Evolution of parton distribution functions in the short-distance factorization scheme

Hervé Dutrieux [ID],[a] Joseph Karpie [ID],[b] Christopher Monahan [ID],[a] Kostas Orginos [ID][a,b] and Savvas Zafeiropoulos [ID][c]

on behalf of the HadStruc collaboration

[a]*Physics Department, William & Mary,*
Williamsburg, VA 23187, U.S.A.
[b]*Thomas Jefferson National Accelerator Facility,*
Newport News, VA 23606, U.S.A.
[c]*Aix Marseille Univ, Université de Toulon,*
CNRS, CPT, Marseille, France

E-mail: hldutrieux@wm.edu, jkarpie@jlab.org, cmonahan@wm.edu,
kostas@wm.edu, savvas.zafeiropoulos@cpt.univ-mrs.fr

ABSTRACT: Lattice QCD offers the possibility of computing parton distributions from first principles, although not in the usual $\overline{MS}$ factorization scheme. Calculations are therefore matched to $\overline{MS}$ using a perturbative procedure which is the source of significant uncertainty within the currently accessible kinematics. We present the possibility of computing the z^2 evolution of non-singlet pseudo-parton distribution functions within the short factorization scheme in a numerically improvable way. The goal is to have tools to evolve a calculation to a scale where perturbative uncertainties are less pronounced. We compare a numerical extraction of the evolution operator from lattice data to the computation of z^2 dependence in perturbation theory. Finally, we discuss how this numerical work may be extended to address the two-scale problem that arises when the Ioffe time range must be made large to extend the reach of the calculation of the pseudo-PDF to smaller values of the momentum fraction.

Contents

1 Introduction

The introduction of the Large-Momentum Effective Theory (LaMET) by X. Ji [1, 2] opened a new window for calculations of light-cone correlation functions using lattice quantum chromo-dynamics (QCD). Numerous works have implemented both LaMET and many alternatives, such as the pseudo-distribution approach at the center of this paper [3–5] and the current-current correlator approach [6–11]. We refer the reader to the recent reviews [12–14] for a panorama of this field of research. The fundamental principle underlying these approaches is that equal-time non-local matrix elements in Euclidean space — with space-like separation of the constituent fields — contain the same soft physics as light-cone distributions, although they differ by their regularization of collinear divergences [8, 10, 15]. Evolution in z^2 of pseudo-distributions in this so-called "short-distance factorization" scheme (SDF) offers the remarkable possibility of accessing non-perturbative evolution from first principles.

Perturbation theory provides significant inputs for evolution in SDF: the matching kernels that relate SDF to $\overline{MS}$ for parton distribution functions (PDFs) are available to

order $\mathcal{O}(\alpha_s^2)$ [16, 17], and PDF evolution equations in $\overline{MS}$ are known exactly to order $\mathcal{O}(\alpha_s^3)$ [18–20]. However, due to the perturbative divergence of the strong coupling at Λ_{QCD}, matching derived from perturbation theory can induce a sizeable theoretical uncertainty even at quite small values of z^2. In principle, provided the data is obtained in a range of scales where universality of evolution is relevant, one could compute on the lattice the evolution in z^2 to an arbitrary high degree of accuracy, bound to become more precise than perturbation theory and its systematic uncertainties. The possibility to extract this object would in fact be useful to determine a range of validity of the leading-twist dominance in the data. In this paper, we describe a strategy to extract numerically the evolution operator. Hopefully, the ever-growing set of lattice calculations should allow this object to become extremely precise in the future. We demonstrate in particular that the form we choose for the evolution operator — which generalizes the operator product expansion (OPE) at small values of z^2 — allows us to place a constraint on the z^2 dependence of the pseudo-distributions, and therefore improve their extraction with a much lesser degree of model dependence compared to imposing $\overline{MS}$ perturbative results. Furthermore, one can then assess the degree of compatibility of the two approaches.

Our approach is not only useful to work with lattice calculations in a regime where perturbation theory has severe shortcomings. We believe the general principles that we develop may prove useful to extend the range in Ioffe time accessible to current calculations. Before going further, it is useful to retrace some of the difficulties that lattice computations of non-local matrix elements face. As for any lattice calculation with a position-space cut-off a (i.e. a space-time lattice) and a finite volume L, the physical box size must be much larger than the inverse of the hadronic scale Λ_{QCD}^{-1} to avoid finite volume effects, and the lattice spacing must be much smaller than Λ_{QCD}^{-1} to avoid discretization errors. Thus, we have the scale hierarchy:

$$a \ll \Lambda_{QCD}^{-1} \ll L. \tag{1.1}$$

Similarly, the hadron momentum P and the separation z between space-like separated fields must obey:

$$a \ll P^{-1}, \ z \ll L. \tag{1.2}$$

On the other hand, probing the parton distributions in SDF with a fine resolution in x requires large Ioffe times, $\nu = z \cdot P$, which can be achieved by either large momentum, P, or large separations, z, or both, challenging the hierarchy expressed in eq. (1.2). For large z, the increasing presence of higher twist effects and the breakdown of perturbation theory prevent a clean use of these data in our current state of understanding of the matrix element [21], although the ratio renormalization in SDF reduces some higher twist effects [5, 22]. The major source of improvement in terms of kinematic range on the lattice — which is desirable both for the pseudo- and quasi-distribution formalism — resides therefore in an access to larger momenta P. However, two obvious issues plague the quest for larger momenta: the requirement of finer lattices, and the issue of the exponential decrease of the signal-to-noise ratio.

A similar problem arises for heavy quark physics in lattice QCD, where the lattice spacing must be much smaller than the inverse of the heavy quark mass [23]. A solution arises through the relaxation of the requirement that $L \gg \Lambda_{QCD}^{-1}$. Working in volumes smaller than the hadronic scale allows a considerable alleviation of the computational burden, but requires a careful connection to the infinite volume limit, obtained thanks to the step-scaling function technique [24]. Step-scaling is also used to determine the running of the strong coupling constant in lattice QCD over a range of scales spanning several orders of magnitude [25, 26]. We will explore at the end of this paper the possibility to compute parton distributions in small volumes with larger hadon boost and therefore wider ranges of Ioffe time. Inspired by our work on the numerical evolution operators in SDF, we will propose a step-scaling technique to recover the hadronic parton distribution in infinite volume.

In order to keep the discussion simple and highlight the main points of our work, in this paper we concentrate on the flavor non-singlet unpolarized PDF case. The outline of the paper is as follows. First, we briefly introduce the SDF formalism in section II. Then we derive non-singlet PDF evolution operators in SDF from $\overline{MS}$ perturbative results at all orders in subsection III.A. Finding an appropriate truncation and evaluating some of the systematic effects it generates is discussed in subsection III.B. We perform a numerical application with the perturbative ingredients in subsection III.C. After that, we extract in section IV a non-perturbative SDF evolution operator from actual lattice data for the first time by several methods, and discuss the results in view of the previous perturbative calculations. Finally, in section V, we describe a program to access larger momenta thanks to small volume calculations of non-local matrix elements.

2 PDFs in the short-distance factorization approach

PDFs can be probed through the computation of objects in which quark fields are displaced at space-like separations rather than by light-cone intervals. In this work, we will focus on a non-singlet quark PDFs. The basic building block in this paradigm is the matrix element (in the case of unpolarized quark PDFs of the nucleon):

$$\mathcal{M}^{\alpha}(z, P) \equiv \langle P | \bar{\psi}(0) \frac{\lambda_3}{2} \gamma^{\alpha} \hat{E}(0, z; A) \psi(z) | P \rangle, \qquad (2.1)$$

where λ_3 is a non-singlet flavor projection and $\hat{E}(0, z; A)$ is the $0 \to z$ straight Wilson line gauge link formed by the gauge field A_μ in the fundamental representation of SU(3). The external hadronic states $|P\rangle$ carry momentum P. A Lorentz covariant decomposition of this matrix element yields:

$$\mathcal{M}^{\alpha}(z, P) = P^{\alpha} M(\nu, z^2) + z^{\alpha} N(\nu, z^2), \qquad (2.2)$$

where we have introduced the quantity $\nu = z \cdot P$ known as Ioffe time. In the pseudo-distribution approach, one uses $z = (0, 0, 0, z_3)$, $\alpha = 0$ and the momentum $P = (P_0, 0, 0, P_3)$ such that $\mathcal{M}^0(z, P) = P^0 M(\nu, z^2)$, and forms the Lorentz invariant ratio:

$$\mathfrak{M}(\nu, z^2) \equiv \frac{\mathcal{M}^0(z, P)}{\mathcal{M}^0(z, 0)} \frac{\mathcal{M}^0(0, 0)}{\mathcal{M}^0(0, P)}, \qquad (2.3)$$

which is finite in the continuum limit, requires no renormalization, and is directly related to the PDF. The fact that the soft physics contained in the z^2-dependent matrix element is similar to that of ordinary PDFs up to higher-twist contributions allows one to interpret these matrix elements as PDFs expressed within another factorization scheme, the SDF. The approach is formalized through a non-local OPE, in which short distance contributions are computed perturbatively, as discussed, e.g., in [4, 8, 10, 15].

As a consequence, the matrix elements computed on the lattice can be related to the Fourier transform of $\overline{MS}$ PDFs by:

$$\mathfrak{M}(\nu, z^2) = \int_0^1 d\alpha \, \mathcal{C}(\alpha, z^2 \mu^2, \alpha_s(\mu^2)) \mathcal{Q}(\alpha\nu, \mu^2) + z^2 \mathcal{B}(\nu, z^2), \tag{2.4}$$

where $\alpha_s(\mu^2)$ is the strong coupling. $\mathcal{C}(\alpha, z^2\mu^2, \alpha_s(\mu^2))$ is called the matching kernel, computed in perturbation theory, and $\mathcal{Q}(\nu, \mu^2)$ is the (normalized) Ioffe-time distribution (ITD) [27], defined by:

$$\mathcal{Q}(\nu, \mu^2) \equiv \int_{-1}^1 dx \, e^{i\nu x} q(x, \mu^2) \Big/ \int_{-1}^1 dx \, q(x, \mu^2), \tag{2.5}$$

where $q(x, \mu^2)$ is the $\overline{MS}$ PDF. Notice that the ITD is a complex quantity, whose real part probes the x-even part of $q(x, \mu^2)$ and the imaginary part probes the x-odd part of $q(x, \mu^2)$. The term $z^2 \mathcal{B}(\nu, z^2)$ in eq. (2.4) captures additional corrections to the leading order expression, which vanish in the limit of $z^2 = 0$. In particular, because the reduced function by construction satisfies $\mathfrak{M}(0, z^2) = 1$, one can see that $\mathcal{B}(0, z^2) = 0$ and

$$\int_0^1 d\alpha \, \mathcal{C}(\alpha, z^2 \mu^2, \alpha_s(\mu^2)) = 1. \tag{2.6}$$

The expansion to fixed order $\mathcal{O}(\alpha_s)$ of the matching kernel $\mathcal{C}(\alpha, z^2\mu^2, \alpha_s(\mu^2))$ yields [4, 15, 28]:

$$\mathcal{C}(\alpha, z^2\mu^2, \alpha_s(\mu^2)) = \delta(1-\alpha) - \frac{\alpha_s(\mu^2)}{2\pi} \left[\ln\left(-z^2\mu^2 \frac{e^{2\gamma_E+1}}{4}\right) B_1(\alpha) + D(\alpha) \right] + \mathcal{O}(\alpha_s^2), \tag{2.7}$$

where γ_E is the Euler-Mascheroni constant, and $B_1(\alpha)$ and $D(\alpha)$ are given by:

$$B_1(\alpha) = C_F \left[\frac{1+\alpha^2}{1-\alpha} \right]_+ , \quad D(\alpha) = C_F \left[4\frac{\ln(1-\alpha)}{1-\alpha} - 2(1-\alpha) \right]_+ . \tag{2.8}$$

Here we follow the two prescriptions,

$$f(x) = \int_0^1 d\alpha \, f(\alpha x) \delta(1-\alpha), \tag{2.9}$$

and

$$G(\alpha)_+ \equiv G(\alpha) - \delta(\alpha) \int_0^1 d\alpha' \, G(\alpha'). \tag{2.10}$$

The expansion of the matching kernel up to order $\mathcal{O}(\alpha_s^2)$ can be found, e.g., in [16, 17].

3 Evolution operators in perturbation theory

In this section, we will derive evolution equations in the SDF scheme based on the perturbative matching kernels. For convenience, we will only consider a non-singlet PDF so as to not have to deal with the complication of gluon mixing.

3.1 All-order expression

The DGLAP differential evolution equation of unpolarized non-singlet PDFs reads [29–31]:

$$\frac{d}{d\ln\mu^2}\,q(x,\mu^2) = \int_x^1 \frac{d\alpha}{\alpha}\,B(\alpha,\alpha_s(\mu^2))\,q\left(\frac{x}{\alpha},\mu^2\right),\tag{3.1}$$

where $B(\alpha,\alpha_s(\mu^2))$ is known as the DGLAP splitting function and admits a perturbative expansion as:

$$B(\alpha,\alpha_s(\mu^2)) = \sum_{n=1}^{\infty}\left(\frac{\alpha_s(\mu^2)}{2\pi}\right)^n B_n(\alpha).\tag{3.2}$$

The expression of $B_1(\alpha)$ has already been given in eq. (2.8), obtained by expanding the matching kernel relating z^2-dependent matrix elements to μ^2-dependent PDFs. We clarify the reason the matching kernel contains the expansion of the DGLAP splitting function later on.

Performing the integration of the differential evolution equation (3.1), we introduce the DGLAP integrated evolution operator from μ_0^2 to μ_1^2, denoted $\mathcal{E}(\alpha;\mu_0^2,\mu_1^2)$, such that:

$$q(x,\mu_1^2) = \int_x^1 \frac{d\alpha}{\alpha}\,\mathcal{E}(\alpha;\mu_0^2,\mu_1^2)\,q\left(\frac{x}{\alpha},\mu_0^2\right).\tag{3.3}$$

Anticipating the concepts presented at the end of this paper, we will refer to $\mathcal{E}(\alpha;\mu_0^2,\mu_1^2)$ also as the "step-scaling" function in the $\overline{MS}$ scheme, as it characterizes scale dependence in a "step" from μ_0^2 to μ_1^2. For brevity, we will from now on use the symbolical notation for the DGLAP convolution:

$$[f(\mathrm{param}_f)\otimes g(\mathrm{param}_g)](x) \equiv \int_x^1 \frac{d\alpha}{\alpha}\,f(\alpha,\mathrm{param}_f)\,g\left(\frac{x}{\alpha},\mathrm{param}_g\right).\tag{3.4}$$

It is trivial to verify that $f\otimes g = g\otimes f$, and that $1_\otimes : \alpha \mapsto \delta(1-\alpha)$ is the identity element $(f\otimes 1_\otimes = f)$. We will also use the notation $f^{\otimes n}$ to denote the repeated convolution of f with itself n times, and $f^{\otimes -1}$ the inverse for the convolution defined by:

$$f^{\otimes -1}\otimes f = 1_\otimes.\tag{3.5}$$

There exist some non-trivial considerations on the existence and properties of this inverse, which we will address in the next section. To exemplify the use of the notation, the DGLAP integrated evolution equation (3.4) now reads:

$$q(\mu_1^2) = \mathcal{E}(\mu_0^2,\mu_1^2)\otimes q\left(\mu_0^2\right),\tag{3.6}$$

and $\mathcal{E}(\mu_0^2,\mu_1^2)^{\otimes -1}$ represents the DGLAP backward evolution operator, that is $\mathcal{E}(\mu_1^2,\mu_0^2)$.

The Fourier transform of the DGLAP differential evolution equation (3.1) gives the differential evolution equation of the ITD [27]:

$$\frac{d}{d\ln\mu^2}\, \mathcal{Q}(\nu,\mu^2) = \int_0^1 d\alpha\, B(\alpha,\alpha_s(\mu^2))\, \mathcal{Q}(\alpha\nu,\mu^2)\,. \tag{3.7}$$

Applying the Fourier transform has changed the nature of the convolution compared to the one we have dubbed by $\otimes$, although the splitting function B is unchanged. Among the obvious differences, the integral runs from 0 to 1 instead of x to 1. Consequently, the evolution of ITDs probes the entire range of the evolution operator, regardless of the value of the Ioffe time ν, whereas evolution of momentum-dependent PDFs only probes the evolution operators for $\alpha > x$. Therefore, extracting ITDs provides, in principle, a better setting for the study of evolution than working in x space — although obtaining a good resolution requires as always to access large values of Ioffe time.

Observing that the z^2-dependent matrix element, whose relation to the ITD is stated in eq. (2.4), has no μ^2 dependence, we write, assuming z^2 is sufficiently small that power corrections are suppressed:

$$0 = \frac{d}{d\ln\mu^2}\mathfrak{M}(\nu,z^2) = \int_0^1 d\alpha\, \frac{d}{d\ln\mu^2}\mathcal{C}(\alpha,z^2\mu^2,\alpha_s(\mu^2))\, \mathcal{Q}(\alpha\nu,\mu^2)$$
$$+ \int_0^1 d\beta\, \mathcal{C}(\beta,z^2\mu^2,\alpha_s(\mu^2)) \int_0^1 d\alpha\, B(\alpha,\alpha_s(\mu^2))\, \mathcal{Q}(\alpha\beta\nu,\mu^2) + \mathcal{O}(z^2)\,, \tag{3.8}$$

with the leading power term giving:

$$\int_0^1 d\beta\, \mathcal{C}(\beta,z^2\mu^2,\alpha_s(\mu^2)) \int_0^\beta \frac{d\alpha}{\beta} B\left(\frac{\alpha}{\beta},\alpha_s(\mu^2)\right) \mathcal{Q}(\alpha\nu,\mu^2) =$$
$$\int_0^1 d\alpha\, \mathcal{Q}(\alpha\nu,\mu^2) \int_\alpha^1 \frac{d\beta}{\beta} \mathcal{C}(\beta,z^2\mu^2,\alpha_s(\mu^2)) B\left(\frac{\alpha}{\beta},\alpha_s(\mu^2)\right)\,. \tag{3.9}$$

The latter term contains the DGLAP convolution, and the combination of eqs. (3.8) and (3.9) gives therefore that the renormalization group equation of the matching kernel is exactly the opposite of the ordinary DGLAP equation:

$$\frac{d}{d\ln\mu^2}\mathcal{C}(z^2\mu^2,\alpha_s(\mu^2)) = -B(\alpha_s(\mu^2))\otimes\mathcal{C}\left(z^2\mu^2,\alpha_s(\mu^2)\right)\,. \tag{3.10}$$

This relation already expressed in various forms in the literature (see e.g. [32–34]) explains the observation made earlier that the matching kernel contains the terms of the expansion of the DGLAP splitting function. Due to the sign flip compared to the ordinary differential equation, the μ^2 dependence of the matching kernel is dictated by the DGLAP backward evolution operator (compare to eq. (3.6)):

$$\mathcal{C}(z^2\mu^2,\alpha_s(\mu^2)) = \mathcal{E}(\mu^2,\mu'^2)\otimes\mathcal{C}(z^2\mu'^2,\alpha_s(\mu'^2))\,. \tag{3.11}$$

In short, the matching can be performed at an intermediate scale μ'^2 and evolved later without any consequences, because of a trade-off between the matching and the DGLAP
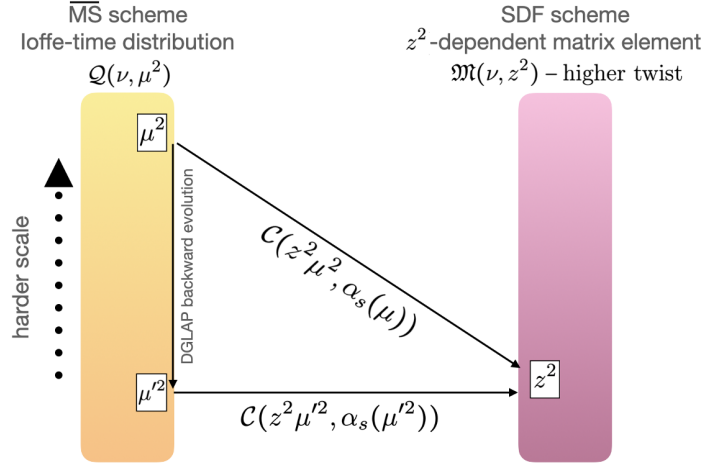
Figure 1. The matching to any scale in $\overline{MS}$ can be performed by dividing the operation into matching to an intermediate scale and evolution to the desired final scale as long as all orders are considered.

evolution that occurs if all orders are considered. The physical content of eq. (3.11) is very intuitive, and summarized in figure 1. Of course, if one uses a finite truncation, then the scale at which the matching is done actually matters.

We have reviewed the evolution equations of the PDF (3.6), Ioffe-time distribution (3.7) and matching kernel (3.11). We are now ready to derive the final result of this subsection, namely the evolution equation with respect to z^2 of the matrix element.

Expressing the matrix element (with the subtraction of the higher-twist corrections) as a function of itself at a different value of z^2, some algebra gives:

$$\mathfrak{M}(\nu, z_1^2) = \int_0^1 d\alpha \, \mathcal{C}(\alpha, z_1^2 \mu^2, \alpha_s(\mu^2)) \mathcal{Q}(\alpha\nu, \mu^2), \tag{3.12}$$

$$= \int_0^1 d\alpha \, \mathfrak{M}(\alpha\nu, z_0^2) \int_\alpha^1 \frac{d\beta}{\beta} \mathcal{C}\left(\frac{\alpha}{\beta}, z_1^2 \mu^2, \alpha_s(\mu^2)\right) \mathcal{C}^{\otimes -1}(\beta, z_0^2 \mu^2, \alpha_s(\mu^2)). \tag{3.13}$$

The SDF step-scaling function is therefore given by:

$$\Sigma(z_0^2, z_1^2) \equiv \mathcal{C}(z_1^2 \mu^2, \alpha_s(\mu^2)) \otimes \mathcal{C}^{\otimes -1}(z_0^2 \mu^2, \alpha_s(\mu^2)). \tag{3.14}$$

The SDF step-scaling function is independent of μ^2 if all orders are considered. To see this, we notice that introducing an intermediate scale $(\lambda z^2)^{-1}$ where $\lambda \equiv -e^{2\gamma_E + 1}/4$ gives a simplified form of the matching kernel without logarithmic dependence in eq. (2.7):

$$\mathcal{C}_0(z^2) \equiv \mathcal{C}\left(\frac{1}{\lambda}, \alpha_s((\lambda z^2)^{-1})\right) = 1_\otimes - \frac{\alpha_s((\lambda z^2)^{-1})}{2\pi} D + \mathcal{O}(\alpha_s^2). \tag{3.15}$$

We explore in more depth the perturbative expansion of the SDF matching kernel, how it relates to the DGLAP splitting function at higher orders, and the question of the scale setting
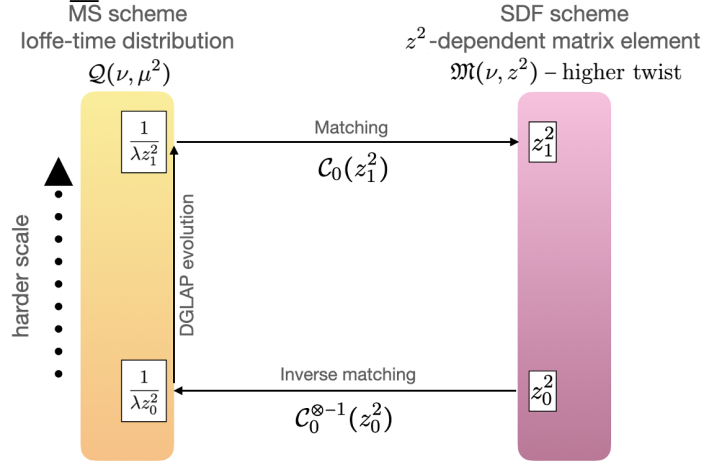
Figure 2. The evolution in SDF depending on z^2 is derived from the $\overline{MS}$ evolution by a back-and-forth matching procedure.

subtended by this coefficient λ in appendix A. Then eq. (3.11) gives:

$$\Sigma(z_0^2, z_1^2) = \mathcal{E}\left(\mu^2, \frac{1}{\lambda z_1^2}\right) \otimes \mathcal{C}_0(z_1^2) \otimes \mathcal{E}\left(\frac{1}{\lambda z_0^2}, \mu^2\right) \otimes \mathcal{C}_0^{\otimes -1}(z_0^2), \tag{3.16}$$

$$= \mathcal{E}\left(\frac{1}{\lambda z_0^2}, \frac{1}{\lambda z_1^2}\right) \otimes \mathcal{C}_0(z_1^2) \otimes \mathcal{C}_0^{\otimes -1}(z_0^2). \tag{3.17}$$

The latter expression shows how the SDF step-scaling function can be computed perturbatively to any order from the $\overline{MS}$ one. The content of eq. (3.17) is again intuitive, and summarized in figure 2.

