Parton distributions from lattice data

Parton distributions from lattice data:
the nonsinglet case

Krzysztof Cichy[1], Luigi Del Debbio[2] and Tommaso Giani[2]

[1] *Faculty of Physics, Adam Mickiewicz University,*
Umultowska 85, 61-614 Pozna n, Poland
[2] *The Higgs Centre for Theoretical Physics, The University of Edinburgh,*
Peter Guthrie Tait Road, Edinburgh EH9 3FD, United Kingdom

Abstract

We revise the relation between Parton Distribution Functions (PDFs) and matrix elements computable from lattice QCD, focusing on the quasi-Parton Distribution Functions (qPDFs) approach. We exploit the relation between PDFs and qPDFs in the case of the unpolarized isovector parton distribution to obtain a factorization formula relating the real and imaginary part of qPDFs matrix elements to specific nonsinglet distributions, and we propose a general framework to extract PDFs from the available lattice data, treating them on the same footing as experimental data. We implement the proposed approach within the `NNPDF` framework, and we study the potentiality of such lattice data in constraining PDFs, assuming some plausible scenarios to assess the unknown systematic uncertainties. We finally extract the two nonsinglet distributions involved in our analysis from a selection of the available lattice data.

Contents

1 Introduction

A precise understanding of the structure of the proton, encoded in the Parton Distribution Functions (PDFs), is required to make predictions and analyses in collider physics. The PDFs sets currently used in phenomenology studies are extracted from global QCD analysis over experimental data [1–5]. The increasing number and precision of the experiments included in these fits, together with the development of robust fitting methodologies, are pushing the determination of PDFs to a new level of accuracy.

PDFs are formally defined as matrix elements of renormalized operators in QCD. The computations of PDFs from first principle using lattice QCD has been considered a great challenge for a long time. However, recent theoretical developments [6, 7] have shown how lattice QCD simulations of certain spatial correlation functions between two boosted nucleon states can give access to the so-called quasi-PDFs. These functions are then directly related to PDFs by means of a factorization theorem, just like high-energy processes cross sections are.

Following these novel ideas, a number of publications have appeared (see Refs. [8, 9] for recent reviews), from both lattice and high-energy communities, addressing the main theoretical problems of the new lattice approach: the definition and renormalization of the non-local operators involved in the lattice simulation [10–20], the proof of the factorization theorem between PDFs and quasi-PDFs [7, 21–25], the computation of the matching coefficients relating lattice-computable quantities to PDFs in different renormalization schemes [21, 22, 24–33]. Also, data coming from first lattice QCD simulations have started appearing and have gotten into already a relatively advanced stage over the last few years [13, 30–46]. This gives an idea of what PDFs from the lattice might look like, not only for nonsinglet quark PDFs of the nucleon, but also for the pion PDF and distribution amplitude, as well as for the gluon PDF of the nucleon. Given the general interest shown by the community, a quick improvement in the technologies involved in such lattice simulations is to be expected in the next few years. A great quantity of increasingly precise lattice data is then likely to be available in the near future, requiring detailed studies about the possible impact they might have on the overall precision of PDFs determination.

Despite the increasing number of numerical results becoming available, an optimal strategy for reconstructing the PDFs from these data has not been entirely addressed yet. For example, in Ref. [47] a series of possible approaches to tackle the problem of incomplete and discretized Fourier transform has been presented, showing some interesting and promising results. In this

work, we exploit the factorization of quasi-PDFs in PDFs and perturbatively computable coefficients to extract nonsinglet distributions from the data of Refs. [30,45], setting a general strategy within the NNPDF framework, which allows to systematically extract parton distributions from the available lattice data. The discussion is focused on the quasi-PDFs case, but it can be extended to include data coming from different lattice approaches, like for example pseudo-PDFs [48–50], fictitious heavy/light quark [51,52] or current-current correlators [22,24,53], paving the way to a first global lattice QCD analysis [24].

The structure of the paper is the following. In Sec. 2 we revise the theoretical basis of the problem and we set up the notation. In Sec. 3 we describe the lattice data considered in this work, together with the the corresponding systematic uncertainties. We consider different possible scenarios for the latter, studying their impact on the final result. We then work out a factorization formula relating such data to the two nonsinglet distributions T_3 and V_3, defined in Sec. 2. In Sec. 4 we describe the fitting settings, revising the main feature of the NNPDF framework and finally in Sec. 5 we present our results. We conclude in Sec. 6.

2   From Parton Distributions to quasi-Parton Distributions

In this section, we revise the formal definition of PDFs and quasi-PDFs as matrix elements of non-local operators in QCD. We recall the main theoretical framework concerning the definition, renormalization and matching between spatial and light-cone quantities. For a broad overview of several aspects of quasi-PDFs, we refer to the recent extensive review [8], and references therein.

The bare unpolarized quark PDF is defined as [54,55]

$$f_{\mathrm{q}}^{(0)}(x) = \int \frac{d\xi^-}{4\pi} e^{-ixP^+\xi^-} \langle \mathrm{P}|\bar{\psi}_{\mathrm{q}}^{(0)}(\xi^-) \gamma^+ U(\xi^-,0) \psi_{\mathrm{q}}^{(0)}(0)|\mathrm{P}\rangle, \qquad (2.1)$$

where $|\mathrm{P}\rangle$ denotes a hadronic state with momentum $P^\mu = (P^0, 0, 0, P^z)$, and $P^\pm = \frac{(P^0 \pm P^z)}{\sqrt{2}}$ are light-cone coordinates. The index q identifies the parton under investigation. For instance, in a theory where we only consider the four lightest quarks, we have $q = u, d, s, c$. The momentum carried by the parton is $k^\mu = xP^\mu$, $\psi_{\mathrm{q}}^{(0)}$ is the bare quark field operator and the Wilson line U is given by

$$U(\xi^-, 0) = \mathrm{P} \exp\left(-ig \int_0^{\xi^-} d\eta^- A^{(0)+}(\eta^-)\right). \qquad (2.2)$$

An analogous definition can be given for the gluon bare PDFs, denoted as $f_g^{(0)}(x)$. The superscripts (0) in the above expressions identify bare fields: the matrix elements that enter in the definition of $f_{\mathrm{q}}^{(0)}$ are ultraviolet divergent, and therefore need to be renormalized. Renormalized parton distributions are usually defined by minimal subtraction, and the relation between the bare and the renormalized quantities is given by

$$f_a^{(0)}(x) = \sum_b \int_x^1 \frac{dy}{y} \, \mathrm{Z}_{ab}\left(\frac{x}{y}, \mu\right) f_b(y, \mu^2), \qquad (2.3)$$

where the indices a and b run over all the parton types (gluon and flavors of quarks) and μ denotes the renormalization scale introduced by the minimal subtraction scheme [54].

Focusing on the quark PDFs for now, the renormalized distributions introduced above have a compact support given by the interval $[-1, 1]$. For phenomenological applications, it is customary to define the PDFs on the interval $[0, 1]$, and to introduce independent functions for the quarks and the antiquarks, which we denote $q(x, \mu^2)$ and $\bar{q}(x, \mu^2)$ respectively. The relation between f_q, q and $\bar{q}$ is

$$f_{\mathrm{q}}(x, \mu^2) = \begin{cases} q(x, \mu^2), & \text{if } x > 0, \\ -\bar{q}(-x, \mu^2), & \text{if } x < 0. \end{cases} \qquad (2.4)$$

It is useful to consider the symmetrised and antisymmetrised combinations of f_q in the interval $x \in [0,1]$:

$$f_q^{\text{sym}}(x,\mu^2) \ = f_q(x,\mu^2) + f_q(-x,\mu^2)\,, \tag{2.5}$$

$$f_q^{\text{asy}}(x,\mu^2) \ = f_q(x,\mu^2) - f_q(-x,\mu^2)\,. \tag{2.6}$$

It can be readily shown that

$$f_q^{\text{sym}}(x,\mu^2) = q(x,\mu^2) - \bar{q}(x,\mu^2) = q^-(x,\mu^2)\,, \tag{2.7}$$

$$f_q^{\text{asy}}(x,\mu^2) = q(x,\mu^2) + \bar{q}(x,\mu^2) = q^+(x,\mu^2)\,. \tag{2.8}$$

where q^+ and q^- are defined by the equations above. The flavor decomposition can be rewritten by collecting the quark fields in a vector, $e.g.$ $\psi = (\psi_u, \psi_d, \psi_s, \psi_c)$, and defining the following nonsinglet bare PDFs:

$$f_A^{(0)}(x) \ = \int \frac{d\xi^-}{4\pi} e^{-ixP^+\xi^-} \langle P|\bar{\psi}^{(0)}(\xi^-)\lambda_A \gamma^+ U(\xi^-,0)\psi^{(0)}(0)|P\rangle\,, \tag{2.9}$$

where $A = 3, 8, 15$, and we have used the Gell-Mann matrices

$$\lambda_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \lambda_8 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \lambda_{15} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -3 \end{pmatrix}. \tag{2.10}$$

In this notation f_3 corresponds to $f^{u-d} = f_u - f_d$, $f_8 = f^{u+d-2s}$, and so on. The symmetrised and antisymmetrised combinations map directly into the so-called *evolution basis* for the PDFs that is normally used in phenomenological studies, see $e.g.$ Ref. [56] for a detailed definition of the flavor decomposition. More specifically, we have:

$$f_3^{\text{asy}} = u^+ - d^+ = T_3\,, \tag{2.11}$$

$$f_3^{\text{sym}} = u^- - d^- = V_3\,, \tag{2.12}$$

$$f_8^{\text{asy}} = u^+ + d^+ - 2s^+ = T_8\,, \tag{2.13}$$

$$f_8^{\text{sym}} = u^- + d^- - 2s^- = V_8\,, \tag{2.14}$$

$$f_{15}^{\text{asy}} = u^+ + d^+ + s^+ - 3c^+ = T_{15}\,, \tag{2.15}$$

$$f_{15}^{\text{sym}} = u^- + d^- + s^- - 3c^- = V_{15}\,. \tag{2.16}$$

We will return to these relations below when we discuss the factorization of the quasi-PDFs. The renormalized PDFs $f_a\left(x,\mu^2\right)$ – or equivalently $q(x,\mu^2)$ and $\bar{q}(x,\mu^2)$ – are currently extracted from global fits of experimental data, see $e.g.$ Refs. [57,58] for recent reviews.

Eq. (2.1) defines the PDFs in terms of matrix elements of QCD operators computed between hadron states, which makes their universal and non-perturbative nature explicit. Their direct computation through lattice QCD simulations looks therefore a very appealing target. Unfortunately, the real time dependence of PDFs makes it impossible to compute these matrix elements on the lattice, where the theory is defined on a Euclidean space with imaginary time. To overcome this problem, a new approach was proposed in Ref. [6]: the matrix element appearing in Eq. (2.1), defined along a light-cone direction, is replaced by a correlator defined along a purely spatial direction. The resulting quantity is called a (bare) quasi-PDF. Denoting by Γ a generic Dirac structure and by the suffix A the specific nonsinglet distribution we want to consider, we can introduce the *Ioffe time distributions* [48,59], defined as the matrix element between nucleon states with momentum P

$$M_{\Gamma,A}^{(0)}(\xi,P) = \langle P|\mathcal{M}_{\Gamma,A}^{(0)}(\xi)|P\rangle\,, \tag{2.17}$$

with

$$\mathcal{M}_{\Gamma,A}^{(0)}(\xi) = \bar{\psi}^{(0)}(\xi)\lambda_A \Gamma\, U(\xi,0)\psi^{(0)}(0)\,. \tag{2.18}$$

The vector bilocal operator obtained for $\Gamma = \gamma^\mu$ can be decomposed in terms of two form factors depending on Lorentz invariants:

$$M^{(0)}_{\gamma^\mu,A}(\xi,P) = 2P^\mu \, \mathrm{h}^{(0)}_{\gamma^\mu,A}\left(\xi \cdot P, \xi^2\right) + \xi^\mu \, \tilde{\mathrm{h}}^{(0)}_{\gamma^\mu,A}\left(\xi \cdot P, \xi^2\right).$$

The light-cone PDF in Eq. (2.9) is obtained by taking the Fourier transform with respect to ξ^- of a Ioffe time distribution with $\Gamma = \gamma^+$, $\xi = (0, \xi^-, 0_\perp)$ and $P = (P^+, 0, 0_\perp)$, given by

$$M^{(0)}_{\gamma^+,A}(\xi,P) = \langle \mathrm{P}|\bar\psi^{(0)}\left(\xi^-\right) \lambda_A \gamma^+ U\left(\xi^-,0\right) \psi^{(0)}(0)|\mathrm{P}\rangle = 2P^+ \, \mathrm{h}^{(0)}_{\gamma^+,A}\left(\xi^- P^+, 0\right). \qquad (2.19)$$

Choosing instead a pure spatial direction $\xi = (0,0,0,z)$, and taking the Fourier transform with respect to z, we obtain the definition of the quasi-PDF

$$\tilde{f}^{(0)}_A(x,P_z) = \int_{-\infty}^\infty \frac{dz}{4\pi} \, e^{ix\mathrm{P}_z z} \, M^{(0)}_{\Gamma,A}(z,P). \qquad (2.20)$$

Choosing for example $\Gamma = \gamma^0$ we get

$$\tilde{f}^{(0)}_A(x,P_z) = 2P_0 \int_{-\infty}^\infty \frac{dz}{4\pi} \, e^{ix\mathrm{P}_z z} \, \mathrm{h}^{(0)}_{\Gamma,A}\left(zP_z, z^2\right), \qquad (2.21)$$

which will be the case considered in this work. As in the case of standard PDFs (which from now on we will call *light-cone* PDFs), the matrix elements defining the quasi-PDFs contain UV divergences, and need to be renormalized. The perturbative renormalization of a bilocal operator, as the one appearing in Eq. (2.21), is one of the theoretical problems of this novel approach. It was shown in Ref. [16], that the UV behaviour of the quasi-PDFs is very different from that of PDFs, and that the position space operator appearing in Eq. (2.21) can be multiplicatively renormalized, according to

$$\mathrm{h}_{\Gamma,A}\left(zP_z, z^2, \mu\right) = Z_A(z) \, e^{\delta m|z|} \mathrm{h}^{(0)}_{\Gamma,A}\left(zP_z, z^2\right), \qquad (2.22)$$

where the exponential factor $e^{\delta m|z|}$ reabsorbs the power divergence from the Wilson line, and the position-dependent factor $Z_A(z)$ takes care of the remaining UV logarithmic divergences. Importantly, the quasi-PDFs retain a dynamical dependence on the hadron momentum P, unlike PDFs, which are defined to be invariant under Lorentz boosts. Also, their support is defined to be the full real axis.

