

# Unraveling generalized parton distributions through Lorentz symmetry and partial DGLAP knowledge

P. Dall’Olio<sup>1</sup>, F. De Soto<sup>2</sup>, C. Mezrag<sup>3</sup>, J. M. Morgado Chávez<sup>3,\*</sup>, H. Moutarde<sup>3</sup>,  
J. Rodríguez-Quintero<sup>1,3</sup>, P. Sznajder<sup>4</sup> and J. Segovia<sup>2</sup>

<sup>1</sup>*Departamento Ciencias Integradas, Centro de Estudios Avanzados en Física, Matemáticas y Computación, Facultad Ciencias Experimentales, Universidad de Huelva, Huelva 21071, Spain*

<sup>2</sup>*Departamento Sistemas Físicos, Químicos y Naturales, Universidad Pablo de Olavide, 41013 Sevilla, Spain*

<sup>3</sup>*Irfu, CEA, Université Paris-Saclay, 91191 Gif-sur-Yvette, France*

<sup>4</sup>*National Centre for Nuclear Research (NCBJ), 02-093 Warsaw, Poland*



(Received 23 January 2024; accepted 9 April 2024; published 10 May 2024)

Relying on the polynomiality property of generalized parton distributions, which roots on Lorentz covariance, we prove that it is enough to know them at vanishing and low skewness within the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi region to obtain a unique extension to their entire support up to a  $D$  term. We put this idea in practice using two methods: reconstruction using artificial neural networks and finite-elements methods. We benchmark our results against standard models for generalized parton distributions. In agreement with the formal expectation, we obtain very a accurate reconstruction for a maximal value of the skewness as low as 20% of the longitudinal momentum fraction. This result might be relevant for reconstruction of generalized parton distribution from experimental and lattice QCD data, where computations are, for now, restricted in skewness.

DOI: [10.1103/PhysRevD.109.096013](https://doi.org/10.1103/PhysRevD.109.096013)

## I. INTRODUCTION

Introduced in the 1990s [1–3], generalized parton distributions (GPDs) are today at the core of hadron structure studies, both experimentally and theoretically. This enthusiasm can be understood, as GPDs encode the multidimensional structure of hadrons [4,5] (2D in the transverse plane and 1D in longitudinal momentum space). They are also connected to the energy-momentum tensor, allowing in principle to extract the contributions of quarks and gluons to the total angular momentum [6], and also the contributions to the pressure and shear forces within the hadrons [7].

Therefore, since the early 2000s, attempts to extract GPDs from experimental data have been performed (see the examples of [8,9]). However, the task remains hard for multiple reasons. First, GPDs are connected to experimental data of deep exclusive processes, like deep virtual Compton scattering (DVCS), which are much harder to measure than inclusive ones, connected to regular parton distribution functions (PDFs). Further, the connection between experimental data and GPDs in standard deep exclusive processes like DVCS, reveals itself noninvertible

at fixed scale and mathematically ill posed when evolution is turned on [10,11]. The situation might be better for processes involving the production of an additional particle in the final state (see Refs. [12–16]).

The difficulty in extracting GPDs comes also from the fact that they have to obey a number of theoretical constraints whose fulfillment is *a priori* not granted using generic modeling techniques. Among those, let us mention polynomiality [17,18] and positivity [19–22]. A way to fulfill systematically both at the same time was introduced in [23] based on the so-called covariant extension [24,25], and yielded the first evaluation of experimental feasibility of pion DVCS [26]. The covariant extension technique was originally developed to recover the Efremov-Radyushkin-Brodsky-Lepage (ERBL) kinematic region from the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) one, in a unique way, up to a  $D$ -term.<sup>1</sup> However, as we will show in this paper, its application is in fact more general and allows one to uniquely continue a GPD known solely in the vanishing and low- $\xi$  region to the entire kinematic domain. This application is important as it allows reconstruction for GPDs based on collider data (see for instance [27]) supplemented<sup>2</sup> with lattice QCD data. Moreover, current lattice QCD calculations are able to extract GPDs. However,

\*Corresponding author: jose-manuel.morgadochavez@cea.fr

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI. Funded by SCOAP<sup>3</sup>.

<sup>1</sup>The details about the ambiguities generated by the  $D$  term can be found in Ref. [24] and Sec. II.

<sup>2</sup>The reader can refer to [28] for the impact of combining experimental and lattice data.

if the entire  $x$  range of GPDs is accessible in a single lattice QCD calculation, separated evaluations become necessary when targeting changing values of  $\xi$  or  $t$  (see for instance [29–32]). Given the cost of such simulations, it would be desirable to find a way of relying on a minimal knowledge of GPDs in their variable  $\xi$ , being able to extend them to their entire kinematic domain. We find our formalism, capable of extending GPDs from a limited region in  $\xi$  to the entire support, to have the potential to become useful in this context.

In Sec. II, we start by recalling the mathematics behind the covariant extension, highlighting that both an incomplete  $x$  and  $\xi$  domain are allowed. In Sec. III, we check the actual feasibility using two different numerical techniques, artificial neural networks (ANNs) and finite-elements method (FEM), on two models. The first one is derived from light-front wave function computations [25] and the second one is the phenomenological parametrization by Goloskokov and Kroll [33]. Finally in Sec. IV, we conclude.

## II. POLYNOMIALITY AND INVERSION OF THE RADON TRANSFORM

In this work we show that the sole knowledge of GPDs over part of their DGLAP region is enough for their determination over their entire support. Importantly, this result is completely general and does not show any relation with the structure of the particular GPDs or the underlying physics. On the contrary, the only two elements needed for its derivation are: (1) the representation of GPDs through the Radon transform of double distributions; and (2) the support theorem [34] for the Radon transform. The former claim directly follows from Lorentz symmetry. The later is only related to the formal definition of the Radon transform. Thus, in the following we introduce the necessary ingredients that allow us to combine these two elements to end up showing that, indeed, a partial knowledge of a GPD within its support constrains it all.

### A. GPD definition and polynomiality

GPDs are defined as a light-front projection of a nondiagonal hadronic matrix element of a bilocal operator [1–3], and they are usually expressed in terms of three kinematic variables; namely, the light-front momentum fraction,  $x$ , the skewness or light-front momentum fraction transfer,  $\xi$ , and the squared momentum transfer,  $t$ . Concerning  $x$  and  $\xi$ , the kinematic domain of GPD defines its support with  $x, \xi \in [-1, 1]$ ; the region  $|x| \geq |\xi|$  ( $|x| \leq |\xi|$ ) being denominated DGLAP (ERBL) where the GPD can be expanded in Fock space only involving states with the same number  $N$  (different number  $N$  and  $N + 2$ ) of particles [17].

This being given, it can be proven for any  $m$ -order GPD's Mellin moment that Lorentz symmetry implies [17,18,35]

$$\int_{-1}^1 dx x^m H(x, \xi, t) = \sum_{k=0}^{m+1} c_k^{(m)}(t) \xi^k, \quad (2.1)$$

where  $H(x, \xi, t)$  denotes the unpolarized quark GPD, which we use here as an illustration.<sup>3</sup> This property is widely dubbed *polynomiality* and has both a deep origin and important implications. Particularly, if the GPD is known for all  $x$  and  $t$  in a given compact range of  $\xi$ , one can calculate the lhs of Eq. (2.1) in there and, *a priori*, determine the polynomial coefficients in the rhs for any  $m$ -order Mellin moment. It becomes thus clear that, rooting on Lorentz covariance and on the uniqueness of Mellin moments, the GPD appears defined in its entire support from its knowledge within a restricted kinematic domain. Moreover, the GPD reconstruction can be supplementarily constrained by profiting the highly nontrivial interplay between the structure of the DGLAP and ERBL regions expressed by the polynomiality property in Eq. (2.1). In the following, we shall elaborate further on this, and exploit it, capitalizing especially on the related mathematical literature devoted to the incomplete data problem in *computerized tomography* [37].

To the purpose of this work, ultimately capitalizing on the result from Eq. (2.1), the momentum transfer  $t$  can be fixed and taken to be a constant parameter all throughout the analysis; any conclusion would then be irrespective of its particular value. Thus, for the sake of simplicity, it will be considered implicit from now on and omitted in the notation. Furthermore, time reversal symmetry [36,38] can be also invoked and seen to impose that any sensible GPD is an even function of  $\xi$ , thus entailing that no odd power is allowed in Eq. (2.1):  $c_{2n+1}^{(m)} = 0$  for any integer  $n$ ; and, particularly,  $c_{n+1}^{(n)} = 0$  for any even integer  $n$ . In the following we shall concentrate on the  $\xi \geq 0$  region.

### B. The Radon transform representation of GPDs

Then, as it has been carefully discussed in Ref. [24] (see also references therein), given an odd function  $D(z)$  with support  $z \in [-1, 1]$ , defined by its Mellin moments  $\int_{-1}^1 dz z^m D(z) = c_{m+1}^{(m)}$ , one can prove that any  $m$ -order Mellin moment of  $H(x, \xi) - \text{sgn}(\xi)D(x/\xi)$  is a polynomial of degree  $m$  on  $\xi$ ; otherwise said, it fulfills the Ludwig-Helgason consistency condition entailing that it is in the range of a Radon transform  $\mathcal{R}$  [39],

$$\begin{aligned} H(x, \xi) - \text{sgn}(\xi)D\left(\frac{x}{\xi}\right) &= \mathcal{R}F(x, \xi) \\ &= \int_{\Omega} d\beta d\alpha \delta(x - \beta - \alpha\xi) F(\beta, \alpha), \end{aligned} \quad (2.2)$$

<sup>3</sup>In the case of polarized or transversity GPDs, the polynomiality property as expressed in Eq. (2.1) must be modified accordingly [36].

where  $\Omega = \{(\beta, \alpha) \in \mathbb{R}^2 / |\alpha| + |\beta| \leq 1\}$  is the support for the distribution  $F$  whose Radon transform  $\mathcal{R}F$  is in the physical domain of  $(x, \xi)$ .  $D$  is an odd function in its single argument, as it clearly comes from its definition in terms of Mellin moments which are zero for even order. Strictly speaking, the Ludwig-Helgason condition and the Radon transform are expressed<sup>4</sup> in terms of the distance  $s$  and the polar angle  $\phi$  (e.g., see Ref. [40]), which can be mapped into the usual GPD variables  $x$  and  $\xi$  through the transformations  $x = s / \cos \phi$  and  $\xi = \tan \phi$ , such that the  $m$ -order Mellin moment of  $[H - \text{sign}(\xi)D] / \cos \phi$  is an homogeneous polynomial of degree  $m$  with terms  $\cos^{m-k} \phi \sin^k \phi$ . This change from mathematical to physical variables has no consequence for our purposes.

One can then introduce a second distribution  $G(\beta, \alpha) = \delta(\beta)D(\alpha)$  such that  $\mathcal{R}G(\beta, \alpha) = D(x/\xi)/|\xi|$ . Consequently, the GPD  $H$  is given as  $\mathcal{R}F(x, \xi) + \xi \mathcal{R}G(x, \xi) = H(x, \xi)$ , with  $F$  ( $G$ ) being even (odd) in  $\alpha$ . Then, owing to its fulfilling condition (2.1), any sensible GPD can be represented as the Radon transform of two two-dimensional distributions  $F$  and  $G$ , widely named double distributions (DDs) [23,24,41,42]. However, as exposed in Refs. [41,42], the pair of DDs ( $F$ ,  $G$ ) does not offer a unique representation for the GPD  $H(x, \xi)$ . Instead, there exists a family of transformations, usually called *scheme* transformations, which make it possible to redefine  $F$  and  $G$  for the same GPD. A judicious choice of the transformation (see, for instance, Appendix B of Ref. [24]) allows for the two DDs to rely on one single DD  $h$ ,

$$\begin{aligned} H(x, \xi) &= (1-x)\mathcal{R}h(x, \xi) \\ &= (1-x) \int_{\Omega} d\beta d\alpha \delta(x - \beta - \alpha\xi) h(\beta, \alpha), \end{aligned} \quad (2.3)$$

which corresponds to the so-called  $P$  scheme [43]. At this point, it is worthwhile highlighting that, if  $(1-x)\mathcal{R}h(x, \xi)$  fulfills the polynomiality condition (2.1),

$$\tilde{H}(x, \xi) = (1-x)\mathcal{R}h(x, \xi) + \text{sign}(\xi)D^-(x/\xi) \quad (2.4)$$

also does, where  $D^-(z)$  is any odd function in its single argument, with support  $z \in [-1, 1]$ . The latter is important for our purposes of facing the inverse problem, as  $H$  and  $\tilde{H}$  only differ by the contribution from the line  $\beta = 0$  within the DD support, which impacts only on the GPD kinematic domain  $|x| < |\xi|$ .

Importantly, sticking to a particular DD scheme does not spoil the general character of the construction: One can always rely on the known set of DD-scheme transformations to express the GPD in any other picture. Capitalizing on this freedom, in the following, we will argue on top of DD schemes involving one single DD,  $h$ .

<sup>4</sup>They admit a direct geometric interpretation representing the central case to which important mathematical literature has been thus far devoted [37].