By differentiating eq. (3.17) with respect to $\ln z_1^2$, one can derive SDF splitting functions:

$$\frac{d}{d\ln z_1^2}\Sigma(z_0^2, z_1^2) = \frac{d}{d\ln z_1^2}\mathcal{E}\left(\frac{1}{\lambda z_0^2}, \frac{1}{\lambda z_1^2}\right) \otimes \mathcal{C}_0(z_1^2) \otimes \mathcal{C}_0^{\otimes -1}(z_0^2)$$

$$+ \mathcal{E}\left(\frac{1}{\lambda z_0^2}, \frac{1}{\lambda z_1^2}\right) \otimes \frac{d}{d\ln z_1^2}\mathcal{C}_0(z_1^2) \otimes \mathcal{C}_0^{\otimes -1}(z_0^2). \tag{3.18}$$

The first term produces the opposite of the ordinary DGLAP splitting function:

$$\frac{d}{d\ln z_1^2}\mathcal{E}\left(\frac{1}{\lambda z_0^2}, \frac{1}{\lambda z_1^2}\right) = -B(\alpha_s((\lambda z_1^2)^{-1})) \otimes \mathcal{E}\left(\frac{1}{\lambda z_0^2}, \frac{1}{\lambda z_1^2}\right), \tag{3.19}$$

whereas the second term can be reshaped by noticing that:

$$\mathcal{E}\left(\frac{1}{\lambda z_0^2}, \frac{1}{\lambda z_1^2}\right) \otimes \mathcal{C}_0^{\otimes -1}(z_0^2) = \Sigma(z_0^2, z_1^2) \otimes \mathcal{C}_0^{\otimes -1}(z_1^2). \tag{3.20}$$

Therefore, we obtain:

$$\frac{d}{d\ln z_1^2}\Sigma(z_0^2, z_1^2) = \left\{-B(\alpha_s((\lambda z_1^2)^{-1})) + \frac{d}{d\ln z_1^2}\mathcal{C}_0(z_1^2) \otimes \mathcal{C}_0^{\otimes -1}(z_1^2)\right\} \otimes \Sigma(z_0^2, z_1^2). \tag{3.21}$$

Alternatively, using the running of the strong coupling defined by:

$$\frac{d}{d\ln\mu^2}\alpha_s(\mu^2) = \beta(\alpha_s(\mu^2)), \tag{3.22}$$

we find:

$$\frac{d}{d\ln z_1^2}\Sigma(z_0^2,z_1^2)=\left\{-B(\alpha_s((\lambda z_1^2)^{-1}))-\beta(\alpha_s((\lambda z_1^2)^{-1}))\frac{d}{d\alpha_s}\mathcal{C}_0(z_1^2)\otimes\mathcal{C}_0^{\otimes-1}(z_1^2)\right\}\otimes\Sigma(z_0^2,z_1^2).\tag{3.23}$$

The bracketed result of eq. (3.23) is the SDF splitting function. This result closely resembles the "physical anomalous dimensions" derived in [35–37] for the Q^2 dependence of the structure functions of deep inelastic scattering (DIS) — or the results obtained for the renormalization relating dimensionally regularized quantities to $\overline{MS}$ (see, for instance, eq. (9) of [38]). This does not come as a surprise since $\mathfrak{M}(\nu,z^2)$ is scheme-independent. Therefore, the SDF step-scaling and splitting functions are independent of the scheme used to compute them in perturbation theory, provided all orders are considered. Notably, scheme independence can be seen by recognizing that $\beta(\alpha_s)\,d/d\alpha_s$ is scheme-invariant and that the scheme dependence of the first term cancels that of the second term. However, any truncation of the perturbative expansion introduces scheme dependence.

3.2 Performing the perturbative truncation

Eqs. (3.17) and (3.23) give, respectively, the SDF step-scaling function and the SDF splitting function at any desired order in perturbation theory. To derive the SDF perturbative evolution, we only need to find a consistent expression for $\mathcal{C}_0^{\otimes-1}$. For instance, if one works at order $\mathcal{O}(\alpha_s)$ for the matching kernel (3.15), one would be tempted to use:

$$\mathcal{C}_0^{\otimes-1}(z_0^2)=\left(1_\otimes-\frac{\alpha_s(\mu_0^2)}{2\pi}D\right)^{\otimes-1},\tag{3.24}$$

where $\mu_0^2=(\lambda z_0^2)^{-1}$. However, that this expression defines an actual distribution is doubtful. Working in Mellin space allows one to pinpoint the potential issue. The Mellin moments of the matching kernel are defined as:

$$c_n(z^2)\equiv\int_0^1 d\alpha\,\alpha^{n-1}\mathcal{C}_0\left(\alpha,z^2\right).\tag{3.25}$$

It is well-known and easy to verify that the Mellin transform converts the DGLAP convolution $\otimes$ into an ordinary product of Mellin moments, and that, provided $\mathcal{C}_0^{\otimes-1}$ exists, its Mellin moments are simply $1/c_n$. An issue arises obviously if $c_n=0$. We refer the interested reader to appendix B for more details on the consequences of this pole in the Mellin moments of $\mathcal{C}_0^{\otimes-1}$. The intuitive conclusion of appendix B is that one can obtain a workable definition of $\mathcal{C}_0^{\otimes-1}$ provided that one truncates it equally to the same perturbative order, as done e.g. in [39, 40]. Here:

$$\mathcal{C}_0^{\otimes-1}(z_0^2)=1_\otimes+\frac{\alpha_s(\mu_0^2)}{2\pi}D.\tag{3.26}$$

Then the direct relation between the SDF and $\overline{MS}$ step-scaling functions (3.17) yields:

$$\Sigma(z_0^2,z_1^2)=\mathcal{E}(\mu_0^2,\mu_1^2)\otimes\left(1_\otimes-\frac{\alpha_s(\mu_1^2)}{2\pi}D\right)\otimes\left(1_\otimes+\frac{\alpha_s(\mu_0^2)}{2\pi}D\right).\tag{3.27}$$

A contribution of order $\mathcal{O}(\alpha_s^2)$ results in that $\Sigma(z_0^2, z_0^2) \neq 1_\otimes$, so not applying any evolution does not correspond to the identity operator. To preserve $\Sigma(z_0^2, z_0^2) = 1_\otimes$, a further truncation is required yielding the final form of the SDF step-scaling function matched at fixed order $\mathcal{O}(\alpha_s)$. We will be using this form for our numerical applications:

$$\Sigma(z_0^2, z_1^2) = \mathcal{E}(\mu_0^2, \mu_1^2) \otimes \left(1_\otimes + \frac{\alpha_s(\mu_0^2) - \alpha_s(\mu_1^2)}{2\pi} D \right). \tag{3.28}$$

3.2.1 Alternative truncation choices

In the $\overline{MS}$ scheme a fixed-order truncation at the level of the step-scaling function is sub-optimal, and a fixed-order truncation at the level of the splitting function is preferred. When integrating the differential equation with the truncated splitting function, one obtains a step-scaling function with renormalization group improvement — that is the resummation of the dominant logarithms of the ratio of scales, classified as LL (leading logarithms resummed), NLL (next-to-leading logarithms), and so on. Eq. (3.28) presents a factorized form of the step-scaling function, with an evolution and a matching term. It is straightforward to implement the $\overline{MS}$ renormalization group improvement on $\mathcal{E}$. On the other hand, the truncation we have suggested is of fixed-order at the level of the matching term. Since we have computed the SDF splitting function, we could try to introduce a renormalization group improvement of the matching part as well, using eq. (3.23), instead of working directly with the step-scaling functions (3.17).

Applying our $\mathcal{O}(\alpha_s)$ truncated $\mathcal{C}_0$ and $\mathcal{C}_0^{\otimes -1}$ in eq. (3.23) yields:

$$\frac{d}{d \ln z_1^2} \Sigma(z_0^2, z_1^2) = \left\{ -B(\alpha_s(\mu_1^2)) + \frac{\beta(\alpha_s(\mu_1^2))}{2\pi} D \otimes \left[1_\otimes + \frac{\alpha_s(\mu_1^2)}{2\pi} D \right] \right\} \otimes \Sigma(z_0^2, z_1^2). \tag{3.29}$$

However, this expression lends itself to multiple different perturbative truncations depending on the ingredients one chooses to use. Indeed, as $\beta(\alpha_s) = -\beta_0 \alpha_s^2 + \ldots$, the one-loop direct matching kernel D only enters at order $\mathcal{O}(\alpha_s^2)$ in eq. (3.29). The one-loop inverse matching kernel only enters at order $\mathcal{O}(\alpha_s^3)$. If one used a fixed-order truncation of eq. (3.29) to order $\mathcal{O}(\alpha_s^3)$ for instance, one would be manipulating simultaneously the three-loop DGLAP splitting function, two-loop direct matching kernel and one-loop inverse matching kernel, in a curious mixture from a physical point of view.

Instead, we propose to illustrate the similarities and differences between Eqs (3.29) and (3.28) by expressing both with the same physical ingredients, namely one-loop DGLAP splitting function ($\mathcal{O}(\alpha_s)$), matching kernel ($\mathcal{O}(\alpha_s)$), and beta function ($\mathcal{O}(\alpha_s^2)$).

Let us first review the effect of renormalization group improvement at one-loop on the $\overline{MS}$ step-scaling function. The LL evolution operator, $\mathcal{E}_{LL}(\mu_0^2, \mu_1^2)$ is given by the solution of:

$$\begin{cases} \dfrac{d}{d \ln \mu_1^2} \mathcal{E}_{LL}(\mu_0^2, \mu_1^2) = \dfrac{\alpha_s(\mu_1^2)}{2\pi} B_1 \otimes \mathcal{E}_{LL}(\mu_0^2, \mu_1^2), \\ \dfrac{d}{d \ln \mu_1^2} \alpha_s(\mu_1^2) = -\beta_0 \alpha_s^2(\mu_1^2). \end{cases} \tag{3.30}$$

This system of differential equations yields:

$$\mathcal{E}_{LL}(\mu_0^2, \mu_1^2) = \sum_{k=0}^{+\infty} \frac{1}{k!} B_1^{\otimes k} \left(\frac{1}{2\pi \beta_0} \ln \left(\frac{\alpha_s(\mu_0^2)}{\alpha_s(\mu_1^2)} \right) \right)^k. \tag{3.31}$$

The result looks more familiar when expressed in terms of Mellin moments. If γ_n and $e_n(\mu_0^2, \mu_1^2)$ are respectively the Mellin moments of B_1 and $\mathcal{E}_{LL}(\mu_0^2, \mu_1^2)$ defined analogous to eq. (3.25), then:

$$e_n(\mu_0^2, \mu_1^2) = \left(\frac{\alpha_s(\mu_0^2)}{\alpha_s(\mu_1^2)} \right)^{\gamma_n/(2\pi\beta_0)}. \tag{3.32}$$

A crucial observation is that, although B_1 is mathematically a distribution defined as a plus-prescription according to eq. (2.8), $\mathcal{E}_{LL}(\mu_0^2, \mu_1^2)$ is an ordinary function as long as $\mu_1 > \mu_0$: the $\overline{MS}$ renormalization group improvement has the effect of removing the plus-prescription that appears in the splitting function. This effect is not evident as $\mathcal{E}_{LL}(\mu_0^2, \mu_1^2)$ is written as an infinite sum of convolutions of B_1, as seen in eq. (3.31).

As we will show below, $\mathcal{E}_{LL}(\mu_0^2, \mu_1^2)$ is not only an ordinary function, but a positive one for $\mu_1 > \mu_0$. In the context of the probabilistic interpretation of the PDF $q(x, \mu^2)$ as the number density of partons carrying a momentum fraction x at scale μ^2, one could likewise interpret $\mathcal{E}_{LL}(\alpha, \mu_0^2, \mu_1^2)$ as the probability of finding a parton with momentum fraction αx at scale μ_1^2 inside a parton with momentum fraction x at scale μ_0^2. Indeed, we remind that the probability density of the product of two independent probability densities f and g restricted to the interval $[0,1]$ is precisely $f \otimes g$. This picture has been used to derive the DGLAP evolution equation [41]. Note, however, that the probabilistic interpretation may appear problematic, as it is established that renormalization may spoil the positivity of PDFs at low $\overline{MS}$ scales [42].

Aside from probabilistic considerations, the existence of evolution operators, whether in μ^2 or z^2, can be tied to the validity of the OPE that defines the moments of (pseudo-) PDFs [22] at small z^2. Provided that the moments exist, in Mellin space the DGLAP convolution reduces to:

$$q_n(\mu_1^2) = e_n(\mu_0^2, \mu_1^2) q_n(\mu_0^2), \tag{3.33}$$

where $q_n(\mu^2)$ are the Mellin moments of $q(x, \mu^2)$. Eq. (3.33) defines the evolution operator straightforwardly as the inverse Mellin transform of $e_n(\mu_0^2, \mu_1^2) = q_n(\mu_1^2)/q_n(\mu_0^2)$.[1] However, one still needs to demonstrate that the operator defined in this fashion is restricted to a support $\alpha \in [0,1]$, which can be done to all orders in perturbation theory thanks to the use of partonic on-shell states (see e.g. [48]).

Let us mention another perspective, related to the arguments of growth of higher-twists and breaking of perturbation theory mentioned in the introduction, on why the evolution operators $\mathcal{E}$ and Σ are limited to small z^2. The factorization scale of PDFs is akin to a cut-off in the $k_\perp$ integration of transverse momentum of partons inside the hadron. One can introduce

---

[1] An enlightening parallel could be drawn with the Radon transform and the polynomiality property of generalized parton distributions (GPDs). The polynomiality property implies a sophisticated dependence of GPDs on their variables (x, ξ) [43, 44]. However, if one applies an inverse Radon transform to GPDs to express them under the form of "double distributions" [45], the polynomiality property becomes trivially verified [46, 47]. In the end, the polynomiality property is converted into a requirement of well-definition of the inverse Radon transform with additional requirements on the support of the distributions. In our case, the DGLAP relation implies a sophisticated dependence of PDFs on their variables (x, μ^2) that can be converted into a requirement of a well-defined inverse Mellin transform with again some support properties. That the Mellin transform is applicable is ultimately guaranteed by the validity of the OPE.

a "primordial" (straight-link) transverse momentum distribution (TMD) $\mathcal{F}(x, k_\perp)$ [3, 49, 50] whose Fourier transforms give the pseudo-ITD:

$$\mathfrak{M}(\nu, z_\perp^2) = \int dx \int d^2 k_\perp e^{i\nu x + i k_\perp \cdot z_\perp} \mathcal{F}(x, k_\perp^2). \tag{3.34}$$

With $z^2 = 0$, one can write schematically the PDF evolution equation as:

$$\frac{dq(x, \mu^2)}{d\ln \mu^2} = \frac{d}{d\ln \mu^2} \int_0^{\mu^2} d^2 k_\perp \mathcal{F}(x, k_\perp^2). \tag{3.35}$$

The large $k_\perp$ region of the TMD can be computed in perturbation theory — in renormalizable theories, it varies as $1/k_\perp^2$, which integrated against $d^2 k_\perp$ gives the usual logarithmic scale dependence at large μ^2 or small z^2 of the PDFs. On the other hand, the small $k_\perp$ (soft) regime of the TMD is deeply non-perturbative and tied to the confinement and finite size of the proton. Typical models of the soft TMD component have a very different form compared to the hard tail, such as a Gaussian dependence in $k_\perp$. The evolution equation is therefore fundamentally different at small $k_\perp$ or large z, and there are no first-principle grounds that the step-scaling definition we have used would hold at large z — except maybe for the hope that the cancellation of z^2 effects induced by the ratio renormalization results in a trivial evolution at large z.

Finally, note that regardless of the probabilistic interpretation attributed to $\mathcal{E}_{LL}(\alpha, \mu_0^2, \mu_1^2)$, it is a positive ordinary function, which allows one to consider evolution to a higher scale as a simple reweighting of the parton distribution at an initial scale. This forms the basis of the study led in [51] which deals with the question "which region of the parton distribution at initial scale contributes the most to which region at final scale?".

We will only show that $\mathcal{E}_{LL}$ is an ordinary function in the main text, the positivity being demonstrated in appendix C. That $\mathcal{E}_{LL}$ does not present the same distribution-like features as B_1 can be broadly understood as a consequence of the fact that $\mathcal{E}_{LL}$ results from the integration of B_1, and therefore smooths out the singular behavior of the latter. The argument can be made rigorous in Mellin space, where the distribution-like features, which are localized at $\alpha = 1$, are mirrored in the large n behavior of the Mellin moments. Indeed, the Mellin moments of an ordinary integrable function whose support is restricted to $[0, 1]$ tend to 0 as $n \to +\infty$. A Dirac delta at $\alpha = 1$ induces Mellin moments which tend to a non-zero constant as $n \to +\infty$, whereas a plus-prescription $\mathcal{O}(1/(1-\alpha))_+$ induces moments diverging logarithmically at large n. That is the case for the Mellin moments of B_1, known as the leading order (LO) DGLAP anomalous dimensions, γ_n:

$$\gamma_n \overset{n \to +\infty}{=} -2 C_F \ln(n) + C_F \left(\frac{3}{2} - 2\gamma_E\right) + \mathcal{O}\left(\frac{1}{n}\right). \tag{3.36}$$

Therefore, the large n behavior of the Mellin moments of $\mathcal{E}_{LL}(\mu_0^2, \mu_1^2)$ reads:

$$e_n(\mu_0^2, \mu_1^2) = \left(\frac{\alpha_s(\mu_0^2)}{\alpha_s(\mu_1^2)}\right)^{\gamma_n/(2\pi\beta_0)} \overset{n \to +\infty}{\sim} A n^{-\frac{C_F}{\pi\beta_0} \ln\left(\frac{\alpha_s(\mu_0^2)}{\alpha_s(\mu_1^2)}\right)}, \tag{3.37}$$

where

$$A = \left(\frac{\alpha_s(\mu_0^2)}{\alpha_s(\mu_1^2)}\right)^{C_F(3-4\gamma_E)/(4\pi\beta_0)}. \tag{3.38}$$

When $\mu_1 > \mu_0$, we have $\alpha_s(\mu_0^2)/\alpha_s(\mu_1^2) > 1$, so $e_n(\mu_0^2, \mu_1^2)$ falls off as a negative power of n. A straightforward inverse Mellin transform implies a behavior of $\mathcal{E}_{LL}(\alpha, \mu_0^2, \mu_1^2)$ for $\alpha \to 1$ dominated by the ordinary function:

$$\mathcal{E}_{LL}(\alpha, \mu_0^2, \mu_1^2) \overset{\alpha \to 1}{\sim} (-\ln(\alpha))^{-1 + \frac{C_F}{\pi \beta_0} \ln\left(\frac{\alpha_s(\mu_0^2)}{\alpha_s(\mu_1^2)}\right)} \times \frac{A}{\Gamma\left(\frac{C_F}{\pi \beta_0} \ln\left(\frac{\alpha_s(\mu_0^2)}{\alpha_s(\mu_1^2)}\right)\right)}, \qquad (3.39)$$

where Γ is the ordinary Gamma function. The LL $\overline{MS}$ step-scaling function therefore only presents an integrable power divergence at $\alpha = 1$ when $\mu_1 > \mu_0$. On the contrary, if $\mu_1 = \mu_0$, the step-scaling function reduces to the identity $1_\otimes$, which is a Dirac delta at $\alpha = 1$. If $\mu_1 < \mu_0$, then $e_n(\mu_0^2, \mu_1^2)$ increases at large n as a power of n. This shows that the LL backward DGLAP evolution operator presents distribution-like features in the vicinity of $\alpha = 1$, which explains the observation that backward evolution is considerably noisier than the forward one: renormalization group improvement does not always smooth out the distribution-like features.

The large n behavior of the Mellin moments of the matching kernel, presented in more details in eq. (B.3), diverges as a power of $\ln(n)$ as the matching kernel contains a plus-prescription:

$$c_n(z^2) \overset{n \to +\infty}{\sim} -\frac{\alpha_s(\mu^2)}{\pi} C_F \ln^2(n). \qquad (3.40)$$

Therefore, the SDF step-scaling function derived in eq. (3.28) has Mellin moments $\sigma_n(z_0^2, z_1^2)$ that satisfy:

$$\sigma_n(z_0^2, z_1^2) \overset{n \to +\infty}{\sim} e_n\left(\mu_0^2, \mu_1^2\right) \frac{\alpha_s(\mu_0^2) - \alpha_s(\mu_1^2)}{\pi} C_F \ln^2(n). \qquad (3.41)$$

If $z_0^2 > z_1^2$, e_n falls off like a negative power of n, and the $\ln^2(n)$ term added by the matching kernels does not change that general behavior: the SDF step-scaling function remains an ordinary function. An inverse Mellin transform gives that the large α behavior of the SDF step-scaling function is multiplied by $\ln^2(1-\alpha)$ compared to the LL $\overline{MS}$ step-scaling function.

On the other hand, eq. (3.29) yields:

$$\begin{cases} \dfrac{d}{d\ln\mu_1^2} \Sigma(z_0^2, z_1^2) = \left\{ \dfrac{\alpha_s(\mu_1^2)}{2\pi} B_1 + \beta_0 \dfrac{\alpha_s^2(\mu_1^2)}{2\pi} D + \beta_0 \dfrac{\alpha_s^3(\mu_1^2)}{(2\pi)^2} D^2 \right\} \otimes \Sigma(z_0^2, z_1^2), \\ \dfrac{d}{d\ln\mu_1^2} \alpha_s(\mu_1^2) = -\beta_0 \alpha_s^2(\mu_1^2). \end{cases} \qquad (3.42)$$

The solution, which one may verify by differentiating the result and observing that it obeys the previous system, is obtained as:

$$\Sigma(z_0^2, z_1^2)$$
$$= \mathcal{E}_{LL}(\mu_0^2, \mu_1^2) \otimes \left(\sum_{k=0}^{+\infty} \frac{1}{k!} D^{\otimes k} \left(\frac{\alpha_s(\mu_0^2) - \alpha_s(\mu_1^2)}{2\pi} \right)^k \right) \otimes \left(\sum_{k=0}^{+\infty} \frac{1}{k!} D^{\otimes 2k} \left(\frac{\alpha_s^2(\mu_0^2) - \alpha_s^2(\mu_1^2)}{8\pi^2} \right)^k \right), \qquad (3.43)$$

or noting d_n the Mellin moments of D:

$$\sigma_n(z_0^2, z_1^2) = \left(\frac{\alpha_s(\mu_0^2)}{\alpha_s(\mu_1^2)} \right)^{\gamma_n/(2\pi\beta_0)} \exp\left(d_n \frac{\alpha_s(\mu_0^2) - \alpha_s(\mu_1^2)}{2\pi} + d_n^2 \frac{\alpha_s^2(\mu_0^2) - \alpha_s^2(\mu_1^2)}{8\pi^2} \right). \qquad (3.44)$$

By expanding eq. (3.43) to order $\mathcal{O}(\alpha_s)$, we obtain:

$$\Sigma(z_0^2, z_1^2) = \mathcal{E}_{LL}(\mu_0^2, \mu_1^2) \otimes \left(1 + \frac{\alpha_s(\mu_0^2) - \alpha_s(\mu_1^2)}{2\pi} D + \mathcal{O}(\alpha_s^2)\right), \tag{3.45}$$

which is exactly the form of eq. (3.28): one finds naturally that the fixed-order truncation of the matching gives the same result as the renormalization group improved one up to higher order corrections. However, the dominant term at large n of the Mellin moments of the step-scaling function is now (see eq. (B.3) for more details):

$$e_n(\mu_0^2, \mu_1^2) \exp\left(d_n \frac{\alpha_s(\mu_0^2) - \alpha_s(\mu_1^2)}{2\pi} + d_n^2 \frac{\alpha_s^2(\mu_0^2) - \alpha_s^2(\mu_1^2)}{8\pi^2}\right)$$

$$\stackrel{n \to \pm\infty}{=} e_n(\mu_0^2, \mu_1^2) \exp\left(C_F^2 \ln^4(n) \frac{\alpha_s^2(\mu_0^2) - \alpha_s^2(\mu_1^2)}{2\pi^2} + \mathcal{O}(\ln^3(n))\right). \tag{3.46}$$

Here e_n tends to zero only as $\exp(K \ln(n))$, with $K < 0$ when $n \to +\infty$ (see eqs. (3.36) and (3.38)), which is not fast enough to counter the divergence of the last factor in eq. (3.46). Therefore, the SDF step-scaling function derived through the splitting function has lost the appealing characteristic of being an ordinary function and the renormalization group improvement applied to the matching part enhances the distribution-like behavior (as in the case of the backward evolution in $\overline{MS}$). Therefore, we will stick in the rest of this work with the form expressed in eq. (3.28).

Let us draw a short conclusion from this discussion. Dealing with several perturbative objects, namely the evolution and the matching terms either at the level of the step-scaling or splitting functions, the possibilities of perturbative truncation at a given order are multiple, and give different results due to the handling of higher order terms. We have confronted several possible truncations, and selected the one we found to have the most desirable properties. However, it must be clear that all methods are essentially valid, the difference between them vanishing in the limit $\alpha_s \to 0$, or when all orders are taken into account. We will show in the next section, devoted to a numerical application, that we are in neither scenario. The uncertainty introduced by the perturbative matching will motivate us to advocate for a new technique later in the document.

