et al. (NNPDF), Eur. Phys. J. C **82**, 428 (2022), 2109.02653.
  - [6] A. Courtoy, J. Huston, P. Nadolsky, K. Xie, M. Yan, and C. P. Yuan, Phys. Rev. D **107**, 034008 (2023), 2205.10444.
  - [7] L. Kotz, A. Courtoy, P. Nadolsky, F. Olness, and M. Ponce-Chavez, Phys. Rev. D **109**, 074027 (2024), 2311.08447.
  - [8] L. Kotz, A. Courtoy, P. Nadolsky, and M. Ponce-Chavez, Phys. Rev. D **112**, L071502 (2025), 2505.13594.
  - [9] L. Kotz, A. Courtoy, T. J. Hobbs, P. Nadolsky, F. Olness, M. Ponce-Chavez, and V. Purohit (2025), 2507.22969.
  - [10] L. A. Harland-Lang, T. Cridge, and R. S. Thorne, Eur. Phys. J. C **85**, 316 (2025), 2407.07944.
  - [11] M. N. Costantini, M. Madigan, L. Mantani, and J. M. Moore, JHEP **12**, 064 (2024), 2404.10056.
  - [12] J. Gao and P. Nadolsky, JHEP **07**, 035 (2014), 1401.0013.
  - [13] I. Novikov et al., Phys. Rev. D **102**, 014040 (2020), 2002.02902.
  - [14] K. Kovarik, P. M. Nadolsky, and D. E. Soper, Rev. Mod. Phys. **92**, 045003 (2020), 1905.06957.
  - [15] R. Anwar, M. Hamilton, and P. M. Nadolsky (2019), 1901.05511.
  - [16] S. Carrazza, J. I. Latorre, J. Rojo, and G. Watt, Eur. Phys. J. C **75**, 474 (2015), 1504.06469.
  - [17] S.-M. Hu, A.-S. Xiong, J. Xu, F.-S. Yu, and J.-X. Yu (2025), 2508.05171.
  - [18] A. Ablat et al., Eur. Phys. J. Plus **139**, 1063 (2024), 2408.04020.
  - [19] F. Santambrogio, *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*, vol. 87 of *Progress in Nonlinear Differential Equations and Their Applications* (Birkhuser Cham, 2015), 1st ed., ISBN 978-3-319-20828-2, URL <https://link.springer.com/book/10.1007/978-3-319-20828-2>.
  - [20] C. Villani, *Optimal Transport: Old and New*, vol. 338 of *Grundlehren der mathematischen Wissenschaften* (Springer Berlin, Heidelberg, 2009), 1st ed., ISBN 978-3-540-71050-9, URL <https://link.springer.com/book/10.1007/978-3-540-71050-9>.
  - [21] P. T. Komiske, E. M. Metodiev, and J. Thaler, Phys. Rev. Lett. **123**, 041801 (2019), 1902.02346.
  - [22] P. T. Komiske, E. M. Metodiev, and J. Thaler, JHEP **07**, 006 (2020), 2004.04159.
  - [23] D. Ba, A. S. Dogra, R. Gambhir, A. Tasissa, and J. Thaler, JHEP **06**, 195 (2023), 2302.12266.
  - [24] R. D. Ball, J. Cruz-Martinez, L. Del Debbio, S. Forte, Z. Kassabov, E. R. Nocera, J. Rojo, R. Stegeman, and M. Ubiali (NNPDF) (2022), 2211.12961.
  - [25] R. Gambhir, M. LeBlanc, and Y. Zhou, in *39th Annual Conference on Neural Information Processing Systems: Includes Machine Learning and the Physical Sciences (ML4PS)* (2025), 2509.19431.
  - [26] A. Garcia Caffaro, I. Mout, and C. Shimmin (2025), 2510.06314.
  - [27] C.-T. group (Joey Huston), *On the way to ct25*, Slides presented at PDF4LHC+PHYSTAT 2025 (2025), URL [https://indico.cern.ch/event/1553776/contributions/6647087/attachments/3149496/5592474/CT\\_update\\_PDF4LHC\\_oct7\\_25.pdf](https://indico.cern.ch/event/1553776/contributions/6647087/attachments/3149496/5592474/CT_update_PDF4LHC_oct7_25.pdf).
  - [28] R. D. Ball et al. (PDF4LHC Working Group), J. Phys. G **49**, 080501 (2022), 2203.05506.
  - [29] M. T. M. Emmerich and A. H. Deutz, Natural Computing **17**, 585 (2018).
  - [30] M. Ehrgott, *Multicriteria Optimization*, Lecture Notes in Economics and Mathematical Systems, Vol. 491 (Springer Berlin Heidelberg, 2005), ISBN 978-3-540-21398-7.