The interest in quasi-PDFs comes from the potential to relate them to light-cone PDFs; factorization allows us to rewrite the quasi-PDFs as a convolution of the light-cone PDFs with a coefficient function that can be computed in perturbation theory, up to corrections that are suppressed by inverse powers of P_z. As verified at 1-loop order in Ref. [21], and proved in Refs. [24, 25] using the OPE, renormalized quasi-PDFs share the same IR behaviour as the renormalized light-cone PDFs. It follows that they can be written as

$$\tilde{f}_A\left(x,\mu^2\right) = \int_{-1}^1 \frac{dy}{|y|} \, C_A\left(\frac{x}{y}, \frac{\mu}{|y|P_z}, \frac{\mu}{\mu'}\right) f_A\left(y,\mu'^2\right) + \mathcal{O}\left(\frac{M^2}{P_z^2}, \frac{\Lambda^2_{\mathrm{QCD}}}{P_z^2}\right), \qquad (2.23)$$

where the terms $\mathcal{O}\left(\frac{M^2}{P_z^2}, \frac{\Lambda^2_{\mathrm{QCD}}}{P_z^2}\right)$ include the power corrections suppressed by the hadron momentum. The functions C_A, usually called matching coefficients, depend on the choice of the renormalization scheme, and on the kind of quasi-PDF under consideration. The first matching expressions, for all Dirac structures, were derived in Ref. [21], using a simple transverse momentum cutoff scheme. In later works, matching coefficients were derived that relate the quasi-PDFs in different renormalization schemes to light-cone PDFs in the $\overline{\mathrm{MS}}$ scheme. The matching from $\overline{\mathrm{MS}}$ quasi-PDFs was first considered in Ref. [28], both for non-singlet and singlet quark PDFs, as well as for gluons. Even though one can choose operators for the latter that do not mix with singlet quark quasi-PDFs under renormalization [19, 20], mixing under matching is inevitable. No mixing of the flavour nonsinglet sector with flavour singlet or gluon sectors occurs, as stated in Eq. (2.23). Ref. [28] did not, however, address the known issue of self-energy corrections,

exhibiting a logarithmic UV divergence. This was resolved in Ref. [25] by adding terms outside of the plus prescription in the matching coefficient. As noticed in Ref. [30], such prescription for renormalizing this divergence violates vector current conservation, *i.e.* the integral of the matched PDF is different from the integral of the input quasi-PDF, and not necessarily equal to 1 over the whole integration range. As a remedy, a modified matching expression, which is given explicitly in Eq. (A.1) of App. A, was proposed in Ref. [30]. It consists in resorting to pure plus functions when subtracting the logarithmic divergence in self-energy corrections. However, this is an additional subtraction with respect to the minimal subtraction of the $\overline{\text{MS}}$ scheme and thus, defines a modified $\overline{\text{MS}}$ scheme, the so-called M$\overline{\text{MS}}$ scheme. As such, it requires the quasi-PDF to be expressed in this modified scheme. The expression for the conversion of $\overline{\text{MS}}$-renormalized matrix elements to the M$\overline{\text{MS}}$ scheme was worked out in Ref. [45] and we refer to it for the details of the procedure. Nevertheless, this modification is numerically very small, as also shown in Ref. [45]. An alternative modification of the $\overline{\text{MS}}$ scheme that guarantees vector current conservation was derived in an updated version of Ref. [25]. This defines the so-called "ratio" scheme. In this scheme, only pure plus functions are used, like in the M$\overline{\text{MS}}$ scheme, but the modification is done also for the "physical" region of $0 < \xi < 1$ (in the notation of Eq. (A.1)). Thus, the expected numerical effect of this modification is larger, as shown explicitly in Ref. [45]. For this reason, we choose to use the M$\overline{\text{MS}}$ procedure, with details of the lattice computation of the bare matrix elements and the renormalization in the M$\overline{\text{MS}}$ scheme outlined in the next section. Yet another possibility of performing the matching consists in directly relating the quasi-PDFs in the intermediate RI-type scheme to $\overline{\text{MS}}$ light-cone PDFs. This was proposed in Ref. [29] for the unpolarized case. Obviously, such procedure is equivalent to the one adopted here, with possibly different systematic effects. All of the discussed papers considered the matching to only first order in perturbation theory. It remains an important direction to derive the two-loop formulae, which would allow to estimate the size of perturbative truncation effects in the conversion from the intermediate lattice scheme to $\overline{\text{MS}}$ and in the matching itself.

3 Nonsinglet distributions from quasi-PDFs Matrix Elements

In this section, we describe the lattice data used in this work, presenting briefly the quasi-PDFs matrix elements (MEs) computed in Refs. [30, 45]. Using the results recalled in Sec. 2, we show that we can factorize such matrix elements into two nonsinglet distributions and a perturbatively computable coefficient, just as if they were experimental data for high-energy cross sections.

3.1 Lattice data

The field of nucleon isovector $(u - d)$ quasi-PDFs has matured in recent years. Exploratory studies for all types of collinear PDFs – unpolarized, helicity and transversity – were performed in 2014-2016 [34–37]. They used lattice ensembles with non-physical pion masses and the results had unsubtracted divergences, due to the lack of a well-defined renormalization procedure. The latter was proposed and applied for the first time in Refs. [12, 13], utilizing a variant of the regularization-independent momentum subtraction scheme (RI'-MOM) [60]. Moreover, another major progress for unpolarized PDFs was the identification of a lattice-induced mixing between the bilinear operator used in the first exploratory studies, which was defined using γ_z to determine the Dirac structure, and the scalar bilinear operator (in spin space) [12]. Even though in principle it is possible to compute the matrix elements of the latter and a mixing renormalization matrix to properly subtract the divergences [13], this is bound to lead to much deteriorated precision, due to the rather poor signal for the scalar operator. Instead, it is preferable to define the quark bilinear using the γ_0 Dirac matrix, since this procedure does not give rise to mixing. Moreover the quasi-PDF computed with it converges faster in powers of $1/P_z^2$ to the light-cone PDF, as argued in Ref. [61]. Summarising in just one sentence, we could say that the major progresses with respect to the early works for unpolarized quasi-PDFs came from: (1) change of the Dirac structure in order to avoid the mixing, (2) non-perturbative renormalization procedure, (3) simulations at the physical pion mass. Matrix elements corresponding to such setup were computed in Refs. [30, 41, 45] and they are briefly described below. For a recent review of other available results for quasi-PDF matrix elements, see *e.g.* Ref. [8].

The data used in this work to illustrate the impact of lattice calculations on phenomenological fits were computed by the Extended Twisted Mass Collaboration (ETMC)[1]. They used one ensemble of gauge field configurations with two degenerate light quarks [62] with masses chosen to reproduce the physical value of the pion mass ($m_\pi \approx 130$ MeV, *i.e.* slightly below the actual physical value). The lattice spacing is $a = 0.0938(3)(2)$ fm [63] and the lattice has $48^3 \times 96$ sites, corresponding to the spatial extent L of around 4.5 fm and $m_\pi L = 2.98$. ETMC calculated bare quasi-PDF matrix elements for the unpolarized, helicity and transversity cases, but we concentrate only on the unpolarized one. The lattice data are available for three nucleon boosts, $P_z = 6\pi/L, 8\pi/L$ and $10\pi/L$ (0.83 GeV, 1.11 GeV and 1.38 GeV in physical units) and for four values of the temporal separation between the nucleon creation and annihilation operators, $t_s/a=8, 9, 10, 12$ (0.75, 0.84, 0.94, 1.13 fm). As shown in Refs. [30, 45], there are signs of convergence in the nucleon momentum (the largest two momenta give compatible results), indicating that the boost is already enough to suppress higher-twist effects below statistical precision. Moreover, as pointed out in Ref. [45], excited-states contamination at the largest source-sink separation is small, *i.e.* the single-state fits at this t_s are compatible with two-state fits including all four values of t_s. Hence, for the purpose of this study, we consider only the data at the largest nucleon boost and at the largest source-sink separation. They are shown in Fig. 3.1.
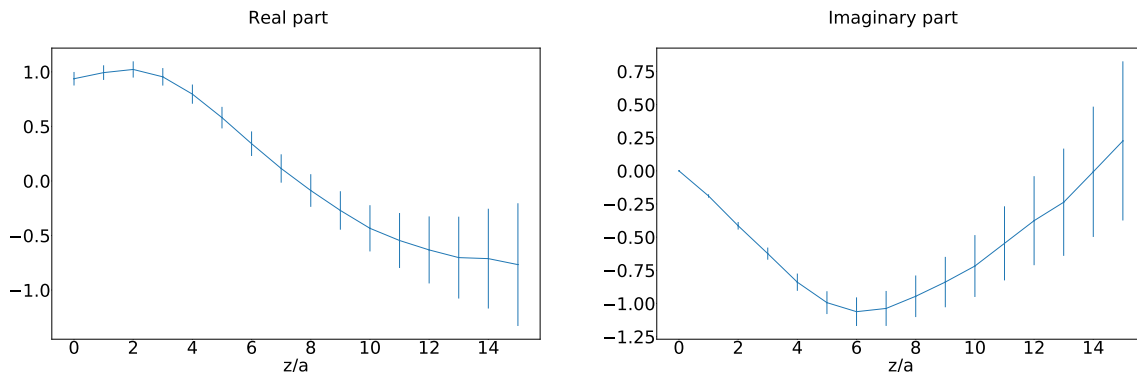

Figure 3.1. Real (left) and imaginary (right) part of the quasi-PDF ME for the data used in this work, computed in Refs. [30, 45]. The error band displayed accounts only for statistical uncertainty. The separation z is expressed in units of the lattice spacing a.

The bare lattice data contain two types of divergences. First of all, there are standard logarithmic divergences with respect to the regulator, *i.e.* terms that behave like $\log(a\mu)$. Additionally, for non-zero Wilson line lengths, further power-like divergences appear. They resum into a multiplicative exponential factor, $e^{\delta m|z|/a}$, where δm is operator-independent. The desired renormalization scheme for the final results is the $\overline{\text{MS}}$ scheme of dimensional regularization. However, obviously, the latter is impossible on a lattice, restricted to integer dimensions. Thus, the usage of an intermediate lattice renormalization scheme is required. In Ref. [13], it was proposed to use an RI'-type prescription. The renormalization conditions are enforced on the amputated vertex functions of operators with different Wilson line lengths z, setting them to their tree-level values. A similar renormalization condition is applied for the quark propagator. This results in a set of matrix elements renormalized in the RI' scheme. Thus, a perturbative conversion from the RI' to the $\overline{\text{MS}}$ scheme is needed. Such a conversion was derived in Ref. [12] to one-loop order and was applied to the RI'-renormalized matrix elements. As we discussed in the previous section, to guarantee vector current conservation, we use a modified $\overline{\text{MS}}$ scheme, termed the M$\overline{\text{MS}}$ scheme. Thus, another perturbative conversion of the $\overline{\text{MS}}$-renormalized matrix elements is required, according to the formula given in Ref. [45]. After this conversion, renormalized matrix elements in the M$\overline{\text{MS}}$ scheme are the starting point of the current analysis.

It is important to emphasize that despite having numerical evidence for the smallness of the effects of the nucleon momentum and of excited states, matrix elements from lattice studies come with a variety of other systematic effects. We discuss them in the next subsection. For more details about the lattice computation of the matrix elements, we refer the reader to Ref. [45].

---

[1]Until 2018 known as the European Twisted Mass Collaboration.

3.2 Systematics in matrix elements of quasi-PDFs

A proper investigation of systematic effects in matrix elements evaluated in lattice QCD simulations is a difficult task, necessitating dedicated efforts. Such efforts consist in simulating with varied parameter values, such as the lattice spacing, the lattice volume, or the temporal separation between the source and the sink in nucleon three-point functions. Moreover, unrelated to the lattice regularization, there are theoretical uncertainties intrinsic to the quasi-distribution approach whose impact should also be assessed [2]. For an extensive review of these different uncertainties, we refer to Refs. [8, 9], while a discussion of the systematic effects in the ETMC quasi-PDFs computation can be found in Ref. [45]. The latter contains the analysis of the effects investigated so far and a discussion of directions that need to be pursued to fully quantify all the relevant systematics.

Here, we briefly summarize the conclusions reached up to the present stage. It is important to emphasize that while the impact of some systematics is already known to a reasonable degree, reliable estimates of certain types of effects are still largely unknown. Nevertheless, rough assessments can be made even in the case of missing lattice data, by looking at the behaviour of related observables such as the average quark momentum fraction or nucleon charges that have a long history of evaluations on the lattice [66–70]. This allows us to build scenarios describing the potential impact of the systematics on the matrix elements of quasi-PDFs. We consider three scenarios where the systematic effect is a given percentage of the central value of the matrix element and three further ones where it is a given additive shift. We always exclude from the analysis the imaginary part of the matrix element at $z = 0$, equal to 0 by antisymmetry with respect to the sign change of z.

Cut-off effects. One of the most obvious systematic effects in lattice computations comes from the finite value of the lattice spacing, a, *i.e.* the ultraviolet cut-off imposed for the regularization of the theory. While a proper investigation of this uncertainty requires explicit simulations at a few values of the lattice spacing, which are still missing for quasi-PDFs, we may assume that discretization effects are not excessive. This expectation is based on two indirect, but related premises. First, one of the manifestations of large cut-off effects is the violation of the continuum relativistic dispersion relation, which is, however, not observed in the lattice data in Ref. [45]. Second, the first moment of the unpolarized $u - d$ PDF gives the quark momentum fraction $\langle x \rangle_{u-d}$. This quantity was intensively investigated on the lattice and we may take the typical size of discretization effects found in such studies. Looking at a summary plot including data from different lattice groups, such as Fig. 12 from Ref. [67], we see that cut-off effects at lattice spacings comparable to the one of the present work are typically at the 5-15% level in a fixed lattice setup (same discretization, pion mass, volume etc.). Thus, we investigate 6 plausible choices for the magnitude of cut-off effects: 10%, 20%, 30% of the matrix element and additive effects of 0.1, 0.2 and 0.3.