### C. The inverse Radon transform problem

Let us at this point consider a GPD for which its knowledge is restricted to a given kinematic domain. Then, rooting on a proof of uniqueness for a DD whose Radon transform reproduces the GPD on such a restricted domain, the inversion of the Radon transform can be featured as a sensible procedure for a covariant extension from restricted to full GPD kinematic ranges. So this has been suggested and illustrated in Refs. [24,25,40], where GPDs known within their DGLAP kinematic domain were successfully extended to the ERBL region through the Radon transform of the DD previously obtained by inversion.

More generally, the problem of inverting the Radon transform has been exhaustively studied in the mathematical context of computerized tomography [37]; for which, in most realistic scenarios, the Radon transform is known only for hyperplanes belonging to a particular subset of the whole domain.<sup>5</sup> Although explicit inversion formulas remain mostly not available, several theorems prove the uniqueness of the inverse function, depending on the properties of the functions involved and given the data of the Radon transform corresponding to specific subsets of hyperplanes. This implies that if a particular function is found, whose Radon transform reproduces the data known for the subset of hyperplanes for which the theorems apply, then that function is guaranteed to be unique. The relevant theorem on which the GPD covariant extension relies is the support theorem by Boman and Todd-Quinto [34] that, for the sake of completeness, will be reproduced below:

*Theorem.* Let us consider the Radon transform of a generalized function  $f \in \mathcal{E}'(\mathbb{R}^n)$  along the hyperplane  $z \cdot \theta = s$  ( $z \in \mathbb{R}^n$ ,  $\theta \in S^{n-1}$ , and  $s \in \mathbb{R}$ )

$$\mathcal{R}f(\theta, s) = \int_{z \cdot \theta = s} dz f(z). \quad (2.5)$$

Let  $W$  be an open unbounded connected subset of  $S^{n-1} \times \mathbb{R}$ , such that  $\mathcal{R}f(\theta, s) = 0$  for  $(\theta, s) \in W$ . Then  $f = 0$  on  $\cup \{z \cdot \theta = s | (\theta, s) \in W\}$ .

This theorem applies to  $f \in \mathcal{E}'(\mathbb{R}^n)$ , i.e., to distributions with compact support. It can therefore be applied to the Radon transform of a DD, whose domain is the compact skewed square  $\Omega$ . For this two-dimensional case, the argument of the function corresponds to the DD variables, i.e.,  $z \equiv (\beta, \alpha)$  and  $\theta \in S^1 \equiv (\cos \phi, \sin \phi)$ , which

<sup>5</sup>In computerized tomography, this is the case when the particular geometry of the probing beam only allows to scan any two-dimensional section of the problem object along a particular set of lines. This is also the case in the context of the GPD covariant extension, where the original GPD data, expressed as the Radon transform of a DD, are only known for those lines belonging to the DGLAP region.

correspond to  $(\theta, s)$ , the Radon transform variables<sup>6</sup> usually employed in mathematical literature and which convert to the standard GPD variables  $(x, \xi)$  as it was made explicit above. In this way, the line  $z \cdot \theta = s$  corresponds to  $x - \beta - \alpha\xi = 0$ .

The above theorem guarantees that the knowledge of a GPD on certain subsets of its domain is enough to uniquely determine the double distribution whose Radon transform generates that GPD. What follows below aims at making transparent this implication, exhibiting that it guarantees the uniqueness of the distribution  $F(\beta, \alpha)$  from Eq. (2.2), if  $\mathcal{R}F(x, \xi)$  is fixed by the knowledge of  $H(x, \xi)$  on a subset of the DGLAP region.

Notice first that the lhs of Eq. (2.2) also includes a term  $D(x/\xi)$  which, anyhow, has to be evaluated outside its support for any DGLAP kinematic configuration and, hence, vanishes. However, this  $D$  term raises a well-known ambiguity as, despite the uniqueness of  $F(\beta, \alpha)$ , the GPD  $H(x, \xi)$  remains undetermined within the ERBL region by the impact of such a term which, as argued above, corresponds to a contribution with support only on the line  $\beta = 0$ . The same is true for DDs in any scheme and, particularly, is made apparent by  $P$ -scheme Eqs. (2.3) and (2.4), which exhibit how DGLAP data cannot unambiguously fix the ERBL GPD. On the other hand, since  $H(x, -\xi) = H(x, \xi)$ , we can consider  $\xi \geq 0$ , implying  $\phi \in [0, \frac{\pi}{2})$ ; and expose that the whole DGLAP region  $|x| > \xi$ , also expressed as  $|s| > \sin \phi$ , is not a connected set. However, this makes no limitation for the application of the theorem, because it can be separately applied for the two disjoint, open connected regions  $\{|x| > \xi\} \cap \{x > 0\}$  (particle contribution) and  $\{|x| > \xi\} \cap \{x < 0\}$  (antiparticle contribution).

Indeed, without lack of generalization, we can focus on the region  $x > 0$  and consider the following subsets (see the right plots of Fig. 1):

$$\mathcal{O}_\lambda = \{|x| > \xi\} \cap \{x > 0\} \cap \{0 \leq \xi < \lambda x\}, \quad (2.6)$$

where  $\lambda \in (0, 1]$ . One should realize that  $\{\beta > 0\} = \cup \{z \cdot \theta = s | (\theta, s) \in \mathcal{O}_\lambda\}$ . Namely, the set of lines on the  $(\beta, \alpha)$  plane corresponding to  $\mathcal{O}_\lambda$ , i.e., the lines  $\alpha = (x - \beta)/\xi$ , with  $\xi \in (0, \lambda x)$ , sweep the half plane  $\beta > 0$  irrespective of the value of  $\lambda$  (this is graphically illustrated by the right panel of Fig. 1). The support theorem therefore implies that if the GPD  $H(x, \xi)$  is uniquely given on the DGLAP subset  $\mathcal{O}_\lambda$ , the DD is uniquely fixed on the half plane  $\beta > 0$ . By the same argument, one can uniquely fix the DD on the  $\beta < 0$  half plane by knowing the GPD on the corresponding  $\mathcal{O}_\lambda$  subsets with  $x < 0$ .

<sup>6</sup>Notice that although  $s$  represents the distance between the hyperplane and the origin, the Radon transform is defined also for negative values of  $s$ , since it satisfies  $\mathcal{R}f(\theta, s) = \mathcal{R}f(-\theta, -s)$ .

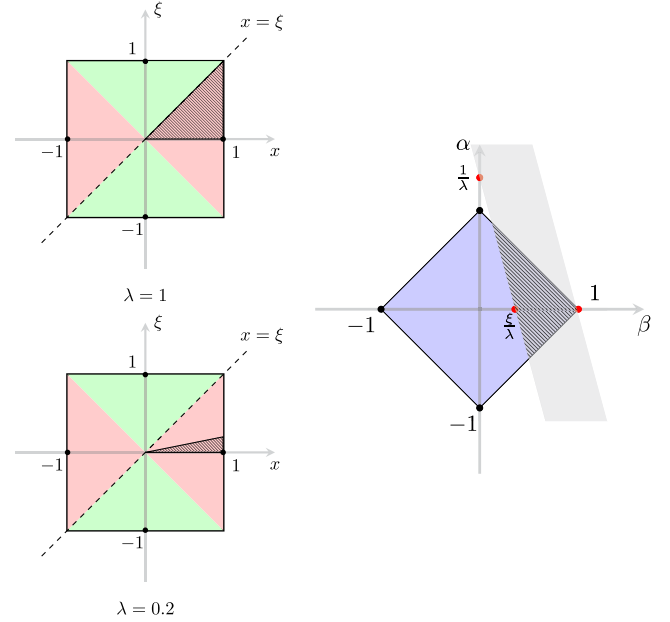


FIG. 1. Graphic illustration. Left panel: the square being a graphical representation of the domain of GPDs, green triangles correspond to the ERBL region and red areas to the DGLAP one. The hatched area represents, for different values of  $\lambda$  in (2.6), the portion of the overall GPD domain which is taken as input in the reconstruction procedure. Right panel: the skewed square being the domain over which DDs have support, the Radon transform Eq. (2.3) can be interpreted as an integration over straight lines,  $\alpha = (x - \beta)/\xi$ , with  $\alpha_0 = x/\xi$  and  $\beta_0 = x$  as  $\alpha$  and  $\beta$  intercept, respectively. The support theorem guarantees the uniqueness for the DD within the region swept by all the straight lines such that  $(x, \xi) \in \mathcal{O}_\lambda$ ; namely,  $\xi/\lambda < x \leq 1$  with  $0 \leq \xi < \lambda$ . Then, for any fixed  $\xi$ , one is left with parallel lines of slope  $1/\xi$  bounded by  $1/\lambda < \alpha_0 \leq 1/\xi$  and, correspondingly,  $\xi/\lambda < \beta_0 \leq 1$ ; schematically displayed in the plot by the shaded area between the two parallels crossing the bounding intercepts (red circles). Therefore, when varying  $\xi$  from 0 to  $\lambda$ , the entire DD domain ( $\beta > 0$ ) is covered.

Accordingly, for any  $\lambda \in (0, 1]$ , the DGLAP region  $x > 0$  fixes uniquely the DD for all  $\beta > 0$  and, separately, the DGLAP region  $x < 0$  does the same for all  $\beta < 0$ . Furthermore, one can simply consider a positive  $\alpha$  in both cases, owing to the even parity of the DDs  $F$  (or  $h$  in this scheme). On the other hand, as discussed above, the axis  $\beta = 0$  remains undetermined by the single implementation of DGLAP data. The choice of  $\lambda = 1$  entails that the GPD is uniquely determined, with the exception of the Radon transform of terms with support only on  $\beta = 0$ , by its whole DGLAP knowledge.<sup>7</sup> More importantly and not thus far highlighted, is that the same is formally true for  $0 < \lambda < 1$ . Namely, the GPD

<sup>7</sup>The covariant extension from DGLAP to ERBL kinematic ranges introduced and discussed in Refs. [24,25,40,44] relies on this remarkable output.

covariant extension can be made from its knowledge on proper subsets of the DGLAP region where the skewness parameter is restricted to an arbitrarily small range, i.e.,  $\xi \in [0, \lambda x]$ , with a positive and arbitrarily small value of  $\lambda$ . Beyond its formal interest by itself, this is very relevant for physical purposes, since in actual experimental setups one usually has access only to small values of  $\xi$ ; and can be also very helpful for the completion and GPD reconstruction from lattice data.

In summary, we have shown that the polynomiality property of GPDs is guaranteed *if and only if* a Radon transform representation of GPDs is possible. From that point on, relying on the Boman and Tod-Quinto theorem, we showed that the sole knowledge of GPDs within certain subsets of their DGLAP regions is enough to fix the corresponding double distributions. Given this general statement, a limited knowledge of DGLAP GPDs always allows to find the corresponding double distribution through the inversion of the Radon transform. The latter finally allows to construct the remnant of the GPD kinematic domain. In the following sections we will numerically examine the soundness of this conclusion by trying to invert the GPD Radon transform having as inputs its data on these restricted subsets  $\mathcal{O}_\lambda$  and subsequently reconstructing the GPD on its whole domain.

### III. PROOF OF CONCEPT AND FEASIBILITY

If a kinematic completion of GPDs is formally possible, a proof of concept and feasibility yet remains to be studied. We therefore assess this problem by following two approaches: (1) applying the FEM discussed and tested in Refs. [23,24,40,45], Sec. III A; and (2) using ANNs for the extraction of DDs, Sec. III B. The main purpose for this double implementation, with two very different methods to invert the Radon transform, stems from our aim of discarding systematic effects related to a particular inversion procedure. To this goal, we shall test the results obtained with both methods against two benchmarking models for GPDs<sup>8</sup> (Table I): The *algebraic model* for the pion GPD first constructed in [25] and the renowned *Goloskokov-Kroll model* [33,46–48]. A thorough inter-comparison of the two approaches, as they apply to our problem, is out of the scope of the present work.

#### A. Finite-elements method (FEM)

A suitable numerical approach to the Radon transform inverse problem is to proceed through discretization and interpolation in the space of DDs, which in the following will be referred to as FEM strategy. This methodology was first presented in [24] and further elaborated in [23,45]

<sup>8</sup>For further details about these two models see Appendixes A 1 and A 2.

TABLE I. Syllabus—Summary of the models and strategies explored in this study.

Method	Model	
	Algebraic	Goloskokov-Kroll
FEM	Sec. III A 1	Sec. III A 2
ANN	Sec. III B 1	Sec. III B 2

where its usefulness in the kinematic completion of GPDs was exposed, setting the ground for a pioneering exploratory study of DVCS on pions [23]. As an outcome of these studies, a C++ implementation of this tool is already available within the PARTONS framework [49]. However, this work employs a slight extension of that version. We thus find it advantageous to briefly expose in the following the main ideas behind this program. The interested reader can find detailed discussions in [24,26,40,45].