3.3 A perturbative numerical application

For this preliminary numerical exploration of PDF evolution in the SDF scheme, we will revisit the results of the isovector proton PDF published in [52], obtained with a lattice spacing of $a = 0.094\,\mathrm{fm}$. The space-like separation z is an integer multiple of the lattice spacing a. It is customary to exclude non-local operators separated by only one lattice spacing, $z = a$, for which discretization errors are expected to be large. The smallest distances we will consider are therefore $z_1 = 2a = 0.188\,\mathrm{fm}$ and $z_0 = 3a = 0.282\,\mathrm{fm}$. To give an account of the uncertainty in the scale fixing procedure, we will consider scales given by $\mu^2 = (\lambda z^2)^{-1}$ with $\lambda \in [-2, -0.5]$,[2] which yields:

$$\mu_1 \in [0.74, 1.48]\ \mathrm{GeV} \quad \text{and} \quad \mu_0 \in [0.49, 0.99]\ \mathrm{GeV}. \tag{3.47}$$

---

[2] We have derived the formalism so far with $\lambda = -e^{2\gamma_E + 1}/4$ which guarantees the simple form of the matching in eq. (3.15). For arbitrary values of λ, D should be replaced by $D + \ln\left(-\frac{e^{2\gamma_E + 1}}{4\lambda}\right) B_1$ (see appendix A).
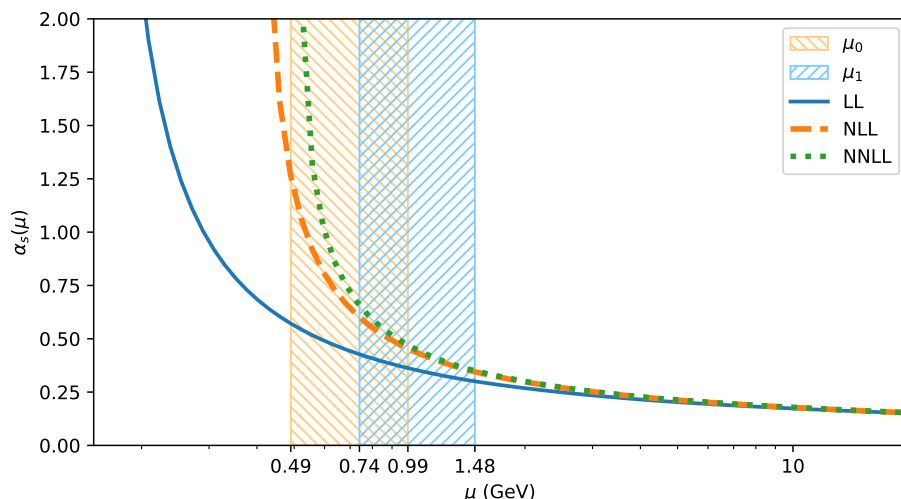
Figure 3. Running of the strong coupling as a function of the scale at LL, NLL and NNLL.

The lower range of these $\overline{MS}$ scales is sufficiently low that instabilities in the perturbation theory could be expected. Indeed, using the PDG world average of $\alpha_s(M_Z = 91.19 \text{ GeV}) = 0.118$ [53] in a variable flavor number scheme with threshold crossings at pole masses of the charm $m_c = 1.4 \text{ GeV}$ and bottom $m_b = 4.5 \text{ GeV}$, the running of the strong coupling computed with the APFEL++ evolution code [54, 55] is shown on figure 3. This plot demonstrates that going to higher order in perturbation theory is not a tool to access lower scales, as the divergence of α_s appears earlier. That the renormalization group improvement makes SDF matching diverge as z increases and gives a very different picture compared to the fixed-order truncation is well established, see e.g. ref. [21] or [56], where the renormalization group improvement can only be applied for $z \leq 3a$ for $a = 0.093 \text{ fm}$. On the other hand, when the perturbative expansion converges (say above 1 GeV), going to higher order reduces the systematic uncertainty [17].

The situation is critical and unreliable at the low scale μ_0 at NNLL. We will therefore limit ourselves to a comparison of the LL and NLL results. We extract from APFEL++ the LL and NLL step-scaling functions in the $\overline{MS}$ scheme, between μ_0 and μ_1 in figure 4. The edges of the bands for perturbative quantities represent the results obtained for $\lambda = -0.5$ and $\lambda = -2$. At LL, there is only one non-singlet splitting function. In contrast, at NLL there are two components to the non-singlet splitting function, which are responsible for the independent evolution of the respective flavor asymmetries:

$$(u - \bar{u}) - (d - \bar{d}) \quad \text{and} \quad (u + \bar{u}) - (d + \bar{d}). \tag{3.48}$$

The first asymmetry corresponds to the real part of the isovector matrix element, whereas the second corresponds to the imaginary part. In figure 4 we have plotted the NLL evolution operator of the real part of the isovector matrix element. The evolution operator of the imaginary part is essentially the same at this order in perturbation theory.
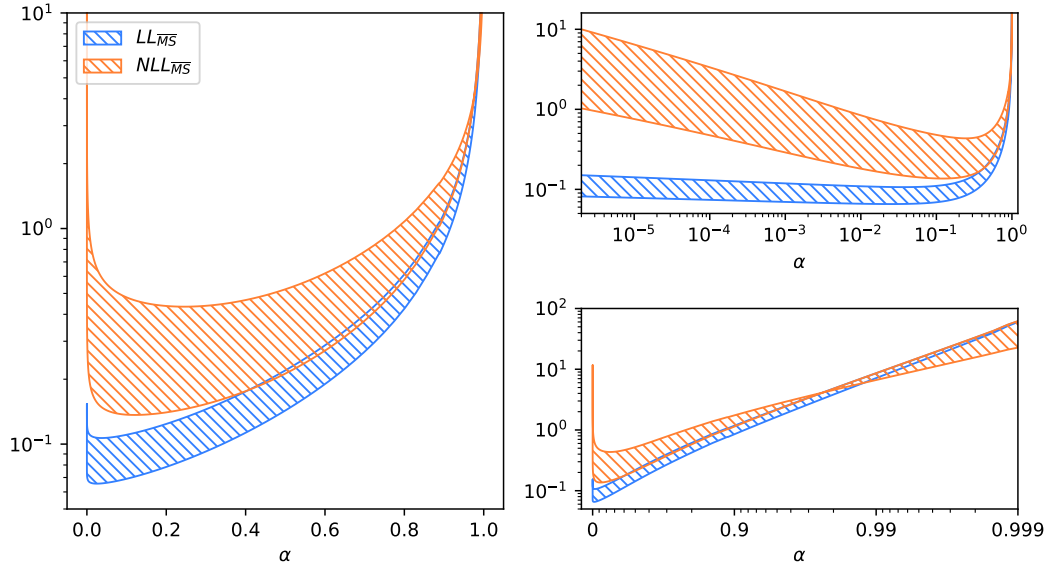
Figure 4. LL (hatched blue) and NLL (hatched orange) non-singlet $\overline{MS}$ step-scaling functions. The lower edge of the band around $\alpha = 0.5$ represents evolution with $\mu^2 = -2/z^2$, that is from $\mu_0 = 0.99\,\text{GeV}$ to $\mu_1 = 1.48\,\text{GeV}$. The upper edge at $\alpha = 0.5$ represents evolution with $\mu^2 = -0.5/z^2$ (so stronger evolution effects as the scale is lower) from $\mu_0 = 0.49\,\text{GeV}$ to $\mu_1 = 0.74\,\text{GeV}$. A zoom on the behavior at small α (upper right) and large α (lower right) is displayed.

A few comments on figure 4 are in order:

1. We have demonstrated in the previous section that the (forward) LL $\overline{MS}$ step-scaling function is an ordinary positive function. We observe that the numerical extraction of the step-scaling function also converges towards an ordinary positive function at NLL, allowing us to present the results as simple curves in the figure.

2. Both the LL and NLL $\overline{MS}$ step-scaling functions diverge at small and large α. The NLL step-scaling function has a stronger divergence at small α compared to the LL, and inversely at large α. This corresponds to the fact that the NLL step-scaling function deviates more than the LL one from the identity $1_\otimes = \delta(1-\alpha)$: NLL radiates more small momentum fraction partons than LL. At small α, the difference is of several orders of magnitude, a testament to the poor perturbative convergence achieved at those scales. The scale uncertainty of the NLL curves is larger than that of the LL as a consequence of the stronger divergence of α_s.

3. We have already derived, in eq. (3.39), the behavior of the LL $\overline{MS}$ step-scaling function at large α through an analysis of the large n behavior of the anomalous dimensions γ_n. Applying the formula to the case $\lambda = -1$ (i.e. $\mu_0 = 0.70\,\text{GeV}$ and $\mu_1 = 1.05\,\text{GeV}$) yields:

$$\mathcal{E}_{LL}(\alpha, \mu_0^2, \mu_1^2) \overset{\alpha \to 1}{\sim} 0.148(-\ln(\alpha))^{-0.865}. \tag{3.49}$$

We can proceed in a similar way at small α by studying the Mellin moments in the limit where $n \to 0$. We find for the anomalous dimensions:

$$\gamma_n \overset{n \to 0}{=} C_F\left(\frac{1}{n} + \frac{1}{2}\right) + \mathcal{O}(n). \tag{3.50}$$

Hence

$$e_n(\mu_0^2,\mu_1^2) = \left(\frac{\alpha_s(\mu_0^2)}{\alpha_s(\mu_1^2)}\right)^{\gamma_n/(2\pi\beta_0)} \overset{n\to 0}{\sim} B\left(\frac{\alpha_s(\mu_0^2)}{\alpha_s(\mu_1^2)}\right)^{C_F/(2n\pi\beta_0)}, \qquad (3.51)$$

where

$$B = \left(\frac{\alpha_s(\mu_0^2)}{\alpha_s(\mu_1^2)}\right)^{C_F/(4\pi\beta_0)}. \qquad (3.52)$$

The inverse Mellin transform of this term is not as straightforward as the one we performed for the large α behavior. It is useful to introduce the confluent hypergeometric limit function $_0F_1(;2;x)$ defined by its series expansion:

$$_0F_1(;2;x) = \sum_{n=0}^{\infty} \frac{x^n}{n!(n+1)!}, \qquad (3.53)$$

and to notice that the Mellin transform of $_0F_1(;2;-a\ln(\alpha))\Theta(1-\alpha)$ gives:

$$\frac{e^{a/n}-1}{a}, \qquad (3.54)$$

where Θ is the Heaviside step function. In the limit of $n\to 0$, we may neglect the term $-1/a$, and obtain therefore that the small α behavior of the LL $\overline{MS}$ step-scaling function is dominated by:

$$\mathcal{E}_{LL}(\alpha,\mu_0^2,\mu_1^2) \overset{\alpha\to 0}{\sim} B\frac{C_F}{2\pi\beta_0}\ln\left(\frac{\alpha_s(\mu_0^2)}{\alpha_s(\mu_1^2)}\right)\,_0F_1\left(;2;-\frac{C_F}{2\pi\beta_0}\ln(\alpha)\ln\left(\frac{\alpha_s(\mu_0^2)}{\alpha_s(\mu_1^2)}\right)\right). \qquad (3.55)$$

For the values of μ_0 and μ_1 we consider here, we obtain:

$$\mathcal{E}_{LL}(\alpha,\mu_0^2,\mu_1^2) \overset{\alpha\to 0}{\sim} 0.0701\,_0F_1(;2;-0.0677\ln(\alpha)). \qquad (3.56)$$

In figure 5 we represent the result of the matching at order $\mathcal{O}(\alpha_s)$ of both the LL and NLL $\overline{MS}$ step-scaling functions following eq. (3.28). A first consequence of the matching is a significant reduction of the contribution for $\alpha < 0.9$, to the point that the matched step-scaling becomes compatible with zero or slightly negative in this region. That evolution in the SDF scheme might not populate — or at least considerably less than the $\overline{MS}$ evolution — the small momentum fraction domain is an interesting finding. In particular, the reasoning developed in [51], which used the fact that evolution ended up dominating the small x behavior of the parton distributions to produce a perturbative modelling of this region, may be inapplicable in the SDF scheme.

More strikingly, for very large values of α, the matched functions can dip to extremely negative values. The smaller the value of $|\lambda|$ (so the higher the perturbative scale), the stronger the negative dip. LL also exhibits a stronger dip than NLL. Only for values of α typically larger than $1-10^{-9}$ is the matched LL step-scaling function in line with the predicted asymptotic behavior — namely $\ln^2(1-\alpha)C_F(\alpha_s(\mu_0^2)-\alpha_s(\mu_1^2))/\pi$ times the LL $\overline{MS}$ large α behavior (3.41).

To understand this very stark oscillation at very large α, recall that the $\overline{MS}$ step-scaling functions are positive functions whose weight is mostly gathered around $\alpha = 1$. As such, one
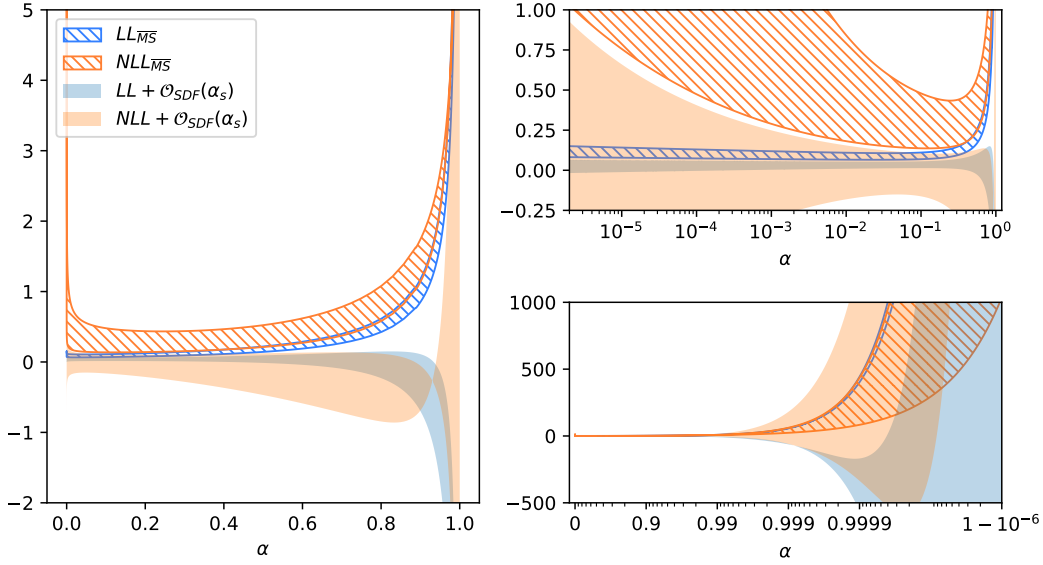
Figure 5. The same LL (hatched blue) and NLL (hatched orange) $\overline{MS}$ step-scaling functions as on figure 4. The uniformly colored bands are the $\mathcal{O}(\alpha_s)$ matching to the SDF scheme of both the LL and NLL step-scaling functions. The edges of the band are made of the result for $\mu^2 = -0.5/z^2$ and $\mu^2 = -2/z^2$.

could consider them akin to a smeared Dirac delta function. The higher the perturbative scale, the less evolution effects are felt and the closer the $\overline{MS}$ step-scaling is to a Dirac delta function. We have noticed already that LL step-scaling functions are also closer to a Dirac delta than the NLL ones. The SDF step-scaling function (3.28) is the convolution of this smeared Dirac delta with a term which contains a plus-prescription, the distribution D. A plus-prescription distribution convoluted with a smeared Dirac delta exhibits large oscillations at large α, which become bigger as the smeared Dirac delta approaches an actual Dirac delta. In that limit, the convolution would become equal to the plus-prescription itself, which cannot be represented as an ordinary function since it would exhibit oscillations of infinite amplitude localized at $\alpha = 1$.

However spectacular this strong oscillation is, it is barely relevant to our study for two main reasons. The first is that it is likely an artifact of the perturbative truncation. A fixed-order truncation of the $\overline{MS}$ step-scaling function presents a "naked" plus-prescription — an infinitely narrow oscillation of infinite amplitude. One needs the resummation of the full leading logarithmic expansion to ensure the $\overline{MS}$ step-scaling function is an ordinary well-behaved function. Although we use renormalization group improvement for the $\overline{MS}$ step-scaling function, we used a fixed-order truncation of the non-logarithmic matching part (see e.g. eq. (3.28) and the discussion below). Renormalization group improvement of the non-logarithmic part through the SDF splitting function did not actually prove beneficial (see eq. (3.46) and discussion). The discussion of the last section demonstrates that oscillating features at large α are very dependent on the handling of missing higher orders. Therefore, one could legitimately worry that there is little physical insight to draw from the pointwise
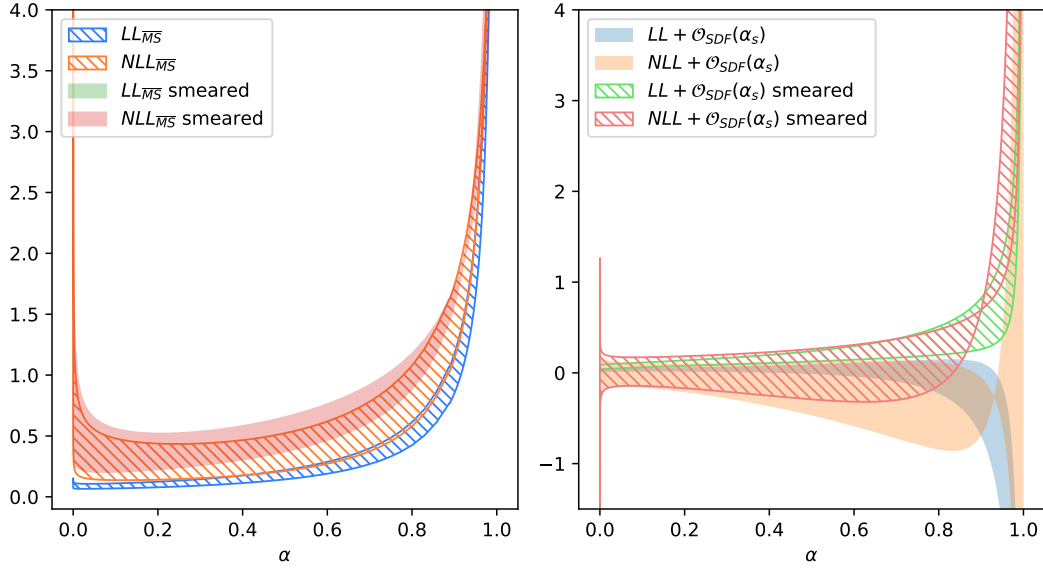
Figure 6. On the left, the effect of smearing the $\overline{MS}$ step-scaling functions with the kernel $S(\alpha)$ (3.57). There is little influence on the overall shape of the function. On the right, the smearing of the matched step-scaling functions. This time, the effect is dramatic, entirely erasing the sharp negative fluctuation at very large α.

value of the evolution operator, at least at large α. This is indeed the case, but in practice, this is unlikely to cause a problem: what matters is only what can be measured, as we explain below. A resummation of threshold logarithms (see e.g. [33, 57]) could be envisioned, but the data we are working with is in fact not sensitive to these kind of effetcs.

More fundamentally, the second reason why these oscillations are mostly not relevant is that the step-scaling function is only ever accessed when convoluted with a PDF / ITD / matrix element, which are smooth functions. If the ITD only exhibits sizeable variations on a scale $\Delta\nu$, then the information content in the ITD over the interval $[0, \nu_{\max}]$ can be roughly summarized by $\nu_{\max}/\Delta\nu$ points. If one views the extraction of the step-scaling function as a linear system where we have about $\nu_{\max}/\Delta\nu$ good measurements, then we can also only characterize the evolution operator on a similar number of points, so at a resolution level of $\Delta\nu/\nu_{\max}$.

This argument is of course only qualitative. One could argue that by measuring arbitrarily well the ITD on any interval in Ioffe-time, one could extract equally well the step-scaling regardless of the value of $\nu_{\max}$ or the level of smoothness $\Delta\nu$. However, in practice, $\nu_{\max}/\Delta\nu$ represents an empirical threshold between the resolution which can be reasonably reached, and the one which becomes computationally highly challenging.

In fact, current lattice data probably do not reach the threshold of $\Delta\nu/\nu_{\max}$. If $\nu = z_3 \cdot P_z$, the closest $\Delta\nu$ we can probe at fixed value of z^2 is given by a variation of one lattice unit of the hadron momentum. Then the lattice proxy of $\Delta\nu/\nu_{\max}$ is given by the inverse of the number of available hadron momenta, which is typically of the order of ~ 0.1. Some of the matched step-scaling functions in figure 5 show a strong negative dip up to $\alpha \sim 0.9999$,

followed by an extremely sharp trend reversal. We would need a resolution of the order 10^{-4} to be effectively sensitive to this kind of feature of the step-scaling function. Therefore, the data we collect do not allow us to access the step-scaling function in the full complexity of its α-dependence, but rather its convolution with some unknown smearing kernel of width $\Delta\alpha \sim 0.1$. As an example, we depict on figure 6 the effect of convoluting all the step-scaling functions with an additional smearing kernel:

$$S(\alpha) = \frac{(-\ln(\alpha))^{-0.9}}{\Gamma(0.1)} \, . \tag{3.57}$$

The smearing has the general effect of taming the behavior at large α, and displacing some weight of the step-scaling function towards lower values of α. The general shape of the $\overline{MS}$ step-scaling functions is only weakly altered by the smearing. On the other hand, the smearing has entirely suppressed the sharp negative oscillation of the SDF-matched functions at large α. Therefore, we do not expect that these oscillations will play any role in evolving realistic data points. The displacement of the curves when convoluted with the smearing kernel represents a typical systematic uncertainty we will be facing in any extraction of the step-scaling function due to the fact that we do not know precisely what the smearing kernel is aside from an estimate of its resolution.

4 Fitting a non-perturbative step-scaling function to lattice data

In principle, the lattice matrix elements already published should give us a first hint of non-perturbative evolution in the SDF scheme, which we can compare to the perturbative application we have performed in the previous section. The leading order DGLAP evolution was first used in [5], where numerical evidence suggested that the quenched lattice QCD results for $\mathfrak{M}(\nu, z^2)$ indeed follow this evolution in z^2. More generally, a wealth of numerical analyses have shown that for $z < 0.2 \sim 0.3\,\mathrm{fm}$, various truncations of the perturbative matching give rather similar results, mostly compatible with the phenomenological knowledge of unpolarized PDFs [40, 52, 58–75].

Data published in [52] with 2+1 flavors of Wilson clover quarks with a lattice spacing of $a = 0.094\,\mathrm{fm}$, volume of $32^3 \times 64$, pion mass of $358\,\mathrm{MeV}$ and 349 gauge configurations are shown on figure 7. We present the two values $z_0 = 3a$ and $z_1 = 2a$, and the different Ioffe times are obtained by varying the momentum P of the hadron. The small value of z should guarantee a reduction of higher-twist effects and offers scales where we can compare to the perturbative results at LL and NLL.

As the hadron momentum increases, the signal-to-noise ratio decreases. To counter this phenomenon, for $P \geq 4 \times \frac{2\pi}{L}$, the hadron interpolating operators are "phased" [76, 77]: their spatial extension is modulated by an oscillation which increases their coupling to the ground state of the boosted hadron. The effect is striking on figure 7. The fifth red point, where the phasing kicks in, has a significantly smaller uncertainty than the fourth point as witnessed by the tightening of the red band. The same behavior is also visible for the fifth blue point (the rightmost one — the first blue point is hidden at $\nu = 0$ by the first red one) compared to the fourth blue point. This gives however the unfortunate consequence that the distribution at the initial scale is considerably less constrained than at the final
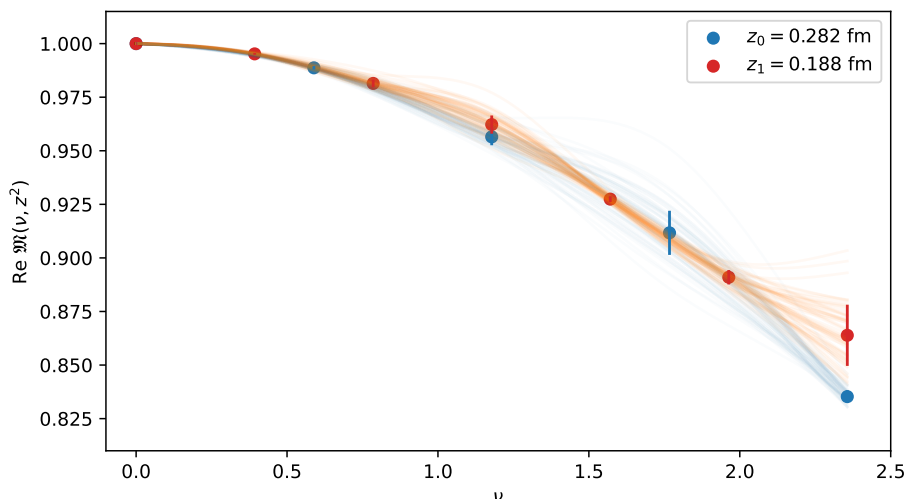
Figure 7. The real part of the isovector proton matrix element $\mathfrak{M}(\nu, z^2)$ published in [52]. The light curves represent 60 cubic splines interpolations of jack-knife samples of the matrix elements.

scale for $\nu \in [1.2, 2]$, and the opposite is true for $\nu \geq 2$, to some extent obfuscating the z^2 dependence of the matrix elements.

4.1 Methodology

Our strategy to extract an empirical step-scaling function from the lattice data is first to propose a parametric form of $\Sigma(\alpha; z_0^2, z_1^2)$, and then to fit it to the lattice data:

$$\mathfrak{M}(\nu, z_1^2) = \int_0^1 d\alpha \, \Sigma(\alpha; z_0^2, z_1^2) \mathfrak{M}(\alpha\nu; z_0^2). \tag{4.1}$$

As discussed in section III.B, the existence of Σ can be inferred at small z^2 from the validity of the OPE and all order perturbative results. Let us note that if there are significant higher twist contributions in the lattice data, they will be encompassed in some way in the extracted Σ, whose existence or support we might not be able to guarantee. This limits us to only studying the small z^2 domain.