Finite volume effects (FVE). Another natural source of uncertainty in all lattice simulations is the finite size of the box, L, which acts as an infrared regularization. Similarly to discretization effects, a robust investigation of these effects necessitates running the computations for a few values of the lattice size. However, the difference with respect to the lattice spacing effects, typically linear in a or a^2 in the asymptotic scaling regime, is that FVE are usually suppressed as $\exp(-m_\pi L)$, where m_π is the pion mass. This leads to typically $\mathcal{O}(1-5\%)$ effects in hadron structure observables if $m_\pi L \geq 3$. For the matrix elements used in this work, $m_\pi L \approx 3$ – thus, the reasonable assumption about the size of FVE is approx. 5%. In addition to these "standard" FVE of lattice computations, it has been recently pointed out that the usage of a spatially extended operator, including a Wilson line, may lead to additional FVE [71]. The intuition behind this is that further FVE may appear when the Wilson line has non-negligible size with respect to the lattice length in the boost direction. The analysis of Ref. [71] pertains to a toy scalar theory and predicts a FVE of the form $\exp(-M(L-z))$ (possibly with a polynomial amplifying prefactor), with M being the analogue of the mass of the investigated hadron in the quasi-PDF approach. Given that the nucleon mass is at the physical point around 7 times larger than the pion mass, that would lead to totally irrelevant effects, since the maximum considered z is more than 3 times smaller than L. However, it can not be excluded that in QCD, the form of

---

[2]Note that theoretical uncertainties can be included in global fits of PDFs as detailed in Ref. [64, 65]

this FVE can be more severe, e.g. $\exp(-m_\pi(L-z))$. With the physical m_π and $z_{\max} \approx L/3$, this could lead to the amplification of FVE from $\mathcal{O}(5\%)$ to even above 10% at large z. We remark that ETMC has investigated FVE in the renormalization functions for the matrix elements and found no sign of excessive FVE coming at large z (total FVE not larger than around 3%) [45]. We investigate 3 scenarios for fixed percentage effects: constant FVE of 2.5% and 5%, as well as z-dependent ones of the form $\exp(-3 + 0.062z/a)\%$, where 0.062 is the pion mass value for the present ensemble, expressed in lattice units. In addition, we consider 3 shifts: 0.025, 0.05 and $\exp(-3 + 0.062z/a)$.

Excited states contamination. One of the key uncertainties in nucleon structure calculations is whether the ground state hadron state is isolated. If the temporal separation between the interpolating operators creating the nucleon and annihilating it is too small, uncontrolled excited states contamination may appear, leading to a bias in the results. In the context of quasi-PDFs, an important aspect is that this contamination strongly depends on the boost, causing a delicate interplay between the need of large momentum, required for robust matching to light-cone distributions, and excited states contamination, larger for high boost. Thus, a careful analysis is needed to ensure ground state dominance. Such an analysis was performed for the matrix elements used in this work [45]. The conclusion that we use for the present case is that these matrix elements are safe against excited states effects at the level of their statistical precision. In this way, we choose three values of uncertainty from excited states: 5%, 10% and 15%. When the renormalized matrix elements are close to zero, the relative uncertainty can be larger and thus, we consider also three additive scenarios with magnitude 0.05, 0.1 and 0.15.

Truncation effects. The perturbative ingredients of the quasi-PDF approach are of two kinds. One of them is related to the fact that the lattice approach works in integer dimensions and thus, dimensional regularization of the $\overline{\text{MS}}$ scheme is impossible. Instead, as discussed above, a non-perturbative renormalization programme has been proposed by ETMC [13], utilizing a variant of the regularization-independent momentum subtraction scheme (RI'-MOM). The renormalization correlators obtianed in this way can then be translated perturbatively to the $\overline{\text{MS}}$ scheme and finally to the M$\overline{\text{MS}}$ scheme, using formulae derived in Refs. [12, 45]. These formulae are currently available to the one-loop level and thus subject to a truncation effect from unknown higher orders. A manifestation of this effect is the fact that the Z-factors have a non-vanishing imaginary part even after conversion to $\overline{\text{MS}}$, where they should be purely real. To evaluate the impact of this uncertainty, we compare the renormalized matrix elements with the ones obtained from applying only the real parts of the Z-factors. We find that the matrix elements obtained by this alternative procedure are compatible with the actual ones within statistical uncertainties, with relatively larger effects observed for small z/a in the imaginary part (up to $\mathcal{O}(5\%)$) and intermediate z/a in the real part (the real part is small there – thus, the observed absolute effects of around 0.2 can be a large percentage of the value). Apart from the scheme conversion truncation effects, the necessary perturbative ingredient of the approach is the matching between quasi-PDFs and light-cone distributions, also known to one loop [21, 25, 45]. Without knowing the two-loop formulae, it is difficult to estimate their size. We emphasize that, thus, the extension of the one-loop conversion, evolution and matching formulae to (at least) two loops is mandatory to properly estimate truncation effects. With two-loop equations at hand, one can e.g. apply standard methods of estimating truncation effects through scale variations. For now, we observe that comparing the quasi-distribution and the resulting light-cone PDF, the numerical magnitude of the matching factor can be significant and thus, the higher order effects may be sizable. The "natural" size of such truncation effects is of $\mathcal{O}(\alpha_s^2)$, which amounts to around 10% at the renormalization scale we consider. However, they are rather uncertainties of the procedure, so they can not be translated to uncertainties of the matrix elements. These uncertainties are the analogue of the theoretical uncertainties that come from a truncated perturbative expansion in the description of observables in phenomenological fits of the PDFs, and can be treated as mentioned in the footnote above. Finally, we consider 6 scenarios for truncation effects pertinent to matrix elements (*i.e.* originating from the perturbative uncertainty in Z-factors): 10%, 20%, 30% of the central value of the matrix element, as well as shifts of 0.1, 0.2 and 0.3.

Higher twist effects. For the current analysis we decide to ignore the effect of higher twists, *i.e.* the presence of power-like corrections to the factorization formula. At this preliminary stage,

we are not concerned by their effects, but a more precise phenomenological analysis should definitely take those into account. In particular, it should be kept in mind that, as argued in Refs. [25, 72], power corrections can be enhanced for both $x \to 0$ and $x \to 1$, limiting the quasi-PDFs approach in the small-x and large-x regime. We will come back to this point in the conclusions.

Other effects. Apart from the systematics mentioned above, there are some other effects that potentially affect the results. One of them is the usage of a setup including two degenerate light quarks. However, this effect, working in the isospin limit instead of taking into account the different masses and electric charges of the light quarks, is expected to be much below the level of the current precision – of the order of the proton-neutron mass splitting, *i.e.* at the per mille level. A similar magnitude can be expected for the contribution of the neglected sea quark loops from heavier quarks. Such effects can at present be safely ignored and will become important only when aiming at an $\mathcal{O}(1\%)$ precision or better.

Scenario	Cut-off	FVE	Excited states	Truncation
S1	10%	2.5%	5%	10%
S2	20%	5%	10%	20%
S3	30%	$e^{-3+0.062z/a}\%$	15%	30%
S4	0.1	0.025	0.05	0.1
S5	0.2	0.05	0.1	0.2
S6	0.3	$e^{-3+0.062z/a}$	0.15	0.3

Table 3.1. Scenarios of the impact of different systematic effects in the renormalized matrix elements of quasi-PDFs. Percentage values for scenarios S1-S3 should be understood as a given fraction of the central value of the matrix element, while absolute values for S4-S6 are shifts independent from the matrix element.

Final scenarios. In the end, we define 6 scenarios of possible impact of systematic effects, summarized in Tab. 3.1. Scenarios S1-S3 include uncertainties that are a fixed percentage of the central value of the matrix element, while for S4-S6, the uncertainties are additive shifts independent from the value of the matrix element. Scenarios S1, S4 can be considered as the most "optimistic" ones. More realistic estimates of uncertainties are included in S2 and S5. Finally, S3 and S6 are "pessimistic", *i.e.* assume largest plausible estimates of the various systematic effects.

3.3 From parton distributions to lattice observables

In this work, we aim at describing the data presented in the previous subsection; further studies with more data and a more detailed treatment of systematic errors are deferred to future publications. Hence, we specialize our discussion to the case of the unpolarized isovector parton distribution. Following the notations introduced above, the parton distribution f_3 is defined as

$$f_3 \left(x, \mu^2\right) = \begin{cases} u\left(x, \mu^2\right) - d\left(x, \mu^2\right) , & \text{if } x > 0 \\ -\bar{u}\left(-x, \mu^2\right) + \bar{d}\left(-x, \mu^2\right) , & \text{if } x < 0 \end{cases} \tag{3.1}$$

where the support is given by $x \in [-1, 1]$. The factorization theorem in Eq. (2.23) becomes

$$\tilde{f}_3 \left(x, \mu, P_z\right) = \int_{-1}^{+1} \frac{dy}{|y|} C_3 \left(\frac{x}{y}, \frac{\mu}{|y|P_z}\right) f_3 \left(y, \mu^2\right) , \tag{3.2}$$

where the quasi-PDF is the one given by $\Gamma = \gamma^0$ and the explicit expression of the matching coefficients is given in App. A. Starting from the definition of quasi-PDFs given in Eq. (2.21), it is clear that the lattice ME is given by the inverse Fourier transform of Eq. (3.2), which yields an equation relating the light-cone PDFs on the right hand side to the lattice observable:

$$h_{\gamma^0, 3} \left(zP_z, z^2, \mu\right) = \int_{-\infty}^{\infty} dx \, e^{-i(xP_z)z} \int_{-1}^{+1} \frac{dy}{|y|} C_3 \left(\frac{x}{y}, \frac{\mu}{|y|P_z}\right) f_3 \left(y, \mu^2\right) . \tag{3.3}$$

Since C_3 is purely real, we can split the above complex identity into two real equations, relating the real and imaginary part of the ME $\mathrm{h}_{\gamma^0,3}(z)$ to the light-cone distribution f_3. For the purpose of this work, we introduce two lattice observables, denoted by $\mathcal{O}_{\gamma^0}^{\mathrm{Re}}(z,\mu)$ and $\mathcal{O}_{\gamma^0}^{\mathrm{Im}}(z,\mu)$, defined as

$$\mathcal{O}_{\gamma^0}^{\mathrm{Re}}(z,\mu) \equiv \mathrm{Re}\left[\mathrm{h}_{\gamma^0,3}\left(zP_z,z^2,\mu^2\right)\right] = \int_{-\infty}^{\infty} dx\, \cos\left(xP_z z\right) \int_{-1}^{+1} \frac{dy}{|y|} C_3\left(\frac{x}{y}, \frac{\mu}{|y|P_z}\right) f_3\left(y,\mu^2\right),$$
(3.4)

$$\mathcal{O}_{\gamma^0}^{\mathrm{Im}}(z,\mu) \equiv \mathrm{Im}\left[\mathrm{h}_{\gamma^0,3}\left(zP_z,z^2,\mu^2\right)\right] = -\int_{-\infty}^{\infty} dx\, \sin\left(xP_z z\right) \int_{-1}^{+1} \frac{dy}{|y|} C_3\left(\frac{x}{y}, \frac{\mu}{|y|P_z}\right) f_3\left(y,\mu^2\right),$$
(3.5)

where we have only included z and μ in the arguments of $\mathcal{O}_{\gamma^0}^{\mathrm{Re}}$ and $\mathcal{O}_{\gamma^0}^{\mathrm{Im}}$ in order to simplify the notation – since we are working here with only one value of P_z there is little advantage in keeping all the arguments. The explicit expression of C_3 contains plus distributions. Making them explicit we can write the equations above as

$$\mathcal{O}_{\gamma^0}^{\mathrm{Re}}(z,\mu) = \int_0^1 dx\, \mathcal{C}_3^{\mathrm{Re}}\left(x,z,\frac{\mu}{P_z}\right) V_3(x,\mu) = \mathcal{C}_3^{\mathrm{Re}}\left(z,\frac{\mu}{P_z}\right) \circledast V_3\left(\mu^2\right),$$
(3.6)

$$\mathcal{O}_{\gamma^0}^{\mathrm{Im}}(z,\mu) = \int_0^1 dx\, \mathcal{C}_3^{\mathrm{Im}}\left(x,z,\frac{\mu}{P_z}\right) T_3(x,\mu) = \mathcal{C}_3^{\mathrm{Im}}\left(z,\frac{\mu}{P_z}\right) \circledast T_3\left(\mu^2\right)$$
(3.7)

where V_3 and T_3 are the nonsinglet distributions defined by

$$V_3(x) = u(x) - \bar{u}(x) - \left[d(x) - \bar{d}(x)\right],$$
(3.8)
$$T_3(x) = u(x) + \bar{u}(x) - \left[d(x) + \bar{d}(x)\right],$$
(3.9)

where, for simplicity, the μ dependence has been omitted. The equations above relate the position space matrix elements computable on the lattice with the collinear PDFs. Similar expressions were worked out in Ref. [73] in the context of the pion distribution amplitude. The proof of Eqs. (3.6), (3.7) does require some care, and it is fully worked out in App. A. The coefficients $\mathcal{C}_3^{\mathrm{Re,Im}}$ are related to the real and imaginary part of the Fourier transform of the matching coefficient C_3 appearing in Eq. (3.2). Since the latter is defined in terms of plus distributions, the computation is quite involved, and the explicit expression of $\mathcal{C}_3^{\mathrm{Re,Im}}$ is obtained in App. A, making the action of the plus distributions explicit before taking the Fourier transform. A discussion about the convergence of the integrals involved is also reported there. The above results show how fixed z matrix elements defining the quasi-PDF in position space give direct access to two independent nonsinglet distributions, through the integration of the parton distribution over its full support with a perturbatively computable coefficient. We will denote this operation as $\circledast$.

It is useful at this point to recall the form of the QCD factorization formula for the DIS nonsinglet structure function. Considering $F_2^{\mathrm{NS}}\left(x,Q^2\right) \equiv F_2^p\left(x,Q^2\right) - F_2^d\left(x,Q^2\right)$, where F_2^p and F_2^d are the structure functions of the proton and of the deuteron respectively, we have

$$F_2^{\mathrm{NS}}\left(x,Q^2\right) = \int_x^1 \frac{dy}{y} \mathcal{C}_3^{\mathrm{DIS}}\left(\frac{x}{y}, \frac{Q^2}{\mu^2}, \alpha_s\right) T_3\left(y,\mu^2\right) = \mathcal{C}_3^{\mathrm{DIS}}\left(\frac{Q^2}{\mu^2}, \alpha_s\right) \otimes T_3\left(\mu^2\right)$$
(3.10)

Comparing Eqs. (3.6), (3.7) with Eq. (3.10), we see that the lattice observables introduced above can be treated on the same footing as experimental data for DIS structure functions, as they are related to the nonsinglet distributions through a convolution with a coefficient that can be computed in perturbation theory. However, the form of such convolution, denoted by $\circledast$, is quite different from the one appearing in the DIS case, denoted by $\otimes$: the former involves a DIS-like convolution first, to go from the PDFs to quasi-PDFs, followed by an integration over the full x-range to go to position space. This suggests that this kind of convolution, if implemented in a PDFs fit, may constrain the output much more than what the standard DIS convolution can do.