The starting point of the FEM approach is to introduce a discretization of the DD domain,  $\Omega = \cup_e \Omega_e$ , and embed in it a set of interpolating polynomials such that the DD may be approximated within the discretized domain. The accuracy of this approximation is essentially determined by the choice of the interpolating functions as well as the mesh itself. Inspired by the method of finite elements widespread in, e.g., the study of partial differential equations, we choose to perform piecewise polynomial interpolation,<sup>9</sup> allowing to approximate the DD as

$$h_{\text{FEM}}(\beta, \alpha) = \sum_e P_e(\beta, \alpha), \quad (3.1)$$

where  $P_e(\beta, \alpha)$  are the polynomial interpolants chosen for each element  $\Omega_e$ , which can in turn be expressed as

$$P_e(\beta, \alpha) = \begin{cases} \sum_{k \in \Omega_e} v_k(\beta, \alpha) h_k, & \text{for } (\beta, \alpha) \in \Omega_e, \\ 0, & \text{otherwise,} \end{cases} \quad (3.2)$$

where  $k$  labels the items of a set of interpolation nodes  $n_j \equiv (\beta_j, \alpha_j)$  appropriately distributed over  $\Omega_e$ ; and where  $v_j(\beta, \alpha)$  represents the Lagrange interpolating polynomials associated to the node  $n_j$ . In this work we take them to be degree-two polynomials.<sup>10</sup> Finally,  $h_j$  is the value of the DD at the interpolation node  $n_j$ , so we may write the integral problem of Eq. (2.3) as

<sup>9</sup>See Appendix C in [45] and references therein for a detailed description of piecewise polynomial interpolation in two dimensions applied in this context.

<sup>10</sup>This feature constitutes the main upgrade of our implementation with respect to that of Refs. [23,24,40,45], where degree zero and one interpolating polynomials were employed.

$$\begin{aligned}
H(x, \xi) &= \sum_j \left[ \int_{\Omega_c} d\beta d\alpha \delta(x - \beta - \alpha\xi) v_j(\beta, \alpha) \right] h_j \\
&= \sum_j \mathcal{R} v_j(x, \xi) h_j.
\end{aligned} \tag{3.3}$$

Importantly, because the interpolating basis functions  $v_j$  are simple second order polynomials, the integrals can be evaluated in closed form, turning the original problem into a set of algebraic equations connecting the DD and its associated GPD. Furthermore, choosing pairs  $(x_i, \xi_i)$ , a system of algebraic equations can be built to represent the Radon transform problem:

$$\begin{aligned}
\begin{pmatrix} H(x_1, \xi_1) \\ \vdots \\ H(x_N, \xi_N) \end{pmatrix} &= \begin{pmatrix} \mathcal{R} v_1(x_1, \xi_1) & \cdots & \mathcal{R} v_n(x_1, \xi_1) \\ \vdots & \ddots & \vdots \\ \mathcal{R} v_1(x_N, \xi_N) & \cdots & \mathcal{R} v_n(x_N, \xi_N) \end{pmatrix} \\
&\times \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}.
\end{aligned} \tag{3.4}$$

In particular, one may choose the sampling points  $(x_i, \xi_i)$  within the DGLAP region, where the GPD is assumed to be known; even better, on a subset of that domain such as the one considered in Eq. (2.6). When doing so, a system of algebraic equations where the only unknowns are the values of the DD at the interpolation nodes arises. As explained in [23] its solution can be found using a least squares minimization,

$$\mathbf{h}_{\text{FEM}} = \left( \mathcal{R}^T \mathcal{R} \right)^{-1} \mathcal{R}^T \mathbf{H}|_{\text{DGLAP}}, \tag{3.5}$$

where boldface letters represent the vectors (matrices) of Eq. (3.4) in compact form. Thereupon, one may finally reconstruct the DD according to Eq. (3.1) [or Eq. (3.2)] and employ it to evaluate the GPD within the ERBL region through a Radon transform (in its discretized version):

$$\mathbf{H}|_{\text{ERBL}} = \tilde{\mathcal{R}} \mathbf{h}_{\text{FEM}} = \tilde{\mathcal{R}} \left( \mathcal{R}^T \mathcal{R} \right)^{-1} \mathcal{R}^T \mathbf{H}|_{\text{DGLAP}}, \tag{3.6}$$

where  $\tilde{\mathcal{R}}$  is a matrix representation of the Radon transform integral operator built following the procedure described before and evaluated for  $(x_i, \xi_i)$  chosen within the ERBL region.

In this way, a FEM-inspired strategy allows to find a solution for the inverse Radon transform problem. This approach being grounded on the theoretical construction of Sec. II, the covariant extension of GPDs from the DGLAP to the ERBL region turns feasible. Nonetheless,

even if the input DGLAP GPD is known to arbitrary precision,<sup>11</sup> the procedure is affected by uncertainties. Provided that the solution to our problem exists (Sec. II), the construction of this section reveals (1) the choice of interpolating polynomials, (2) the mesh definition (in particular its size), and (3) the number/distribution of sampling lines  $(x_i, \xi_i)$  in Eq. (3.4) as the main sources of possible deviation from the actual solution. In practice, we have tested three types of polynomial interpolants: zero, first, and second order polynomials. Optimization in the sense of trading performance of the implementation for accuracy of the solution lead us to use degree-two polynomials for all calculations. With regard to the mesh, a *Delaunay* triangulation was built over the DD domain [23,24], the average size of the elements being fixed to a model-specific optimal value found as to balance precision (expected for finer meshes) and stability (precluded in the same direction) [50]. Coincidentally, the meshes revealing optimal for the study of the algebraic and the Goloskokov-Kroll (GK) models were found to be identical and made up from 17 elements of average area 0.03 (arbitrary units). Finally, the number  $N$  of sampling lines was set according to the criterion of [23] where it was argued that the actual answer reveals in the least squares solution [Eq. (3.5)] when  $N_{\text{sample}}$  is taken to be at least 4 times the number of elements in the mesh. Here we took  $N_{\text{sample}} = 6n$ . These three sets of parameters being thus fixed, the solution to the inverse Radon transform problem can be found. From that point on there remains one single parameter whose effect on the solution has to be assessed: the distribution of sampling lines within the (restricted) DGLAP region. Here we follow once again the prescription of [23] and decide to draw them randomly following a uniform distribution  $x, \xi/x \in [0, 1]$ . Therefore, the distribution of lines leading to the solution of the inverse Radon transform varies among runs. In lack of a criterion to fix them, such constitutes the main source of statistical uncertainty in our calculation. Here we decide to estimate it through the production of replicas following a strategy similar to that

<sup>11</sup>Of course that situation is unlikely to be encountered in a realistic scenario. As an illustration, one can consider the case of study where the input DGLAP GPD is extracted from simulations on the lattice, thus each point being affected by the corresponding errors. In that case, one must properly account for error propagation from the DGLAP to the ERBL region as driven by Eq. (3.6). However, the extraction of GPDs from the lattice as well as their experimental assessment are still at an early stage; and more importantly the sources of this kind of errors have little to do with our approach. On the other hand, our numerical solution of an ill-posed problem such as the inverse Radon transform itself introduces systematic errors for which we may account without regard to the nature of the feeding distribution. They are precisely the sources and magnitudes of these last uncertainties which we assess in this work.

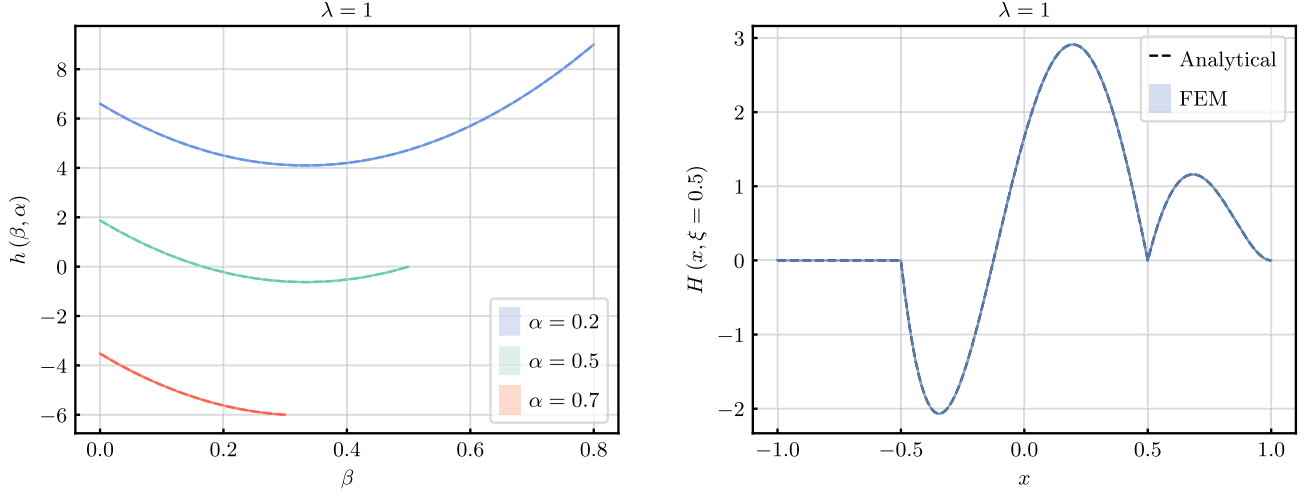


FIG. 2. (FEM) Algebraic model. The left panel shows the DD  $h_{\text{FEM}}(\beta, \alpha)$  obtained through the procedure described in Sec. III A (solid lines), displayed as a function of the  $\beta$  variable for three different values of  $\alpha = \{0.2, 0.5, 0.7\}$  ( $0 \leq \beta \leq 1 - \alpha$ ). For comparison, the exact analytical result of Eq. (A5) is drawn through dashed lines. The sampling of the DGLAP region corresponds to  $\xi \in [0, \lambda x]$  with  $\lambda = 1$  (the entire domain). The right plot displays the GPD  $H(x, \xi)$  at the illustration value  $\xi = 0.5$ , obtained as the Radon transform of the DD shown in the left, compared to the analytical expression, Eqs. (A4) and (A6).

in [51]: We generate 250 sets of sampling lines and solve our problem for each of them. Thus, one is given with the corresponding number of solutions to our problem, say DDs, called *replicas*. Then, for a given point, say the interpolation nodes, we check for outliers following a  $5\sigma$  rule [52]. Finally, the replicas identified as outliers in any of the points are removed from the uncertainty estimate, which is given by the standard deviation of the population of replicas. Accordingly, all our results (Figs. 2–5) are given as a solid line, representing the mean value; and the corresponding one-sigma band.

### 1. Algebraic model

Employing the general setup described in the previous paragraphs, the algebraic model can be extended from the DGLAP region, cf. Eq. (A4), to the ERBL domain. The case  $\xi = 0.5$  is used to illustrate our findings. Figures 2 and 3 show the results obtained as the solution to our problem using the region  $\mathcal{O}_\lambda$  as the input domain. Three values of  $\lambda$  have been explored, ranging for  $\lambda = 1$  (corresponding to the entire DGLAP region) down to  $\lambda = 0.5$  and 0.2. In all cases and for the sole sake of validation, we compare with the known shapes of the algebraic-model GPDs. The agreement between the actual GPD and the reconstructed form is astonishing, showing that indeed the FEM strategy allows to perform the covariant extension of GPDs from the DGLAP to the ERBL region with a minimal knowledge of the input GPD and no prior knowledge of the double distribution. These results will agree with those found in Sec. III B using the ANN implementation, supporting the thesis of this study.

If the accuracy of the solution to problem is remarkably good, we can still get a better grasp about it by looking at

the reconstructed DD  $h_{\text{FEM}}(\beta, \alpha)$  against the expectation, Eq. (A5): left panels of Figs. 2 and 3. Again, both are essentially indistinguishable. The reason for that resides in the structure of the DD generating the algebraic model: a second order polynomial in the kinematic variables, Eq. (A5); just as they are the interpolants  $P_\epsilon(\beta, \alpha)$  introduced within the FEM strategy. In that way, the exact solution is accidentally put by construction in the range of the discrete Radon transform of Eq. (3.4), meaning that in the absence of noise, the exact result can be found. In modern terms, this situation resembles that found in Bayesian reconstruction when a default model is introduced as prior information. In the language of Bayes theorem, such a default model represents the most probable answer in the absence of any data. When the chosen default model indeed represents the input data, the functional space is strongly dumped, and the chances of finding an accurate solution grow. Nonetheless, it is worth highlighting at this point that our FEM algorithm is not designed specifically to solve the inverse Radon transform problem in the case of the algebraic model. Accordingly, it is not that in order to achieve this solution the algorithm needed from prior knowledge about the polynomial structure of the DD; instead, we only understood the accuracy of the results by looking at the actual solution afterwards. Moreover, the kernel of the FEM algorithm employs a general-purpose interpolation routine which does not change between models.