If we can only guarantee that this empirical model describes the z^2 dependence of the data at small z^2 based on general observations from perturbation theory, the reader may wonder why not just use perturbative computations directly. As we have underlined before, the perturbative truncation comes with sizeable uncertainties at the value of z^2 we are concerned with, and it is easy to envision that lattice calculations will outperform those uncertainties. From a more general point of view, perturbative expansions are "asymptotic" and their actual convergence is a difficult matter. On a practical level, higher-order perturbation theory is very cumbersome to conduct. For all these reasons we hope to ultimately extract more precise evolution operators from the data rather than from perturbation theory. Finally, it may be that we can demonstrate the universality of our numerical evolution operator by studying various hadrons up to values of z^2 where $\alpha_s(1/z^2)$ is fairly large. The existence of a universal evolution operator in a fully non-perturbative framework may allow us to define a range of

validity of factorization without reference to perturbation theory — where factorization is understood as the general premise of a separation of scales between a universal short-distance behavior and a long-distance hadronic one.

We only have at our disposal a discrete set of Ioffe-time values $(\nu_{i,0})_i$ at the initial scale z_0^2 to compute the integral of eq. (4.1). Therefore, we perform a cubic spline interpolation of the initial-scale matrix element samples. Denoting by $(\nu_{j,1})_j$ the set of Ioffe-time values at the final scale z_1^2, we can write schematically the fit as the optimization of a function T_Σ of the initial-scale matrix elements:

$$\mathfrak{M}(\nu_{j,1}, z_1^2) = T_\Sigma\left(\nu_{j,1}, z_1^2; \left(\mathfrak{M}(\nu_{i,0}; z_0^2)\right)_i\right), \tag{4.2}$$

where T_Σ contains both the parametric dependence of the step-scaling function Σ, and the effect of the cubic spline interpolation on the discrete set of matrix elements at initial scale $\left(\mathfrak{M}(\nu_{i,0}; z_0^2)\right)_i$.

The presence of largely correlated uncertainties within the jack-knife samples of the matrix elements must be accounted for in the fit. At a fixed value of P, the points at z_0^2 and z_1^2 are correlated in excess of 97%. On the other hand, at a fixed value of z^2, neighboring points in Ioffe time (so in P) are typically correlated at the level of 50%. It drops to about 20% with one more separation in Ioffe time, and to less than 10% with one more.

Eq. (4.2) represents a typical fit of the form $y(t) = f(t, x)$ where we want to learn the parameters of f, but there are uncertainties in the value of x which are furthermore correlated with the uncertainties of y. For such an errors-in-variables model, we will use the total least square method. We assume that there exist true values $\mathfrak{M}^*(\nu, z^2)$ of the matrix elements, and that the jack-knife samples arise from a measurement of $\mathfrak{M}^*(\nu, z^2)$ with *classical* errors. This means that although there is correlation between the measurement of $\mathfrak{M}(\nu, z^2)$ at various kinematics, there is no correlation assumed between the uncertainty of the measurement and the true value $\mathfrak{M}^*(\nu, z^2)$. We form a vector of fit residuals defined by:

$$r = \begin{pmatrix} \left(\mathfrak{M}^*(\nu_{i,0}, z_0^2) - \mathfrak{M}(\nu_{i,0}, z_0^2)\right)_i \\ \left(T_\Sigma\left(\nu_{j,1}, z_1^2; (\mathfrak{M}^*(\nu_{i,0}; z_0^2))_i\right) - \mathfrak{M}(\nu_{j,1}, z_1^2)\right)_j \end{pmatrix}. \tag{4.3}$$

The vector of residuals contains both the discrepancy of the true values to the measured ones, and quantifies the ability of the model to reproduce the data. Then introducing the empirical covariance matrix evaluated from the full set of jack-knife samples:

$$\Omega = \begin{pmatrix} \mathrm{Cov}[\mathfrak{M}(\nu_{i,0}, z_0^2), \mathfrak{M}(\nu_{j,0}, z_0^2)]_{i,j} & \mathrm{Cov}[\mathfrak{M}(\nu_{i,0}, z_0^2), \mathfrak{M}(\nu_{j,1}, z_1^2)]_{i,j} \\ \mathrm{Cov}[\mathfrak{M}(\nu_{i,0}, z_0^2), \mathfrak{M}(\nu_{j,1}, z_1^2)]_{i,j} & \mathrm{Cov}[\mathfrak{M}(\nu_{i,1}, z_1^2), \mathfrak{M}(\nu_{j,1}, z_1^2)]_{i,j} \end{pmatrix}, \tag{4.4}$$

we form a goodness-of-fit measure:

$$\chi^2 = r^T \Omega^{-1} r, \tag{4.5}$$

where r^T denotes the transpose of r. To give an account of uncertainty, we study the distribution of parameters in the vicinity of the minimal χ^2. This distribution reflects the

aleatoric uncertainty of the data, but not the epistemic uncertainty introduced by the specific choice of the parametric form.

In practice we face an ill-defined inverse problem. The step-scaling function might be in full generality a distribution with an arbitrary dependence on α. However, as we have discussed in the previous section, since we only have a restricted kinematic access to the PDFs which phenomenology tells us are smooth functions, we are only sensitive to the step-scaling function smeared on some resolution of the order of $\Delta\alpha \sim 0.1$. The fact that we can really only extract the smeared step-scaling function is actually a blessing to some extent, as demonstrated by our toy smearing on figure 6: there, the smeared perturbative SDF step-scaling function exhibits a fairly generic shape similar to its $\overline{MS}$ counterparts, i.e. an ordinary function with a simple divergence at $\alpha = 0$ and $\alpha = 1$ once the large oscillation at very large α is erased by smearing. At first, we hope to circumvent the ill-posedness by finding a generic parameterization of the smeared step-scaling function that we can fully constrain from the data, but it is clear that this object can only be used to evolve the PDF at the same level of resolution as the one used for its extraction. That we have done a good job at extracting the smeared step-scaling function within the uncertainties of the data will be identified by the fact that we obtain a good reproduction of the z^2 dependence of the data, that is a χ^2 (4.5) per fitted points close to one (called χ^2 per n_{pts} in the following). We will be considering another, less model-dependent approach later on.

We propose the following parameterization of the step-scaling function:

$$\Sigma(\alpha; z_0^2, z_1^2) = A\alpha^{-\delta}(1+r\alpha) + b(-\ln(\alpha))^{-\eta}\ln^2(1-\alpha) + \sigma\alpha(1-\alpha). \quad (4.6)$$

There are six parameters: $A, \delta, r, b, \eta, \sigma$. The first term produces a divergence when $\alpha \to 0$, the second term when $\alpha \to 1$, and the third term only contributes to the intermediate α range. One will notice that the asymptotic behaviors of eq. (4.6) when $\alpha \to 0$ and 1 do not correspond to those we have analytically computed for the LL $\overline{MS}$ case (see eqs. (3.55) and (3.39)). The large α behavior of the LL $\overline{MS}$ step-scaling function only diverges as $b(-\ln(\alpha))^{-\eta}$, and the small α behavior was determined to be of the general form $A_0F_1(;2;-\delta\ln(\alpha))$, which is less divergent than any term $A\alpha^{-\delta}$ where $\delta > 0$. But although eq. (4.6) does not have the same asymptotical behavior than the true LL $\overline{MS}$ step-scaling function, it is able to produce a very satisfactory account on a compact interval $[\alpha_0, \alpha_1] \subset (0,1)$. We demonstrate in figure 8 that our parametric form is able to reproduce satisfactorily the various perturbative step-scalings we have derived before on a large range of α when we use the smearing of the extremely sharp variations of the SDF-matched step-scaling functions.

4.2 Test of the methodology

To exemplify our fitting methodology, we first work in a controlled setting with synthetic data. We start with a simple toy valence PDF at $\mu_0 = 0.49\,\text{GeV}$, normalized so that its integral from 0 to 1 equals 1:

$$q(x, \mu_0^2) = \frac{35}{32}x^{-0.5}(1-x)^3. \quad (4.7)$$

We evolve this PDF to the scale $\mu_1 = 0.74\,\text{GeV}$ using the exact LL DGLAP evolution, and extract the values of the ITD (2.5) on the kinematics shown by the datapoints on figure 7.
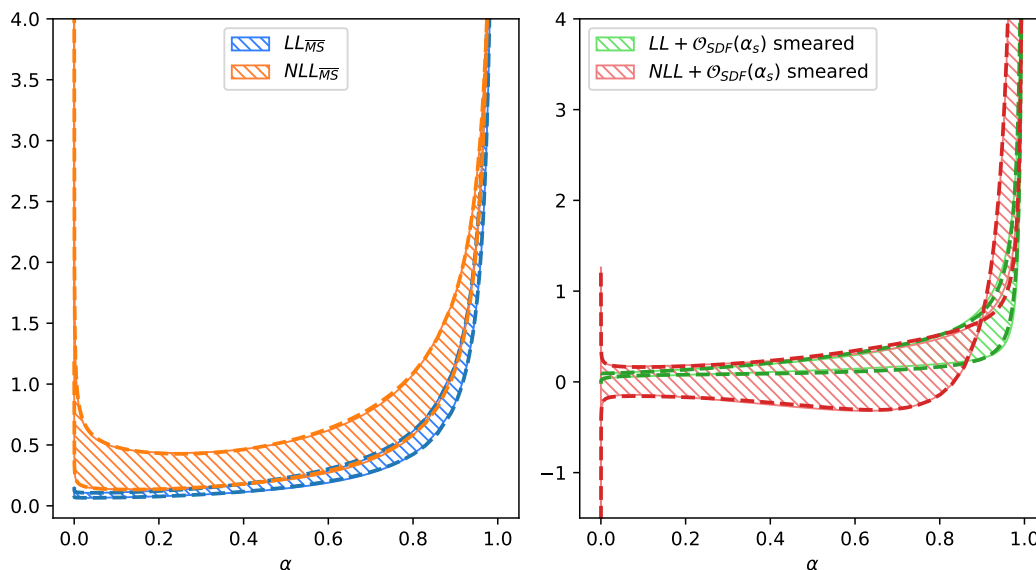
Figure 8. The bands are the same as in figure 6. The dotted lines represent fits with the functional form of eq. (4.6) on the edges of the bands. The agreement is very satisfactory. Since the asymptotic behaviors are not the same, going to extremely small or large values of α would show discrepancies, but those are not meaningful numerically due to the resolution smearing induced by the convolution.

Then we fluctuate the true data according to the covariance matrix of the lattice data (4.4). The result is depicted in figure 9. Although the true values of the ITD at final scale (red stars) are strictly above their values at initial scale (blue stars), the uncertainty makes it far less visible at the level of the noisy data. Nonetheless, how well can we extract the LL $\overline{MS}$ step-scaling function from this noisy dataset?

Exploring the 10-parameter space (6 parameters for the step-scaling function and 4 parameters for the unknown true values of the ITD at initial scale) with a non-linear χ^2 proves rather challenging. We start from 2000 random initial sets of widely distributed parameters and run independent stochastic gradient optimizations to find good fit candidates. In fact, this will provide a series of local minima, and hopefully the global one among them. We achieve a smallest χ^2 per n_{pts} of 0.82,[3] which demonstrates an excellent ability to explain the data with the parametric form of the step-scaling function and within the uncertainty.

To produce an account of uncertainty, we select first the 8 local minima with a χ^2 per n_{pts} below 1, and then the 32 ones with a χ^2 per n_{pts} below 3. We explore the vicinity of these local minima to produce as large a sampling as possible of the distribution of parameters which offers either a χ^2 per n_{pts} below 1 or 3.[4] The result with a threshold of 1 is shown as

---

[3]Since the data at Ioffe time $\nu = 0$ is exactly 1 with no uncertainty, to practically perform a fit, we have to grant it a standard deviation — or enforce analytically the reproduction of that point. We choose the precise uncertainty of 0.0001, which ensures that the integral of the step-scaling function is equal to 1 ± 0.0001.

[4]For this part of the study, we present uncertainties computed from a fixed threshold in χ^2 per fitted points. Linear propagation of uncertainty invites to use one unit of the full χ^2 (not divided by the number of points or degrees of freedom) around its (unique) minimum as the standard measure of uncertainty. However, in the highly non-linear case we are facing, we find this uncertainty quantification both overly optimistic and difficult
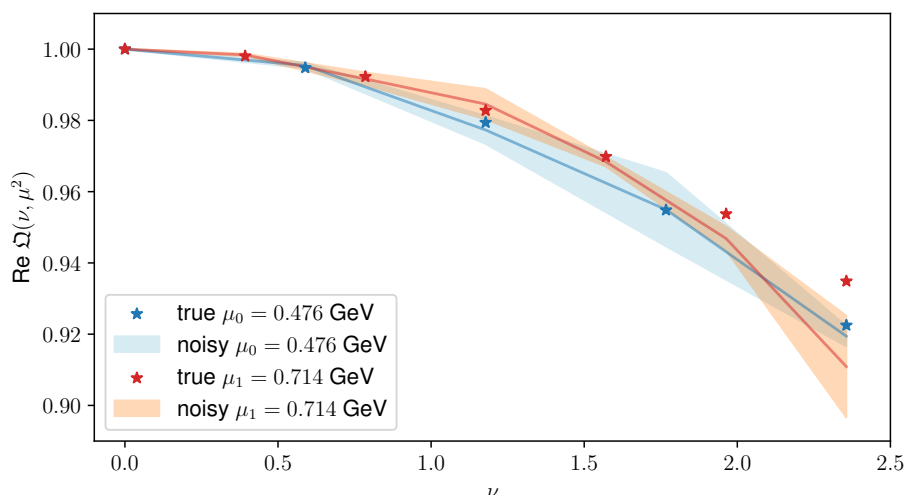
Figure 9. Real part of the Ioffe-time distribution associated to the PDF (4.7). The stars represent the true values of the Ioffe-time distribution, and the bands the result of their reshuffling with the covariance matrix of the actual lattice data shown of figure 7.

the hashed band on figure 10, whereas the result for a threshold of 3 is depicted as the dotted grey lines. In fact, exploring the χ^2 in the vicinity of the local minima is rather useless. The plot would be exactly the same if we had only plotted the distribution of local minima: the χ^2 varies so abruptly that a change of the parameters by less than 0.1% is often enough to see the χ^2 increase by several units. Therefore, the crucial part of the evaluation of the uncertainty is to attempt to determine the largest possible amount of local minima.

In spite of the significant noise in the data, the fit allows us to reconstruct the LL $\overline{MS}$ step-scaling function in good agreement with its true value. The small α domain is poorly constrained as one would expect considering the limited range in Ioffe time where the data are available. The smeared LL $\overline{MS}$ step-scaling function that we have introduced in figure 6 is also mostly compatible with the extraction, a fortiori if a threshold on χ^2 per n_{pts} of 3 is used. We expected this result as we have explained that we only have an access to the step-scaling function with a limited resolution.

In figure 11, we show the "true" values of the ITD obtained thanks to the fit, with both the thresholds in χ^2 per n_{pts} of 1 and 3. The correction induced by the fit reduces significantly the uncertainty and discrepancy introduced by the noise. For instance, the last point at the higher scale μ_1 is correctly displaced by more than 1σ compared to its noisy value, and its uncertainty is considerably reduced compared to the original data for a threshold of 1. The second to last point at initial scale sees its uncertainty reduced by a factor 3 compared to the original data. As we will see later on, it is the requirement of self-consistent evolution in z^2, encompassed in the form of the convolution of eq. (4.1), that is responsible for the correction of aberrations in the data, more than the parametric form chosen for Σ. These corrections would otherwise require significant computational time to obtain from better

---

to interpret. We prefer to use the intuitive absolute goodness of fit as a criterion. In the later results with a quadratic χ^2, we will use the standard result of linear propagation of uncertainty.
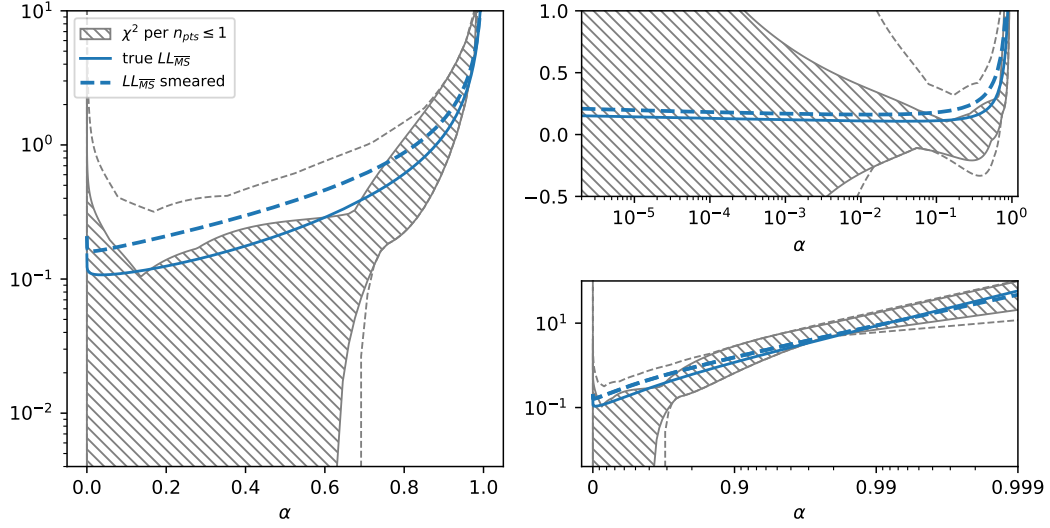
Figure 10. The hatched band results from an exploration of the fits whose χ^2 per n_{pts} is less than 1. The dotted grey lines represent the uncertainty with a χ^2 per $n_{pts} \leq 3$. This extraction of the step-scaling function is in good agreement with the actual LL $\overline{MS}$ step-scaling function which has been used to produce the data, and also mostly compatible with its smeared version.
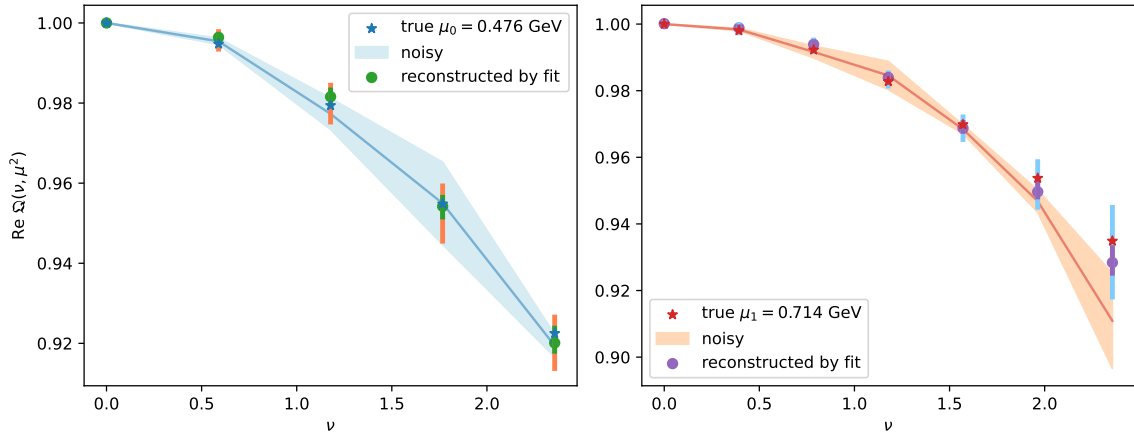

Figure 11. The stars and bands are the same as on figure 9. The green points with error bars on the left represent the fitted result of the true ITD value with a χ^2 per d.o.f. threshold of 1 at initial scale $\mathcal{Q}^*(\nu, \mu_0^2)$. The larger orange error bars stands for a threshold of 3. The purple points with purple and light blue error bars on the right represent the convolution of the fitted step-scaling function with the fitted true ITD values at initial scale $T_\Sigma(\nu, \mu_1^2, (\mathcal{Q}^*(\nu, \mu_0^2)))$.
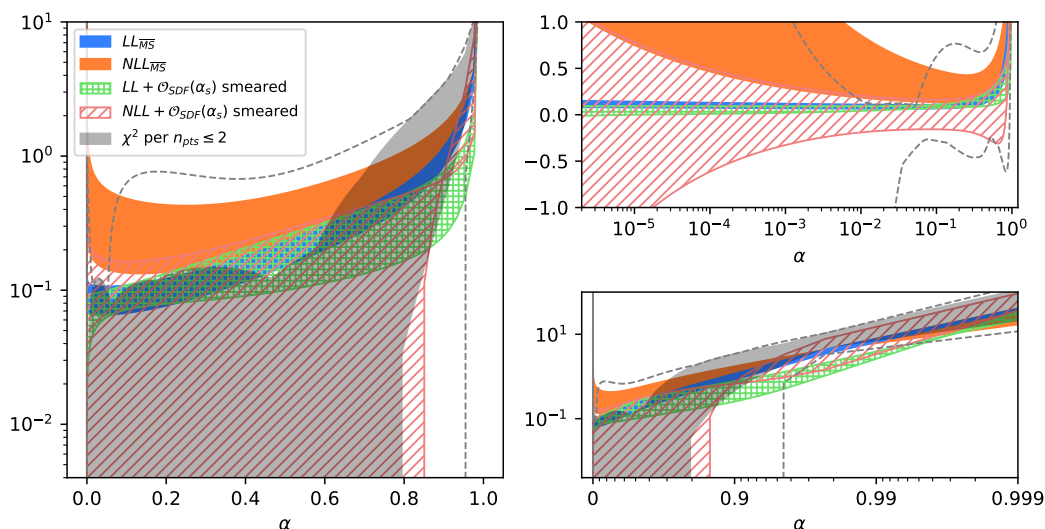
Figure 12. Empirical extraction of the SDF step-scaling function from the real part of the lattice data. The shadow band contains all fit results whose χ^2 per n_{pts} is less than 2, the dotted grey lines less than 3.

lattice calculations. This underlines how crucial a fine understanding of the evolution effect is to the effort of extraction of parton distributions from lattice QCD.

4.3 Extraction from actual lattice data with a parametric form

The success of the previous test motivates us to apply the same technique to the actual lattice data. The results are shown in figures 12 and 13 for the real part. The smallest χ^2 per n_{pts} that we have found for the real part is 1.65, which shows a satisfactory level of explanation of the data. This time, we will therefore compare a threshold of 2 (9 local minima found) and 3 (32 local minima found).

Overall, the extracted SDF step-scaling function is highly compatible with most inputs from perturbation theory, except the NLL $\overline{MS}$ evolution if one uses the most stringent threshold in χ^2. The SDF step-scaling function extracted directly from the lattice data shows compatibility with zero as soon as $\alpha < 0.8$ as the matched and smeared NLL calculation, and a very similar large α behavior to all perturbative derivations. One should remember that the uncertainty depicted for perturbative derivations is the theoretical uncertainty tied to scale fixing, whereas the uncertainty of the lattice SDF step-scaling function is tied to the uncertainty of the lattice data and the choice of parametric form. That both coincide so well for the matched and smeared NLL term is therefore rather coincidental. The "true" values of the matrix elements are shown in figure 13. The matrix elements corrected by the requirement of self-consistent z^2 evolution show a much reduced effect of evolution compared to that typically produced by evolution in the $\overline{MS}$ scheme, in line with the conclusions we drew from the study of the perturbative matching. We observe that, as in our toy model, the last point at higher scale has been significantly moved, and the uncertainty of some data points largely reduced by our fitting procedure.
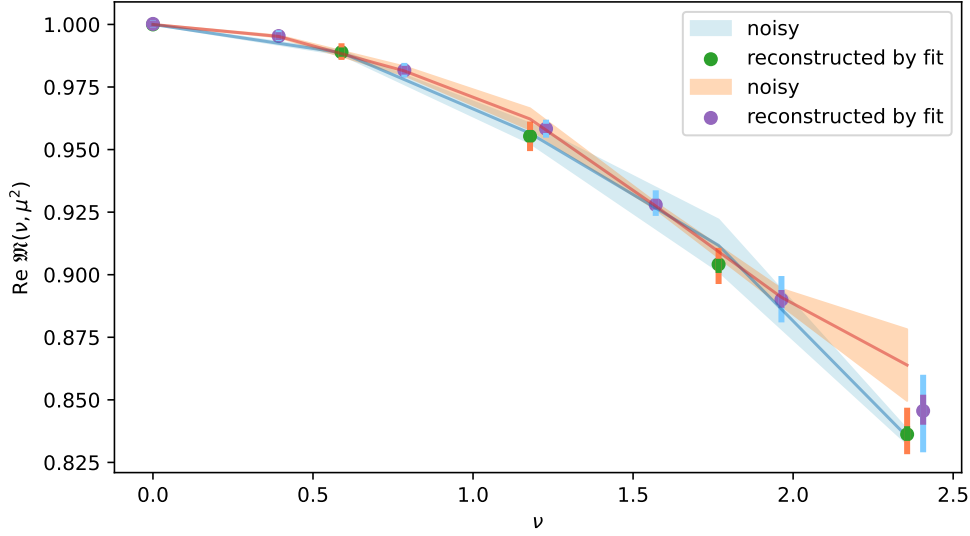
Figure 13. Real part of the isovector proton matrix element $\mathfrak{M}(\nu, z^2)$ displaced by our fitting methodology. The two colored error bars stand for a tolerance parameter of 2 and 3. The fourth and seventh purple points have been slightly displaced to avoid collision with the green points. The fit correction produces data with a smaller evolution effect than the — slightly erratic — effect present in the original data.
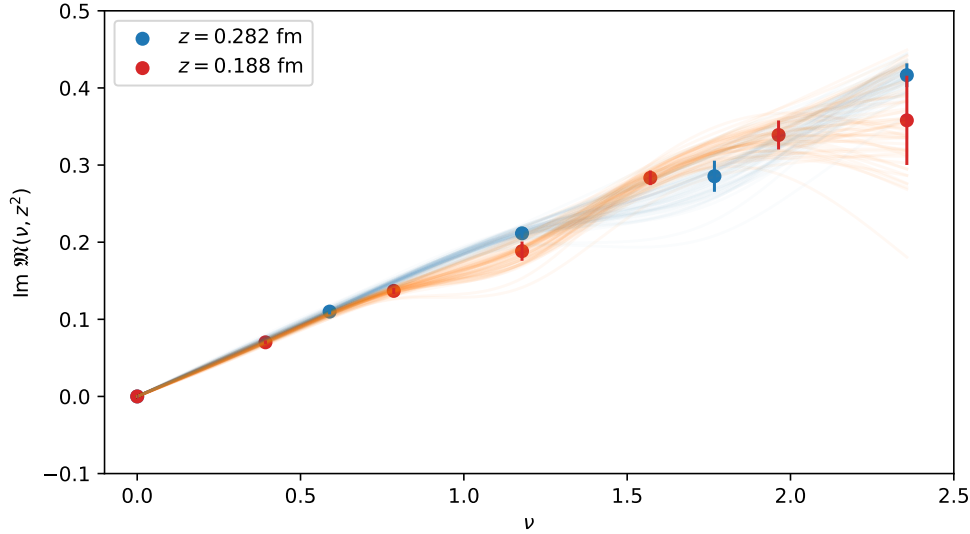

Figure 14. Imaginary part of the isovector proton matrix element $\mathfrak{M}(\nu, z^2)$ published in [52].