11

4 Neural network fits

In this section, we set up a neural net fit based on the results presented in Sec. 3. The implementation is done within the NNPDF framework. We begin by briefly recalling the main features of such fits, focusing on the parametrization of the parton distributions, the minimization and cross-validation algorithms and the Monte Carlo replicas approach, referring to the specific NNPDF publications for more details. We then describe in detail the implementation of Eqs. (3.6), (3.7), and in particular the construction of FastKernel tables for the lattice observables. Once the observables can be computed using a FastKernel table, their inclusion in a PDF fit within the NNPDF environment becomes straightforward.

4.1 Fitting strategy and general settings

Given an ensemble of data whose values can be computed as a convolution of a perturbative Wilson coefficient and the PDF using some factorization theorem, the PDFs can be extracted from a minimum-χ^2 fit to these data. It is worthwhile to stress once again that in this respect there is no difference between the lattice data and experimental data. In order to define the χ^2 a parametrization of the functional form is required; the minimization is then performed as a function of the parameters that define such functional form. In our case, and more generally in the NNPDF approach, the functional form of the PDFs is given by a single-layer neural network, and the parameters are the weights and the biases of the neural network [74]. The parton distributions independentely parametrized are the gluon and the singlet distribution, (g, Σ), which mix under evolution, and the nonsinglet distributions given by $(V, T_3, V_3, T_8, V_8, T_{15})$, whose definition is given for example in Ref. [75]. As discussed in the previous sections, the lattice data we consider here give access only to T_3 and V_3. Since the mixing with other flavors does not occur neither in the matching nor in the DGLAP evolution, our fits yield results for these two distributions only. The parametrization of their x-dependence at the fitting scale $\mu_0 = 1.65$ GeV is

$$V_3\left(x, \mu_0^2\right) \propto x^{\alpha_V} (1 - x)^{\beta_V} \, \mathrm{NN}_V\left(x\right) , \tag{4.1}$$

$$T_3\left(x, \mu_0^2\right) \propto x^{\alpha_T} (1 - x)^{\beta_T} \, \mathrm{NN}_T\left(x\right) , \tag{4.2}$$

where NN denotes a neural net, and the parameters α and β are additional free parameters, which will be fitted during the training. As extensively discussed in several NNPDF publications, see for example Refs. [75, 76], the function multiplying the neural net improves the stability and the speed of the minimization procedure, without introducing a bias in the result.

Having chosen a parametrization for the PDFs, the optimal fit is determined by varying its free parameters in order to minimize some figure of merit, representing the fit quality. This is given by the χ^2 function, defined as

$$\chi^2\left(\theta\right) = \frac{1}{N_{\mathrm{dat}}} \sum_{i,j} \left(\mathcal{O}\left(z_i\right) - \mathcal{O}^{\mathrm{th}}\left(z_i, \theta\right)\right) \left[\mathrm{Cov}^{-1}\right]_{\mathrm{ij}} \left(\mathcal{O}\left(z_j\right) - \mathcal{O}^{\mathrm{th}}\left(z_j, \theta\right)\right) , \tag{4.3}$$

where $\mathcal{O}\left(z_i\right)$ denotes the measured lattice observable and $\mathcal{O}^{\mathrm{th}}\left(z_i, \theta\right)$ is the corresponding theoretical prediction, in terms of the matching coefficient and the parametrized parton distribution, given in Eqs. (3.6), (3.7). We denote by θ the set of free parameters entering the neural net and the preprocessing term. The covariance matrix entering the χ^2 definition is defined to take into account the distribution of the experimental data entering the fit and their correlations. It is used for both the χ^2 definition and the generation of Monte Carlo replicas, as described below. In the specific case of lattice data, it has to be constructed using the information about the statistical and systematic uncertainties that arise in a lattice QCD simulation, following the standard NNPDF proceedure for the treatment of correlated uncertainties, see for example Refs. [75, 76]. The explicit expression used here reads

$$\mathrm{Cov}_{ij} = \sigma_{i,s}^2 \, \delta_{ij} + \sum_k \sigma_{i,k} \sigma_{j,k}$$

where i and j run over the lattice points and $\sigma_{i,s}$, $\sigma_{i,k}$ are respectively the total statistical and the set of systematical uncertainties of the i-th lattice point. The different sources of systematics are described in Sec. 3.2, and the scenarios we consider here are summarized in Tab. 3.1. The total uncertainty for a single lattice point can be obtained combining in quadrature all the different uncertainties.

The minimization of the χ^2 is performed through the CMA algorithm [77, 78], employing a cross-validation technique to avoid overfitting. In this method, the available data are split in two sets. The first, the training set, is used for the minimization of the error function, while the second, the validation set, does not enter the fitting proceedure. At each iteration of the minimization algorithm, the error function between the theory predictions from the neural net and the data is computed for both the training and validation set. At an early stage of the training, both these quantities are expected to decrease. However, towards the end of the training, while the error function over the training set will keep decreasing, the same value computed over the validation data will reach a minimal value, and eventually it may even start increasing. This is a signal of overfitting, and the point in parameter space yielding the minimal value of the validation error is the one taken as the fit result. It is important to notice that such procedure is even more important when a small amount of data is available, like in the present case, since the less data there are the easier it is to fit their statistical noise.

In the `NNPDF` framework a Monte Carlo approach is implemented in order to get a reliable estimate of the PDF error. In this method an ensemble of $N_{\rm rep}$ artificial data is generated for each experimental point, assuming a multigaussian distribution given by the experimental covariance matrix. $N_{\rm rep}$ independent fits are performed, generating a Monte Carlo ensemble of PDFs that faithfully reproduces the statistical features of the original experimental (or lattice) data. The Monte Carlo method therefore propagates the error from the data to the PDFs set in a natural way, without the need of any assumption on the way error propagation happens.

4.2 FastKernel implementation

One of the primary issues in PDFs fits is the computational time required to obtain the theoretical prediction for the experimental data entering the definition of the χ^2 in Eq. (4.3): the parton distributions have to be evolved from the fitting scale up to the observable scale, and then they have to be convoluted with the correct coefficient function. As seen in Sec. 3.3, in the case of lattice observables the integration of the parton distributions over their full support is needed. This makes the form of the convolution $\circledast$ more complicated than the one we usually have for other observables. Therefore, despite the fact that in this work we will be using a small number of lattice data, we find it useful to implement Eqs. (3.6), (3.7) by means of FastKernel tables, introduced and validated in Refs. [76, 79] in the context of global QCD fits, and currently used within the `NNPDF` code to obtain all the required theoretical predictions in a global fit. From a practical point of view, this also allows us to treat the lattice data on exactly the same footing as the experimental ones, allowing a smooth and natural way to introduce them in a parton distributions fit. In the following we briefly revise this approach, referring to the original publication quoted above for more details.

The lattice observables $\mathcal{O}_{\gamma^0}^{\rm Re,Im}\left(z,\mu^2\right)$ are determined at a given renormalization scale μ^2. They can be written in terms of the nonsinglet distributions at a given reference scale μ_0^2, by first evolving the parton distribution up to the scale μ^2, and then convoluting it with the coefficents $\mathcal{C}_3^{\rm Re,Im}$ defined in Eqs. (3.6), (3.7) and worked out in App. A. For the nonsinglet distributions considered in this work, the evolution is given by

$$T_3\left(x,\mu^2\right) = \int_x^1 \frac{dy}{y}\, {\rm K}_3^{(+)}\left(\frac{x}{y},\alpha_s,\alpha_s^0\right) T_3\left(y,\mu_0^2\right)\,, \tag{4.4}$$

$$V_3\left(x,\mu^2\right) = \int_x^1 \frac{dy}{y}\, {\rm K}_3^{(-)}\left(\frac{x}{y},\alpha_s,\alpha_s^0\right) V_3\left(y,\mu_0^2\right)\,. \tag{4.5}$$

where the kernels ${\rm K}^{(\pm)}$ are obtained by solving the DGLAP evolution equations in the nonsinglet sector. For V_3 and T_3 we have two different nonsinglet evolution kernels, denoted by ${\rm K}_3^{(-)}$ and ${\rm K}_3^{(+)}$ respectively. They are currently available from a number of public codes, however one way

they can be worked out and implemented is summarised in the App. B. Eqs. (4.4) and (4.5) can be rewritten expressing the parton distribution in terms of an interpolation basis [76], for instance for the case of T_3

$$T_3\left(x,\mu_0^2\right) = \sum_\beta T_3\left(x_\beta,\mu_0^2\right)\mathcal{I}^{(\beta)}\left(x\right) + \mathcal{O}\left[\left(x_{\beta+1}-x_\beta\right)^p\right], \tag{4.6}$$

where p is the lowest order neglected in the interpolation. In other words, the interpolating functions act by picking up the value of the PDF at some point x_β of a predefined x-grid. Substituting in the evolution equation Eq. (4.4) we get

$$T_3\left(x_\alpha,\mu^2\right) = \sum_\beta \mathcal{K}_{\alpha\beta}^{(+)} T_3\left(x_\beta,\mu_0^2\right). \tag{4.7}$$

with

$$\mathcal{K}_{\alpha\beta}^{(+)} = \int_{x_\alpha}^1 \frac{dy}{y}\, \mathrm{K}^{(+)}\left(\frac{x_\alpha}{y},\alpha_s,\alpha_s^0\right)\mathcal{I}^{(\beta)}\left(y\right). \tag{4.8}$$

The interpolation basis used at the initial scale can also be used to interpolate the parton distributions at the scale μ^2 appearing in Eqs. (3.6), (3.7). For the imaginary part of the lattice observable we get

$$\mathcal{O}_{\gamma^0}^{\mathrm{Im}}\left(z,\mu\right) = \sum_\alpha C_{z\alpha}^{\mathrm{Im}} T_3\left(x_\alpha,\mu^2\right), \tag{4.9}$$

with

$$C_{z\alpha}^{\mathrm{Im}} = \int_0^1 dx\, \mathcal{C}_3^{\mathrm{Im}}\left(x,z,\frac{\mu}{P_z}\right)\mathcal{I}^{(\alpha)}\left(x\right). \tag{4.10}$$

Putting together Eqs. (4.7) and (4.9) we get

$$\mathcal{O}_{\gamma^0}^{\mathrm{Im}}\left(z,\mu\right) = \sum_\beta \mathcal{H}_{z\beta}^{\mathrm{Im}} T_3\left(x_\beta,\mu_0^2\right), \tag{4.11}$$

where

$$\mathcal{H}_{z\beta}^{\mathrm{Im}} = \sum_\alpha C_{z\alpha}^{\mathrm{Im}} \mathcal{K}_{\alpha\beta}^{(+)}. \tag{4.12}$$

Eq. (4.12) defines the FastKernel table which enters the computation of the χ^2 during the fit. It connects the parton distribution at the fitting scale to the lattice observable, taking into account the QCD evolution, the matching and the Fourier trasnform, expressing them through a single matrix vector multiplication. Clearly a similar set of equations defines a FastKernel table that yields the real part of the lattice observable, $\mathcal{O}_{\gamma^0}^{\mathrm{Re}}$, as a function of the valence parton distribution V_3.

5 Results

Let us now proceed to presenting and discussing our numerical results. First, we study the way the available lattice data might constrain the parton distributions in a fit, by mean of closure tests: fake data for the real and imaginary part of the ME are generated according to Eqs. (3.6), (3.7) using as input a chosen PDFs set. The fitting code is then run over these pseudo-data, using exactly the same setting used in a common fit. By comparing the output of such fits with the known input PDFs sets, we can assess the accuracy we may expect to get from the current knowledge of the lattice data and their systematics.

Then we present results for fits run over the data presented in Sec. 3.1, studying the 6 different scenarios for the treatment of the systematic errors described in Sec. 3.2 and summarized in Tab. 3.1. The results presented here have been produced using the NNPDF fitting code [1] and the ReportEngine software [80].

5.1 Closure tests

As shown in Sec. 3.3, we can relate PDFs to lattice observables through the matching convolution of Eq. (3.2) followed by a Fourier transform. As already pointed out at the end of Sec. 3.3, the resulting convolution ⊛ is quite different from the one entering standard QCD fits. In this section, we assess how much this operation together with the available lattice data from Refs. [30,45] are able to constrain parton distributions in a fit, running some preliminary closure tests. For a detailed description of the closure test procedure, we refer to Sec. 4 of Ref. [81]. We generate pseudo-data corresponding to the data of Ref. [30] using NNPDF31_nlo_as_0118 as our input PDFs set, and we run the fitting code over them. The outcome of the closure test fit is then used to assess how well the input PDFs can be reconstructed starting from the 16 position space ME points and their uncertainties.

In order to get an idea of the impact on the fit of the statistical and systematic ME errors, we consider three different scenarios: first we generate fake data assuming no systematic uncertainties and a small uncorrelated statistical uncertainty for each point, constant for all of them and of the order of the smallest real one. From the results of this closure test we can estimate the real constraining power of the convolution ⊛, assuming an ideal scenario where all the systematics are under control and the statistical error is kept small. Second, we repeat the exercise but using the real statistical uncertainties, to assess how much the real statistics of the current simulations affect the conclusions of the previous case. Finally we look at the effect of the systematics, considering as a specific example the scenario S2 of Table 3.1. The three cases are summarized in Table 5.1 and the results are shown in Figs. 5.1, 5.2, 5.3.

Closure test	Statistics	Systematics
CT1	fake	-
CT2	real	-
CT3	real	S2

Table 5.1. Closure tests with different choices of the statistical and systematic error. The results for each option above is summarised in the plots below.

Looking at the results for CT1, Fig. 5.1, it is worth stressing that the lattice data entering the fit are just 16 for the real part and 15 for the imaginary part of the matrix element. Just half of them are actually used in the training procedure, while the other ones are used to build the validation set. In a standard NLO QCD global fit, like the one used as input PDF here, the number of points entering the analysis is $\mathcal{O}(4000)$. Fig. 5.1 shows how good the convolution ⊛ is in constraining the PDFs, assuming an ideal scenario where all the systematics are under control, and the statistics are kept small. Looking at the results for CT2 and CT3 in Figs. 5.2 and 5.3, it is clear how big the impact of the statistical and systematic uncertainties of the ME is on the PDFs error: in both cases the input PDFs set is reconstructed within 1-sigma level, with some tension for V_3 at medium x in the first case. The PDFs error is increasingly big, becoming huge when the full systematics are considered. Fig. 5.3 shows what we may expect in a real life scenario.

To sum up, the results from CT1 show how promising this kind of lattice data might be in constraining PDFs. On the other hand, the results from CT2 and CT3 highlight the importance of having a good control over both the statistical and systematical uncertainties in the lattice simulations of the ME. It is worth noticing, however, that the overall error band of the reconstructed PDFs, even in presence of the full systematic errors, would surely be reduced when new data are available.