### 2. Goloskokov-Kroll model

With a proof of concept for the usefulness of the FEM construction in solving the inverse Radon transform problem in a particularly favorable case, challenging our

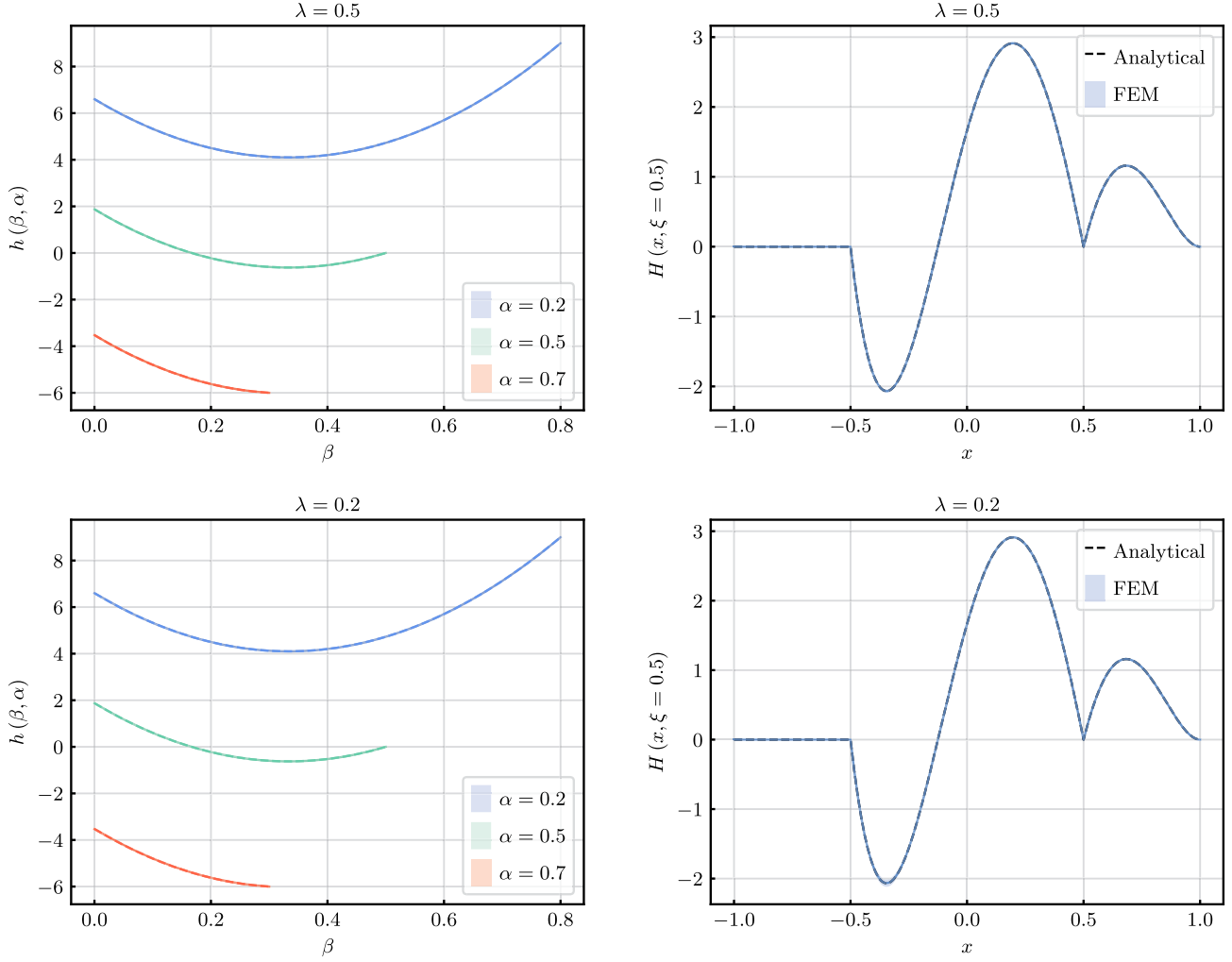


FIG. 3. (FEM) Algebraic model. The same as Fig. 2 with a DGLAP sampling given by  $\xi \in [0, \lambda x]$  and  $\lambda = 0.5$  (top) and  $0.2$  (bottom).

implementation in more complicated scenarios becomes crucial. The case of the GK model constitutes an ideal test ground for three reasons: first, as remarked in Sec. A 2, the GK model is widely employed in phenomenological analyses of GPDs. Second, the coincidence between the basis functions and the actual expression observed in the algebraic model is no longer found. Third, the DD giving rise to the GK model exhibits an integrable divergence at low values of  $\beta$ , challenging even more our implementation. It is precisely this later feature that has triggered a slight modification of the construction employed in the case of the algebraic model, redefining the approximating DD,  $h_{\text{FEM}}$  as

$$h_{\text{FEM}}(\beta, \alpha) = r(\beta) \sum_e P_e(\beta, \alpha), \quad r(\beta) \equiv \frac{1}{\sqrt{\beta}}, \quad (3.7)$$

so as to smooth the divergent behavior of  $h_{\text{GK}}$  when applying the FEM reconstruction.

In practice we have found this redefinition to be key in stabilizing the problem's solution. Because the behavior of

the underlying DD is generally unknown, this observation might seem a curse of the present approach. However, we believe that to be only apparent. One main reason is that, although DDs must be generally assumed to be unknown, we can rely on the behavior of the input DGLAP GPD to *infer* information about that of the structure of the underlying DDs. As a simple illustration, one expects the PDFs to exhibit a diverging behavior in the limit  $x \rightarrow 0$ . One may argue, for instance using the Goloskokov-Kroll model, that such low- $x$  divergences translate into a similar behavior of the DD as  $\beta$  approaches zero. Moreover, these singularities are expected to be integrable in the case of valence distributions [as those being considered here, thus triggering our choice  $r(\beta) = 1/\sqrt{\beta}$ ]. A second argument is of a practical nature. Indeed, the relevance of the smoothing suggested in Eq. (3.7) can be traced back to the structure of the interpolating basis functions: Here they are simple degree-two polynomials which are unlikely to accurately piecewise approximate rapidly growing functions such as  $h_{\text{GK}}$  in the limit of small  $\beta$ . In contrast, the approximating functions employed in our ANN approach (where no



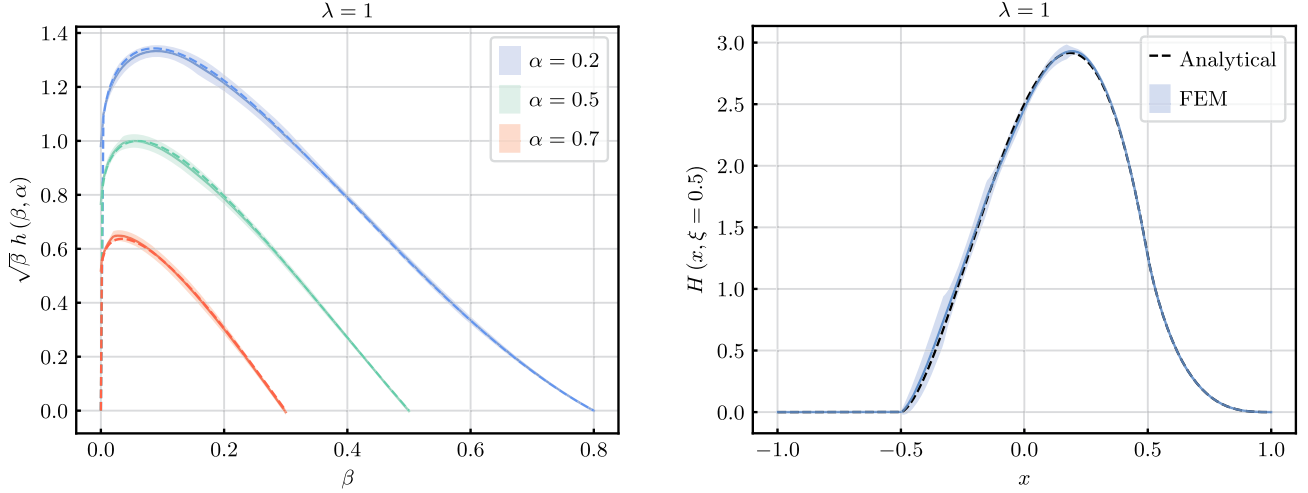


FIG. 4. (FEM) GK model. The left panel shows the DD  $h_{\text{FEM}}(\beta, \alpha)$  obtained through the procedure described in Sec. III A (solid lines), displayed as a function of the  $\beta$  variable for three different values of  $\alpha = \{0.2, 0.5, 0.7\}$  ( $0 \leq \beta \leq 1 - \alpha$ ). For comparison, the exact analytical result of Eq. (A12) is drawn through dashed lines. The sampling of the DGLAP region corresponds to  $\xi \in [0, \lambda x]$  with  $\lambda = 1$  (the entire domain). The right plot displays the GPD  $H(x, \xi)$  at the illustration value  $\xi = 0.5$ , obtained as the Radon transform of the DD shown in the left, compared to the analytical expression, Eqs. (A13) and (A14).

special attention was paid to this feature) are more complicated functions of the kinematic variables  $(\beta, \alpha)$ , thus hinting a modification of the interpolants  $P_e$  to be a general possible solution to this problem. Of course, there is no reason not to upgrade the interpolating polynomials  $P_e$  to more sophisticated structures capable of accounting for these kinds of effects; however, we prefer here to keep things simple and proceed using Eq. (3.7). We highlight that, if we do take advantage of prior knowledge here, it is not at the level of the DD but at the level of the input GPD.

This being clarified, we can operate as before and explore the covariant extension of the GK model from the DGLAP to the ERBL region. Again we use  $\xi = 0.5$  as an illustration for our results; we study three possible sampling regions  $\mathcal{O}_\lambda$ , with  $\lambda = 1, 0.5, 0.2$ , as well. The corresponding results are shown in the right panels of Figs. 4 and 5. Again, the overall fidelity of the reconstructed GPD is noticeable. However, this time we see a growing deterioration, not that substantial at the level of the mean value but at that of the uncertainties, as the area covered by the input data set decreases. In particular, deviations seem to be more apparent in the region  $-\xi < x < 0$ . This can be traced back to numerical instabilities triggered by the smoothing factor  $\sqrt{\beta}$  introduced in Eq. (3.7). Indeed, integration over  $\beta$  translates [by means of the delta distribution in Eq. (2.2)] into  $\sqrt{x - \alpha\xi}$ , which shall be bounded from below. However, for values  $x < 0$ , if numerical precision languishes, the argument might show very small while nonzero values, that being at the origin of the observed impaired behavior.

Further insights can be obtained by again looking at the “intermediate” DD (left panels of Figs. 4 and 5). The results show general good agreement, especially when the entire

DGLAP region is given as input information. However, the reconstructed signal rapidly deteriorates as  $\lambda$  decreases. Nevertheless, those deviations wash out after integration to reach the GPD domain, similarly to what will be seen in Sec. III B. This is indeed a very noticeable feature, as the quantity relevant for phenomenology and the one in lattice QCD is, indeed, the integrated DD; i.e., the GPD.

## B. Artificial neural networks (ANNs)

An entirely different approach to tackle the inverse Radon transform problem consists in fitting the DD by means of an ANN with one hidden layer. The advantages of this strategy are twofold: First, the universal approximation theorem [53] guarantees that this simple network architecture is able to approximate, at arbitrary precision for sufficiently large width, any compactly supported continuous function, therefore providing an unbiased parametrization of the DD; second, the optimization algorithms that are used to train the network come equipped with regularization methods, introduced to avoid overfitting, which are particularly useful in this context to overcome the ill posedness of the Radon transform inverse problem [54,55].

The DD  $h(\beta, \alpha)$  is therefore approximated by the final output of an ANN consisting of three layers: an input stratum, where the two DD variables  $\beta$  and  $\alpha$  are input features; a hidden layer,  $L = 1$ , consisting of a number  $N$  of neurons; and finally, the output ( $L = 2$ ) layer, made from a single neuron (see Fig. 6). The number of neurons in the intermediate layer,  $N$ , is tuned in each calculation to optimize the results. As a function of the input variables, the network’s output is denoted  $h_{\text{ANN}}(\beta, \alpha)$ . The explicit functional form, in terms of the weights ( $w_i$ ) and bias ( $b_i$ ) parameters is

