Direct calculation of Parton Distribution Functions (PDFs) on the lattice

Dissertation

zur Erlangung des akademischen Grades

DOCTOR RERUM NATURALIUM
(Dr. rer. nat.)

im Fach: Physik
Spezialisierung: Theoretische Physik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von M.Sc. Floriano Manigrasso

Präsident (komm.) der Humboldt-Universität zu Berlin: Prof. Dr. Peter Frensch

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät: Prof. Dr. Elmar Kulke

Gutachter*innen:

1. Prof. Dr. Agostino Patella
2. Prof. Dr. Karl Jansen
3. Prof. Dr. Haralambos Panagopoulos

Tag der mündlichen Prüfung: 26/04/2022



Department of Physics

Direct calculation of Parton Distribution Functions (PDFs) on the lattice

Floriano Manigrasso

A Dissertation Submitted to the University of Cyprus,

Humboldt University of Berlin and University of Rome "Tor Vergata"

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

February, 2022

I declare that I have completed the thesis independently using only the aids and tools specified. I have not applied for a doctor's degree in the doctoral subject elsewhere and do not hold a corresponding doctor's degree. I have taken due note of the Faculty of Mathematics and Natural Sciences PhD Regulations, published in the Official Gazette of Humboldt-Universität zu Berlin no. 42/2018 on 11/07/2018.

Floriano Manigrasso
June 15, 2022

Abstract

In this work, we address a number of crucial steps in order to evaluate the nucleon unpolarized helicity and trasversity parton distribution functions within the framework of lattice QCD. Discretization artifacts are investigated using a $N_f = 2 + 1 + 1$ gauge ensemble of Wilson twisted mass fermions simulated at a pion mass of approximately $M_\pi = 370$ MeV. The unpolarized and helicity parton distribution functions show a non-negligible dependence on the lattice spacing, with the continuum extrapolation producing a better agreement with phenomenology.

The direct computation of the Fourier transform using discrete lattice data may introduce artifacts and we, therefore, use a new data-driven method based on Gaussian process regression, the so-called Bayes-Gauss Fourier transform to overcome the limitations of the discrete Fourier transform. We find that this data-driven approach can drastically reduce the artifacts introduced by the discretization of the Fourier transform, however, the final effect on the light-cone PDFs is small. This finding suggests that the deviations seen between the current lattice QCD results and phenomenological determinations cannot be attributed solely to the discretization of the Fourier transform.

Furthermore, we present results of the first *ab initio* calculation of the individual up, down, and strange unpolarized, helicity and transversity parton distribution functions for the proton. The analysis is performed on an $N_f = 2+1+1$ twisted mass clover-improved fermion ensemble simulated at a pion mass of 260 MeV. We employ the hierarchical probing algorithm to evaluate the disconnected quark loops, allowing us to obtain non-zero results for the disconnected isoscalar contribution and the strange quark matrix elements. In particular, the results for the strange quark distributions, which are poorly constrained from global fit analysis, provide valuable information on the structure of the nucleon.

Περίληψη

Σε αυτήν την εργασία, εξετάζουμε έναν αριθμό κρίσιμων βημάτων προκειμένου να αξιολογήσουμε τη μη πολωμένη, ελικότητα και εγκάρσια κατανομή νουκλεονίων στο πλαίσιο της Κβαντικής Χρωμοδυναμικής (ΚΧΔ) στο πλέγμα. Ανεπιθύμητες συνεισφορές προεχρομένες από την διακριτοποίηση διερευνώνται χρησιμοποιώντας σύνολα $N_f = 2 + 1 + 1$ φερμιονίων συνεστραμμένης μάζας Ωιλσον που προσομοιώνονται σε μάζα πιονίου περίπου $M_\pi = 370$ MeV. Οι συναρτήσεις μη πολωμένης και ελικοειδούς κατανομών δείχνουν μια μη αμελητέα εξάρτηση από το a, με την προέκταση συνεχούς να παράγει μια καλύτερη συμφωνία με τη φαινομενολογία.