We conduct a similar study for the imaginary part of the lattice data at the same scales. At $\nu = 0$, the imaginary part of the isovector data is zero, due to the x symmetries of this non-singlet distribution. This means that we lose one constraint on the step-scaling function, namely that its integral over α should be equal to one. The data is depicted on figure 14. The effect of z^2 evolution is here again fairly difficult to discern clearly, due to the noise in the data.
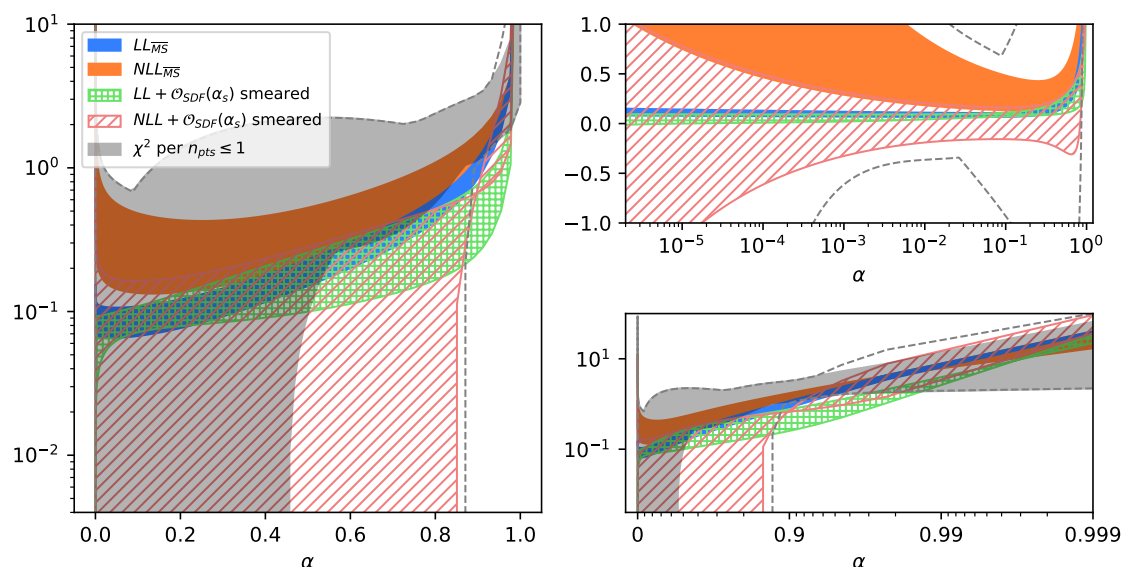
Figure 15. Empirical extraction of the SDF step-scaling function from the imaginary part of the lattice data.

The smallest χ^2 per n_{pts} we find is 0.54, that is significantly smaller that with both our previous studies of the real part (ideal LL $\overline{MS}$ case and actual lattice data). The data is easier to explain, and the χ^2 not such an abrupt function as for the real case, to the point that all 22 local minima we find correspond to a χ^2 per $n_{pts} \leq 1$. We will use a tolerance parameter of 1 and 3. The result is shown on figures 15 and 16.

Overall, the imaginary data is less constrained by evolution, can be fit much more easily and shows larger uncertainties both at small and large α. Less stark behavior at large α are more favored by the fit, but also much larger contributions at small α. The results of the "corrected" lattice data on figure 16 show very little evolution effect.

Notice that we have theoretically no expectation that the real and imaginary parts of the isovector matrix elements evolve in a similar way. However, there is no difference at LL in $\overline{MS}$, and the difference at NLL is so minute that it is not visible on the plot. Unfortunately, the larger uncertainties of the step-scaling function extracted from the imaginary part do not allow any conclusive statement on a difference with the evolution of the real part.

4.4 Bayesian reconstruction of the step-scaling function

As we have already mentioned, determining the SDF step-scaling function from lattice data is an ill-posed integral inverse problem, which tends to have significant modeling bias when studying fixed functional forms. There exists an alternative class of solutions to the parametric forms studied in the previous sections to this sort of inverse problem based upon Bayes's theorem, such as the Maximum Entropy Method [78] and Bayesian Reconstruction [79]. These approaches are similar in theme, but vary in the specific choice of prior distribution given via Bayes's Theorem. In work on lattice calculation of PDFs, these approaches have been used to determine the PDF from lattice matrix elements [80, 81].
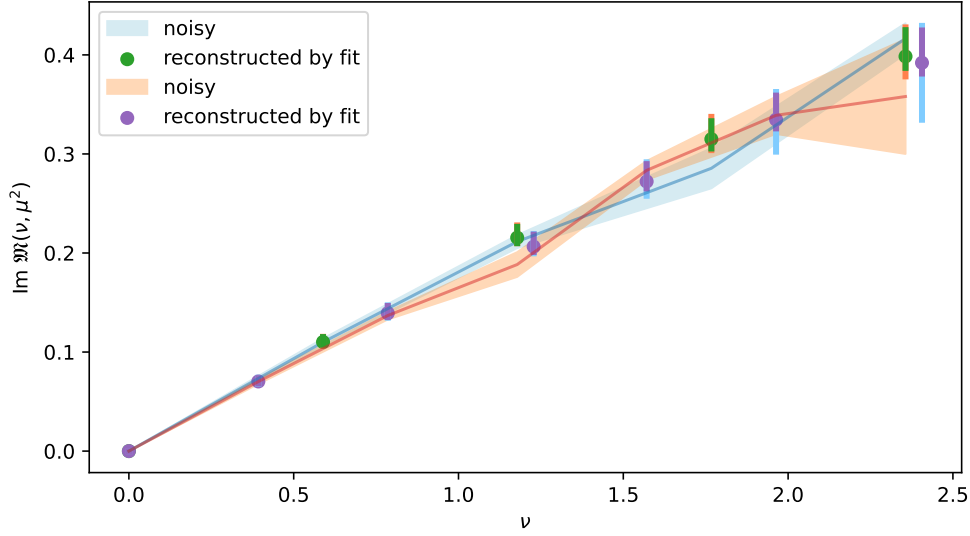
Figure 16. Imaginary part of the isovector proton matrix element $\mathfrak{M}(\nu, z^2)$ displaced by our fitting methodology. The error bars are noticeably dissymmetric as the non-linear fitting procedure produces non-Gaussian uncertainty distributions. The dot corresponds to the mean of the distribution, whereas the error bar to all values below the threshold in χ^2.

The goal of each of these approaches is to determine the most probable function whose integrals correspond to a given set of data and some prior information. The most probable function is given by the expectation value:

$$\langle \Sigma(\alpha; z_0^2, z_1^2) \rangle = \int D\Big[\Sigma(\alpha; z_0^2, z_1^2) \Big] \Sigma(\alpha; z_0^2, z_1^2) P\Big[\Sigma(\alpha; z_0^2, z_1^2) | \mathfrak{M}(z_0^2), \mathfrak{M}(z_1^2), I \Big], \qquad (4.8)$$

where the posterior distribution $P\Big[\Sigma(\alpha; z_0^2, z_1^2) | \mathfrak{M}(z_0^2), \mathfrak{M}(z_1^2), I \Big]$ is defined by Bayes's theorem as:

$$P\Big[\Sigma(\alpha; z_0^2, z_1^2) | \mathfrak{M}(z_0^2), \mathfrak{M}(z_1^2), I \Big] \propto \exp\left[-\frac{\chi^2}{2} + u S\big[\Sigma(\alpha; z_0^2, z_1^2) \big] \right]. \qquad (4.9)$$

The χ^2 term represents the likelihood of the observed data knowing the value of the step-scaling function, and is defined in the same manner as in eq. (4.5) including the interpolation of $\mathfrak{M}^*(z_0^2)$ with a cubic spline. The $S[\Sigma]$ term represents the prior enforced on the step-scaling function, which distinguishes the various approaches of this type. The hyperparameter u controls the relative importance of the goodness of fit and the prior. The hyperparameter can be varied to study its effect or can be integrated away. The values of the function Σ on a discrete grid in α become the model parameters and the integral whose inverse is desired and those in S are evaluated numerically. In this work, the trapezoid method is used on various grids. Due to the fact that no evolution corresponds to a step-scaling which is a Dirac delta at $\alpha = 1$, high resolution in the large α region is required to accurately reproduce Σ. The choice of this grid, the integration method, and the choice of S will collectively define a model with some specific bias. By varying the choices, the model dependence can be studied.

The goal of the prior information is to bias the result away from unphysical results. In the case of spectral functions, known to be positive definite, it is desirable to require the function to be positive. This feature was therefore baked into the Bayesian Reconstruction prior information in [79]. Without physical guidance many "reasonable" choices must be studied, though as always with Bayesian methods the choice of prior information can dramatically change the result. Hence what is "reasonable" may end up debatable.

In this work, we shall use the Quadratic Difference Ratio (QDR) which is defined by the prior function:

$$\log\left[P[\Sigma|I]\right] = u S_{\mathrm{QDR}}[\Sigma] = -u\int_0^1 d\alpha\, \frac{\left(\Sigma(\alpha)-h(\alpha)\right)^2}{2\sigma^2(\alpha)} \tag{4.10}$$

where h is a prior model and σ a prior weight function. This prior will attempt to push the step-scaling function towards the prior model, but through the prior weight it allows Σ in different regions of α to differ from the prior model by a varying amount. If the hyperparameter u is integrated away, then the magnitude of the weight function becomes trivialized and only its shape dictates which regions of α must agree more or less than other regions. Without integrating u, its size as well as σ determine how restrictive this prior distribution is.

For the first example of this approach, we take the seemingly simple choices of:

$$u = 1 \qquad h(\alpha) = 0 \qquad \sigma(\alpha) = 1 \tag{4.11}$$

for the prior distribution. The function is parameterized on a grid of 1000 equally spaced points. The set of Σ_i on that grid that maximizes the posterior is shown in figure 17. The minimum of twice the negative log posterior is given by $\chi^2 = 3.76$ and $2uS = 9.20$. It should be noted that the χ^2 reported has not been reduced by the number of datapoints, so this value represents significant agreement with the data. The errors are determined by the inverse of the Hessian near the minimum since both S and χ^2 are quadratic in Σ_i and $\mathfrak{M}^*(z_0^2)$. Some level of tension with the prior was expected, as for instance the prior does not respect the basic requirement that the integral of the step-scaling function should be close to 1. On the other hand, the χ^2 is significantly less than the number of data points, so the data is well reproduced. The reconstructed step-scaling function is globally compatible with the parametric fit result presented in figure 12, except at very large α where the flat prior penalizes an attempt at divergence. Although the uncertainty obtained thanks to this Bayesian reconstruction is typically quite larger than that of the parametric fit, the correction effect on the matrix elements shown on the right panel of figure 17 is similar to the one derived with the most stringent statistical tolerance in the parametric fit of figure 13.

Using the same flat prior, we introduce another grid in α consisting of 1000 unevenly spaced points. Specifically, the points between $[10^{-4}, 0.1]$ and $[0.9, 1-10^{-12}]$ are evenly spaced logarithmically, while those in the middle are evenly spaced linearly. There are twice as many points in the upper α region than the other two. The results are shown in figure 18. The minimum of twice the negative log posterior is very similar: $\chi^2 = 3.76$ and $2uS = 9.24$. Since the grid is denser for $\alpha \leq 0.1$ and $\alpha \geq 0.9$ compared to the previous case, more fluctuations away from the prior are allowed for the same prior penalty in eq. (4.10) as these fluctuations occur on a shorter range in α. This generates a much larger uncertainty in the extraction of the
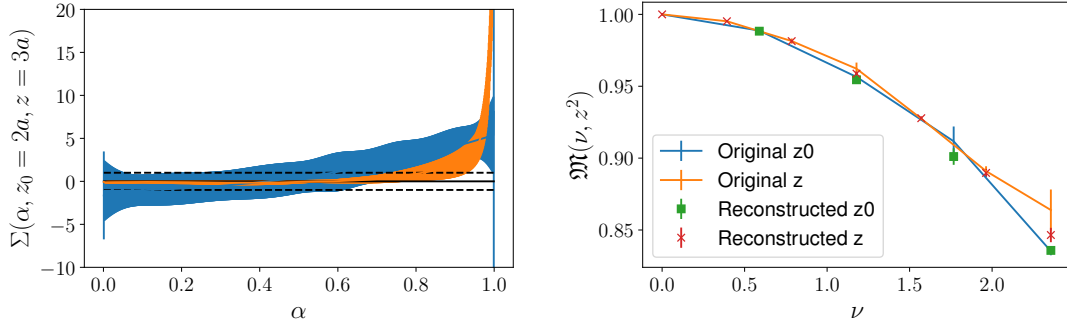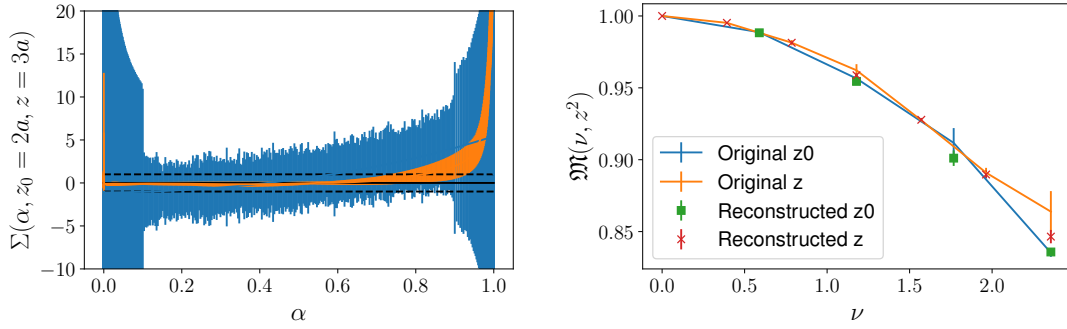
Figure 17. (Left) The step-scaling function from Bayesian Reconstruction (blue) and from the analysis of section 4.3 (orange). The grid consists of 1000 evenly spaced points in α. The prior distribution is defined by $u = 1$, $h(\alpha) = 0$, and $\sigma(\alpha) = 1$. The prior model (solid) and width (dashed) are shown in black. (Right) The reproduction of the data sets.
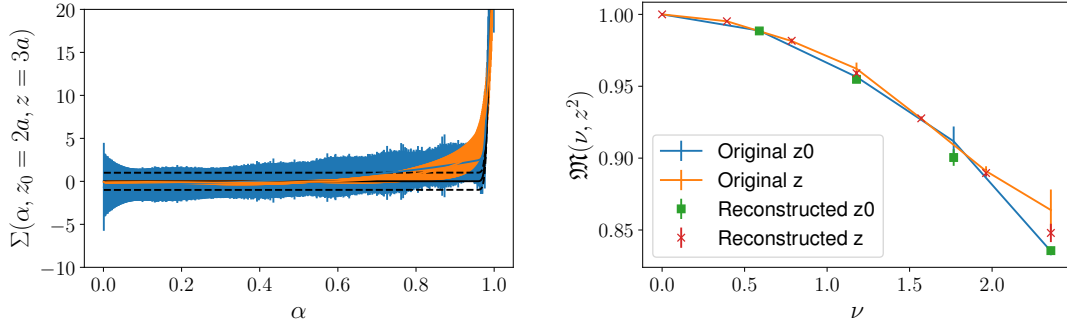

Figure 18. (Left) The step-scaling function from Bayesian Reconstruction (blue) and from the analysis of section 4.3 (orange). The grid consists of 1000 unevenly spaced points in α. The prior distribution is defined by $u = 1$, $h(\alpha) = 0$, and $\sigma(\alpha) = 1$. The prior model (solid) and width (dashed) are shown in black. (Right) The reproduction of the data sets.

step-scaling function as the smoothness of the reconstructed function decreases. Remarkably, this has very little effect on the correction of the matrix elements. Indeed, the matrix elements result from a convolution with the spline-interpolated data, so as we have already argued, they are not sensitive to the short-range fluctuations of the extraction of the step-scaling function.

There is little reason to believe that the prior model $h = 0$ will generate a useful bias to obtain the physics we are interested in. For instance, it contains potentially incorrect limits for $\alpha = 0, 1$ and does not produce an overall integral close to 1. In order to reproduce the Dirac-like behavior as $\alpha \to 1$, the prior distribution could be a sharply peaked Gaussian

$$h(\alpha) = N(1-\alpha, w) = \frac{\exp\left[-\frac{(1-\alpha)^2}{2w^2}\right]}{w\sqrt{2\pi}} \tag{4.12}$$

which simulates $\frac{1}{2}\delta(1-\alpha)$ for small w. Figures 19 and 20 show the results of this analysis with $w = 0.01$. The same grids in α were used again. In both scenarios, the sharply peaked Gaussian prior pulls upward the large α region. The minima of these fits are $\chi^2 = 3.28$ and
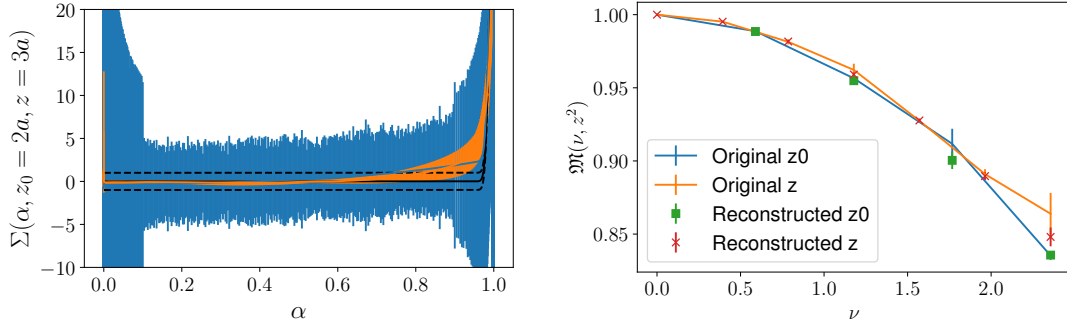
Figure 19. (Left) The step-scaling function from Bayesian Reconstruction (blue) and from the analysis of section 4.3 (orange). The grid consists of 1000 evenly spaced points in α. The prior distribution is defined by $u = 1$, $h(\alpha) = N(1-\alpha, w = 0.01)$, and $\sigma(\alpha) = 1$. The prior model (solid) and width (dashed) are shown in black. (Right) The reproduction of the data sets.

Figure 20. (Left) The step-scaling function from Bayesian Reconstruction (blue) and from the analysis of section 4.3 (orange). The grid consists of 1000 unevenly spaced points in α. The prior distribution is defined by $u = 1$, $h(\alpha) = N(1-\alpha, w = 0.01)$, and $\sigma(\alpha) = 1$. The prior model (solid) and width (dashed) are shown in black. (Right) The reproduction of the data sets.

$2uS = 2.94$ for the even grid and $\chi^2 = 3.24$ and $2uS = 1.88$ for the uneven grid. The data is slightly better reproduced than with the previous prior model while the prior term has decreased significantly. The correction of the matrix elements is once again unbothered by the change from even to uneven grid, and very close to the results obtained with the parametric fit or the flat prior.

A final prior model and width to test would be the results from the parameterized model in section 4.3. The model is set to the central value and the width is given by the larger of the upper or lower bands in the χ^2 per $n_{pts} < 2$ error estimation. The results are shown in figure 21. This analysis has again a slightly better agreement to the data and a much better agreement to the prior model. For the even grid, $\chi^2 = 3.06$ and $2uS = 0.046$ and for the uneven grid, $\chi^2 = 3.04$ and $2uS = 0.044$. Since this prior model was itself obtained through a careful fit on the data, it is unsurprising that the prior distribution is well satisfied.

In all these cases, the errors have been quite large on the resulting step-scaling function. This is due to the fact that there is no requirement of smoothness in the QDR prior defined
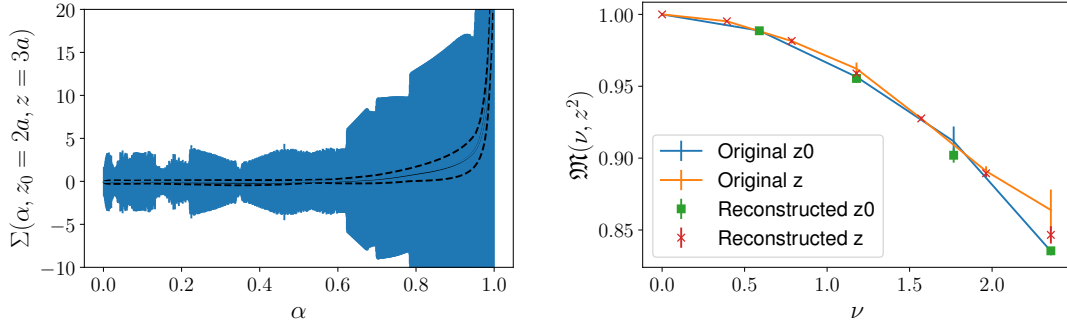
Figure 21. (Left) The step-scaling function from Bayesian Reconstruction (blue). The grid consists of 1000 evenly spaced points in α. The prior distribution is defined by $u=1$, and the fit results of section 4.3. The prior model (solid) and width (dashed) are shown in black. (Right) The reproduction of the data sets.
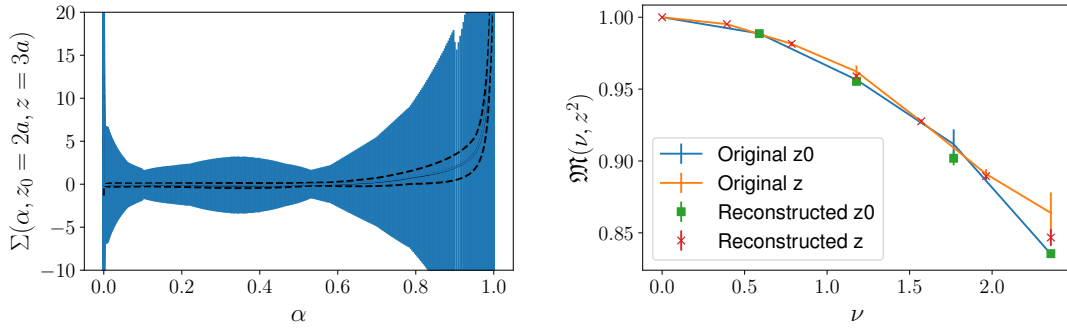

Figure 22. (Left) The step-scaling function from Bayesian Reconstruction (blue). The grid consists of 1000 unevenly spaced points in α. The prior distribution is defined by $u=1$, and the fit results of section 4.3. The prior model (solid) and width (dashed) are shown in black. (Right) The reproduction of the data sets.

in eq. (4.10): the reconstructed step-scaling function can vary abruptly from one value of the grid to the next. Using the inverse Hessian of the first fit with $h=0$ and $\sigma=1$, we produce samples of the step-scaling function from the multivariate Gaussian distribution. The samples fluctuate quite wildly as shown in figure 23. Instead one could consider a prior which does not bias the central value at all but instead biases the derivatives to be smooth.