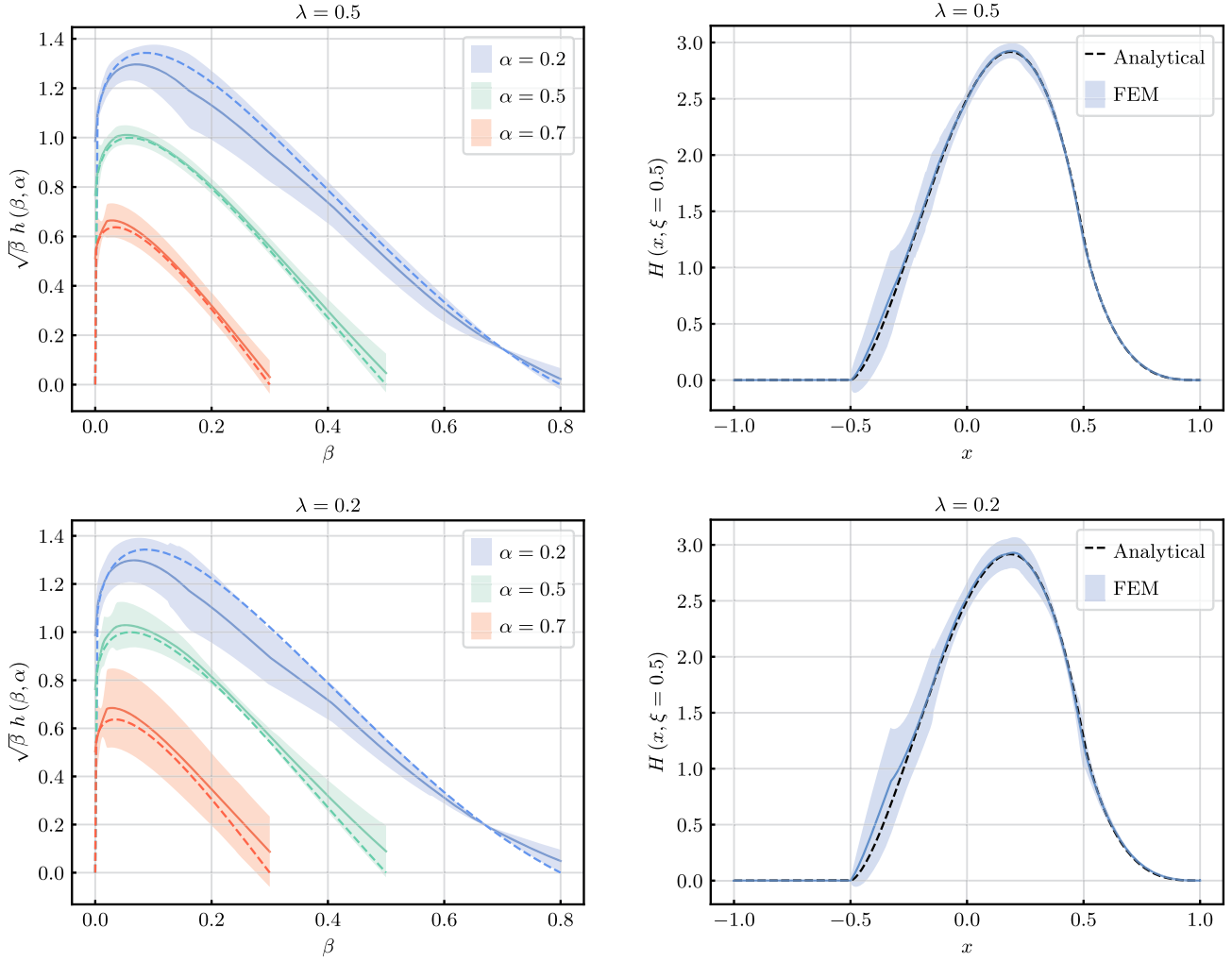


FIG. 5. (FEM) GK model. The same as Fig. 2 with a DGLAP sampling given by  $\xi \in [0, \lambda x]$ , with  $\lambda = 0.5$  (top) and 0.2 (bottom).

$$\begin{aligned}
 h_{\text{ANN}}(\beta, \alpha) &\equiv o^{(2)} = \sum_{i=1}^N w_i^{(2)} o_i^{(1)} + b^{(2)} \\
 &= \sum_{i=1}^N w_i^{(2)} \left[ \sigma \left( w_{\beta i}^{(1)} \beta + w_{\alpha i}^{(1)} \alpha + b_i^{(1)} \right) \right. \\
 &\quad \left. + \sigma \left( w_{\beta i}^{(1)} \beta - w_{\alpha i}^{(1)} \alpha + b_i^{(1)} \right) \right] + b^{(2)}, \quad (3.8)
 \end{aligned}$$

where a rescaled variable  $\alpha' = \alpha/1 - |\beta|$  has been introduced, which takes values between  $-1$  and  $1$  (since  $|\alpha| + |\beta| \leq 1$ ) which is supposed to better perform as an input to the sigmoid activation function  $\sigma(x) = (1 + e^{-x})^{-1}$  which we employ in this work.

With this setup the total number of learning parameters is  $4N + 1$ , divided between  $3N$  weights and  $N + 1$  biases. Notice that in Eq. (3.8) the traditional hidden output has been augmented with an identical term with  $w_\alpha \rightarrow -w_\alpha$ , as it was already implemented in [51], which explicitly enforces the parity condition  $h_{\text{ANN}}(\beta, -\alpha) = h_{\text{ANN}}(\beta, \alpha)$ .

The network is trained using GPD data. Actually, the inputs for the algorithm are the GPD variables  $(x_i, \xi_i)$ , randomly chosen with a uniform probability distribution from the DGLAP region. Each pair of values  $(x_i, \xi_i)$  corresponds to a line  $\alpha = -\beta/\xi_i + x_i/\xi_i$  in the  $\beta$ - $\alpha$  plane, along which the Radon transform has to be evaluated (integrating  $h_{\text{ANN}}(\beta, \alpha)$ ) to produce a predicted GPD value  $\hat{H}(x_i, \xi_i)$  that is to be compared with the true one  $H(x_i, \xi_i)$ . The network depicted in Fig. 6 may therefore be considered as embedded in a larger one, with two extra layers: An input layer that for each  $(x_i, \xi_i)$  outputs the batch of  $(\beta_k, \alpha_k)$  values positioned along the corresponding line<sup>12</sup>; and a final output layer that combines the batch of the  $h_{\text{ANN}}(\beta_k, \alpha_k)$  values to produce  $\hat{H}(x_i, \xi_i)$ .

<sup>12</sup>The size of the batch is the number of nodes whose values depend on the quadrature technique, which we use of numerical integration.

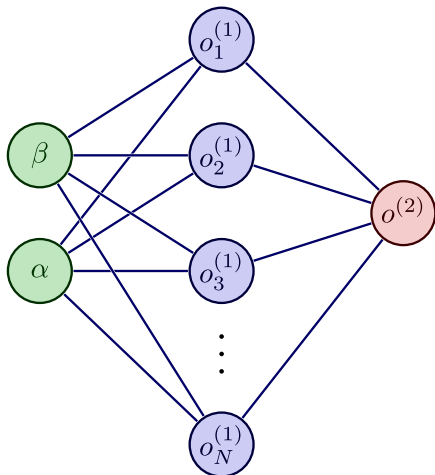


FIG. 6. ANN architecture with one hidden layer that parametrizes the DD  $h(\beta, \alpha)$ . Here  $o_i^{(L)}$  stem for output of the  $i$ th neuron in the  $L$ th layer of the network. The zeroth layer is implicitly identified as the input one. As explained in the text, the activation function of each neuron is chosen to be a sigmoid. Further, the number of neurons in the hidden layer is tuned in each calculation. An optimal solution for the algebraic model is found using 100 neurons in the hidden layer while for the Goloskokov-Kroll model we employed 25.

The network parameters are updated at each iteration, using the adaptive gradient descent ADAM algorithm<sup>13</sup> [56] to minimize the loss function, here chosen to be the *mean squared error* (MSE):

$$\text{MSE} = \frac{1}{N_{\text{sample}}} \sum_{i=1}^{N_{\text{sample}}} \left( \hat{H}(x_i, \xi_i) - H(x_i, \xi_i) \right)^2. \quad (3.9)$$

A *dropout* regularization is implemented [57], where hidden neurons are randomly turned off with a fixed probability rate. It is worth noticing that the ill posedness of a problem like the Radon transform inversion has been commonly treated using a *Tikhonov* regularization. It has been shown, however, that these two approaches are essentially equivalent [54,55], the dropout regularization being more effective for deeper networks.

The reader can note that the number of neurons and the choice of activation function play here a similar role compared to the number of cells discretizing space and the choice of interpolating polynomials, in the FEM approach. Yet, one can expect that the convergence and smoothness properties might be different between the two approaches.

<sup>13</sup>In contrast, an alternative implementation could have employed genetic algorithms for the optimization task [51].

### 1. Algebraic model

In the case of the algebraic model, the ANN has been designed with  $N = 100$  hidden neurons, tuned by trial and error, and trained with a sample of  $N_{\text{sample}} = 10^4$  GPD data points. Without loss of generality, both entries  $x$  and  $\xi$  are considered positive: The  $x$  values are generated within the interval  $[0, 1]$ , which is the  $x$  domain for the quark sector in the DGLAP region, and the associated  $\xi$  may remain also positive while covering the entire DGLAP region owing to the parity of the GPD. Still, for the inversion of the Radon transform (2.3), the subset  $\mathcal{O}_\lambda$  in Eq. (2.6) is fixed by restricting the skewness parameter to the interval  $[0, \lambda x]$ , with  $0 < \lambda \leq 1$ , where the covariant extension is formally achievable, as discussed in Sec. II.

However, despite the support theorem guarantees the uniqueness of the inverted Radon transform from  $\mathcal{O}_\lambda$  with any  $\lambda > 0$ , in practice, the noise in the obtained DD increases as the range of the skewness parameter becomes smaller. This is exhibited in the left panels of Figs. 7 and 8, where three outcomes for the DD function are plotted, corresponding to the values of  $\lambda = \{1, 0.5, 0.2\}$ . The values of  $h(\beta, \alpha)$  are shown as functions of the  $\beta$  variable for three different values of  $\alpha$ . The error bands associated with the ANN results correspond to a standard deviation from the mean value estimated by training the network from scratch on 50 independent trials (replicas).

The errors increase as  $\lambda$  decreases, in particular closer to the boundary  $\beta = 1 - \alpha$ , where the DD is discontinuous. These somewhat large deviations of the numerical predictions from the exact analytical expression become milder, however, once the DD is integrated over the lines in order to get the GPD predictions. This is shown in the right panels of Figs. 7 and 8, where the results displayed correspond to the particular value of  $\xi = 0.5$ , chosen for illustrative purposes. This is noteworthy especially because the main goal of the current application of the support theorem is the GPD reconstruction; namely, the extension of GPD's knowledge from a restricted DGLAP domain to that on its entire support, particularly the full ERBL region. Although achieved by inferring the functional form of the DD, it is actually its Radon transform to be of primary interest here.

The kinematic restriction for the GPD data of the training set  $\mathcal{O}_\lambda$  has been pushed down to a value as low as  $\lambda = 0.2$  in FEM. In this case, the quality of the results for both the inverted DD and the derived GPD is still good for  $\lambda = 0.5$ , as it is shown in Fig. 8, but it becomes strongly degraded for lower values of  $\lambda$ , these results not being then displayed. The main reason for this will be discussed below and comes as a drawback of the GPD behavior for the algebraic model approached with ANNs, presumably requiring further optimization. The GK model, in exchange, will be seen to admit, also for  $\lambda = 0.2$ , a very reliable GPD from a DD approximated with ANNs.

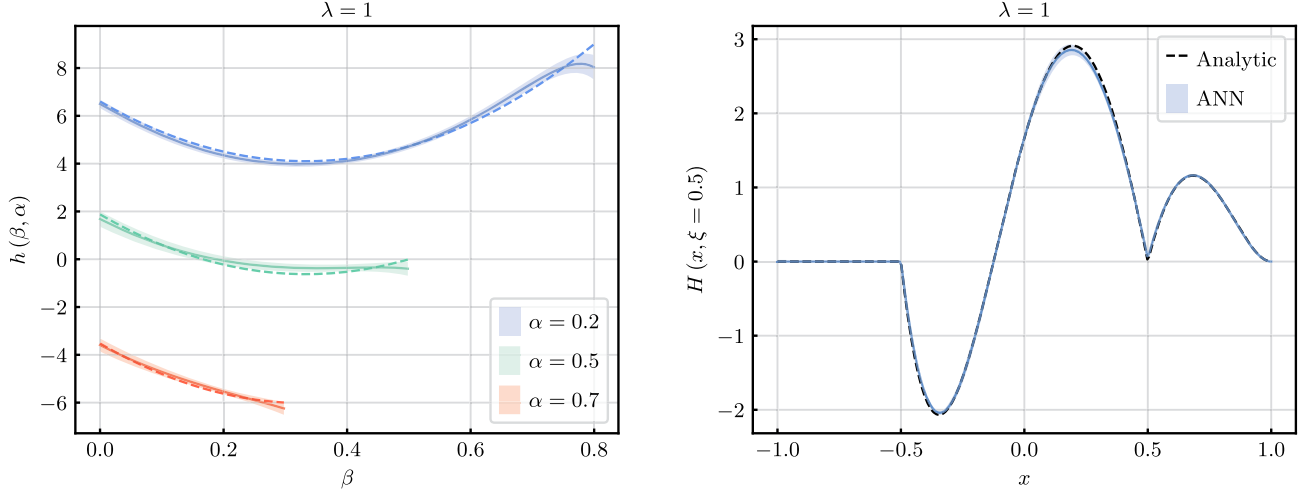


FIG. 7. (ANNs) Algebraic model. The left panel shows the DD  $h_{\text{ANN}}(\beta, \alpha)$  obtained through the procedure described in Sec. III B (solid lines), displayed as a function of the  $\beta$  variable for three different values of  $\alpha = \{0.2, 0.5, 0.7\}$  ( $0 \leq \beta \leq 1 - \alpha$ ). For comparison, the exact analytical result of Eq. (A5) is drawn through dashed lines. The sampling of the DGLAP region corresponds to  $\xi \in [0, \lambda x]$  with  $\lambda = 1$  (the entire domain). The right panel displays the GPD  $H(x, \xi)$  at the illustration value  $\xi = 0.5$ , obtained as the Radon transform of the DD shown in the left, compared to the analytical expression, Eqs. (A4) and (A6).