5.2 Fit results

In this section, we present our results for fits ran over the data from Refs. [30,45], described in Sec. 3.1. As mentioned before, we consider 6 different scenarios for the treatment of the systematic errors, summarized in Table 3.1. We show results for "optimistic" (S1,S4), "realistic" (S2,S5) and "pessimistic" (S3,S6) scenarios, the difference between the elements of each couple
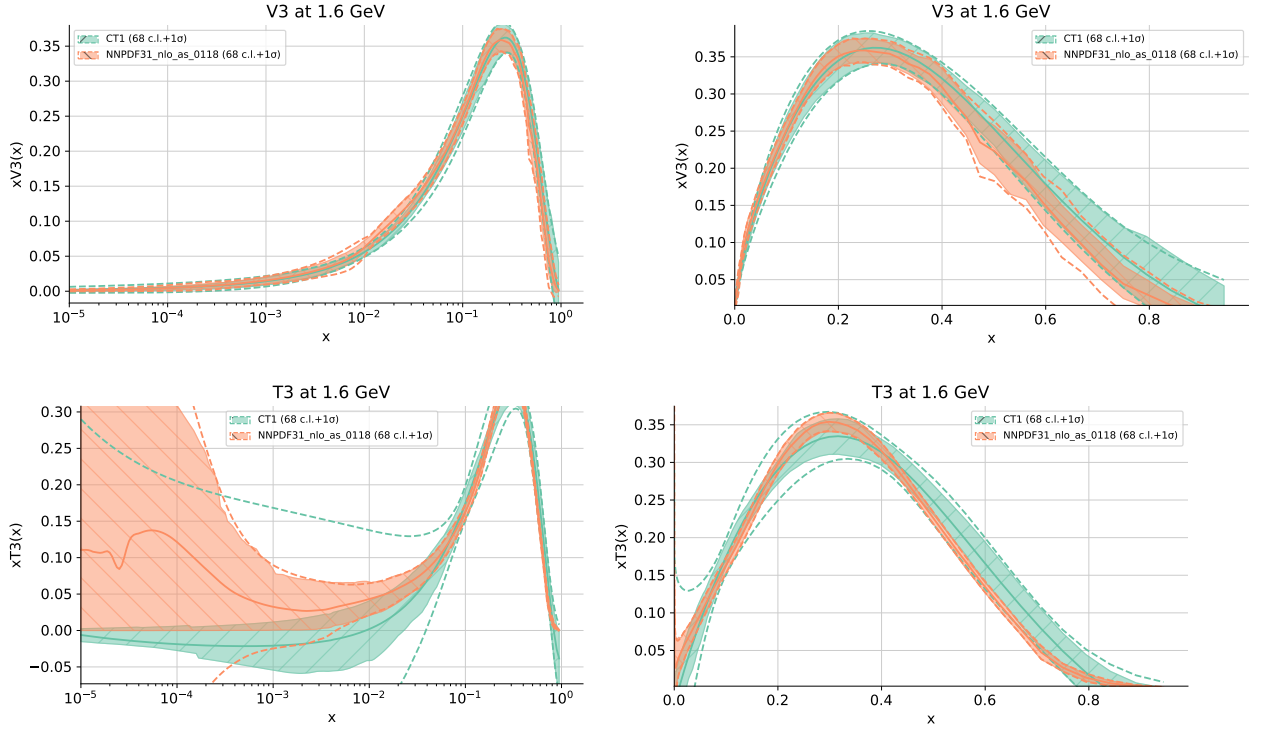
Figure 5.1. Closure test fit with fixed small statistical error and no systematics (CT1) compared to the input PDFs set. V_3 (top line) and T_3 (lower line) combinations in linear and logarithmic scale are shown. The input PDFs set is fully reconstructed within 1-sigma level, getting PDFs with an error band comparable to the input one.
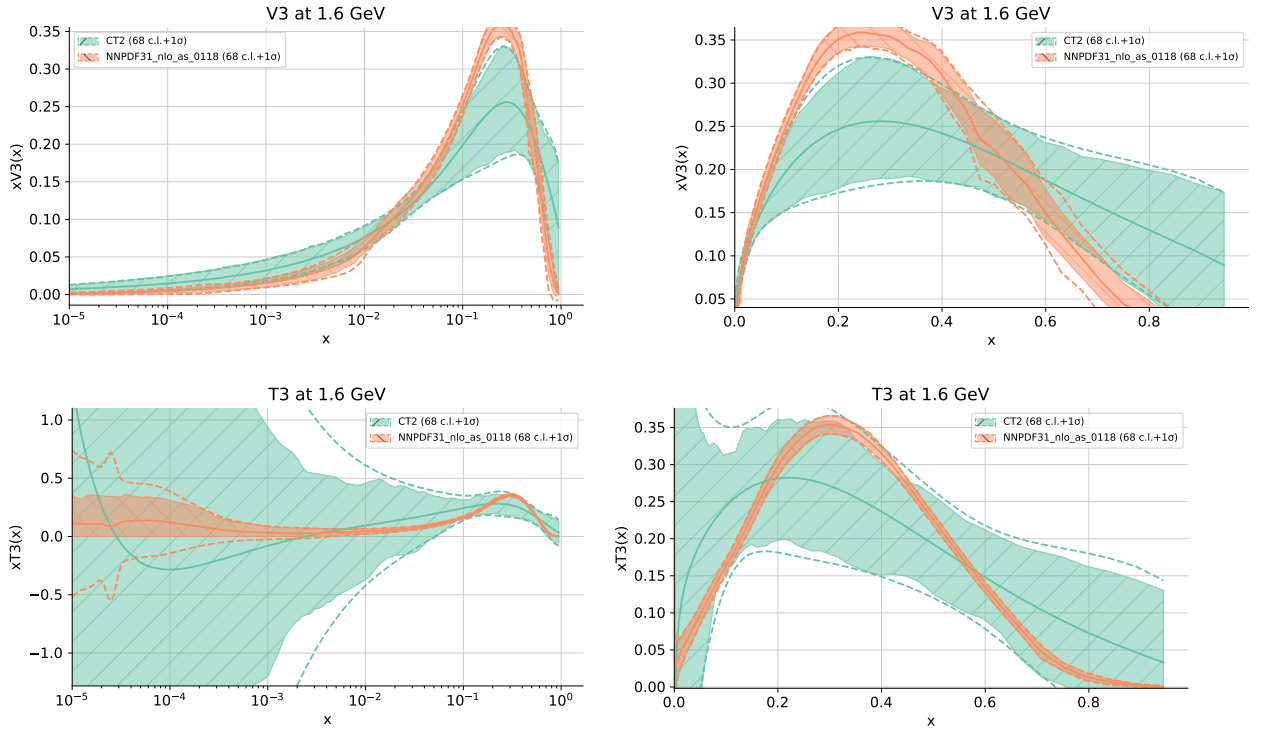

Figure 5.2. Closure test fit with real statistical error and no systematics (CT2) compared to the input PDFs set. Top line: V_3 combination in linear and logarithmic scale. V_3 (top line) and T_3 (lower line) combinations in linear and logarithmic scale are shown. The error band of the reconstructed set is way bigger than the one of the input PDFs, showing a non negligible impact of the current statistics over the final PDFs error.

being the nature of the systematic errors: an additive shift given by a percentage of the ME for the first, a constant shift for all the ME points for the second one.

The results of the fit for the two optimistic scenarios are shown in Fig. 5.4. S1 is sligthly more conservative than S4, but overall there is not much difference between them. The situation
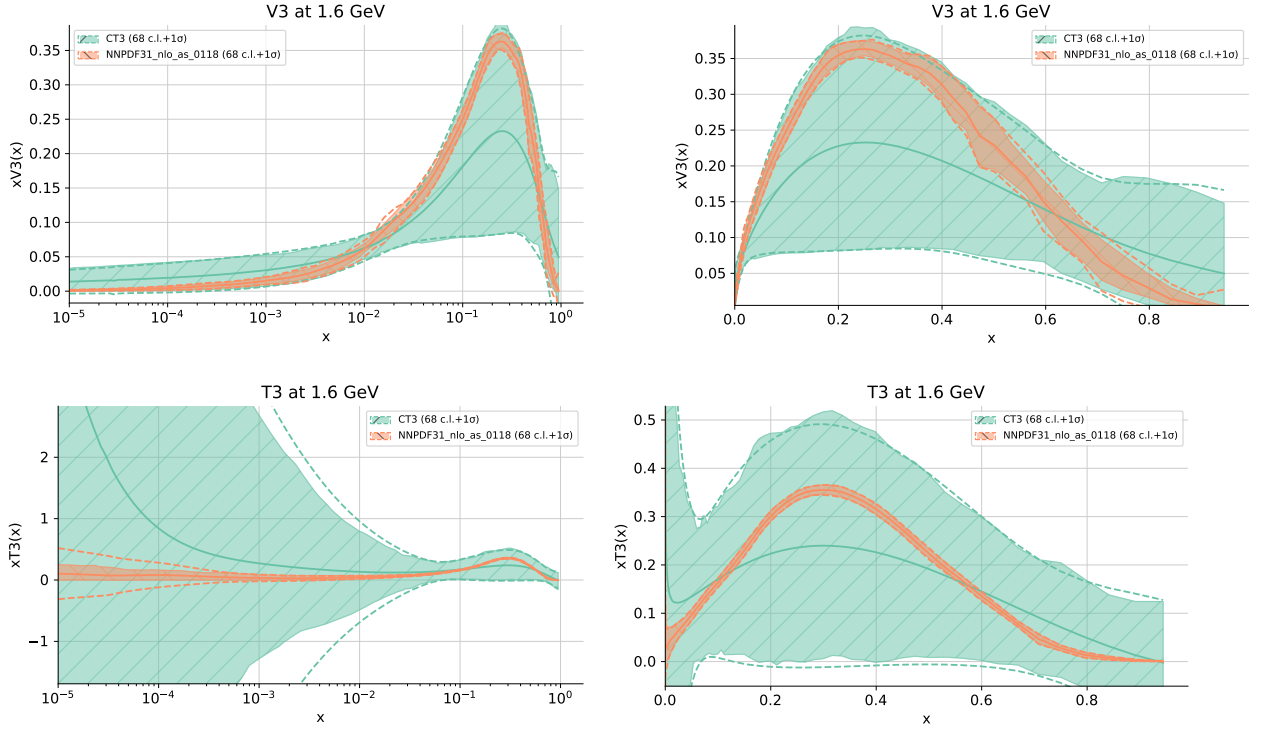
Figure 5.3. Closure test fit with real statistical and systematic error (CT3) compared to the input PDFs set. V_3 (top line) and T_3 (lower line) combinations in linear and logarithmic scale are shown. The errors of the reconstructed PDFs are huge.

changes for the more realistic scenarios (Fig. 5.5), where S2 is much more conservative than S5. In the former case the tension with `NNPDF31_nlo_0118` is smaller than what we observe in the previous scenarios, due to the increase in the error band and to a slight shift of the central replica of the fit. Similar comments can be made for the most pessimistic scenarios, shown in Fig. 5.6, having S3 with a huge error band and a more remarkable shift of the central replica towards the one of NNPDF31. Overall, we notice how, when the systematics are given by a percentage of the ME, we get qualitatively different results moving from one scenario to the other, while in the case we consider constant shifts there is no much difference between different cases.
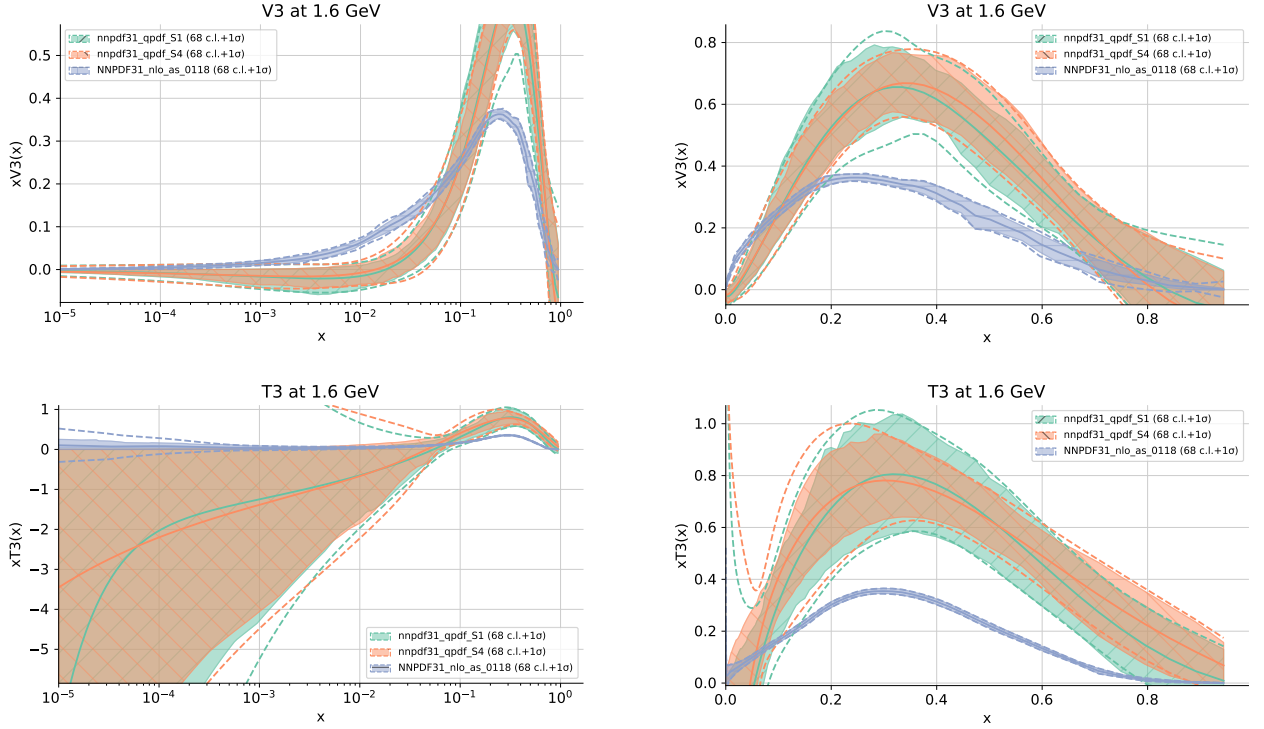
Figure 5.4. S1 vs. S4: S1 results are sligthly more conservative than the S4 ones, but overall there is no significant difference between the two optimistic scenarios.
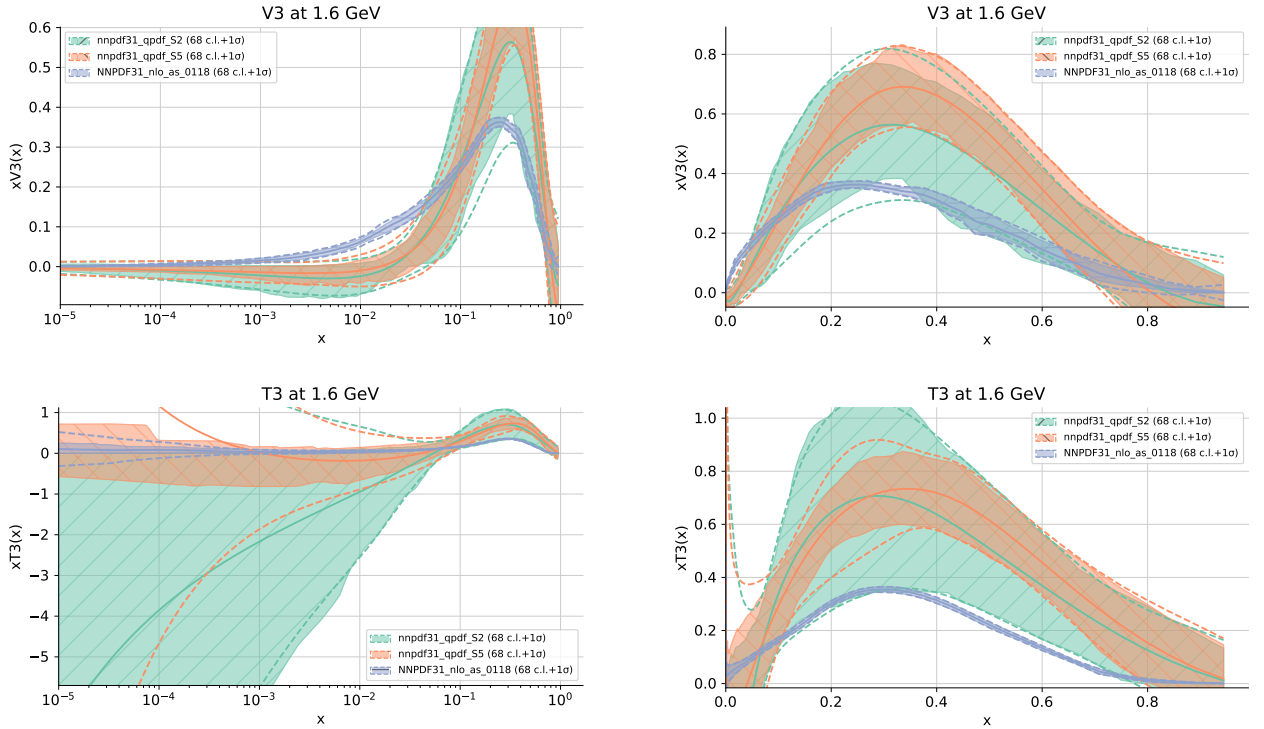