Consider instead a smoothness based prior such as:

$$S = \int_0^1 d\alpha \, \frac{\alpha(1-\alpha)}{2} \left(\Sigma'(\alpha)\right)^2 = \sum_i \frac{\alpha_i(1-\alpha_i)}{2} \frac{(\Sigma(\alpha_{i+1})-\Sigma(\alpha_i))^2}{\alpha_{i+1}-\alpha_i} \qquad (4.13)$$

This prior is intended to keep the function to be relatively smooth in the middle α region while not disfavoring the potential divergences are $\alpha=0$ or 1. The fit results are shown in figure 24 where the variance is smaller than the QDR case. As can be seen in figure 23 this corresponds to sampled step-scaling functions that are noticeably smoother than the QDR case. The posterior at the minimum value is given by $\chi^2=3.82$ and $2uS=7.56$. While the
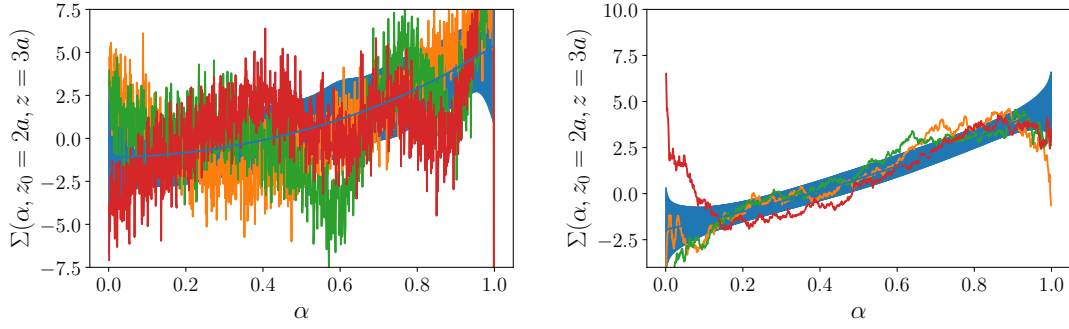
Figure 23. Probable step-scaling functions (red/orange/green) are sampled from a multivariate normal distribution with the full central value and covariance (blue). (Left) Results from the QDR prior with $h = 0$ and $\sigma = 1$. The function is allowed to vary wildly from one point to the next as represented by the large error band. (Right) Results from the smoothness prior. The functions vary significantly less from one point to the next resulting in the smaller variance.
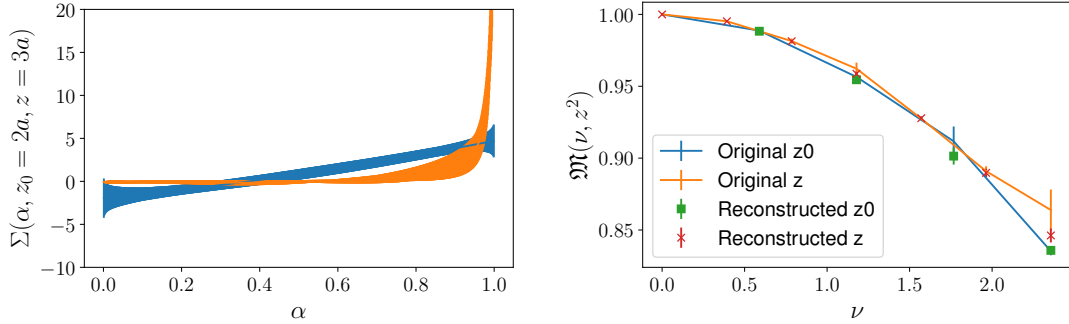

Figure 24. (Left) The step-scaling function from Bayesian Reconstruction (blue) and from the analysis of section 4.3 (orange). The grid consists of 1000 evenly spaced points in α. The prior distribution is defined by $u = 1$ and the smoothness function in eq. (4.4). (Right) The reproduction of the data sets.

size of the prior terms are not directly comparable, the χ^2 is comparable to the previous cases, and the effect of correction on the matrix elements again very similar.

We have shown that the Bayesian reconstruction methods can be utilized to determine the step-scaling function from lattice QCD data. With each of our different choices of prior distribution, the data are well reproduced by step-scaling functions that share the same general features, also common with the parametric fit. Among those, a small α contribution mostly compatible with zero, and an increase at large α. Although we have exhibited a large sensitivity of the reconstruction to the value of hyperparameters such as the grid density or the prior model, the correction of the matrix elements induced by the fit has shown a great stability. This motivates our claim that it is really the requirement of existence of a step-scaling function along the convolution of eq. (4.1) at small z^2, and the self-consistency that it imposes on the z^2 dependence of the lattice data, that has been the main driver of this correction. Implementing such a correction to the lattice data allows a systematic improvement at the
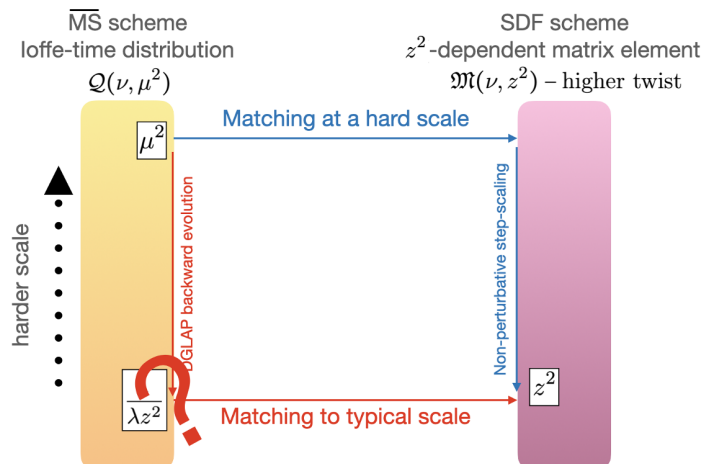
Figure 25. Instead of performing the matching at a low scale $\mu^2 \lambda z^2 = 1$ where α_s is nearly divergent and perturbation theory uncertain, we evolve in a non-perturbative way the matrix element to a smaller value of z^2, where ordinary matching can now be performed safely.

cost of a cheap post-processing with the possibility to limit model dependence considerably compared to using a perturbative matching to achieve the same result.

Finally, this discussion of Bayesian reconstruction has brought again the central aspect of the regularization of the inverse problem. We have addressed it through the angle of the smoothness and smearing throughout this paper. An alternative could be to study many choices of prior distributions to create a Bayesian Model Averaged [82] final result for the step-scaling functions to alleviate the biases from any individual choice. The model averaging could also be augmented by adding results not just from Bayesian Reconstruction but also from parametric models. To accurately and precisely determine the step-scaling function, one must perform this task on a series of ensembles, as we outline in the following section.

5 Lattice calculations of parton distributions in small volumes

At small z^2 the extraction of a step-scaling function in z^2 is possible by imposing self-consistency of the z^2 dependence, which results in a determination of corrected matrix elements with sizable corrections in some cases. A much better improvement could likely be achieved if, instead of using an internal consistency of the lattice data, one used a separate dedicated lattice computation of the step-scaling function. In particular, the control of higher-twist effects — which we have assumed to be negligible here, but could still be present at large Ioffe time for $z \gtrsim 0.2\,\mathrm{fm}$ — could be understood by performing the extraction of the step-scaling function from multiple external states accessible from lattice QCD, such as a pion, a kaon, and even a quark state in a fixed gauge. Then, a range of validity of a truly universal leading-twist step-scaling operator could be determined and used to hunt down higher twist contamination that would appear as discrepancies in the z^2 dependence of the data compared to the expected leading-twist step-scaling function.

This non-perturbative evolution operator could be used to bring lattice calculations to higher scales more amenable to perturbation theory in a self-consistent way, thereby reducing

the size of theoretical uncertainty induced by the perturbative matching as illustrated on figure 25. Additionally if short enough distance scales can be reached, one could invert eq. (3.17) to obtain the $\overline{MS}$ step scaling function from a non-perturbative calculation of the pseudo-PDF step scaling function. This evolution can be used in global analyses of datasets whose hard scales are perhaps too low for accurate pertubative evolution. Small volume calculations, as we will explore below, could be used to reach much smaller values of z^2, at which point one could hope to factorize directly experimental processes without any explicit reference to matching at all. To some extent, the use of non-perturbative evolution to reduce the uncertainty in the matching could be compared to the approach advocated in [83]. There, the authors suggest to determine the Mellin moments of the matching kernel by comparing experimental estimates of moments of the PDF and lattice estimates of the moments of pseudo-PDF in the small z^2 region where the matching is universal. This non-perturbative matching kernel could be used to derive a non-perturbative evolution operator, although the main ingredient is quite different: [83] uses experimental data as a point of reference whereas we suggest to use a dedicated lattice computation fully inscribed within the theoretical framework of the Standard Model which allows for an independent prediction required for BSM studies.

One less obvious use of the discussion of section IV is however its connection to finite volume effects, and the perspectives it opens for a radical new proposal of computations of parton distributions in small volumes. Before highlighting this connection, let us present why we believe that small volume calculations could present a crucial perspective for the program of parton physics on the lattice. Demanding lattice computations by current standards could typically use the following settings with a near-physical pion mass: $a = 0.05\,\mathrm{fm}$ and $L = 128a = 6.4\,\mathrm{fm}$. Then the largest accessible momentum allowed by the Brillouin zone of the lattice is of the order of $P = \frac{2\pi}{L} \times \frac{L}{6a} = \pi/(3a) \approx 4.2\,\mathrm{GeV}$. This happens to also be a momentum at which the issue of signal-to-noise ratio becomes quite critical. The break down of renormalization group improved perturbation theory and the growth of higher twist effects precludes the use of non-local matrix elements with a separation larger than $0.3\,\mathrm{fm}$. If one uses a separation of $z = 0.1\,\mathrm{fm}$ in the hope that it is small enough to effectively remove the theoretical uncertainties in the partonic interpretation of the matrix element, then the largest accessible Ioffe times $z \cdot P$ are only of about 2. This range in Ioffe time (coincidentally, it is roughly the one which we have been working with in this paper — although we worked at smaller momentum and larger z) is quite disappointing considering the efforts required to handle such a fine lattice. Of course, if we do not want to increase z, the only margin of improvement is to go to ever larger momenta and therefore finer lattices. However, both aspects are a serious issue. Just to obtain a range in Ioffe time of about 4 would require momenta of the order of $8\,\mathrm{GeV}$, and lattice spacing of the order of $0.03\,\mathrm{fm}$. The computational cost due to the issue of "critical slowing-down" [84] and the degrading signal-to-noise ratio make this situation unrealistic for what is additionally still a very limited range in Ioffe time. The challenge of accessing large momenta occurs in rather similar terms in the LaMET formalism. For instance, ref. [85] notes that if the validity of the LaMET perturbative matching is limited to $z \approx 0.2\,\mathrm{fm}$, then reliable calculations down to $x = 0.1$ would require momenta in excess of $10\,\mathrm{GeV}$. Efforts to probe larger momenta at fixed lattice size are underway [86], but cannot yet achieve the large momenta necessary.

Requiring very fine lattices, while preserving a volume (much) larger than the hadronic scale (two-scale problem), has been a challenge in several domains of lattice QCD, notably for B-meson physics calculations [24, 87, 88], where the large quark mass plays an analogous role as our requirement of large momentum P. Such calculations are performed by dropping the requirement of infinite volume limit, which allows very fine lattices in ever smaller volumes, while keeping the number of sites constant. This solves both the question of the largest available momentum and the issue of signal-to-noise ratio. In sub-hadronic lattice volumes, hadrons no longer exist, and the uncomfortably light pion no longer dictates the rate of increase of the noise. Instead, it is the scale tied to the finite volume $1/L$ that determines all spectral gaps and the scaling of noise. Therefore, as the volume decreases in physical units while keeping it fixed in lattice units the spectral gaps and the noise scaling will remain roughly constant allowing us to access increasing momentum in physical units.

Every time the volume and spacing of the lattice are divided by two while keeping the physical value of z constant, the accessible range in Ioffe time is multiplied by two. For a relatively similar computational cost as the $a = 0.05$ fm, $L = 128a = 6.4$ fm lattice that we have mentioned before, we could therefore carry out a computation with $a = 0.0063$ fm and $L = 128a = 0.8$ fm. This would give access to a range of Ioffe time of about 16 for $z = 0.1$ fm! The catch, of course, is that in this small volume the matrix element is no longer that of a hadron, and its interpretation in terms of PDFs is impossible without further work. Therefore, one must find a way to understand the nature of the finite volume effects to connect the matrix element computed in small volume to its hadronic counterpart in infinite volume. This is precisely the traditional role of the step-scaling function whose potential in B-physics calculations has been demonstrated in [24, 87, 88].

The possibility to perform small volume computations of the first DGLAP anomalous dimension has been demonstrated in the quenched approximation using the Schrödinger functional method and local matrix elements [89, 90]. We believe that the success of this study, which reproduces the perturbative results at large scales, is a sign that even for non-local matrix elements, with $z \ll L$, the finite volume effects can be propagated through the form of the standard DGLAP convolution. More precisely, we have been working so far in this paper implicitly in the infinite volume and continuum limit, and only focused on the α dependence of the step-scaling function that evolves from one scale z^2 to another. But for $z \ll \Lambda_{QCD}^{-1}, L$, we could likewise introduce a generalized step-scaling function accounting for changes in volume:

$$\Sigma(\alpha; z_0^2, z_1^2; L_0, L_1). \tag{5.1}$$

The general principle of the computation of the PDF in small volume could then be the following:

1. Compute the matrix element at $z = 0.1$ fm in a small volume lattice $L = 128a = 0.8$ fm. The accessible Ioffe time range could be up to 16.

2. Compute the matrix element in a larger volume by a factor 2. The accessible Ioffe time range is reduced to 8. By using a similar procedure to that explained in section IV of this paper, compute the generalized step-scaling function $\Sigma(\alpha; z_0 = z_1 = 0.1$ fm$; L_0 = 0.8$ fm$, L_1 = 1.6$ fm$)$ using the data in both volumes on the Ioffe time range up to 8.

3. Repeat the previous step by doubling the volume up to $L = 6.4$ fm, each time computing the step-scaling function on a shrinking Ioffe time range.

4. Use the computed step-scaling functions to evolve the initial computation in the full glory of its Ioffe time range up to 16 all the way to the large volume.

One will notice that the main "trick" is that the step-scaling function is independent of the Ioffe-time range if z^2 is small enough, and can therefore be computed even though the range in Ioffe time decreases as one approaches the infinite volume limit. One could argue that since the step-scaling functions to the largest volumes are computed on much reduced Ioffe time ranges, they will be less precise and will add systematic uncertainty when evolving the large Ioffe time range of the small volume matrix elements. We believe, however, that a control of the precision of the step-scaling uncertainty is possible for two main reasons. First, as the volume increases, the effects of small volume should be reduced and the step-scaling function should contain less information — it will be more akin to a simple Dirac delta. Second, if we find the quality of the step-scaling function to be disappointing — and we have tools with reduced model dependence to evaluate that as we have presented in this paper — then we can always increase the density of available Ioffe times within the same limited range thanks to twisted boundary conditions, and therefore obtain a more precise extraction of the step-scaling function. A detailed study of the finite volume step scaling function is beyond the scope of this paper and we leave it for future studies.

6 Conclusions

In this work, we have studied the evolution of the flavor non-singlet unpolarized pseudo-distribution with respect to the scale z^2 and outlined the nature of the step scaling function, which evolves matrix elements from one scale to another. At small enough z^2, where the OPE can be established, we have shown that the existence of this step-scaling function implies a self-consistent effect of z^2 dependence among the lattice data points which can lead to sizeable corrections of fluctuations in the data. The ability to evolve matrix elements from longer to shorter distance scales is fundamentally important for the analysis of parton distributions from lattice QCD. By controlling this evolution properly, one can reach scales where perturbative comparison to the light-cone $\overline{MS}$ parton distributions can be done with reduced perturbative systematic errors. As has been shown this evolution is intimately connected to the standard DGLAP evolution in the $\overline{MS}$ scheme and the matching kernels relating the pseudo-distributions to the $\overline{MS}$ scheme.

Unfortunately the perturbative expansions, and the application of the DGLAP evolution in the $\overline{MS}$ scheme, are potentially insufficient at the scales utilized for modern lattice QCD calculations. We have outlined a proposal to access a larger kinematic range on the lattice, while preserving good control of the matching accuracy. In analogy to approaches used in heavy quark physics in lattice QCD, the step scaling functions, both in scale z^2 and in volume L, can be modelled using the techniques outlined in section 4 on a series of ensembles. By scaling in volume as well as in the invariant scale, the step scaling function can be computed with Schrödinger functional techniques with small enough lattice spacings that perturbative matching to the $\overline{MS}$ scheme can be done with minimal error.

We study the extraction of the step scaling functions to determine the feasibility of our proposal. The step scaling function fit from real lattice data was modelled by a parameterized form. The result reproduced a function quite similar to the known perturbative expansions. To study the model dependence, a non-parameteric form, Bayesian reconstruction, was used to reproduce the step scaling function. The function is determined by its value on a grid over its domain and some interpolation rules. The ill-posed inverse problem is regulated by a choice of prior distribution, several of which were studied. A dedicated study of optimal choices of prior distributions can be performed in the future using mock or real lattice data. Such tests will be key to understanding the model dependence of the final step scaling functions.

Finally, this work can be extended to other parton distributions and their step scaling can be studied similarly. Comparing to the results of a pion matrix element analysis is critical to understanding the systematic errors and range of validity in this approach. Furthermore, the evolution of the nucleon GPD and TMD for example, would be of great interest to current exploratory lattice calculations. Each of these distributions are significantly less well-known from experimental results alone. If lattice QCD data for PDFs are to be compared directly to experimental results, as done in [91, 92], then the evolution of the lattice data from long distance scales must be done properly.

A   Technical appendix on the matching relations

A.1   Perturbative expansion of the matching kernel

Provided a perturbative expansion of the matching kernel exists, it is possible to express it from any z^2 to the $\overline{MS}$ scale $(\lambda z^2)^{-1}$ as:

$$\mathcal{C}_0(z^2) = \sum_{k=0}^{\infty} \left(\frac{\alpha_s((\lambda z^2)^{-1})}{2\pi} \right)^k c_{k,0}(\lambda), \tag{A.1}$$

where $c_{k,0}(\lambda)$ are distributions which depend on the choice of λ. This expression singles out a specific $\overline{MS}$ scale $(\lambda z^2)^{-1}$. The general matching to any scale can be derived exactly following eq. (3.11) as the convolution of $\mathcal{C}_0$ with the backward $\overline{MS}$ step-scaling function. This yields the ordinary presentation of the matching kernel in the literature as a double expansion in terms of $\alpha_s(\mu^2)$ and $\ln(\mu^2 z^2 \lambda)$. If all orders are considered, the intermediate scale $(\lambda z^2)^{-1}$ does not matter. If a perturbative truncation is performed, then this scale becomes a first choice with a simpler matching devoid of logarithms.

If $\lambda = -e^{2\gamma_E + 1}/4$, then eq. (2.7) teaches us that:

$$c_{0,0}(\lambda) = 1_\otimes, \qquad c_{1,0}(\lambda) = -D, \dots \tag{A.2}$$

In general, one has:

$$\mathcal{C}(z^2 \mu^2, \alpha_s(\mu^2)) = 1_\otimes - \frac{\alpha_s(\mu^2)}{2\pi} \left[\ln\left(-z^2 \mu^2 \frac{e^{2\gamma_E + 1}}{4} \right) B_1 + D \right] + \mathcal{O}(\alpha_s^2),$$

$$= 1_\otimes - \frac{\alpha_s(\mu^2)}{2\pi} \left[\ln\left(z^2 \mu^2 \lambda \right) B_1 + \ln\left(-\frac{e^{2\gamma_E + 1}}{4\lambda} \right) B_1 + D \right] + \mathcal{O}(\alpha_s^2), \tag{A.3}$$

so

$$c_{0,0}(\lambda) = 1_\otimes, \qquad c_{1,0}(\lambda) = -D - \ln\left(-\frac{e^{2\gamma_E+1}}{4\lambda}\right)B_1, \dots \tag{A.4}$$

For instance, the derivation is performed with a value of $\lambda = -e^{2\gamma_E}/4$ in [17]. One can verify that their matching relation gives indeed (see the next section for more details):

$$c_{0,0}(\lambda) = 1_\otimes, \qquad c_{1,0}(\lambda) = -D - B_1, \dots \tag{A.5}$$

Let us demonstrate how the double expansion in terms of $\alpha_s(\mu^2)$ and $\ln(\mu^2 z^2 \lambda)$ arises, and how the coefficients in front of the logarithms can be directly obtained as functions of the DGLAP splitting function and the $c_{k,0}(\lambda)$ distributions.

A Taylor expansion of the $\overline{MS}$ step-scaling function in terms of $\ln(\mu^2)$ reads:

$$\mathcal{E}(\mu_0^2, \mu^2) = \sum_{k=0}^{+\infty} \frac{1}{k!} \ln^k\left(\frac{\mu^2}{\mu_0^2}\right) \frac{d^k \mathcal{E}(\mu_0^2, \mu^2)}{d\ln^k(\mu^2)}\bigg|_{\mu^2=\mu_0^2}, \tag{A.6}$$

where the DGLAP differential equation (3.1) and the running of the strong coupling read:

$$\begin{cases} \dfrac{d}{d\ln\mu^2}\mathcal{E}(\mu_0^2, \mu^2) = \left\{\dfrac{\alpha_s(\mu^2)}{2\pi}B_1 + \left(\dfrac{\alpha_s(\mu^2)}{2\pi}\right)^2 B_2 + \dots\right\} \otimes \mathcal{E}(\mu_0^2, \mu^2), \\ \dfrac{d}{d\ln\mu^2}\alpha_s(\mu^2) = -\beta_0\alpha_s^2(\mu^2) - \beta_1\alpha_s^3(\mu^2) - \dots. \end{cases} \tag{A.7}$$

It is easy to see that the differential of order k of $\mathcal{E}(\mu_0^2, \mu^2)$ starts at order $\mathcal{O}(\alpha_s^k)$. For instance:

$$\frac{d^2}{d\ln^2(\mu^2)}\mathcal{E}(\mu_0^2, \mu^2)\bigg|_{\mu^2=\mu_0^2} = \left(\frac{\alpha_s(\mu_0^2)}{2\pi}\right)^2(B_1^{\otimes 2} - 2\pi\beta_0 B_1) + \mathcal{O}(\alpha_s^3). \tag{A.8}$$

Straightforward calculations then gives that the $\overline{MS}$ step-scaling function can be written as:

$$\mathcal{E}(\mu_0^2, \mu^2) = 1_\otimes + \sum_{n=1}^{\infty}\left(\frac{\alpha_s(\mu_0^2)}{2\pi}\right)^n \sum_{k=1}^{n} \ln^k\left(\frac{\mu^2}{\mu_0^2}\right)\mathcal{E}_{n,k}, \tag{A.9}$$

where for instance:

$$\mathcal{E}_{1,1} = B_1, \qquad \mathcal{E}_{2,1} = B_2, \qquad \mathcal{E}_{2,2} = \frac{1}{2}(B_1^{\otimes 2} - 2\pi\beta_0 B_1). \tag{A.10}$$

An important observation is that the terms $\mathcal{E}_{n,n-k}$ which constitute the k-th leading logarithmic expansion ($\mathrm{N}^k\mathrm{LL}$) are only expressed as functions of B_{1+j} and β_j such that $0 \leq j \leq k$.

In fact, we are interested in the backward $\overline{MS}$ step-scaling function in eq. (3.11), so simple calculations give:

$$\mathcal{E}(\mu^2, \mu_0^2) = 1_\otimes + \sum_{n=1}^{\infty}\left(\frac{\alpha_s(\mu_0^2)}{2\pi}\right)^n \sum_{k=1}^{n} \ln^k\left(\frac{\mu^2}{\mu_0^2}\right)\widetilde{\mathcal{E}}_{n,k}, \tag{A.11}$$

where:

$$\widetilde{\mathcal{E}}_{1,1} = -B_1, \qquad \widetilde{\mathcal{E}}_{2,1} = -B_2, \qquad \widetilde{\mathcal{E}}_{2,2} = \frac{1}{2}(B_1^{\otimes 2} + 2\pi\beta_0 B_1). \tag{A.12}$$

Evaluating eq. (A.11) for $\mu_0^2 = (\lambda z^2)^{-1}$ and convoluting with the expansion of $\mathcal{C}_0(z^2)$ yields:

$$\mathcal{C}(z^2\mu^2, \alpha_s(\mu^2)) = 1_\otimes + \sum_{n=1}^\infty \left(\frac{\alpha_s(\mu_0^2)}{2\pi}\right)^n \sum_{k=0}^n \ln^k\left(\mu^2 z^2 \lambda\right) \mathcal{F}_{n,k}, \qquad (A.13)$$

where:

$$\mathcal{F}_{n,0} = c_{n,0}(\lambda), \quad \mathcal{F}_{1,1} = -B_1, \quad \mathcal{F}_{2,1} = -c_{1,0}(\lambda)\otimes B_1 - B_2, \quad \mathcal{F}_{2,2} = \frac{1}{2}(B_1^{\otimes 2} + 2\pi\beta_0 B_1). \tag{A.14}$$

Finally, we expand $\alpha_s(\mu_0^2)$ as a series in $\alpha_s(\mu^2)$ and $\log(\mu^2/\mu_0^2)$:

$$\alpha_s(\mu_0^2) = \sum_{j=0}^\infty \frac{(-1)^j}{j!} \ln^j\left(\frac{\mu^2}{\mu_0^2}\right) \frac{d^j \alpha_s(\mu^2)}{d\log^j(\mu^2)} = \alpha_s(\mu^2) + \beta_0 \alpha_s^2(\mu^2) \ln\left(\frac{\mu^2}{\mu_0^2}\right) + \mathcal{O}(\alpha_s^3), \qquad (A.15)$$

and we obtain the general perturbative expansion of the matching kernel $\mathcal{C}$:

$$\mathcal{C}(z^2\mu^2, \alpha_s(\mu^2)) = 1_\otimes + \sum_{n=1}^\infty \left(\frac{\alpha_s(\mu^2)}{2\pi}\right)^n \sum_{k=0}^n \ln^k(\mu^2 z^2 \lambda) c_{n,k}(\lambda), \qquad (A.16)$$

where:

$$c_{1,1}(\lambda) = -B_1, \quad c_{2,1}(\lambda) = c_{1,0}(\lambda)\otimes(2\pi\beta_0 - B_1) - B_2, \quad c_{2,2}(\lambda) = \frac{1}{2}\left(B_1^{\otimes 2} - 2\pi\beta_0 B_1\right). \tag{A.17}$$

This gives a very simple understanding of the structure of the matching kernel whereas the analytical expressions of $c_{2,1}$ and $c_{2,2}$ already are seriously cumbersome. We check explicitly the validity of the formula for $c_{2,2}(\lambda)$ thanks to the matching kernel at order $\mathcal{O}(\alpha_s^2)$ derived in [17] in the next section. The terms $c_{n,n-k}(\lambda)$ which constitute the N^kLL expansion are only expressed as functions of $c_{j,0}(\lambda)$, B_{1+j} and β_j such that $0 \le j \le k$. In particular, since $c_{0,0}(\lambda) = 1_\otimes$, the LL expansion is independent of λ.