### 2. Goloskokov-Kroll model

In the case of the GK model, as shown in Appendix A 2, the DD function  $h_{\text{GK}}(\beta, \alpha)$  relies on a profile function  $\pi_N(\beta; \alpha)$  as given by Eq. (A7), which fulfills the normalization condition (A9). In the aim of making easier the explicit implementation of this condition, we have chosen this profile to be approximated by  $\pi_{\text{ANN}}$  as it results from the ANN output  $o^{(2)}$  [see Eq. (3.8)] properly normalized. Thus, the GK DD is accordingly approximated by

$$h_{\text{ANN}}(\beta, \alpha) = f(\beta; \mu) \frac{o^{(2)}(\beta, \alpha)}{\int_{-1+|\beta|}^{1-|\beta|} d\alpha' o^{(2)}(\beta, \alpha')}, \quad (3.10)$$

where  $f(\beta, \mu)$  is the parton distribution function given by Eq. (A11). Notice that in Ref. [51], in addition to

implementing this normalization condition, the ANN expression (3.8) was modified in order to impose the vanishing of the output along the border  $\alpha = 1 - \beta$ , a property which is indeed satisfied by the  $h_{\text{GK}}(\beta, \alpha)$  in (A12). We decided, however, not to implement this constraint explicitly and let the network learn it by itself, considering that the vanishing at the boundary is a particular feature of this model, unlike the normalization condition and the parity in the  $\alpha$  variable, which are properties that come from first principles. This decision comes at the cost of a much slower convergence of the optimization algorithm, due to the larger functional space that has to be explored during the training phase. Notice that one is always free to reexpress any distribution in the form (3.10).

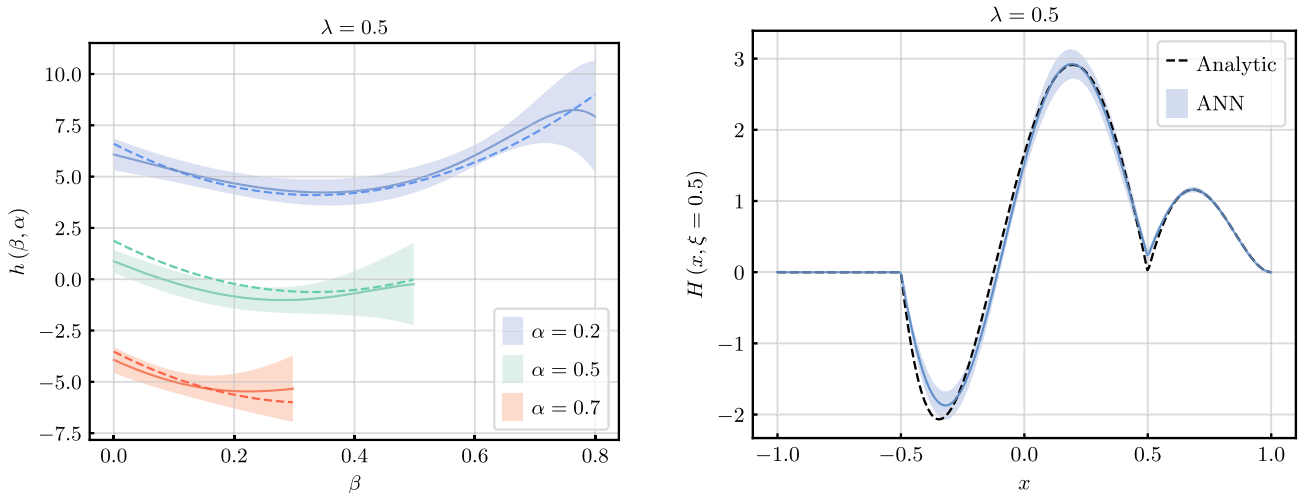


FIG. 8. (ANNs) Algebraic model. The same as Fig. 7 with a DGLAP sampling given by  $\xi \in [0, \lambda x]$  and  $\lambda = 0.5$ .

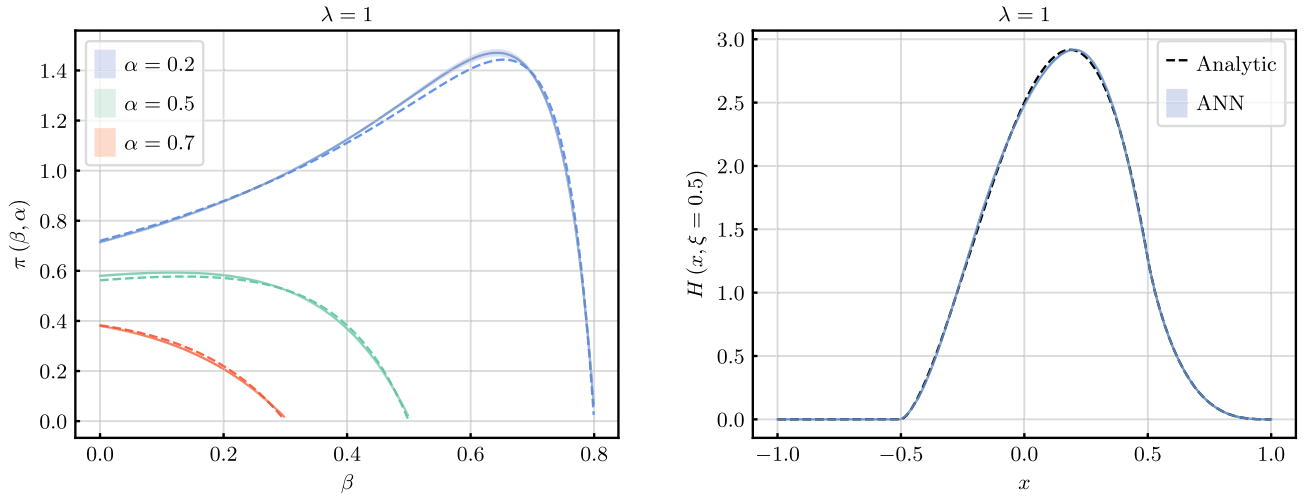


FIG. 9. (ANNs) GK model. The left panel shows the profile function,  $\pi_{\text{ANN}}(\beta, \alpha)$  (solid lines), the direct ANN output which the DD follows immediately from as given in Eq. (3.10), obtained through the procedure described in Sec. III B and displayed as a function of the  $\beta$  variable for three different values of  $\alpha = \{0.2, 0.5, 0.7\}$  ( $0 \leq \beta \leq 1 - \alpha$ ). It compares strikingly well with the exact expressions, Eq. (A10) with  $N = 1$  (dashed lines). The sampling of the DGLAP region corresponds to  $\xi \in [0, \lambda x]$  with  $\lambda = 1$  (the entire domain). The right plot displays the GPD  $H(x, \xi)$  at the illustration value  $\xi = 0.5$ , obtained as the Radon transform of the DD shown in the left, compared to the analytical expression, Eqs. (A13) and (A14).

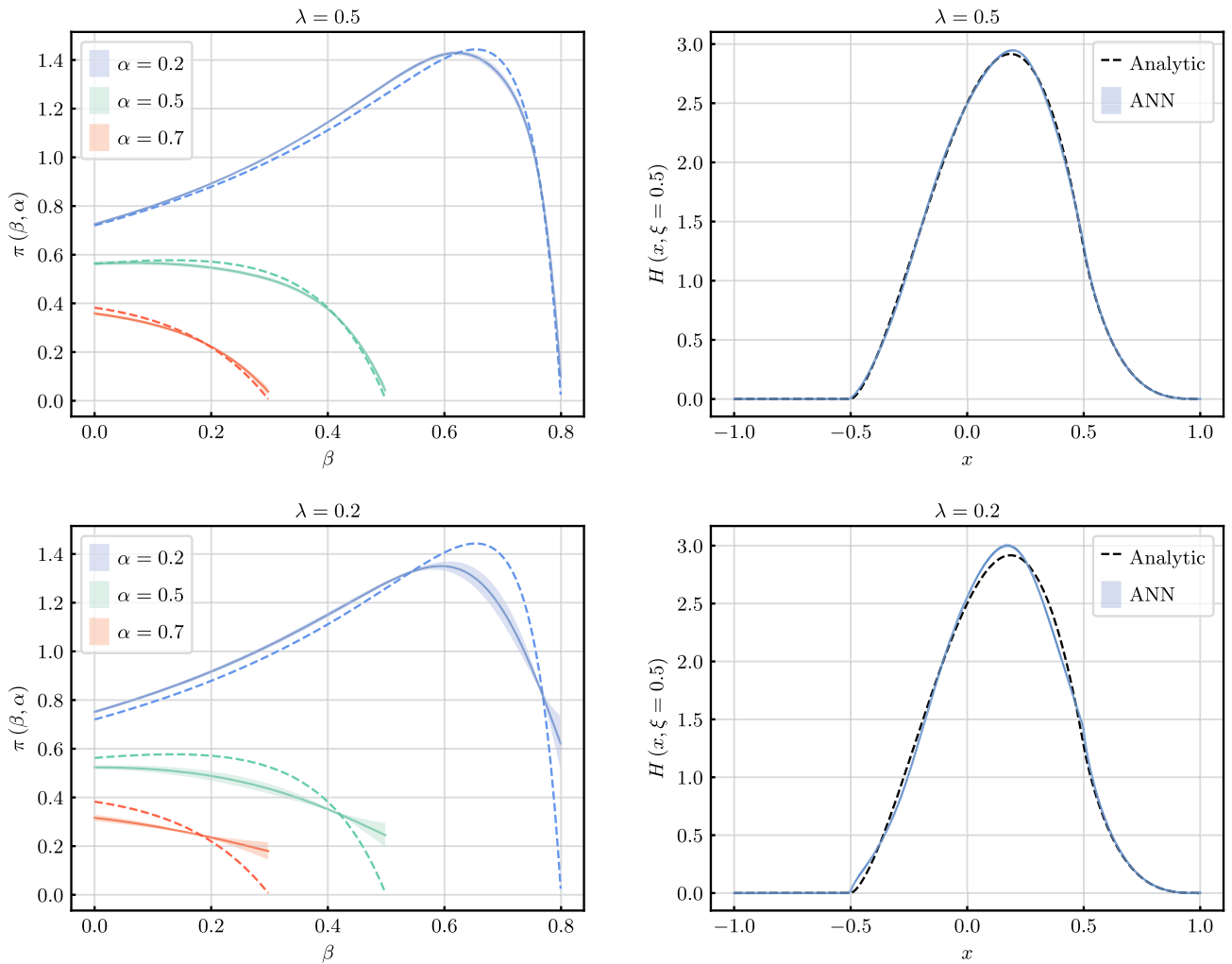


FIG. 10. (ANNs) GK model. The same as Fig. 9 with a DGLAP sampling given by  $\xi \in [0, \lambda x]$ , and  $\lambda = 0.5$  (top),  $0.2$  (bottom).

As opposed to the previous example, the DD function that the network has to learn is now everywhere continuous, since it vanishes at the boundary of its support. This property helps the ANN to better converge to the exact solution, yielding more precise results, as it can be appreciated in Figs. 9 and 10. There, in the left panels, we have chosen to represent the profile function,  $\pi_{\text{ANN}}(\beta, \alpha)$ , the direct output from ANNs which the DD follows immediately from as given in Eq. (3.10). Furthermore, a smaller network was used, with  $N = 25$  hidden neurons, and a smaller training data sampling of  $N_{\text{sample}} = 5 \times 10^3$  GPD values generated in the DGLAP region using Eq. (A13), derived from Eqs. (A11) and (A12). In the case of a kinematic restriction given by  $\lambda = 0.2$ , the prediction for the inverse function is rather imprecise, although the deviation from the exact result of its Radon transform is not substantial (see the bottom panels in Fig. 10).

As stated above, we do not aim at a careful and conclusive comparison of the two different approaches herein followed to invert the Radon transform; namely, ANNs and FEM. Both of them have been implemented instead to avoid drawing biased conclusions about our methodology that may stem for the suitability of a particular implementation to the models chosen as an illustration. To this aim, we have independently tuned the setups in both ANN and FEM approaches such that they deliver good results when they compare to the exact ones. In particular, we have presented results in which the ratio of ANN to FEM input data is roughly 100.

Notwithstanding, some exploratory studies in the context of approximating the GK model through the ANN approach evidence that still acceptable results, while poorer, could be achieved by reducing the number of input data to 250. However, similar studies around the algebraic model seem to point out an issue arising from the use of nonpolynomial activation functions, requiring a large number of neurons to reproduce the polynomial behavior of the algebraic DD. The use of extra neurons being associated to the need of more training data, this observation suggests a limitation of the ANN-based method. One could instead consider different activation functions, better adapted to a polynomial DD, but this would be an *ad hoc* improvement relying on *prior* knowledge of the DD. On the other hand, techniques of data augmentation which are commonly used in machine learning applications could be also implemented. Indeed, we have triggered an analysis with 250 data points and used standard interpolation procedures to generate enough additional data to invert the Radon transform for the algebraic model GPD, thus getting results similar to those obtained with FEM. Nevertheless, this analysis neglects the impact of errors and noise, which will become also strongly amplified by the data augmentation. A careful and systematic analysis, considering noise and ANNs optimization would be needed to perform a reliable comparison with FEM. This is anyhow out of the scope of the present work.

#### IV. SUMMARY AND CONCLUSIONS

In this work we have demonstrated that, at fixed Mandelstam  $t$ , Lorentz covariance as encoded through the renowned polynomiality imposes a strong enough constraint so that the sole knowledge of GPDs at zero and very-low skewness is enough to have them determined over their entire kinematic range. The formal proof of this result relies entirely on a theorem by Boman and Todd-Quinto [34], which we have revisited here and made it more transparent for the field of GPDs by its formulation in the appropriate language.