Figure 5.5. S2 vs. S5: S2 results are more conservative than the S5 ones, showing also a small shift of the replica 0 towards the light-cone PDFs. Overall, S2 results are comptible with `NNPDF31_nlo_0118` within 1-sigma level.
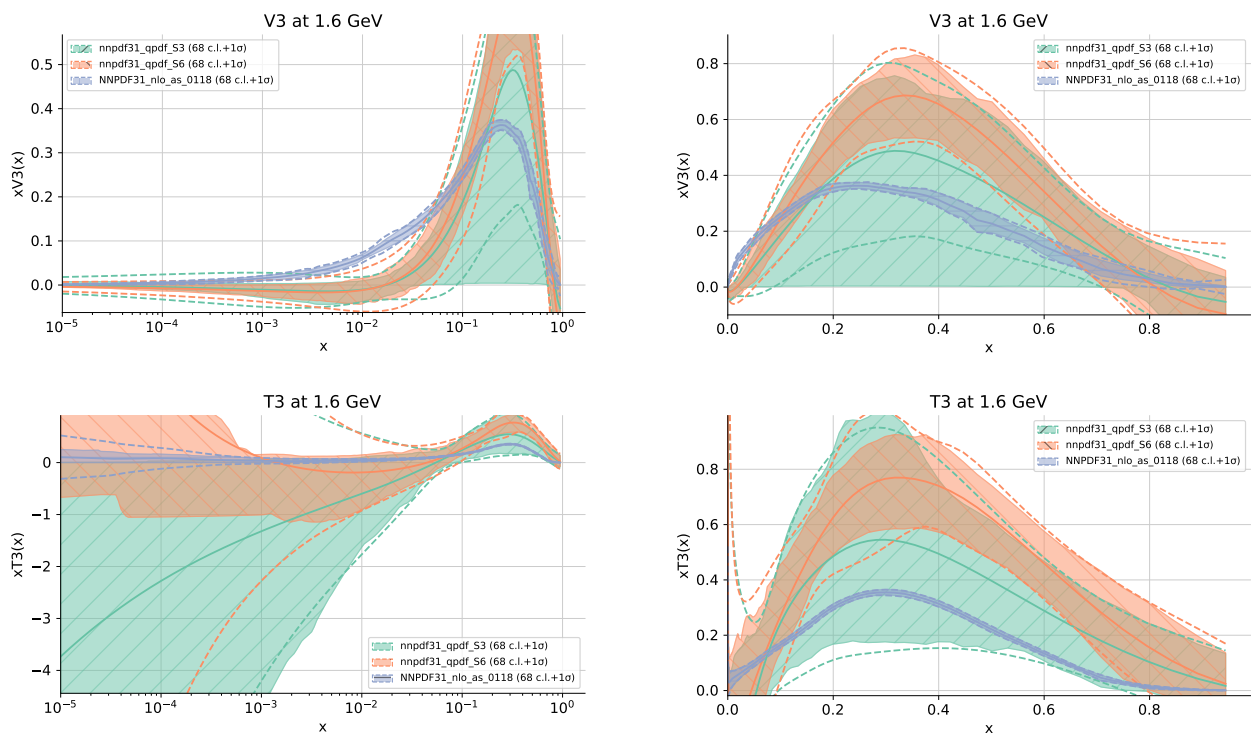
Figure 5.6. S3 vs. S6: S3 results are extremely conservative, while those for S6 do not show a qualitative difference with respect to S4 and S5.

6 Conclusions and future work

In this work, we have used the factorization of quasi-PDFs in order to relate the unpolarized isovector parton distribution to well-defined matrix elements computable on the lattice. Using some of the currently available lattice data, we have used such result to extract the nonsinglet distributions V_3 and T_3 within the NNPDF framework, studying also different possible scenarios for the treatment of the systematic uncertainties from lattice QCD simulations.

Our first results from closure tests show how effective these lattice data might be in constraining PDFs, allowing a consistent determination of the target distribution starting from $\mathcal{O}(15)$ ME points. On the other hand, we show that a consistent treatment of the lattice systematics is extremely important, and how the final result of the fit strongly depends on the specific systematics scenario we consider. Considering the most realistic ones, agreement with the phenomenological PDFs is observed within 1 sigma level, for both the nonsinglet distributions considered here. The error bands are, however, very large with respect to the corresponding phenomenological PDFs, showing again how important the control over the lattice systematics is.

Despite having focused on the quasi-PDFs case, the framework we implemented is general enough to allow the inclusion in the same analysis of data coming from different lattice simulations. One direction for future work might be a similar analysis repeated considering, for example, pseudo-PDFs data rather than quasi-PDFs ones, to see weather the conclusions we got here apply also to data coming from different lattice approaches. A fully global lattice QCD fit, which would be easily implemented within this framework, would then be the next logical step: it would then be interesting to see how much the inclusion of more lattice data in the same analysis affects the error band of the resulting PDFs, and whether or not a better agreement with the PDFs determined from experimental data is reached.

Our approach allows also to combine in the same analysis both experimental and lattice data. In principle we could then study the impact of the available lattice data on an existing PDF set, like NNPDF31. This, however, would require a deeper control and assessment of the lattice systematics involved in the simulations, together with a suitable way of treating the theoretical errors involved in the computation of the theory predictions entering the fits (missing higher orders in the matching coefficents and power correction terms), to avoid biasing the global QCD fit. This will be done in a future work.

Finally, we note how the current analysis can be extended without any additional complications to the other nonsinglet distributions defined in Eq. (2.9), as soon as lattice data for the corresponding matrix elements become available.

A Matching coefficient and lattice convolution

As detailed at the end of Sec. 2, the matching coefficients to be used to relate the data of Refs. [30, 45] to the light-cone PDFs are those expressed in the M$\overline{\text{MS}}$ scheme. Their explicit

expression is given by [30, 45]

$$C_3\left(\xi, \eta\left(\xi\right)\right) = \delta\left(1 - \xi\right) + C_3^{\text{NLO}}\left(\xi, \eta\left(\xi\right)\right)_+,$$

$$C_3^{\text{NLO}}\left(\xi, \eta\left(\xi\right)\right)_+ = \frac{\alpha_s}{2\pi} C_F \begin{cases} \left[\frac{1+\xi^2}{1-\xi} \log \frac{\xi}{\xi-1} + 1 + \frac{3}{2\xi}\right]_{+(1)}^{[1,+\infty]} & \xi > 1 \\ \left[\frac{1+\xi^2}{1-\xi} \log \left[\frac{1}{\eta^2(\xi)} \left(4\xi\left(1-\xi\right)\right)\right] - \frac{\xi(1+\xi)}{1-\xi}\right]_{+(1)}^{[0,1]} & 0 < \xi < 1 \\ \left[-\frac{1+\xi^2}{1-\xi} \log \frac{\xi}{\xi-1} - 1 + \frac{3}{2(1-\xi)}\right]_{+(1)}^{[-\infty,0]} & \xi < 0 \end{cases} .$$

(A.1)

where the superscripts indicate the domain over which the plus prescription acts. The matching coefficients relate the light-cone PDF to the quasi-PDF up to power suppressed terms according to

$$\tilde{f}_3\left(xP_z, \mu^2\right) = \int_{-1}^{1} \frac{dy}{|y|} C_3\left(\frac{x}{y}, \frac{\mu}{yP_z}\right) f_3\left(y, \mu^2\right).$$

(A.2)

In the following, we work out the full expression of the coefficients appearing in Eqs. (3.6), (3.7). Starting from Eq. (A.2) we have

$$\tilde{f}_3\left(x, \mu^2, P_z\right) = \int_{-1}^{1} \frac{dy}{|y|} \delta\left(1 - \frac{x}{y}\right) f_3\left(y, \mu^2\right) + \int_{-1}^{1} \frac{dy}{|y|} C_3^{\text{NLO}}\left(\frac{x}{y}, \frac{\mu}{yP_z}\right)_+ f_3\left(y, \mu^2\right).$$

(A.3)

Let us focus on the next-to-leading order term, making the plus distribution explicit. In order to do so, we find it useful to split the integral in the two contributions for $y < 0$ and $y > 0$. A change of variables, $\frac{x}{y} = \xi$, yields

$$\tilde{f}_3^{\text{NLO}}\left(x, \mu^2, P_z\right) \equiv \int_{-1}^{1} \frac{dy}{|y|} C_3^{\text{NLO}}\left(\frac{x}{y}, \frac{\mu}{yP_z}\right)_+ f_3\left(y, \mu^2\right)$$

$$= \int_{|x|}^{\infty} d\xi \, C_3^{\text{NLO}}\left(\xi, \frac{\mu\xi}{xP_z}\right)_+ \frac{1}{|\xi|} f_3\left(\frac{x}{\xi}, \mu^2\right) +$$

$$+ \int_{-\infty}^{-|x|} d\xi \, C_3^{\text{NLO}}\left(\xi, \frac{\mu\xi}{xP_z}\right)_+ \frac{1}{|\xi|} f_3\left(\frac{x}{\xi}, \mu\right).$$

(A.4)

The plus distribution appearing in the matching coefficients is defined in Ref. [25] and implemented as follows

$$\int_D d\xi \, C\left(\xi, g\left(\xi\right)\right)_+ f\left(\xi\right) = \int_D d\xi \left[C\left(\xi, g\left(\xi\right)\right) f\left(\xi\right) - C\left(\xi, g\left(1\right)\right) f\left(1\right)\right],$$

(A.5)

with $g\left(\xi\right) = \frac{\mu\xi}{xP_z}$ and D representing a generic integration domain, which in our case will be, according to Eq. (A.4), either $(-\infty, -|x|)$ or $(|x|, +\infty)$. It follows

$$\tilde{f}_3^{\text{NLO}}\left(x, \mu^2, P_z\right) = \int_{|x|}^{\infty} d\xi \left[C_3^{\text{NLO}}\left(\xi, \frac{\mu\xi}{xP_z}\right) \frac{f_3\left(\frac{x}{\xi}, \mu^2\right)}{|\xi|} - C_3^{\text{NLO}}\left(\xi, \frac{\mu}{xP_z}\right) f_3\left(x, \mu^2\right)\right]$$

$$+ \int_{-\infty}^{-|x|} d\xi \left[C_3^{\text{NLO}}\left(\xi, \frac{\mu\xi}{xP_z}\right) \frac{f_3\left(\frac{x}{\xi}, \mu^2\right)}{|\xi|} - C_3^{\text{NLO}}\left(\xi, \frac{\mu}{xP_z}\right) f_3\left(x, \mu^2\right)\right].$$

(A.6)

It can be easily verified that the two contributions appearing in the above equation are indeed well defined for every fixed x: the singularity in $\xi = +1$ is cured by the plus prescription, while for $\xi \to \pm\infty$ the matching coefficient behaves like $C\left(\xi\right) \sim \frac{1}{\xi^2}$, which is enough to guarantee the convergence of all the integrals above. For numerical stability we find it useful to avoid the

singularity in $\xi = +1$ introducing a suitable small parameter $\delta \sim 10^{-6}$, and rewriting the above equation as

$$
\begin{aligned}
\tilde{f}_3^{\mathrm{NLO}}\left(x, \mu^2, P_z\right) = & \int_{|x|}^{1-\delta} d\xi\, C_3^{\mathrm{NLO}}\left(\xi, \frac{\mu\xi}{xP_z}\right) \frac{f_3\left(\frac{x}{\xi}, \mu\right)}{\xi} - f_3(x, \mu) \int_{|x|}^{1-\delta} d\xi\, C_3^{\mathrm{NLO}}\left(\xi, \frac{\mu}{xP_z}\right) \\
& + \int_{1+\delta}^{\infty} d\xi\, C_3^{\mathrm{NLO}}\left(\xi, \frac{\mu\xi}{xP_z}\right) \frac{f_3\left(\frac{x}{\xi}, \mu\right)}{\xi} - f_3(x, \mu) \int_{1+\delta}^{\infty} d\xi\, C_3^{\mathrm{NLO}}\left(\xi, \frac{\mu}{xP_z}\right) \\
& - \int_{-\infty}^{-|x|} d\xi\, C_3^{\mathrm{NLO}}\left(\xi, \frac{\mu\xi}{xP_z}\right) \frac{f_3\left(\frac{x}{\xi}, \mu\right)}{\xi} - f_3(x, \mu) \int_{-\infty}^{-|x|} d\xi\, C_3^{\mathrm{NLO}}\left(\xi, \frac{\mu}{xP_z}\right).
\end{aligned}
\tag{A.7}
$$