The conventions of eq. (2.7) use $-\lambda = e^{2\gamma_E + 1}/4 \approx 2.16$. On the other hand, [4, 17] uses $-\lambda = e^{2\gamma_E}/4 \approx 0.79$. Considering that a given choice of λ implies a simplified expression of the matching relation to the scale $\mu^2 = (\lambda z^2)^{-1}$, one could wonder if there is a natural choice to select the value of λ with which to present the results. A similar question has been raised in [4] where a value of $\mu^2 \sim -16/z^2$ ($\lambda = -0.06$) has been argued. A natural way to address the question could be to try and minimize the convolution of a typical ITD with $c_{1,0}(\lambda)$, so as to single out a scale which brings minimal corrections at fixed order $\mathcal{O}(\alpha_s)$. It has been noted in the literature [61] that the terms B_1 and D which intervene in the matching at order $\mathcal{O}(\alpha_s)$ produce typically contributions of opposite signs when convoluted to realistic ITDs. There is an optimal value λ_{opt} for which the convolution of the ITD with $c_{1,0}$ at a given ν will be 0. Figure 26 shows 4 models of ITDs obtained from the cosine transform of $x^a(1-x)^b/B(a+1, b+1)$ where B is the Beta function with varying (a, b). These ITDs are convoluted with B_1 and D. In the range of Ioffe times available to modern lattice QCD, $|\lambda_{opt}| < 0.4$ for all the models given. This value corresponds to scales μ greater than $1/z$ by the factor $(-\lambda_{opt})^{-1/2}$ which is between $[2, 4.5]$ for $\nu < 12$. Figure 27 shows the convolutions of $c_{1,0}$ with $\lambda = -0.1$ and -0.2. These convolutions have reduced magnitude compared to the
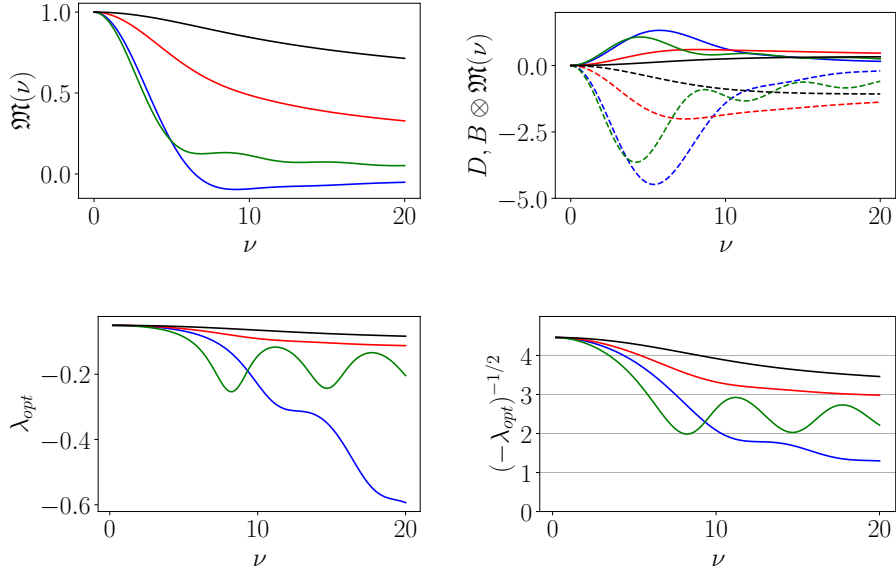
Figure 26. (Upper Left) Models of ITDs similar to those arising from the lattice. (Upper Right) Convolutions of those models with B_1 (solid) and D (dashed). The two convolutions differ in sign and magnitude. (Lower Left) The value of λ_{opt} which will make the convolution with $c_{1,0}$ to be 0. (Lower Right) $(-\lambda_{opt})^{-1/2}$ represents how much larger μ should be than $1/z$ in order to simultaneously cancel the logarithms of those scales and $c_{1,0}$.

convolutions of D in figure 26, which corresponds to $\lambda = -e^{2\gamma_E+1}/4$. There does appear to be a reduction in the magnitude, specifically for the model with the largest convolutions.

Although this choice of scale reduces the typical size of the contribution of order $\mathcal{O}(\alpha_s)$, it does not offer much information on the size of further terms in the expansion. The trade-off is a potentially increase in $c_{2,0}$ which contains terms proportional to $\ln^2\left(-\frac{e^{2\gamma_E+1}}{4\lambda}\right) \approx 9.41$ for $\lambda = -0.1$. Even so a choice of $\lambda > -e^{2\gamma_E+1}/4$ could still improve the perturbative expansion without larger contaminations of neglected terms. For $\lambda = -0.5$ the $c_{1,0}$ convolution's peak is reduced nearly a factor of 2 for some models while the higher order terms increase by the smaller coefficient of $\ln^2\left(-\frac{e^{2\gamma_E+1}}{4\lambda}\right) \approx 2.16$. While the full optimization of λ for NLO and NNLO effects is beyond the extent of this study, it appears clear that varying significantly for $\lambda > -e^{2\gamma_E+1}/4$ has the potential to reduce perturbative effects.

A.2 Observations on the matching relations at order $\mathcal{O}(\alpha_s^2)$

Modifying slightly the matching relation in [17] for our needs, we can write:

$$\langle P|\bar{\psi}(z)\gamma^\nu \hat{E}(0,z;A)\psi(0)|P\rangle = Z^{\overline{MS}}(z^2,\mu^2)\int_{-1}^{1} dx\, q(x,\mu^2) K^\nu(x\nu, z^2, \mu^2). \tag{A.18}$$
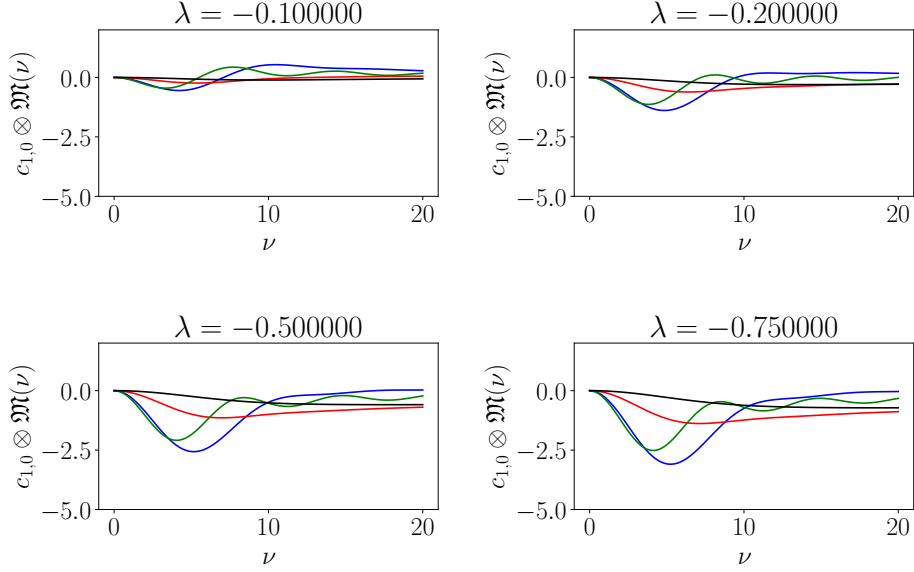
Figure 27. The convolution of the model functions with $c_{1,0}$ for variance λ. This can be compared to $\lambda = -e^{2\gamma_E+1}/4 \approx 2.15$ which are the dashed lines in upper right of figure 26.

On the other hand, we have used in this document the definition, neglecting the higher-twist contributions:

$$
\begin{aligned}
\mathfrak{M}(\nu, z^2) &\equiv \frac{\langle P|\bar{\psi}(z)\gamma^0\hat{E}(0,z;A)\psi(0)|P\rangle}{\langle 0|\bar{\psi}(z)\gamma^0\hat{E}(0,z;A)\psi(0)|0\rangle} \frac{\langle 0|\bar{\psi}(0)\gamma^0\hat{E}(0,z;A)\psi(0)|0\rangle}{\langle P|\bar{\psi}(0)\gamma^0\hat{E}(0,z;A)\psi(0)|P\rangle} \\
&= \int_0^1 d\alpha\, \mathcal{C}(\alpha, z^2, \mu^2)\mathcal{Q}(\alpha\nu, \mu^2)\,.
\end{aligned}
\tag{A.19}
$$

From eq. (A.18), we find that:

$$
\mathfrak{M}(\nu, z^2) = \frac{1}{AK^0(0, z^2, \mu^2)} \int_{-1}^1 dx\, q(x, \mu^2) K^0(x\nu, z^2, \mu^2)\,,
\tag{A.20}
$$

where $A = \int_{-1}^1 dx\, q(x, \mu^2)$. Remembering that we have defined the Ioffe-time distribution through eq. (2.5):

$$
\mathcal{Q}(\nu, \mu^2) \equiv \frac{1}{A} \int_{-1}^1 dx\, e^{i\nu x} q(x, \mu^2)\,,
\tag{A.21}
$$

we transform eq. (A.19) into:

$$
\mathfrak{M}(\nu, z^2) = \frac{1}{A} \int_{-1}^1 dx\, q(x, \mu^2) \int_0^1 d\alpha\, e^{ix\alpha\nu}\mathcal{C}(\alpha, z^2, \mu^2)\,,
\tag{A.22}
$$

and the comparison between eqs. (A.20) and (A.22) gives:

$$
\mathcal{C}(\alpha, z^2, \mu^2) = \frac{1}{K^0(0, z^2, \mu^2)} \int_{-\infty}^{+\infty} \frac{dy}{2\pi} e^{-iy\alpha} K^0(y, z^2, \mu^2)\,.
\tag{A.23}
$$

From [17], we obtain that

$$K^0(0, z^2, \mu^2) = 2 + \frac{\alpha_s C_F}{\pi} \left(\frac{3}{2} L + 1 \right) + \mathcal{O}(\alpha_s^2), \tag{A.24}$$

where $L = \ln(-\mu^2 z^2 e^{2\gamma_E + 1}/4)$. On the other hand,

$$\int_{-\infty}^{+\infty} \frac{dy}{2\pi} e^{-iy\alpha} K^0(y, z^2, \mu^2)$$
$$= K^0(0, z^2, \mu^2)\delta(1-\alpha) - \frac{\alpha_s C_F}{\pi} \left(L \left[\frac{1+\alpha^2}{1-\alpha} \right]_+ + \left[\frac{4\ln(1-\alpha)}{1-\alpha} - 2(1-\alpha) \right]_+ \right) + \mathcal{O}(\alpha_s^2). \tag{A.25}$$

The ratio with $K^0(0, z^2, \mu^2)$ gives the expected result (2.7) up to the factor A which arises due to the lost normalization of the PDF in the ratio. We also obtain a correction of order $\mathcal{O}(\alpha_s^2)$ which is a non-trivial consequence of the ratio:

$$\mathcal{C}(\alpha, z^2, \mu^2) = \delta(1-\alpha) - \frac{\alpha_s C_F}{2\pi} \left(L \left[\frac{1+\alpha^2}{1-\alpha} \right]_+ + \left[\frac{4\ln(1-\alpha)}{1-\alpha} - 2(1-\alpha) \right]_+ \right)$$
$$+ \left(\frac{\alpha_s C_F}{2\pi} \right)^2 \left(\frac{3}{2} L^2 \left[\frac{1+\alpha^2}{1-\alpha} \right]_+ + L \left[\frac{1+\alpha^2}{1-\alpha} + \frac{6\ln(1-\alpha)}{1-\alpha} - 3(1-\alpha) \right]_+ \right.$$
$$\left. + \left[\frac{4\ln(1-\alpha)}{1-\alpha} - 2(1-\alpha) \right]_+ \right) + \mathcal{O}(\alpha_s^2). \tag{A.26}$$

The missing term of order $\mathcal{O}(\alpha_s^2)$ is the genuine 2-loop contribution which is particularly cumbersome to express. The leading logarithmic term gives:

$$\frac{\alpha_s^2 C_F}{2\pi^2} L^2 \left[-C_F \left\{ \frac{1-\alpha}{2} + \frac{(1+3\alpha^2)\ln(\alpha)}{4(1-\alpha)} - \frac{(1+\alpha^2)\ln(1-\alpha)}{1-\alpha} \right\} \right.$$
$$\left. - \frac{11 C_A(1+\alpha^2)}{24(1-\alpha)} + \frac{n_f T_F(1+\alpha^2)}{6(1-\alpha)} \right]_+. \tag{A.27}$$

According to the discussion of the previous section in eq. (A.17), the sum of that term and the corresponding correction of order $\mathcal{O}(\alpha_s^2)$ derived in eq. (A.26) should yield exactly $c_{2,2}(\lambda) = \left(B_1^{\otimes 2} - 2\pi\beta_0 B_1 \right)/2$. The sum writes:

$$c_{2,2}(\lambda) = 2 C_F \left[-C_F \left\{ \frac{1-\alpha}{2} + \frac{(1+3\alpha^2)\ln(\alpha)}{4(1-\alpha)} - \frac{(1+\alpha^2)\ln(1-\alpha)}{1-\alpha} - \frac{3(1+\alpha^2)}{4(1-\alpha)} \right\} \right.$$
$$\left. - \frac{11 C_A(1+\alpha^2)}{24(1-\alpha)} + \frac{n_f T_F(1+\alpha^2)}{6(1-\alpha)} \right]_+. \tag{A.28}$$

It is easy to recognize in the last two terms $-\pi\beta_0 B_1$. The first term is more sophisticated, and corresponds to $B_1^{\otimes 2}/2$, as can be observed by comparing the Mellin moments for instance.

B Existence and properties of the inverse for the DGLAP convolution

At order $\mathcal{O}(\alpha_s)$, the Mellin moments of the matching kernel defined in eq. (3.25) read:

$$c_n\left(z^2\right) = 1 - \frac{\alpha_s(\mu^2)}{2\pi} C_F \left(2\psi^{(0)}(n)[2\gamma_E + \psi^{(0)}(n)] - 2\psi^{(1)}(n) + 1 + 2\gamma_E^2 + \frac{\pi^2}{3} - \frac{2}{n} + \frac{2}{(n+1)} \right), \tag{B.1}$$

where $\psi^{(k)}(n)$ is the polygamma function, and $\mu^2 = (\lambda z^2)^{-1}$. The asymptotical behavior of the moments for $n \to 0$ and $n \to +\infty$ yields:

$$c_n\left(z^2\right) \stackrel{n \to 0}{=} 1 - \frac{\alpha_s(\mu^2)}{2\pi} C_F\left(-\frac{2}{n} + 3 - \frac{2\pi^2}{3} + \mathcal{O}(n)\right) \to +\infty\,, \tag{B.2}$$

$$c_n\left(z^2\right) \stackrel{n \to +\infty}{=} 1 - \frac{\alpha_s(\mu^2)}{2\pi} C_F\left(2\ln(n)[\ln(n) + 2\gamma_E] + 1 + 2\gamma_E^2 + \frac{\pi^2}{3} + \mathcal{O}\left(\frac{1}{n}\right)\right) \to -\infty\,. \tag{B.3}$$

Furthermore, we have demonstrated earlier (2.6) that, to all orders,

$$c_1(z^2) = 1\,. \tag{B.4}$$

The derivative of eq. (B.1) is strictly negative for $n \in [1, +\infty)$. The technical demonstration, without interest for the rest of the discussion, is produced at the end of this appendix. As a consequence, eq. (B.1), which spans $(-\infty, 1]$ for $n \in [1, +\infty)$ is bound to vanish for a unique (non-integer) value which we denote $n_0(\alpha_s) > 1$. Therefore, if c_n is truncated to order $\mathcal{O}(\alpha_s)$, $1/c_n$ exhibits a pole at $n = n_0(\alpha_s) > 1$.

In general, the Mellin transform (3.25) can be evaluated for complex values of n. It is known classically that the integral converges in "holomorphy strips" [93] defined in the complex plane generically by $(n_1, n_2) + i\mathbb{R}$, where $n_1, n_2 \in \mathbb{R} \cup \{\pm\infty\}$. The Mellin transform of a distribution is really the pair consisting of the Mellin moments and the holomorphy strip on which they are defined. Let us briefly observe an example borrowed from [93] where the Mellin moments diverge at some value of n (here $n = -2$):

$$c_n = -\frac{1}{n+2}\,. \tag{B.5}$$

There are two possible holomorphy strips, $(-\infty, -2) + i\mathbb{R}$ and $(-2, +\infty) + i\mathbb{R}$. In fact, one can verify easily that the distribution giving rise to the Mellin moments (B.5) in the first holomorphy strip is $\alpha^2 \Theta(\alpha - 1)$ where Θ is the Heaviside step function:

$$\int_0^\infty d\alpha\, \alpha^{n-1} \alpha^2 \Theta(\alpha - 1) = \int_1^\infty d\alpha\, \alpha^{n+1} = -\frac{1}{n+2} \quad \text{if } \mathrm{Re}(n) \leq -2\,. \tag{B.6}$$

On the other hand, the distribution giving rise to the same Mellin moments in the second holomorphy strip is $\alpha^2[\Theta(\alpha - 1) - \Theta(\alpha)]$:

$$\int_0^\infty d\alpha\, \alpha^{n-1} \alpha^2 [\Theta(\alpha - 1) - \Theta(\alpha)] = -\int_0^1 d\alpha\, \alpha^{n+1} = -\frac{1}{n+2} \quad \text{if } \mathrm{Re}(n) \geq -2\,. \tag{B.7}$$

The two distributions which give rise to the same general expression for the Mellin moments, but on two different holomorphy strips are shown on figure 28. It appears clearly that the choice of holomorphy strip changes crucially the properties of the distribution, notably its support.

In the case of the inverse matching kernel, we face a similar issue. In fact, there are infinitely many holomorphic strips available to reconstruct the inverse matching kernel from $1/c_n$, spanning the entire real values of n. But only one of those strips produces an integrable distribution: by definition, if the distribution is integrable, its moment for $n = 1$ is finite.
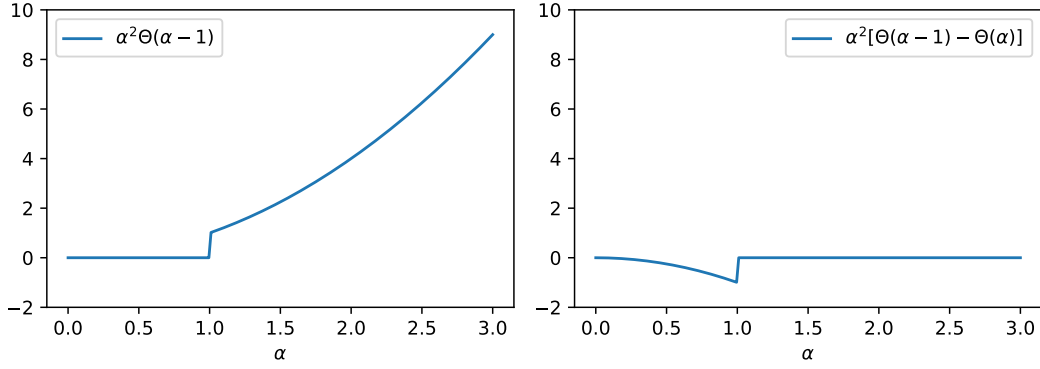
Figure 28. Two distributions whose Mellin moments are $-1/(n+2)$, but defined on different holomorphy strips: (left) $(-\infty, -2) + i\mathbb{R}$, (right) $(-2, +\infty) + i\mathbb{R}$.

Therefore, the holomorphy strip which we will use is $(n', n_0(\alpha_s)) + i\mathbb{R}$, where $n' < 1 < n_0(\alpha_s)$.[5] If the inverse matching kernel exists, then its Mellin moments in the vicinity of $n_0(\alpha_s)$ are:

$$\frac{1}{c_n} \overset{n \to n_0(\alpha_s)}{=\!=} \frac{1}{n - n_0(\alpha_s)} \left(\frac{d}{dn} c_{n_0} \right)^{-1} + \mathcal{O}(1). \tag{B.8}$$

An inverse Mellin transform of the right-hand side of (B.8) on the strip $(n', n_0(\alpha_s)) + i\mathbb{R}$ gives that, *provided the inverse of the matching kernel exists and is integrable*, its general behavior at large α has to be dominated by:

$$\mathcal{C}^{\otimes -1}(\alpha) \overset{\alpha \to +\infty}{\sim} \left(\frac{d}{dn} c_{n_0} \right)^{-1} \alpha^{-n_0(\alpha_s)}. \tag{B.9}$$

Therefore, the only integrable inverse matching kernel we can plausibly reconstruct from $1/c_n$ has a support which extends to arbitrarily large α, unlike the direct matching kernel whose support is limited to $\alpha \in [0, 1]$. This unwanted large α behavior decreases as a power function whose exponent depends on the value of α_s, and vanishes in the limit where $\alpha_s = 0$ which implies $n_0(\alpha_s) = +\infty$.

Schematically speaking, the Mellin moments of the direct matching kernel have the form:

$$c_n = 1 - \alpha_s h(n). \tag{B.10}$$

Hence:

$$\frac{1}{c_n} = 1 + \alpha_s h(n) + \alpha_s^2 h^2(n) + \dots \tag{B.11}$$

The terms of order α_s^2 and beyond are incomplete, since they will receive unknown corrections from higher orders in the perturbative expansion of c_n. Therefore, the issue of determining whether the inverse of the matching kernel truncated at order $\mathcal{O}(\alpha_s)$ exists, and what its support is, is not particularly worrisome *per se*. It should serve however as a warning of potentially complicated effects linked to the perturbative truncation. To recover a well-defined

---

[5]In fact, $n' = 0$ as a consequence of the (non-demonstrated) fact that the derivative of eq. (B.1) is strictly negative on $n \in (0, +\infty)$, and not just on $n \in [1, +\infty)$.

object that can serve as an inverse matching kernel and whose support is restricted to $[0,1]$, it is enough to truncate $\mathcal{C}_0^{\otimes-1}$ (or equivalently $1/c_n$) to order $\mathcal{O}(\alpha_s)$ too, thereby ignoring altogether the higher-order contributions we partially resummed when we considered the exact value of $1/c_n$:

$$\mathcal{C}_0^{\otimes-1}(z^2) = 1_\otimes + \frac{\alpha_s(\mu^2)}{2\pi} D + \mathcal{O}(\alpha_s^2). \tag{B.12}$$

A technical demonstration. We want to demonstrate that the derivative of the Mellin moments of the matching kernel truncated at order $\mathcal{O}(\alpha_s)$ (B.1) is strictly negative for $n \in [1, +\infty)$. In fact, that is also true for $n \in (0, +\infty)$ although the demonstration is more sophisticated in the latter case and of no interest for this discussion. The polygamma function may be defined as:

$$\psi^{(k)}(n) \equiv \frac{d^{k+1}}{dn^{k+1}} \ln\Gamma(n), \tag{B.13}$$

where Γ is the ordinary Gamma function. Then it follows immediately that:

$$\frac{d}{dn}\psi^{(k)}(n) = \psi^{(k+1)}(n), \tag{B.14}$$

and the derivative of eq. (B.1) gives:

$$\frac{d}{dn}c_n = -\frac{\alpha_s(\mu^2)}{2\pi} C_F \left(4\psi^{(1)}(n)(\gamma_E + \psi^{(0)}(n)) - 2\psi^{(2)}(n) + \frac{2}{n^2} - \frac{2}{(n+1)^2} \right). \tag{B.15}$$

In parallel, it is well-known that the polygamma function admits the following integral representation for $k > 0$:

$$\psi^{(k)}(n) = -\int_0^1 dt \frac{t^{n-1}}{1-t} \ln^k(t), \tag{B.16}$$

which gives trivially that $\psi^{(1)}(n) > 0$ and $\psi^{(2)}(n) < 0$ for $n > 0$. Therefore, $\psi^{(0)}(n)$ is a strictly increasing function, and since $\psi^{(0)}(1) = -\gamma_E$, we find for $n \geq 1$,

$$\gamma_E + \psi^{(0)}(n) \geq 0. \tag{B.17}$$

This demonstrates that the derivative (B.15) is strictly negative for $n \geq 1$ and gives the expected result. The result is also true for $n \in (0, 1)$, for instance since

$$\psi^{(0)}(n) - \psi^{(0)}(1) > \frac{\psi^{(2)}(n)}{2\psi^{(1)}(n)}. \tag{B.18}$$

This statement is particularly obvious if $n \geq 1$ since the l.h.s. is positive and the r.h.s. strictly negative, but is more subtle if $n \in (0, 1)$. We will not seek a further demonstration.