We have thus concluded that the knowledge of GPDs for  $x \in [-1, 1]$  and  $\xi \in [0, \lambda x]$  is enough to fully constrain them over their entire  $(x, \xi)$  range. Furthermore, this idea has been implemented in practice through two different algorithms: one based on ANNs, and a second one using a more traditional strategy using FEM. Both approaches have been tested against standard models for GPDs, showing that accurate reconstructions shall be obtained in practice with input information as restricted as to  $\sim 20\%$  of the DGLAP region. Given the increasing interest in the study of GPDs, we believe this approach to be valuable in complementing the limited kinematic range over which we expect outcomes to be obtained from lattice QCD simulations (limited by Ioffe time and hadron momentum accessible, see, e.g., [58]) and experimental data. So far, looking towards a practical application of this methodology, the FEM strategy seems to perform better than the ANN one: The number of inputs required by the ANN approach is 10 times larger than the one needed in the FEM approach. The reason is that in the former case, the required information roughly linearly scales with the size of the network. Nevertheless, optimization of the ANN approach can still be achieved, as well as the use of data augmentation techniques implemented. This work, being intended at drawing the attention to this novel technique and not to a refined kinematic completion of actual data, added to the lack of reliable GPD extractions; we defer such a comparative study, optimization of the numerical techniques, and testing of robustness against noise to a future work.

#### ACKNOWLEDGMENTS

J. M. M. C. thanks V. Bertone, B. Blossier, M. Riberdy, and T. San José for valuable discussions and comments. The work from C. M., J. M. M. C., and H. M. has been supported by the GLUODYNAMICS project funded by “P2IO LabEx (ANR-10-LABX-0038)” in the framework of Investissements d’Avenir (ANR-11-IDEX-0003-01), managed by the Agence Nationale de la Recherche (ANR), France; and by the European Union’s Horizon 2020 research and innovation program under Grant Agreement STRONG 2020—No. 824093. P. D. O. acknowledges financial support from “Junta de Andalucía” through “Programa Operativo FEDER de Andalucía 2014-2020

(PAIDI)” under the Project No. P20\_00764. The work of P. D. O., F. D. S., J. R.-Q., and J. S. was supported by the “Spanish Ministerio de Ciencia e Innovación (MICINN)” through Grants No. PID2019-107844-GB-C22 and No. PID2022-140440-NB-C22.

## APPENDIX: MODELS FOR GENERALIZED PARTON DISTRIBUTIONS

In this work we have illustrated our methodology using two different, well-known models for GPDs: (a) The *algebraic model* [24,25,59] and (b) the *Goloskokov-Kroll model* [33,46–48]. There are two main reasons for this choice. On the one hand, the so-called algebraic model is currently being exploited for the description of pion’s structure (e.g., [23,26]) while the Goloskokov-Kroll model is popular in the study of nucleon GPDs, see, e.g., [9,60] and references therein. We thus find it enlightening to benchmark our approach on the (perhaps) two most paradigmatic scenarios found in the study of hadron structure, pions and nucleons, thus triggering our interest into these two models. A second argument follows from a more technical perspective: If the DD of the algebraic model is a smooth function of its kinematic variables, Sec. A 1, that of the GK model shows an integrable divergence at  $\beta \rightarrow 0$ , Sec. A 2. Therefore we find it appropriate to benchmark our implementations in these two radically different scenarios where numerical accuracy is expected to be challenged in two different manners.

This being said, the two models could have been employed to test our algorithms without any regard to the underlying physics. In fact, we are here simply testing our way of solving the inverse Radon transform problem. However, for the sake of self-consistency, we believe it worthwhile to briefly review their foundations.

### 1. Algebraic model

Capitalizing on a long effort towards the development of models for parton distributions respecting all of the QCD’s basic properties [24,59,61–64], the *algebraic model* for the GPD of valence quarks in a pion was first presented in Ref. [25]. There the authors relied on the overlap representation of GPDs [21,36] to model the DGLAP region in a way consistent with their positivity property [20,22]. To that end, the essential ingredient is a parametrization of the hadron’s light-front wave function (LFWF)  $\Psi(x, \mathbf{k}_\perp; \mu)$ ; which for a two-body system reads [21]

$$H(x, \xi; \mu)|_{\text{DGLAP}} = \sum_\sigma \int \frac{d^2 \mathbf{k}_\perp}{16\pi^3} \Psi_\sigma^* \left( \frac{x-\xi}{1-x}, \mathbf{k}_\perp + \frac{1-x}{1-\xi} \frac{\Delta_\perp}{2}; \mu \right) \times \Psi_\sigma \left( \frac{x+\xi}{1+\xi}, \mathbf{k}_\perp - \frac{1-x}{1+\xi} \frac{\Delta_\perp}{2}; \mu \right), \quad (\text{A1})$$

$\sigma$  being the possible quark-helicity combinations,  $\mathbf{k}_\perp$  the average momentum of the quark pair along the transverse

direction, as defined by the hadron’s (light-cone) momentum,  $p$ , and  $\Delta_\perp$  the transverse components of the four-momentum transfer between quarks.

The light-front wave function giving rise to the algebraic model was obtained in Euclidean space from an appropriate projection of the pion’s Bethe-Salpeter wave function [65], which was in turn constructed using a Nakanishi representation for the Bethe-Salpeter amplitude [66,67],

$$\chi(k, p) = i\mathcal{N}\gamma_5 \int_0^\infty dw \int_{-1}^1 dz \frac{\rho(w, z)M^2}{(k - \frac{1-z}{2}p)^2 + M^2 + w},$$

$$\text{with } \rho(w, z) = \delta(w)(1-z^2). \quad (\text{A2})$$

where  $M$  is a mass scale, supplemented with an ansatz of the form

$$S(k) = (-i\gamma \cdot k + M)/(k^2 + M^2), \quad (\text{A3})$$

for the quark propagator.

Such a parametrization allows for a fully algebraic treatment leading to the evaluation of the two possible light-front wave functions in closed form [59], thus the denomination of the resulting GPD model as “algebraic.” Feeding the overlap representation Eq. (A1) with those LFWFs yields

$$H(x, \xi; \mu)|_{\text{DGLAP}} = 30 \frac{(1-x)^2(x^2 - \xi^2)}{(1 - \xi^2)^2}. \quad (\text{A4})$$

Here, the companion ERBL region was constructed in [25] following the dictations of Lorentz covariance as captured by the renowned polynomiality property and implemented through the (inverse) Radon transform [24,41]. As a result, the underlying DD is found [25]:

$$h_{\text{Alg}}(\beta, \alpha; \mu) = \frac{15}{2} [1 - 3(\alpha^2 - \beta^2) - 2\beta], \quad (\text{A5})$$

while the ERBL GPD turns out to be

$$H(x, \xi; \mu)|_{\text{ERBL}} = \frac{15(1-x)(x^2 - \xi^2)}{2\xi^3(1+\xi)^2} (x + 2x\xi + \xi^2). \quad (\text{A6})$$

As a final remark, notice that no explicit reference has been made to the scale  $\mu$  in the presentation of this model. Its setting is implicit in the expression [Eq. (A1)], where a meson is being represented by two-body LFWFs. A discussion about this subject is definitely outside the scope of the present work; the interested reader can find more information in, e.g., [45] and references therein.

### 2. Goloskokov-Kroll model

The Goloskokov-Kroll (GK) model [33,46–48] is a popular phenomenological parametrization, originally

tailored for the description of deeply virtual meson production (DVMP) but which has also proved to yield good agreement in a leading order description of deeply virtual Compton scattering data<sup>14</sup> [68]. It is built on top of *Radyushkin's double distribution Ansatz* (RDDA) which in turn relays on two simple ideas [69]:

- (i) The profile of a GPD along the  $x$  direction is basically determined by that of  $f(x; \mu) \equiv H(x, \xi = 0; \mu)$ , i.e., the parton distribution function.
- (ii) The shape of a GPD along the  $\xi$  direction characterizes the spread of parton momentum induced by momentum transfer.

A simple way to combine these two assumptions is to model GPDs through DDs of the form

$$h_{\text{RDDA}}^{(N)}(\beta, \alpha; \mu) = f(\beta; \mu) \pi_N(\beta, \alpha), \quad (\text{A7})$$

where  $\pi_N(\beta, \alpha)$  is a profile function. From that point on, Eq. (2.2) gives

$$\begin{aligned} H(x, \xi = 0; \mu) &= \iint_{\Omega} d\beta d\alpha \delta(x - \beta) f(\beta; \mu) \pi_N(\beta, \alpha) \\ &= f(x; \mu) \int_{-1+|x|}^{1-|x|} \pi_N(x, \alpha), \end{aligned} \quad (\text{A8})$$

which produces the desired forward limit if the profile function is normalized as

$$\int_{-1+|x|}^{1-|x|} d\alpha \pi_N(x, \alpha) = 1. \quad (\text{A9})$$

A simple way to fulfill such a requirement is to employ the profile function suggested in [69],

$$\pi_N(\beta, \alpha) = \frac{\Gamma(2N+2)}{2^{2N+1} \Gamma^2(N+1)} \frac{[(1-|\beta|)^2 - \alpha^2]^N}{(1-|\beta|)^{2N+1}}, \quad (\text{A10})$$

which satisfies the condition Eq. (A9)  $\forall N|\text{Re}(N) - 1$ . Strikingly, the  $\alpha$  dependence of such a profile function

TABLE II. Parameters of the *ansatz* Eq. (A11) fitting the CTEQ6m PDF [70] in the range  $10^{-2} \leq \beta \leq 0.5$  and  $\mu_0^2 \equiv 4 \text{ GeV}^2 \leq \mu^2 \leq 40 \text{ GeV}^2$  (taken from [46]).

$\delta$	0.48
$c_0(\mu)$	$1.52 + 0.248 \log(\mu^2/\mu_0^2)$
$c_1(\mu)$	$2.88 - 0.940 \log(\mu^2/\mu_0^2)$
$c_2(\mu)$	$-0.095 \log(\mu^2/\mu_0^2)$

is entirely controlled by a single parameter,  $N$ , producing a rather inflexible modeling.<sup>15</sup>

In [33,46–48] this approach is followed to build the renowned Goloskokov-Kroll model. There, the valence-quark GPD in nucleons is designed choosing  $N = 1$  in the profile function  $\pi_N(\beta, \alpha)$ , which reproduces the expected asymptotic behavior for the quark distribution amplitude [33]; and employing a phenomenological *ansatz* for the parton distribution function,

$$f(\beta; \mu) = \beta^{-\delta} (1 - \beta)^3 \sum_{j=0}^2 c_j(\mu) \beta^{j/2} \quad (\text{A11})$$

with  $\delta$  and  $c_j$  parameters determined from a fit to the CTEQ6m PDF [46,70] (see Table II).

Combining the parametrization above with Radyushkin's DD *ansatz* generates the Goloskokov-Kroll model for the DD of valence quarks within nucleons,

$$\begin{aligned} h_{\text{GK}}(\beta, \alpha; \mu) &\equiv h_{\text{RDDA}}^{(1)}(\beta, \alpha; \mu) \\ &= \frac{3}{4} \left( (1 - \beta)^2 - \alpha^2 \right) \beta^{-\delta} \sum_{j=0}^2 c_j(\mu) \beta^{j/2}. \end{aligned} \quad (\text{A12})$$

For the sake of simplicity, in this work we shall deal with the present model at  $\mu^2 = \mu_0^2 = 4 \text{ GeV}^2$  which leaves us with the GPD

$$\begin{aligned} H(x, \xi; \mu)|_{\text{DGLAP}} &= \frac{3}{4\xi^3} \sum_{j=0}^2 \frac{c_j}{(1 - \xi^2)^{a+1}} \left\{ \frac{\xi^2 - x^2}{a+1} \left[ (x + \xi)^{a+1} (1 - \xi)^{a+1} - (x - \xi)^{a+1} (1 + \xi)^{a+1} \right] \right. \\ &\quad - 2 \frac{\xi^2 - x}{(a+2)(1 - \xi^2)} \left[ (x + \xi)^{a+2} (1 - \xi)^{a+2} - (x - \xi)^{a+2} (1 + \xi)^{a+2} \right] \\ &\quad \left. + \frac{\xi^2 - 1}{(a+3)(1 - \xi^2)^2} \left[ (x + \xi)^{a+3} (1 - \xi)^{a+3} - (x - \xi)^{a+3} (1 + \xi)^{a+3} \right] \right\}, \end{aligned} \quad (\text{A13})$$

<sup>14</sup>For the sake of completeness, one must also point out that such agreement was argued in [8] to be only casual.

<sup>15</sup>In fact, a rapid convergence of the models thus produced through different choices of  $N$  was demonstrated in [61,64].



within the DGLAP region, and

$$H(x, \xi; \mu)|_{\text{ERBL}} = \frac{3}{4\xi^3} \sum_{j=0}^2 c_j \left( \frac{x+\xi}{1+\xi} \right)^{a+1} \left\{ \frac{\xi^2 - x^2}{a+1} + \frac{x+\xi}{1+\xi} \left[ \frac{(x+\xi)(\xi-1)}{a+3} - 2 \frac{\xi^2 - x}{a+2} \right] \right\}, \quad (\text{A14})$$

in the ERBL region, where  $a \equiv j/2 - \delta$ .