In order to obtain the lattice ME, we need to compute the real and imaginary part of the Fourier transform of Eq. (A.3) as shown in Eq. (3.3). Starting from the leading-order contribution, we get

$$
\begin{aligned}
\int_{-\infty}^{\infty} dx\, \cos(xP_z z) \int_{-1}^{1} \frac{dy}{|y|} \delta\left(1 - \frac{x}{y}\right) f_3\left(y, \mu^2\right) \\
= \int_0^1 dy\, \cos(yP_z z)\left(f_3\left(y, \mu^2\right) + f_3\left(-y, \mu^2\right)\right) = \int_0^1 dx\, \cos(xP_z z)\, V_3\left(x, \mu^2\right) \\
= \int_0^1 dx\, \mathrm{A}^{\mathrm{Re, LO}}(xP_z z)\, V_3\left(x, \mu^2\right)
\end{aligned}
\tag{A.8}
$$

where we have integrated in x first, re-expressed the integral $\int_{-1}^1 dy$ as $\int_0^1 dy$, used

$$
f_3(x) + f_3(-x) = f_3^{\mathrm{sym}}(x) = u^-(x) - d^-(x) = V_3(x)
\tag{A.9}
$$

and finally changed variables back to x. Moving now to the next-to-leading order part, we analyse each of the six contributions listed in Eq. (A.7), defining for each lattice observable six integrals to be computed, denoted as $\mathrm{I}_i^{\mathrm{Re}}, \mathrm{I}_i^{\mathrm{Im}}$ $i = 1, .., 6$. Starting from the first contribution to the real part we get

$$
\begin{aligned}
\mathrm{I}_1^{\mathrm{Re}} &= \int_{-\infty}^{\infty} dx\, \cos(xP_z z) \int_{|x|}^{1-\delta} d\xi\, C_3^{\mathrm{NLO}}\left(\xi, \frac{\mu\xi}{xP_z}\right) \frac{f_3\left(\frac{x}{\xi}, \mu\right)}{\xi} \\
&= \int_0^{\infty} dx\, \cos(xP_z z) \int_x^{1-\delta} \frac{d\xi}{\xi} C_3^{\mathrm{NLO}}\left(\xi, \frac{\mu\xi}{xP_z}\right) \left(f_3\left(\frac{x}{\xi}, \mu\right) + f_3\left(-\frac{x}{\xi}, \mu\right)\right) \\
&= \int_0^1 dx\, \cos(xP_z z) \int_{x/(1-\delta)}^1 \frac{dy}{y} C_3^{\mathrm{NLO}}\left(\frac{x}{y}, \frac{\mu}{yP_z}\right) V_3(y, \mu),
\end{aligned}
\tag{A.10}
$$

where in the last line we have changed variables back to $\frac{x}{\xi} = y$. Also, the integration range for x becomes $(0, 1)$, since $x < y < 1$. Renaming variables, we have

$$
\begin{aligned}
\mathrm{I}_1^{\mathrm{Re}} &= \int_0^1 dx \left[\frac{1}{x} \int_0^1 dy\, \Theta\left(x - \frac{y}{1-\delta}\right) \cos(yP_z z) C_3^{\mathrm{NLO}}\left(\frac{y}{x}, \frac{\mu}{xP_z}\right)\right] V_3\left(x, \mu^2\right) \\
&= \int_0^1 dx\, \mathrm{A}_1^{\mathrm{Re, NLO}}\left(x, z, \frac{\mu}{P_z}\right) V_3\left(x, \mu^2\right).
\end{aligned}
\tag{A.11}
$$

Analogously, we find out that the other five contributions can be written as

$$
\mathrm{I}_i^{\mathrm{Re}} = \int_0^1 dx\, \mathrm{A}_i^{\mathrm{Re, NLO}}\left(x, z, \frac{\mu}{P_z}\right) V_3\left(x, \mu^2\right).
\tag{A.12}
$$

22

with

$$A_2^{\text{Re, NLO}} = \cos(xP_z z) \int_x^{1-\delta} d\xi \, C_3^{\text{NLO}} \left(\xi, \frac{\mu}{xP_z} \right), \tag{A.13}$$

$$A_3^{\text{Re, NLO}} = \frac{1}{x} \int_0^\infty dy \, \Theta \left(\frac{y}{1+\delta} - x \right) \cos\left(yP_z z \right) C_3^{\text{NLO}} \left(\frac{y}{x}, \frac{\mu}{xP_z} \right), \tag{A.14}$$

$$A_4^{\text{Re, NLO}} = \cos(xP_z z) \int_{1+\delta}^\infty d\xi \, C_3^{\text{NLO}} \left(\xi, \frac{\mu}{xP_z} \right), \tag{A.15}$$

$$A_5^{\text{Re, NLO}} = -\frac{1}{x} \int_0^\infty dy \, \cos\left(yP_z z \right) C_3^{\text{NLO}} \left(-\frac{y}{x}, \frac{\mu}{xP_z} \right), \tag{A.16}$$

$$A_6^{\text{Re, NLO}} = \cos(xP_z z) \int_{-\infty}^{-x} d\xi \, C_3^{\text{NLO}} \left(\xi, \frac{\mu}{xP_z} \right) \tag{A.17}$$

Collecting all the terms yields Eq. (3.6)

$$\mathcal{O}_{\gamma^0}^{\text{Re}} (z, \mu) = \int_0^1 dx \, \mathcal{C}_3^{\text{Re}} \left(x, z, \frac{\mu}{P_z} \right) V_3 \left(x, \mu^2 \right), \tag{A.18}$$

where

$$\mathcal{C}_3^{\text{Re}} \left(x, z, \frac{\mu}{P_z} \right) = A^{\text{Re, LO}} + A^{\text{Re, NLO}} \tag{A.19}$$

with

$$A^{\text{Re, NLO}} = A_1^{\text{Re, NLO}} - A_2^{\text{Re, NLO}} + A_3^{\text{Re, NLO}} - A_4^{\text{Re, NLO}} - A_5^{\text{Re, NLO}} - A_6^{\text{Re, NLO}}. \tag{A.20}$$

We now turn to the imaginary part of the Fourier transform. The computation is exactly the same as in the previous case, with the only difference that now we have a sin instead of a cos. Because of this, when re-expressing the integral $\int_{-\infty}^\infty dx$ as $\int_0^\infty dx$, we get an additional minus sign, which gives the combination

$$f(x) - f(-x) = f_3^{\text{asy}}(x) = u^+(x) - d^+(x) = T_3(x). \tag{A.21}$$

Therefore, the results for the imaginary part can be obtained from those for the real part simply by replacing cos with sin and V_3 with T_3.

B QCD evolution equations

Let us briefly summarise how the QCD evolution equation is solved for the nonsinglet sector, yielding the evolution kernel that is used in the construction of FastKernel tables. For more details and for the validation of this approach, we refer to Ref. [74] and the subsequent NNPDF publications.

Denoting the nonsinglet distributions V_3 and T_3 with $f^{(-)}$ and $f^{(+)}$ respectively, the QCD evolution equations can be written as

$$\mu^2 \frac{\partial}{\partial \mu^2} f^{(\pm)} \left(x, \mu^2 \right) = \frac{\alpha_s(\mu^2)}{2\pi} \int_x^1 \frac{d\xi}{\xi} P_{qq}^{(\pm)} \left(\frac{x}{\xi}, \alpha_s(Q^2) \right) f^{(\pm)} \left(\xi, \mu^2 \right), \tag{B.1}$$

which in Mellin space becomes

$$\mu^2 \frac{\partial}{\partial \mu^2} f^{(\pm)} \left(N, \mu^2 \right) = \gamma^{(\pm)} \left(N, \alpha_s \right) f^{(\pm)} \left(N, \mu^2 \right). \tag{B.2}$$

The distribution at the scale μ^2 is obtained from the distribution at the scale μ_0^2 by introducing the evolution operator Γ

$$f^{(\pm)} \left(N, \mu^2 \right) = \Gamma^{(\pm)} \left(N, \alpha_s, \alpha_s^0 \right) f^{(\pm)} \left(N, \mu_0^2 \right), \tag{B.3}$$

where $\alpha_s \equiv \alpha_s(\mu^2)$ and $\alpha_s^0 \equiv \alpha_s(\mu_0^2)$. Substituting Eq. (B.3) in Eq. (B.2) and remembering that the dependence of Γ on the scale μ is only through the coupling, we have

$$\beta(\alpha_s) \frac{\partial}{\partial \alpha_s} \Gamma^{(\pm)}(N, \alpha_s, \alpha_s^0) = \gamma^{(\pm)}(N, \alpha_s) \Gamma^{(\pm)}(N, \alpha_s, \alpha_s^0) \ . \tag{B.4}$$

Since the matching coefficients in Eq. (A.1) are known up to NLO ($\mathcal{O}(\alpha_s)$), here we will only consider NLO evolution equations. Then, using

$$\beta(\alpha_s) = \frac{d\alpha_s}{d \log \mu^2} = -\alpha_s^2 \beta_0 - \alpha_s^3 \beta_1 + \mathcal{O}(\alpha_s^4) \tag{B.5}$$

$$\gamma^{(\pm)}(N, \alpha_s) = \frac{\alpha_s}{4\pi} \gamma_0^{(\pm)}(N) + \left(\frac{\alpha_s}{4\pi}\right)^2 \gamma_1^{(\pm)}(N) + \mathcal{O}(\alpha_s^3) \ , \tag{B.6}$$

we can solve Eq. (B.4); the Mellin space expression for the evolution kernel at NLO is

$$\Gamma^{(\pm)}(N, \alpha_s, \alpha_s^0) = 1 + \frac{\alpha_s - \alpha_s^0}{4\pi} \left(\frac{\gamma_1^{(\pm)}(N)}{2\beta_0} - \frac{\beta_1 \gamma_0^{(\pm)}(N)}{2\beta_0^2} \right) \ . \tag{B.7}$$

The solution in the x-space is obtained by computing the inverse Mellin transform of $\Gamma^{(\pm)}(N, \alpha_s, \alpha_s^0)$. Having analytically continued the function $\Gamma(N)$ to the complex plane, the inverse Mellin transform is obtained by computing the contour integral

$$\Gamma^{(\pm)}(x, \alpha_s, \alpha_s^0) = \int_C \frac{dN}{2\pi i} x^{-N} \Gamma^{(\pm)}(N, \alpha_s, \alpha_s^0) \ . \tag{B.8}$$

References

[1] **NNPDF** Collaboration, R. D. Ball et al., *Parton distributions from high-precision collider data, Eur. Phys. J.* **C77** (2017), no. 10 663, [arXiv:1706.00428].

[2] S. Dulat, T.-J. Hou, J. Gao, M. Guzzi, J. Huston, P. Nadolsky, J. Pumplin, C. Schmidt, D. Stump, and C. P. Yuan, *New parton distribution functions from a global analysis of quantum chromodynamics, Phys. Rev.* **D93** (2016), no. 3 033006, [arXiv:1506.07443].

[3] S. Alekhin, J. Blümlein, S. Moch, and R. Placakyte, *Parton distribution functions, α_s, and heavy-quark masses for LHC Run II, Phys. Rev.* **D96** (2017), no. 1 014011, [arXiv:1701.05838].

[4] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, *Parton distributions for the LHC, Eur. Phys. J.* **C63** (2009) 189–285, [arXiv:0901.0002].

[5] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, *LHAPDF6: parton density access in the LHC precision era, Eur. Phys. J.* **C75** (2015) 132, [arXiv:1412.7420].

[6] X. Ji, *Parton physics on a euclidean lattice, Phys. Rev. Lett.* **110** (Jun, 2013) 262002.

[7] X. Ji, *Parton Physics from Large-Momentum Effective Field Theory, Sci. China Phys. Mech. Astron.* **57** (2014) 1407–1412, [arXiv:1404.6680].

[8] K. Cichy and M. Constantinou, *A guide to light-cone PDFs from Lattice QCD: an overview of approaches, techniques and results, Adv. High Energy Phys.* **2019** (2019) 3036904, [arXiv:1811.07248].

[9] C. Monahan, *Recent Developments in x-dependent Structure Calculations, PoS* **LATTICE2018** (2018) 018, [arXiv:1811.00678].

[10] X. Ji and J.-H. Zhang, *Renormalization of quasiparton distribution, Phys. Rev.* **D92** (2015) 034006, [arXiv:1505.07699].

[11] T. Ishikawa, Y.-Q. Ma, J.-W. Qiu, and S. Yoshida, *Practical quasi parton distribution functions*, arXiv:1609.02018.

[12] M. Constantinou and H. Panagopoulos, *Perturbative renormalization of quasi-parton distribution functions*, Phys. Rev. **D96** (2017), no. 5 054506, [arXiv:1705.11193].

[13] C. Alexandrou, K. Cichy, M. Constantinou, K. Hadjiyiannakou, K. Jansen, H. Panagopoulos, and F. Steffens, *A complete non-perturbative renormalization prescription for quasi-PDFs*, Nucl. Phys. **B923** (2017) 394–415, [arXiv:1706.00265].

[14] X. Ji, J.-H. Zhang, and Y. Zhao, *More On Large-Momentum Effective Theory Approach to Parton Physics*, Nucl. Phys. **B924** (2017) 366–376, [arXiv:1706.07416].

[15] X. Ji, J.-H. Zhang, and Y. Zhao, *Renormalization in Large Momentum Effective Theory of Parton Physics*, Phys. Rev. Lett. **120** (2018), no. 11 112001, [arXiv:1706.08962].

[16] T. Ishikawa, Y.-Q. Ma, J.-W. Qiu, and S. Yoshida, *Renormalizability of quasiparton distribution functions*, Phys. Rev. **D96** (2017), no. 9 094019, [arXiv:1707.03107].

[17] J. Green, K. Jansen, and F. Steffens, *Nonperturbative Renormalization of Nonlocal Quark Bilinears for Parton Quasidistribution Functions on the Lattice Using an Auxiliary Field*, Phys. Rev. Lett. **121** (2018), no. 2 022004, [arXiv:1707.07152].

[18] A. V. Radyushkin, *Structure of parton quasi-distributions and their moments*, Phys. Lett. **B788** (2019) 380–387, [arXiv:1807.07509].

[19] J.-H. Zhang, X. Ji, A. Schäfer, W. Wang, and S. Zhao, *Accessing Gluon Parton Distributions in Large Momentum Effective Theory*, Phys. Rev. Lett. **122** (2019), no. 14 142001, [arXiv:1808.10824].