C Positivity of the LL $\overline{MS}$ step-scaling function

To demonstrate the positivity of the LL $\overline{MS}$ step-scaling function, we can show that it satisfies the Hausdorff moment problem [94]. The latter establishes the necessary and sufficient

condition so that a sequence of integer Mellin moments for $n=1$ to infinity corresponds to a unique Borel measure supported on the interval $[0,1]$. The condition is that the sequence of Mellin moments $(e_n)_n$ is completely monotonic, that is for all $n \geq 1$ and $k \geq 0$:

$$(-1)^k \Delta^k e_n \geq 0 \,, \tag{C.1}$$

where Δ^k is the difference operator applied k-times:

$$\Delta e_n = e_{n+1} - e_n \,, \qquad \Delta^2 e_n = e_{n+2} - 2e_{n+1} + e_n \,, \ldots \tag{C.2}$$

We remind that (3.32):

$$e_n(\mu_0^2, \mu_1^2) = \left(\frac{\alpha_s(\mu_0^2)}{\alpha_s(\mu_1^2)} \right)^{\gamma_n/(2\pi\beta_0)} \,, \tag{C.3}$$

and the anomalous dimensions γ_n are given by:

$$\gamma_n = -2\psi^{(0)}(n) - 2\gamma_E + \frac{3}{2} - \frac{1}{n} - \frac{1}{n+1} \,, \tag{C.4}$$

where $\psi^{(0)}$ is the digamma function defined in eq. (B.13).

The theorem 11d in Chapter 4 of [95] gives that the sequence $(e_n)_n$ is completely monotonic if the function e_n of n is a completely monotonic function for $n \geq 1$, a related property expressed in terms of derivatives instead of difference operators:

$$(-1)^k \frac{d^k}{dn^k} e_n \geq 0 \,. \tag{C.5}$$

The demonstration is a straightforward application of the mean value theorem. As $e_n \geq 0$, the property is obviously satisfied for $k = 0$. Differentiating repeatedly e_n as a composite function allows to express its derivatives with respect to those of γ_n for $k \geq 1$:

$$\frac{d^k}{dn^k} e_n = e_n \sum_{m=1}^{k} \frac{1}{(2\pi\beta_0)^m} \ln^m \left(\frac{\alpha_s(\mu_0^2)}{\alpha_s(\mu_1^2)} \right) \sum_{\substack{A=\{i_1,\dots,i_m\} \\ i_j \geq 1, \sum i_j = k}} c_A \prod_{i_j \in A} \frac{d^{i_j}}{dn^{i_j}} \gamma_n \,, \tag{C.6}$$

where

$$c_{\{m\}} = 1 \,, \qquad c_{\{1,\dots,1\}} = 1 \,, \qquad c_{\{2,1\}} = 3 \,, \qquad c_{\{2,1,1\}} = 6 \,, \qquad c_{\{2,2\}} = 3 \,, \qquad c_{\{3,1\}} = 4 \,, \dots \tag{C.7}$$

In spite of its apparent complicated form, eq. (C.6) is simply a Leibniz-rule kind of formula where all possible combinations of the derivatives of γ_n with correct order enter. The precise value of the coefficients c_A is a complicated combinatorial expression which does not matter in this discussion except for the observation that it is a positive number, since no negative signs are ever involved in the differentiation formula. Then, let us express the derivatives of eq. (C.4) for $k \geq 1$:

$$\frac{d^k}{dn^k} \gamma_n = -2\psi^{(k)}(n) - (-1)^k k! \left[\frac{1}{n^{k+1}} + \frac{1}{(n+1)^{k+1}} \right] \,, \tag{C.8}$$

where we have used eq. (B.14). We can rewrite that relation in an integral form inspired from eq. (B.16):

$$(-1)^k \frac{d^k}{dn^k} \gamma_n = 2 \int_0^1 dt \, t^{n-1} (-\ln(t))^k \left[\frac{1}{1-t} - \frac{1}{2} - \frac{t}{2} \right] \,, \tag{C.9}$$

and it is easy to see that the last term is positive for $t \in [0,1]$ (for instance by derivating it yet again) so that $(-1)^k \frac{d^k}{dn^k} \gamma_n \geq 0$ if $k \geq 1$ and $n \geq 0$.

Therefore, if $A = \{i_1, \dots, i_m\}$ is a set of strictly positive integers such that $\sum i_j = k$, then

$$(-1)^k \prod_{i_j \in A} \frac{d^{i_j}}{dn^{i_j}} \gamma_n \geq 0 \,. \tag{C.10}$$

Using this relation in conjunction with eq. (C.6) gives the expected property of eq. (C.5) provided that $\alpha_s(\mu_0^2)/\alpha_s(\mu_1^2) \geq 1$, or in other words if evolution is performed in the forward direction. The property is not verified for the backward evolution operator as we have already observed.

Acknowledgments

References

[1] X. Ji, *Parton Physics on a Euclidean Lattice*, *Phys. Rev. Lett.* **110** (2013) 262002 [arXiv:1305.1539] [INSPIRE].

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

[3] A.V. Radyushkin, *Quasi-parton distribution functions, momentum distributions, and pseudo-parton distribution functions*, *Phys. Rev. D* **96** (2017) 034025 [arXiv:1705.01488] [INSPIRE].

[4] A. Radyushkin, *One-loop evolution of parton pseudo-distribution functions on the lattice*, *Phys. Rev. D* **98** (2018) 014019 [arXiv:1801.02427] [INSPIRE].

[5] K. Orginos, A. Radyushkin, J. Karpie and S. Zafeiropoulos, *Lattice QCD exploration of parton pseudo-distribution functions*, *Phys. Rev. D* **96** (2017) 094503 [arXiv:1706.05373] [INSPIRE].

[6] K.-F. Liu and S.-J. Dong, *Origin of difference between anti-d and anti-u partons in the nucleon*, *Phys. Rev. Lett.* **72** (1994) 1790 [hep-ph/9306299] [INSPIRE].

[7] W. Detmold and C.J.D. Lin, *Deep-inelastic scattering and the operator product expansion in lattice QCD*, *Phys. Rev. D* **73** (2006) 014501 [hep-lat/0507007] [INSPIRE].

[8] V. Braun and D. Müller, *Exclusive processes in position space and the pion distribution amplitude*, *Eur. Phys. J. C* **55** (2008) 349 [arXiv:0709.1348] [INSPIRE].

[9] A.J. Chambers et al., *Nucleon Structure Functions from Operator Product Expansion on the Lattice*, *Phys. Rev. Lett.* **118** (2017) 242001 [arXiv:1703.01153] [INSPIRE].

[10] Y.-Q. Ma and J.-W. Qiu, *Exploring Partonic Structure of Hadrons Using ab initio Lattice QCD Calculations*, *Phys. Rev. Lett.* **120** (2018) 022003 [arXiv:1709.03018] [INSPIRE].

[11] G.S. Bali et al., *Two-current correlations in the pion on the lattice*, *JHEP* **12** (2018) 061 [arXiv:1807.03073] [INSPIRE].

[12] 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] [INSPIRE].

[13] A.V. Radyushkin, *Theory and applications of parton pseudodistributions*, *Int. J. Mod. Phys. A* **35** (2020) 2030002 [arXiv:1912.04244] [INSPIRE].

[14] M. Constantinou et al., *Parton distributions and lattice-QCD calculations: Toward 3D structure*, *Prog. Part. Nucl. Phys.* **121** (2021) 103908 [arXiv:2006.08636] [INSPIRE].

[15] T. Izubuchi et al., *Factorization Theorem Relating Euclidean and Light-Cone Parton Distributions*, *Phys. Rev. D* **98** (2018) 056004 [arXiv:1801.03917] [INSPIRE].

[16] L.-B. Chen, W. Wang and R. Zhu, *Next-to-Next-to-Leading Order Calculation of Quasiparton Distribution Functions*, *Phys. Rev. Lett.* **126** (2021) 072002 [arXiv:2006.14825] [INSPIRE].

[17] Z.-Y. Li, Y.-Q. Ma and J.-W. Qiu, *Extraction of Next-to-Next-to-Leading-Order Parton Distribution Functions from Lattice QCD Calculations*, *Phys. Rev. Lett.* **126** (2021) 072001 [arXiv:2006.12370] [INSPIRE].

[18] S. Moch, J.A.M. Vermaseren and A. Vogt, *The Three loop splitting functions in QCD: The Nonsinglet case*, *Nucl. Phys. B* **688** (2004) 101 [hep-ph/0403192] [INSPIRE].

[19] A. Vogt, S. Moch and J.A.M. Vermaseren, *The Three-loop splitting functions in QCD: The Singlet case*, *Nucl. Phys. B* **691** (2004) 129 [hep-ph/0404111] [INSPIRE].

[20] J. Blümlein, P. Marquard, C. Schneider and K. Schönwald, *The three-loop unpolarized and polarized non-singlet anomalous dimensions from off shell operator matrix elements*, *Nucl. Phys. B* **971** (2021) 115542 [arXiv:2107.06267] [INSPIRE].

[21] X. Ji, *Large-Momentum Effective Theory vs. Short-Distance Operator Expansion: Contrast and Complementarity*, arXiv:2209.09332 [INSPIRE].

[22] 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] [INSPIRE].

[23] G.M. de Divitiis et al., *Heavy quark masses in the continuum limit of quenched lattice QCD*, *Nucl. Phys. B* **675** (2003) 309 [hep-lat/0305018] [INSPIRE].

[24] M. Guagnelli, F. Palombi, R. Petronzio and N. Tantalo, *f(B) and two scales problems in lattice QCD*, *Phys. Lett. B* **546** (2002) 237 [hep-lat/0206023] [INSPIRE].

[25] M. Luscher, P. Weisz and U. Wolff, *A Numerical method to compute the running coupling in asymptotically free theories*, *Nucl. Phys. B* **359** (1991) 221 [INSPIRE].

[26] M. Luscher, R. Sommer, P. Weisz and U. Wolff, *A Precise determination of the running coupling in the SU(3) Yang-Mills theory*, *Nucl. Phys. B* **413** (1994) 481 [hep-lat/9309005] [INSPIRE].

[27] V. Braun, P. Gornicki and L. Mankiewicz, *Ioffe - time distributions instead of parton momentum distributions in description of deep inelastic scattering*, *Phys. Rev. D* **51** (1995) 6036 [hep-ph/9410318] [INSPIRE].

[28] J.-H. Zhang, J.-W. Chen and C. Monahan, *Parton distribution functions from reduced Ioffe-time distributions*, *Phys. Rev. D* **97** (2018) 074508 [arXiv:1801.03023] [INSPIRE].

[29] Y.L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics*, *Sov. Phys. JETP* **46** (1977) 641 [INSPIRE].

[30] V.N. Gribov and L.N. Lipatov, *Deep inelastic e p scattering in perturbation theory*, *Sov. J. Nucl. Phys.* **15** (1972) 438 [INSPIRE].

[31] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, *Nucl. Phys. B* **126** (1977) 298 [ɪɴSPIRE].

[32] C. Monahan and K. Orginos, *Quasi parton distributions and the gradient flow*, *JHEP* **03** (2017) 116 [arXiv:1612.01584] [ɪɴSPIRE].

[33] X. Gao et al., *Origin and resummation of threshold logarithms in the lattice QCD calculations of PDFs*, *Phys. Rev. D* **103** (2021) 094504 [arXiv:2102.01101] [ɪɴSPIRE].

[34] Y. Su et al., *Resumming quark's longitudinal momentum logarithms in LaMET expansion of lattice PDFs*, *Nucl. Phys. B* **991** (2023) 116201 [arXiv:2209.01236] [ɪɴSPIRE].

[35] S. Catani, *Physical anomalous dimensions at small x*, *Z. Phys. C* **75** (1997) 665 [hep-ph/9609263] [ɪɴSPIRE].

[36] J. Blumlein, V. Ravindran and W.L. van Neerven, *On the Drell-Levy-Yan relation to $O(\alpha_s^2)$*, *Nucl. Phys. B* **586** (2000) 349 [hep-ph/0004172] [ɪɴSPIRE].

[37] J. Blümlein and M. Saragnese, *The N3LO scheme-invariant QCD evolution of the non-singlet structure functions F2NS(x,Q2) and g1NS(x,Q2)*, *Phys. Lett. B* **820** (2021) 136589 [arXiv:2107.01293] [ɪɴSPIRE].

[38] V. Bertone et al., *Revisiting evolution equations for generalised parton distributions*, *Eur. Phys. J. C* **82** (2022) 888 [arXiv:2206.01412] [ɪɴSPIRE].

[39] A.V. Radyushkin, *Generalized parton distributions and pseudodistributions*, *Phys. Rev. D* **100** (2019) 116011 [arXiv:1909.08474] [ɪɴSPIRE].

[40] X. Gao et al., *Unpolarized proton PDF at NNLO from lattice QCD with physical quark masses*, *Phys. Rev. D* **107** (2023) 074509 [arXiv:2212.12569] [ɪɴSPIRE].

[41] J.C. Collins and J.-W. Qiu, *A New Derivation of the Altarelli-parisi Equations*, *Phys. Rev. D* **39** (1989) 1398 [ɪɴSPIRE].

[42] J. Collins, T.C. Rogers and N. Sato, *Positivity and renormalization of parton densities*, *Phys. Rev. D* **105** (2022) 076010 [arXiv:2111.01170] [ɪɴSPIRE].

[43] X.-D. Ji, *Off forward parton distributions*, *J. Phys. G* **24** (1998) 1181 [hep-ph/9807358] [ɪɴSPIRE].

[44] A.V. Radyushkin, *Symmetries and structure of skewed and double distributions*, *Phys. Lett. B* **449** (1999) 81 [hep-ph/9810466] [ɪɴSPIRE].

[45] A.V. Radyushkin, *Nonforward parton distributions*, *Phys. Rev. D* **56** (1997) 5524 [hep-ph/9704207] [ɪɴSPIRE].

[46] O.V. Teryaev, *Crossing and radon tomography for generalized parton distributions*, *Phys. Lett. B* **510** (2001) 125 [hep-ph/0102303] [ɪɴSPIRE].

[47] N. Chouika, C. Mezrag, H. Moutarde and J. Rodríguez-Quintero, *Covariant Extension of the GPD overlap representation at low Fock states*, *Eur. Phys. J. C* **77** (2017) 906 [arXiv:1711.05108] [ɪɴSPIRE].

[48] J. Collins, *Foundations of Perturbative QCD*, Cambridge University Press (2011) [DOI:10.1017/cbo9780511975592].

[49] A. Radyushkin, *Nonperturbative Evolution of Parton Quasi-Distributions*, *Phys. Lett. B* **767** (2017) 314 [arXiv:1612.05170] [ɪɴSPIRE].

[50] A.V. Radyushkin, *Structure of parton quasi-distributions and their moments*, *Phys. Lett. B* **788** (2019) 380 [arXiv:1807.07509] [ɪɴSPIRE].

[51] H. Dutrieux, M. Winn and V. Bertone, *Exclusive meets inclusive particle production at small Bjorken xB: How to relate exclusive measurements to PDFs based on evolution equations*, *Phys. Rev. D* **107** (2023) 114019 [arXiv:2302.07861] [INSPIRE].

[52] HADSTRUC collaboration, *Towards high-precision parton distributions from lattice QCD via distillation*, *JHEP* **11** (2021) 148 [arXiv:2107.05199] [INSPIRE].

[53] PARTICLE DATA GROUP collaboration, *Review of Particle Physics*, *PTEP* **2022** (2022) 083C01 [INSPIRE].

[54] APFEL collaboration, *APFEL: A PDF Evolution Library with QED corrections*, *Comput. Phys. Commun.* **185** (2014) 1647 [arXiv:1310.1394] [INSPIRE].

[55] V. Bertone, *APFEL++: A new PDF evolution library in C++*, *PoS* **DIS2017** (2018) 201 [arXiv:1708.00911] [INSPIRE].

[56] S. Bhattacharya et al., *Moments of proton GPDs from the OPE of nonlocal quark bilinears up to NNLO*, *Phys. Rev. D* **108** (2023) 014507 [arXiv:2305.11117] [INSPIRE].

[57] S. Catani, M.L. Mangano, P. Nason and L. Trentadue, *The Resummation of soft gluons in hadronic collisions*, *Nucl. Phys. B* **478** (1996) 273 [hep-ph/9604351] [INSPIRE].

[58] J.-H. Zhang et al., *First direct lattice-QCD calculation of the x-dependence of the pion parton distribution function*, *Phys. Rev. D* **100** (2019) 034505 [arXiv:1804.01483] [INSPIRE].

[59] H.-W. Lin et al., *Proton Isovector Helicity Distribution on the Lattice at Physical Pion Mass*, *Phys. Rev. Lett.* **121** (2018) 242003 [arXiv:1807.07431] [INSPIRE].

[60] C. Alexandrou et al., *Light-Cone Parton Distribution Functions from Lattice QCD*, *Phys. Rev. Lett.* **121** (2018) 112001 [arXiv:1803.02685] [INSPIRE].

[61] B. Joó et al., *Parton Distribution Functions from Ioffe time pseudo-distributions*, *JHEP* **12** (2019) 081 [arXiv:1908.09771] [INSPIRE].

[62] B. Joó et al., *Pion valence structure from Ioffe-time parton pseudodistribution functions*, *Phys. Rev. D* **100** (2019) 114512 [arXiv:1909.08517] [INSPIRE].

[63] H.-W. Lin et al., *Valence-Quark Distribution of the Kaon and Pion from Lattice QCD*, *Phys. Rev. D* **103** (2021) 014516 [arXiv:2003.14128] [INSPIRE].

[64] B. Joó et al., *Parton Distribution Functions from Ioffe Time Pseudodistributions from Lattice Calculations: Approaching the Physical Point*, *Phys. Rev. Lett.* **125** (2020) 232003 [arXiv:2004.01687] [INSPIRE].

[65] L. Del Debbio et al., *Neural-network analysis of Parton Distribution Functions from Ioffe-time pseudodistributions*, *JHEP* **02** (2021) 138 [arXiv:2010.03996] [INSPIRE].

[66] X. Gao et al., *Valence parton distribution of the pion from lattice QCD: Approaching the continuum limit*, *Phys. Rev. D* **102** (2020) 094513 [arXiv:2007.06590] [INSPIRE].

[67] C. Alexandrou et al., *Unpolarized and helicity generalized parton distributions of the proton within lattice QCD*, *Phys. Rev. Lett.* **125** (2020) 262001 [arXiv:2008.10573] [INSPIRE].

[68] HADSTRUC collaboration, *The continuum and leading twist limits of parton distribution functions in lattice QCD*, *JHEP* **11** (2021) 024 [arXiv:2105.13313] [INSPIRE].

[69] M. Bhat et al., *Continuum limit of parton distribution functions from the pseudodistribution approach on the lattice*, *Phys. Rev. D* **106** (2022) 054504 [arXiv:2205.07585] [INSPIRE].

[70] HADSTRUC collaboration, *Transversity parton distribution function of the nucleon using the pseudodistribution approach*, *Phys. Rev. D* **105** (2022) 034507 [arXiv:2111.01808] [INSPIRE].

[71] HADSTRUC collaboration, *Unpolarized gluon distribution in the nucleon from lattice quantum chromodynamics*, *Phys. Rev. D* **104** (2021) 094516 [arXiv:2107.08960] [INSPIRE].

[72] HADSTRUC collaboration, *Toward the determination of the gluon helicity distribution in the nucleon from lattice quantum chromodynamics*, *Phys. Rev. D* **106** (2022) 094511 [arXiv:2207.08733] [INSPIRE].

[73] HADSTRUC collaboration, *Non-singlet quark helicity PDFs of the nucleon from pseudo-distributions*, *JHEP* **03** (2023) 086 [arXiv:2211.04434] [INSPIRE].

[74] J. Delmar et al., *Gluon PDF of the proton using twisted mass fermions*, *Phys. Rev. D* **108** (2023) 094515 [arXiv:2310.01389] [INSPIRE].

[75] JEFFERSON LAB ANGULAR MOMENTUM and HADSTRUC collaborations, *Gluon helicity from global analysis of experimental data and lattice QCD Ioffe time distributions*, *Phys. Rev. D* **109** (2024) 036031 [arXiv:2310.18179] [INSPIRE].

[76] G.S. Bali, B. Lang, B.U. Musch and A. Schäfer, *Novel quark smearing for hadrons with high momenta in lattice QCD*, *Phys. Rev. D* **93** (2016) 094515 [arXiv:1602.05525] [INSPIRE].

[77] C. Egerer, R.G. Edwards, K. Orginos and D.G. Richards, *Distillation at High-Momentum*, *Phys. Rev. D* **103** (2021) 034502 [arXiv:2009.10691] [INSPIRE].

[78] M. Asakawa, T. Hatsuda and Y. Nakahara, *Maximum entropy analysis of the spectral functions in lattice QCD*, *Prog. Part. Nucl. Phys.* **46** (2001) 459 [hep-lat/0011040] [INSPIRE].

[79] Y. Burnier and A. Rothkopf, *Bayesian Approach to Spectral Function Reconstruction for Euclidean Quantum Field Theories*, *Phys. Rev. Lett.* **111** (2013) 182003 [arXiv:1307.6106] [INSPIRE].

[80] J. Karpie, K. Orginos, A. Rothkopf and S. Zafeiropoulos, *Reconstructing parton distribution functions from Ioffe time data: from Bayesian methods to Neural Networks*, *JHEP* **04** (2019) 057 [arXiv:1901.05408] [INSPIRE].

[81] XQCD collaboration, *Towards the nucleon hadronic tensor from lattice QCD*, *Phys. Rev. D* **101** (2020) 114503 [arXiv:1906.05312] [INSPIRE].

[82] W.I. Jay and E.T. Neil, *Bayesian model averaging for analysis of lattice field theory results*, *Phys. Rev. D* **103** (2021) 114502 [arXiv:2008.01069] [INSPIRE].

[83] N. Karthik and R.S. Sufian, *Bayesian-Wilson coefficients in lattice QCD computations of valence PDFs and GPDs*, *Phys. Rev. D* **104** (2021) 074506 [arXiv:2106.03875] [INSPIRE].

[84] M. Lüscher, *Properties and uses of the Wilson flow in lattice QCD*, *JHEP* **08** (2010) 071 [*Erratum ibid.* **03** (2014) 092] [arXiv:1006.4518] [INSPIRE].

[85] Z. Fan et al., *Isovector parton distribution functions of the proton on a superfine lattice*, *Phys. Rev. D* **102** (2020) 074504 [arXiv:2005.12015] [INSPIRE].

[86] X. Gao, W.-Y. Liu and Y. Zhao, *Parton Distributions from Boosted Fields in the Coulomb Gauge*, arXiv:2306.14960 [INSPIRE].

[87] D. Guazzini, R. Sommer and N. Tantalo, *Precision for B-meson matrix elements*, *JHEP* **01** (2008) 076 [arXiv:0710.2229] [INSPIRE].

[88] ETM collaboration, *A Proposal for B-physics on current lattices*, *JHEP* **04** (2010) 049 [arXiv:0909.3187] [INSPIRE].

[89] M. Guagnelli, K. Jansen and R. Petronzio, *Nonperturbative running of the average momentum of nonsinglet parton densities*, *Nucl. Phys. B* **542** (1999) 395 [hep-lat/9809009] [INSPIRE].

[90] M. Guagnelli, K. Jansen and R. Petronzio, *Universal continuum limit of nonperturbative lattice nonsinglet moment evolution*, *Phys. Lett. B* **457** (1999) 153 [`hep-lat/9901016`] [ɪɴSPIRE].

[91] J. Bringewatt et al., *Confronting lattice parton distributions with global QCD analysis*, *Phys. Rev. D* **103** (2021) 016003 [`arXiv:2010.00548`] [ɪɴSPIRE].

[92] JEFFERSON LAB ANGULAR MOMENTUM (JAM) and HADSTRUC collaborations, *Complementarity of experimental and lattice QCD data on pion parton distributions*, *Phys. Rev. D* **105** (2022) 114051 [`arXiv:2204.00543`] [ɪɴSPIRE].

[93] J. Bertrand, P. Bertrand and J. Ovarlez, *Transforms and Applications Handbook*, CRC Press (1995) [`DOI:10.1201/9781315218915`].

[94] F. Hausdorff, *Summationsmethoden und Momentfolgen. I*, *Math. Z.* **9** (1921) 74.

[95] D. Widder, *Laplace Transform (PMS-6)*, Princeton Mathematical Series, Princeton University Press (2015).

[96] D. Stanzione et al., *Frontera: The Evolution of Leadership Computing at the National Science Foundation*, in the proceedings of the *Practice and Experience in Advanced Research Computing (PEARC '20)*, Boston, U.S.A., July 10–14 (2022) [`DOI:10.1145/3311790.3396656`].