- 
- [1] D. Mueller, D. Robaschik, B. Geyer, F.M. Dittes, and J. Hořejši, Wave functions, evolution equations and evolution kernels from light ray operators of QCD, *Fortschr. Phys.* **42**, 101 (1994).
- [2] X. Ji, Deeply virtual Compton scattering, *Phys. Rev. D* **55**, 7114 (1997).
- [3] A. V. Radyushkin, Nonforward parton distributions, *Phys. Rev. D* **56**, 5524 (1997).
- [4] M. Burkardt, Impact parameter dependent parton distributions and off forward parton distributions for  $\zeta \rightarrow 0$ , *Phys. Rev. D* **62**, 071503 (2000); **66**, 119903(E) (2002).
- [5] M. Diehl, Generalized parton distributions in impact parameter space, *Eur. Phys. J. C* **25**, 223 (2002).
- [6] X. Ji, Gauge-invariant decomposition of nucleon spin, *Phys. Rev. Lett.* **78**, 610 (1997).
- [7] M. V. Polyakov, Generalized parton distributions and strong forces inside nucleons and nuclei, *Phys. Lett. B* **555**, 57 (2003).
- [8] K. Kumerički and D. Müller, Description and interpretation of DVCS measurements, *EPJ Web Conf.* **112**, 01012 (2016).
- [9] H. Moutarde, P. Sznajder, and J. Wagner, Border and skewness functions from a leading order fit to DVCS data, *Eur. Phys. J. C* **78**, 890 (2018).
- [10] V. Bertone, H. Dutrieux, C. Mezrag, H. Moutarde, and P. Sznajder, The deconvolution problem of deeply virtual Compton scattering, *Phys. Rev. D* **103**, 114019 (2021).
- [11] E. Moffat, A. Freese, I. Cloët, T. Donohoe, L. Gamberg, W. Melnitchouk, A. Metz, A. Prokudin, and N. Sato, Shedding light on shadow generalized parton distributions, *Phys. Rev. D* **108**, 036027 (2023).
- [12] R. Boussarie, B. Pire, L. Szymanowski, and S. Wallon, Exclusive photoproduction of a  $\gamma\gamma$  pair with a large invariant mass, *J. High Energy Phys.* **02** (2017) 054; **10** (2018) 29.
- [13] G. Duplančić, S. Nabeebaccus, K. Passek-Kumerički, B. Pire, L. Szymanowski, and S. Wallon, Probing chiral-even and chiral-odd leading twist quark generalised parton distributions through the exclusive photoproduction of a  $\gamma\gamma$  pair, *Phys. Rev. D* **107**, 094023 (2023).
- [14] O. Grocholski, B. Pire, P. Sznajder, L. Szymanowski, and J. Wagner, Collinear factorization of diphoton photoproduction at next to leading order, *Phys. Rev. D* **104**, 114006 (2021).
- [15] J.-W. Qiu and Z. Yu, Exclusive production of a pair of high transverse momentum photons in pion-nucleon collisions for extracting generalized parton distributions, *J. High Energy Phys.* **08** (2022) 103.
- [16] J.-W. Qiu and Z. Yu, Extraction of the parton momentum-fraction dependence of generalized parton distributions from exclusive photoproduction, *Phys. Rev. Lett.* **131**, 161902 (2023).
- [17] X. Ji, Off forward parton distributions, *J. Phys. G* **24**, 1181 (1998).
- [18] A. V. Radyushkin, Symmetries and structure of skewed and double distributions, *Phys. Lett. B* **449**, 81 (1999).
- [19] A. V. Radyushkin, Double distributions and evolution equations, *Phys. Rev. D* **59**, 014030 (1999).
- [20] B. Pire, J. Soffer, and O. Teryaev, Positivity constraints for off-forward parton distributions, *Eur. Phys. J. C* **8**, 103 (1999).
- [21] M. Diehl, T. Feldmann, R. Jakob, and P. Kroll, The overlap representation of skewed quark and gluon distributions, *Nucl. Phys.* **B596**, 33 (2001).
- [22] P. V. Pobylitsa, Disentangling positivity constraints for generalized parton distributions, *Phys. Rev. D* **65**, 114015 (2002).
- [23] J. M. Morgado Chávez, V. Bertone, F. De Soto, M. Defurne, C. Mezrag, H. Moutarde, J. Rodríguez-Quintero, and J. Segovia, Pion generalized parton distributions: A path toward phenomenology, *Phys. Rev. D* **105**, 094012 (2022).
- [24] N. Chouika, C. Mezrag, H. Moutarde, and J. Rodríguez-Quintero, Covariant extension of the GPD overlap representation at low Fock states, *Eur. Phys. J. C* **77**, 906 (2017).
- [25] N. Chouika, C. Mezrag, H. Moutarde, and J. Rodríguez-Quintero, A Nakanishi-based model illustrating the covariant extension of the pion GPD overlap representation and its ambiguities, *Phys. Lett. B* **780**, 287 (2018).
- [26] J. M. Morgado Chávez, V. Bertone, F. De Soto, M. Defurne, C. Mezrag, H. Moutarde, J. Rodríguez-Quintero, and J. Segovia, Accessing the pion 3D structure at U.S. and China electron-ion colliders, *Phys. Rev. Lett.* **128**, 202501 (2022).
- [27] H. Dutrieux, M. Winn, and V. Bertone, Exclusive meets inclusive particle production at small Bjorken  $x_B$ : How to relate exclusive measurements to PDFs based on evolution equations, *Phys. Rev. D* **107**, 114019 (2023).
- [28] M. J. Riberdy, H. Dutrieux, C. Mezrag, and P. Sznajder, Combining lattice QCD and phenomenological inputs on generalised parton distributions at moderate skewness, *Eur. Phys. J. C* **84**, 201 (2024).
- [29] Anatoly V. Radyushkin, Generalized parton distributions and pseudodistributions, *Phys. Rev. D* **100**, 116011 (2019).

- [30] A. V. Radyushkin, Double distributions and pseudodistributions, *Phys. Rev. D* **109**, 014514 (2024).
- [31] S. Bhattacharya, K. Cichy, M. Constantinou, J. Dodson, X. Gao, A. Metz, S. Mukherjee, A. Scapellato, F. Steffens, and Y. Zhao, Generalized parton distributions from lattice QCD with asymmetric momentum transfer: Unpolarized quarks, *Phys. Rev. D* **106**, 114512 (2022).
- [32] S. Bhattacharya, K. Cichy, M. Constantinou, J. Dodson, A. Metz, A. Scapellato, and F. Steffens, Chiral-even axial twist-3 GPDs of the proton from lattice QCD, *Phys. Rev. D* **108**, 054501 (2023).
- [33] S. V. Goloskokov and P. Kroll, Vector meson electroproduction at small Bjorken- $x$  and generalized parton distributions, *Eur. Phys. J. C* **42**, 281 (2005).
- [34] J. Boman and E. Todd-Quinto, Support theorems for real-analytic Radon transforms, *Duke Math. J.* **55**, 943 (1987).
- [35] Maxim V. Polyakov and C. Weiss, Skewed and double distributions in pion and nucleon, *Phys. Rev. D* **60**, 114017 (1999).
- [36] M. Diehl, Generalized parton distributions, *Phys. Rep.* **388**, 41 (2003).
- [37] F. Natterer and F. Wübbeling, *Mathematical Method in Image Reconstruction*, Monographs on Mathematical Modeling and Computation (Society for Industrial and Applied Mathematics, New York, 2001).
- [38] A. V. Belitsky and A. V. Radyushkin, Unraveling hadron structure with generalized parton distributions, *Phys. Rep.* **418**, 1 (2005).
- [39] A. Hertle, Continuity of the Radon transform and its inverse on Euclidean space, *Math. Z.* **184**, 165 (1983).
- [40] N. Chouika, Generalized parton distributions and their covariant extension: Towards nucleon tomography, Ph.D. thesis, IRFU, Saclay, DPHN, 2018.
- [41] O. V. Teryaev, Crossing and Radon tomography for generalized parton distributions, *Phys. Lett. B* **510**, 125 (2001).
- [42] B. C. Tiburzi, Double distributions: Loose ends, *Phys. Rev. D* **70**, 057504 (2004).
- [43] P. V. Pobylitsa, Solution of polynomiality and positivity constraints on generalized parton distributions, *Phys. Rev. D* **67**, 034009 (2003).
- [44] N. Chouika, C. Mezrag, H. Moutarde, and J. Rodríguez-Quintero, Concurrent approaches to generalized parton distribution modeling: The pion's case, *EPJ Web Conf.* **137**, 05020 (2017).
- [45] J. M. Morgado Chávez, Generalized parton distributions of the pion: Modeling, evolution and observable implications, Ph.D. thesis, Huelva U., 2022.
- [46] S. V. Goloskokov and P. Kroll, The longitudinal cross-section of vector meson electroproduction, *Eur. Phys. J. C* **50**, 829 (2007).
- [47] S. V. Goloskokov and P. Kroll, The role of the quark and gluon GPDs in hard vector-meson electroproduction, *Eur. Phys. J. C* **53**, 367 (2008).
- [48] S. V. Goloskokov and P. Kroll, An attempt to understand exclusive  $\pi^+$  electroproduction, *Eur. Phys. J. C* **65**, 137 (2010).
- [49] B. Berthou, D. Binosi, N. Chouika, L. Colaneri, M. Guidal, C. Mezrag, H. Moutarde, J. Rodríguez-Quintero, F. Sabatié, P. Sznajder, and J. Wagner, PARTONS: PARTonic tomography of nucleon software. A computing framework for the phenomenology of generalized parton distributions, *Eur. Phys. J. C* **78**, 478 (2018).
- [50] F. Natterer, The finite element method for ill-posed problems, *RAIRO. Analyse numérique* **11**, 271 (1977).
- [51] H. Dutriex, O. Grocholski, H. Moutarde, and P. Sznajder, Artificial neural network modelling of generalised parton distributions, *Eur. Phys. J. C* **82**, 252 (2022); **82**, 389(E) (2022).
- [52] I. F. Ilyas and X. Chu, *Data Cleaning* (Association for Computing Machinery, New York, 2019).
- [53] K. Hornik, M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, *Neural Netw.* **2**, 359 (1989).
- [54] C. M. Bishop, Training with noise is equivalent to Tikhonov regularization, *Neural Comput.* **7**, 108 (1995).
- [55] S. Wager, S. Wang, and P. Liang, Dropout training as adaptive regularization, in *Proceedings of the 27th Annual Conference on Neural Information Processing Systems, NIPS* (2013).
- [56] D. P. Kingma and J. Ba, ADAM: A method for stochastic optimization, in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 2015, Conference Track Proceedings* (2015), edited by Y. Bengio and Y. LeCun.
- [57] S. Nitish, H. Geoffrey, K. Alex, S. Ilya, and S. Ruslan, Dropout: A simple way to prevent neural networks from overfitting, *J. Mach. Learn. Res.* **15**, 1929 (2014).
- [58] K. Cichy *et al.*, Generalized parton distributions from lattice QCD, *Acta Phys. Pol. B Proc. Suppl.* **16**, 7 (2023).
- [59] C. Mezrag, H. Moutarde, and J. Rodríguez-Quintero, From Bethe-Salpeter wave functions to generalised parton distributions, *Few-Body Syst.* **57**, 729 (2016).
- [60] K. Kumericki, S. Liuti, and H. Moutarde, GPD phenomenology and DVCS fitting, *Eur. Phys. J. A* **52**, 157 (2016).
- [61] C. Mezrag, H. Moutarde, and F. Sabatié, Test of two new parametrizations of the generalized parton distribution  $H$ , *Phys. Rev. D* **88**, 014001 (2013).
- [62] C. Mezrag, L. Chang, H. Moutarde, C. D. Roberts, J. Rodríguez-Quintero *et al.*, Sketching the pion's valence-quark generalised parton distribution, *Phys. Lett. B* **741**, 190 (2014).
- [63] C. Mezrag, H. Moutarde, J. Rodríguez-Quintero, and F. Sabatié, Towards a pion generalized parton distribution model from Dyson-Schwinger equations, *arXiv:1406.7425*.
- [64] C. Mezrag, Generalised parton distributions: From phenomenological approaches to Dyson-Schwinger equations, Ph.D. thesis, IRFU, SPhN, Saclay, 2015.
- [65] L. Chang, I. C. Cloet, J. J. Cobos-Martinez, C. D. Roberts, S. M. Schmidt, and P. C. Tandy, Imaging dynamical chiral symmetry breaking: Pion wave function on the light front, *Phys. Rev. Lett.* **110**, 132001 (2013).
- [66] N. Nakanishi, Partial-wave Bethe-Salpeter equation, *Phys. Rev.* **130**, 1230 (1963).
- [67] N. Nakanishi, A general survey of the theory of the Bethe-Salpeter equation, *Prog. Theor. Phys. Suppl.* **43**, 1 (1969).

- [68] P. Kroll, H. Moutarde, and F. Sabatie, From hard exclusive meson electroproduction to deeply virtual Compton scattering, *Eur. Phys. J. C* **73**, 2278 (2013).
- [69] I. V. Musatov and A. V. Radyushkin, Evolution and models for skewed parton distributions, *Phys. Rev. D* **61**, 074027 (2000).
- [70] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, Pavel M. Nadolsky, and W.-K. Tung, New generation of parton distributions with uncertainties from global QCD analysis, *J. High Energy Phys.* **07** (2002) 012.