[20] Z.-Y. Li, Y.-Q. Ma, and J.-W. Qiu, *Multiplicative Renormalizability of Operators defining Quasiparton Distributions*, Phys. Rev. Lett. **122** (2019), no. 6 062002, [arXiv:1809.01836].

[21] X. Xiong, X. Ji, J.-H. Zhang, and Y. Zhao, *One-loop matching for parton distributions: Nonsinglet case*, Phys. Rev. **D90** (2014), no. 1 014051, [arXiv:1310.7471].

[22] Y.-Q. Ma and J.-W. Qiu, *Extracting Parton Distribution Functions from Lattice QCD Calculations*, Phys. Rev. **D98** (2018), no. 7 074021, [arXiv:1404.6860].

[23] R. A. Briceño, M. T. Hansen, and C. J. Monahan, *Role of the Euclidean signature in lattice calculations of quasidistributions and other nonlocal matrix elements*, Phys. Rev. **D96** (2017), no. 1 014502, [arXiv:1703.06072].

[24] Y.-Q. Ma and J.-W. Qiu, *Exploring Partonic Structure of Hadrons Using ab initio Lattice QCD Calculations*, Phys. Rev. Lett. **120** (2018), no. 2 022003, [arXiv:1709.03018].

[25] T. Izubuchi, X. Ji, L. Jin, I. W. Stewart, and Y. Zhao, *Factorization Theorem Relating Euclidean and Light-Cone Parton Distributions*, Phys. Rev. **D98** (2018), no. 5 056004, [arXiv:1801.03917].

[26] X. Ji, A. Schäfer, X. Xiong, and J.-H. Zhang, *One-Loop Matching for Generalized Parton Distributions*, Phys. Rev. **D92** (2015) 014039, [arXiv:1506.00248].

[27] X. Xiong and J.-H. Zhang, *One-loop matching for transversity generalized parton distribution*, Phys. Rev. **D92** (2015), no. 5 054037, [arXiv:1509.08016].

[28] W. Wang, S. Zhao, and R. Zhu, *Gluon quasidistribution function at one loop*, Eur. Phys. J. **C78** (2018), no. 2 147, [arXiv:1708.02458].

[29] I. W. Stewart and Y. Zhao, *Matching the quasiparton distribution in a momentum subtraction scheme*, Phys. Rev. **D97** (2018), no. 5 054512, [arXiv:1709.04933].

[30] C. Alexandrou, K. Cichy, M. Constantinou, K. Jansen, A. Scapellato, and F. Steffens, *Light-Cone Parton Distribution Functions from Lattice QCD*, Phys. Rev. Lett. **121** (2018), no. 11 112001, [arXiv:1803.02685].

[31] C. Alexandrou, K. Cichy, M. Constantinou, K. Jansen, A. Scapellato, and F. Steffens, *Transversity parton distribution functions from lattice QCD*, Phys. Rev. D. (Rapid Communication), in production, (2018) [arXiv:1807.00232].

[32] Y.-S. Liu, J.-W. Chen, L. Jin, H.-W. Lin, Y.-B. Yang, J.-H. Zhang, and Y. Zhao, *Unpolarized quark distribution from lattice QCD: A systematic analysis of renormalization and matching*, arXiv:1807.06566.

[33] Y.-S. Liu, J.-W. Chen, L. Jin, R. Li, H.-W. Lin, Y.-B. Yang, J.-H. Zhang, and Y. Zhao, *Nucleon Transversity Distribution at the Physical Pion Mass from Lattice QCD*, arXiv:1810.05043.

[34] H.-W. Lin, J.-W. Chen, S. D. Cohen, and X. Ji, *Flavor Structure of the Nucleon Sea from Lattice QCD*, Phys. Rev. **D91** (2015) 054510, [arXiv:1402.1462].

[35] C. Alexandrou, K. Cichy, V. Drach, E. Garcia-Ramos, K. Hadjiyiannakou, K. Jansen, F. Steffens, and C. Wiese, *Lattice calculation of parton distributions*, Phys. Rev. **D92** (2015) 014502, [arXiv:1504.07455].

[36] J.-W. Chen, S. D. Cohen, X. Ji, H.-W. Lin, and J.-H. Zhang, *Nucleon Helicity and Transversity Parton Distributions from Lattice QCD*, Nucl. Phys. **B911** (2016) 246–273, [arXiv:1603.06664].

[37] C. Alexandrou, K. Cichy, M. Constantinou, K. Hadjiyiannakou, K. Jansen, F. Steffens, and C. Wiese, *Updated Lattice Results for Parton Distributions*, Phys. Rev. **D96** (2017), no. 1 014513, [arXiv:1610.03689].

[38] J.-H. Zhang, J.-W. Chen, X. Ji, L. Jin, and H.-W. Lin, *Pion Distribution Amplitude from Lattice QCD*, Phys. Rev. **D95** (2017), no. 9 094514, [arXiv:1702.00008].

[39] **LP3** Collaboration, H.-W. Lin, J.-W. Chen, T. Ishikawa, and J.-H. Zhang, *Improved parton distribution functions at the physical pion mass*, Phys. Rev. **D98** (2018), no. 5 054504, [arXiv:1708.05301].

[40] **LP3** Collaboration, J.-H. Zhang, L. Jin, H.-W. Lin, A. Schäfer, P. Sun, Y.-B. Yang, R. Zhang, Y. Zhao, and J.-W. Chen, *Kaon Distribution Amplitude from Lattice QCD and the Flavor SU(3) Symmetry*, Nucl. Phys. **B939** (2019) 429–446, [arXiv:1712.10025].

[41] J.-W. Chen, L. Jin, H.-W. Lin, Y.-S. Liu, Y.-B. Yang, J.-H. Zhang, and Y. Zhao, *Lattice Calculation of Parton Distribution Function from LaMET at Physical Pion Mass with Large Nucleon Momentum*, arXiv:1803.04393.

[42] J.-W. Chen, L. Jin, H.-W. Lin, Y.-S. Liu, A. Schäfer, Y.-B. Yang, J.-H. Zhang, and Y. Zhao, *First direct lattice-QCD calculation of the x-dependence of the pion parton distribution function*, arXiv:1804.01483.

[43] H.-W. Lin, J.-W. Chen, X. Ji, L. Jin, R. Li, Y.-S. Liu, Y.-B. Yang, J.-H. Zhang, and Y. Zhao, *Proton Isovector Helicity Distribution on the Lattice at Physical Pion Mass*, Phys. Rev. Lett. **121** (2018), no. 24 242003, [arXiv:1807.07431].

[44] Z.-Y. Fan, Y.-B. Yang, A. Anthony, H.-W. Lin, and K.-F. Liu, *Gluon Quasi-Parton-Distribution Functions from Lattice QCD*, Phys. Rev. Lett. **121** (2018), no. 24 242001, [arXiv:1808.02077].

[45] C. Alexandrou, K. Cichy, M. Constantinou, K. Hadjiyiannakou, K. Jansen, A. Scapellato, and F. Steffens, *Systematic uncertainties in parton distribution functions from lattice QCD simulations at the physical point*, Phys. Rev. **D99** (2019), no. 11 114504, [arXiv:1902.00587].

[46] T. Izubuchi, L. Jin, C. Kallidonis, N. Karthik, S. Mukherjee, P. Petreczky, C. Shugert, and S. Syritsyn, *Valence parton distribution function of pion from fine lattice*, arXiv:1905.06349.

[47] J. Karpie, K. Orginos, A. Rothkopf, and S. Zafeiropoulos, *Reconstructing parton distribution functions from ioffe time data: from bayesian methods to neural networks*, *Journal of High Energy Physics* **2019** (Apr, 2019) 57.

[48] A. V. Radyushkin, *Quasi-parton distribution functions, momentum distributions, and pseudo-parton distribution functions*, *Phys. Rev.* **D96** (2017), no. 3 034025, [arXiv:1705.01488].

[49] K. Orginos, A. Radyushkin, J. Karpie, and S. Zafeiropoulos, *Lattice QCD exploration of parton pseudo-distribution functions*, *Phys. Rev.* **D96** (2017), no. 9 094503, [arXiv:1706.05373].

[50] J. Karpie, K. Orginos, and S. Zafeiropoulos, *Moments of Ioffe time parton distribution functions from non-local matrix elements*, *JHEP* **11** (2018) 178, [arXiv:1807.10933].

[51] W. Detmold and C. J. D. Lin, *Deep-inelastic scattering and the operator product expansion in lattice QCD*, *Phys. Rev.* **D73** (2006) 014501, [hep-lat/0507007].

[52] V. Braun and D. Mueller, *Exclusive processes in position space and the pion distribution amplitude*, *Eur. Phys. J.* **C55** (2008) 349–361, [arXiv:0709.1348].

[53] R. S. Sufian, J. Karpie, C. Egerer, K. Orginos, J.-W. Qiu, and D. G. Richards, *Pion Valence Quark Distribution from Matrix Element Calculated in Lattice QCD*, *Phys. Rev.* **D99** (2019), no. 7 074507, [arXiv:1901.03921].

[54] J. C. Collins, *INTRINSIC TRANSVERSE MOMENTUM. 1. NONGAUGE THEORIES*, *Phys. Rev.* **D21** (1980) 2962.

[55] J. C. Collins and D. E. Soper, *Parton Distribution and Decay Functions*, *Nucl. Phys.* **B194** (1982) 445–492.

[56] A. Vogt, *Efficient evolution of unpolarized and polarized parton distributions with QCD-PEGASUS*, *Comput. Phys. Commun.* **170** (2005) 65–92, [hep-ph/0408244].

[57] J. Gao, L. Harland-Lang, and J. Rojo, *The Structure of the Proton in the LHC Precision Era*, *Phys. Rept.* **742** (2018) 1–121, [arXiv:1709.04922].

[58] H.-W. Lin et al., *Parton distributions and lattice QCD calculations: a community white paper*, *Prog. Part. Nucl. Phys.* **100** (2018) 107–160, [arXiv:1711.07916].

[59] V. Braun, P. Gornicki, and L. Mankiewicz, *Ioffe - time distributions instead of parton momentum distributions in description of deep inelastic scattering*, *Phys. Rev.* **D51** (1995) 6036–6051, [hep-ph/9410318].

[60] G. Martinelli, C. Pittori, C. T. Sachrajda, M. Testa, and A. Vladikas, *A General method for nonperturbative renormalization of lattice operators*, *Nucl. Phys.* **B445** (1995) 81–108, [hep-lat/9411010].

[61] A. Radyushkin, *Nonperturbative Evolution of Parton Quasi-Distributions*, *Phys. Lett.* **B767** (2017) 314–320, [arXiv:1612.05170].

[62] **ETM** Collaboration, A. Abdel-Rehim et al., *First physics results at the physical pion mass from $N_f = 2$ Wilson twisted mass fermions at maximal twist*, *Phys. Rev.* **D95** (2017), no. 9 094515, [arXiv:1507.05068].

[63] C. Alexandrou and C. Kallidonis, *Low-lying baryon masses using $N_f = 2$ twisted mass clover-improved fermions directly at the physical pion mass*, *Phys. Rev.* **D96** (2017), no. 3 034511, [arXiv:1704.02647].

[64] **NNPDF** Collaboration, R. Abdul Khalek et al., *A First Determination of Parton Distributions with Theoretical Uncertainties*, arXiv:1905.04311.

[65] **NNPDF** Collaboration, R. Abdul Khalek et al., *Parton Distributions with Theory Uncertainties: General Formalism and First Phenomenological Studies*, arXiv:1906.10698.

[66] S. Syritsyn, *Review of Hadron Structure Calculations on a Lattice*, PoS **LATTICE2013** (2014) 009, [arXiv:1403.4686].

[67] M. Constantinou, *Hadron Structure*, PoS **LATTICE2014** (2015) 001, [arXiv:1411.0078].

[68] M. Constantinou, *Recent progress in hadron structure from Lattice QCD*, PoS **CD15** (2015) 009, [arXiv:1511.00214].

[69] C. Alexandrou, *Selected results on hadron structure using state-of-the-art lattice QCD simulations*, in *Proceedings, 45th International Symposium on Multiparticle Dynamics (ISMD 2015): Kreuth, Germany, October 4-9, 2015*, 2015. arXiv:1512.03924.

[70] J. Green, *Systematics in nucleon matrix element calculations*, in *36th International Symposium on Lattice Field Theory (Lattice 2018) East Lansing, MI, United States, July 22-28, 2018*, 2018. arXiv:1812.10574.

[71] R. A. Briceño, J. V. Guerrero, M. T. Hansen, and C. J. Monahan, *Finite-volume effects due to spatially nonlocal operators*, Phys. Rev. **D98** (2018), no. 1 014511, [arXiv:1805.01034].

[72] V. M. Braun, A. Vladimirov, and J.-H. Zhang, *Power corrections and renormalons in parton quasidistributions*, Phys. Rev. **D99** (2019), no. 1 014013, [arXiv:1810.00048].

[73] G. S. Bali et al., *Pion distribution amplitude from Euclidean correlation functions*, Eur. Phys. J. **C78** (2018), no. 3 217, [arXiv:1709.04325].

[74] **NNPDF** Collaboration, L. Del Debbio, S. Forte, J. I. Latorre, A. Piccione, and J. Rojo, *Neural network determination of parton distributions: The Nonsinglet case*, JHEP **03** (2007) 039, [hep-ph/0701127].

[75] **NNPDF** Collaboration, R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, A. Piccione, J. Rojo, and M. Ubiali, *A Determination of parton distributions with faithful uncertainty estimation*, Nucl. Phys. **B809** (2009) 1–63, [arXiv:0808.1231]. [Erratum: Nucl. Phys.B816,293(2009)].

[76] R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali, *A first unbiased global NLO determination of parton distributions and their uncertainties*, Nucl. Phys. **B838** (2010) 136–206, [arXiv:1002.4407].

[77] N. Hansen, *The CMA evolution strategy: A tutorial*, CoRR **abs/1604.00772** (2016) [arXiv:1604.00772].

[78] **NNPDF** Collaboration, V. Bertone, S. Carrazza, N. P. Hartland, E. R. Nocera, and J. Rojo, *A determination of the fragmentation functions of pions, kaons, and protons with faithful uncertainties*, Eur. Phys. J. **C77** (2017), no. 8 516, [arXiv:1706.07049].

[79] V. Bertone, S. Carrazza, and N. P. Hartland, *APFELgrid: a high performance tool for parton density determinations*, Comput. Phys. Commun. **212** (2017) 205–209, [arXiv:1605.02070].

[80] Z. Kassabov, "Reportengine: A framework for declarative data analysis." https://doi.org/10.5281/zenodo.2571601, Feb., 2019.

[81] **NNPDF** Collaboration, R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040, [arXiv:1410.8849].