Ο άμεσος υπολογισμός του μετασχηματισμού Φουριερ χρησιμοποιώντας δεδομένα διακριτού πλέγματος μπορεί να εισάγει λάθη και, ως εκ τούτου, χρησιμοποιήσαμε μια νέα μέθοδο που βασίζεται στην παλινδρόμηση διαδικασίας Γαυσς, τον λεγόμενο μετασχηματισμό Βαψες-Γαυσς Φουριερ για να ξεπεράσουμε τους περιορισμούς του διακριτού μετασχηματισμού Φουριερ. Διαπιστώνουμε ότι αυτή η προσέγγιση μπορεί να μειώσει δραστικά τις επιπτώσεις που εισάγονται από τη διακριτοποίηση του μετασχηματισμού Φουριερ, ωστόσο, το τελικό αποτέλεσμα στις κατανομές κώνου φωτός είναι μικρό. Αυτό το εύρημα υποδηλώνει ότι οι αποκλίσεις που παρατηρούνται μεταξύ των αποτελεσμάτων της ΚΧΔ στο πλέγμα και των φαινομενολογικών προσδιορισμών δεν μπορούν να αποδοθούν αποκλειστικά στη διακριτοποίηση του μετασχηματισμού Φουριερ.

Επιπλέον, παρουσιάζουμε τα αποτελέσματα του πρώτου υπολογισμού των επιμέρους συναρτήσεων κατανομής των υπ, δοων και στρανγε κουώρκς για τη μη πολωμένη, ελικοειδή και εγκάρσια κατανομή για το πρωτόνιο. Η ανάλυση εκτελείται σε ένα σύνολο $N_f = 2+1+1$ φερμιονίων συνεστραμμένης μάζας σε μάζα πιονίου 260 MeV. Χρησιμοποιούμε ιεραρχική ανίχνευση για να αξιολογήσουμε τον αποσυνδεδεμένο βρόχο κουώρκ, επιτρέποντάς μας να λάβουμε μη μηδενικά αποτελέσματα για την αποσυνδεδεμένη ισοβαθμωτή συμβολή και τη συμβολή των παράξενων κουώρκ. Ειδικότερα, τα αποτελέσματα για τις παράξενες κατανομές κουώρκ, οι οποίες είναι ελάχιστα περιορισμένες από την ανάλυση σφαιρικής προσαρμογής, παρέχουν πολύτιμες πληροφορίες για τη δομή του νουκλεονίου.

Zusammenfassung

In dieser Arbeit befassen wir uns mit einer Reihe von entscheidenden Schritten, um die unpolarisierten Helizitäts- und Trasversitäts-Parton-Verteilungsfunktionen der Nukleonen im Rahmen der Gitter-QCD zu bewerten. Diskretisierungsartefakte werden unter Verwendung eines $N_f = 2 + 1 + 1$ Eichensembles von Fermionen mit verdrillter Wilson-Masse untersucht, die bei einer Pionenmasse von ungefähr $M_\pi = 370$ MeV simuliert werden. Die unpolarisierten und Helizitäts Partonverteilungsfunktionen weisen eine nicht vernachlässigbare Abhängigkeit vom Gitterabstand auf, und die Kontinuumsextrapolation ergibt eine bessere Übereinstimmung mit Phänomenologie.

Die direkte Berechnung der Fourier-Transformation mit diskreten Gitterdaten kann Artefakte verursachen. Daher arbeiten wir mit einer neuen datengesteuerten Methode, die auf Gauß-Prozess-Regression basiert, die sogenannte Bayes-Gauß-Fourier Transformation, um die Einschränkungen der diskreten Fourier-Transformation zu überwinden. Wir sind der Meinung, dass dieser datengesteuerte Ansatz die durch die Diskretisierung der Fourier-Transformation eingeführten Artefakte drastisch reduzieren kann, jedoch ist der endgültige Effekt auf die Lichtkegel-PDFs gering. Dieser Befund legt nahe, dass die beobachteten Abweichungen zwischen den aktuellen Gitter-QCD-Ergebnissen und phänomenologischen Bestimmungen nicht allein auf die Diskretisierung der Fourier-Transformation zurückzuführen sind.

Darüber hinaus präsentieren wir die Ergebnisse der ersten ab initio Berechnung der individuellen up, down und strange unpolarisierten, Helizitäts- und Transversitäts Partonverteilungsfunktionen für das Proton. Die Analyse wird an einem durch $N_f = 2{+}1{+}1$ verdrillten Kleeblatt-verbesserten Fermionen-Ensemble durchgeführt, das bei einer Pionenmasse von 260 MeV simuliert wird. Wir verwenden den hierarchischen Sondierungsalgorithmus, um die unzusammenhängenden Quarkschleifen auszuwerten. Dadurch erhalten wir Ergebnisse ungleich Null für den unbegundenen isoskalaren Beitrag und die strange Quark-Matrixelemente. Insbesondere die Ergebnisse für die Strange-Quark-Verteilungen, die durch die Global-Fit-Analyse nur wenig eingeschränkt werden können, liefern wertvolle Informationen über die Struktur des Nukleons.

Acknowledgments

The STIMULATE project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 765048.
The following supercomputing facilities are acknowledged for supporting the computations performed for this thesis. HPC resources of the Cyprus Institute (the Cyclone and Cyclamen machines) and of Temple University, supported in part by the National Science Foundation through major research instrumentation grant number 1625061 and by the US Army Research Laboratory under contract number W911NF-16-2-0189. Computations for this work were carried out in part on facilities of the USQCD Collaboration, which are funded by the Office of Science of the U.S. Department of Energy. This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725. The gauge configurations have been generated by the Extended Twisted Mass Collaboration on the KNL (A2) Partition of Marconi at CINECA, through the Prace project Pra13_3304 "SIMPHYS". This research used resources on the supercomputers JURECA [1] at Jülich Supercomputing Centre and Eagle at Poznań Supercomputing and Networking Center.Calculations were performed using the Grid library [2] and the DD-αAMG solver [3] with twisted mass support [4]. Data analysis and plotting were done using NumPy [5], SciPy [6], and Matplotlib [7].

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Parton distribution functions (PDFs) are key quantities for understanding the structure of hadrons. For example, they encode information on the distribution of spin and momentum among quarks and gluons. Since they are non-perturbative quantities they are difficult to calculate. They are mostly determined through global fits to experimental data of proton scattering. Being light cone matrix elements, they were not calculable within lattice QCD until a new theoretical approach was put forward. It relies on large momentum effective theory (LaMET) and a specific implementation is the so-called quasi-PDF approach [8]. Another similar approach is the so-called *pseudo-PDFs* [9, 10]. The quasi-PDF and pseudo-PDF methodologies allows one to compute parton distributions from first-principle lattice QCD simulations. Our study is mainly focused on the computation of nucleon PDFs employing the quasi-PDFs methodology, however, all the approaches related with the computation of PDFs within lattice QCD share similar challenging aspects, both computational and theoretical. The quasi-PDFs methodology allows a quantitative prediction of nucleon parton distributions [11–20], as well as interesting insights on the Δ^+ baryon [21] and the pion meson [22–26]. Moreover, recently, the formalism is extended to generalized parton distributions (GPDs) and exploratory studies are undertaken [27–32]. Furthermore, the formalism can be applied to transverse-momentum dependent parton distribution functions (TMDs) [33–44].

The work presented in this thesis, is based on Refs. [45–48] and has been carried out within the Extended Twisted Mass Collaboration (ETMC). It covers different aspects of the lattice computation of PDFs employing the quasi-PDf approach, and will be presented in Chaps. 5, 7 and 6. In Chap. 2, we give a brief theoretical introduction to the parton model, the definition of parton distribution functions and their moments and an overview about the state-of-the-art regarding the phenomenological determination of the nucleon

PDFs. In Chap. 3, we summarize the most important theoretical aspects of quantum chromodynamics (QCD) and introduce lattice QCD (LQCD), a non-perturbative approach to the solution of QCD in a discretized space-time. The computation of PDFs in the LQCD framework is not straightforward: different approaches are developed over the years to directly compute parton distributions, among which quasi-PDFs that is discussed in the last section of Chap. 3. In Chap. 4, we examine in depth the technical aspects regarding the computation of the matrix elements required to evaluate the quasi-PDF operator and the signal improvement techniques.

A crucial aspect in the evaluation of PDFs in lattice QCD is controlling the various sources of systematic effects. For instance, the discretization of space implies that we replace the continuous Fourier transform (FT) with a discrete FT (DFT). Such replacement introduces in certain circumstances artifacts to the resulting PDFs, constituting an additional source systematic effect. Therefore, we develop a new data-driven procedure, namely the Bayes-Gauss FT (BGFT), presented in Chap. 5 allowing us to overcome the limitations of the DFT for PDF reconstruction when employing the quasi-PDF approach.

Due to technical reasons, the lattice community focused in the last years on the evaluation of the isovector $u - d$ contributions to the nucleon PDFs. However, the development of new computational techniques, together with the continuous advancement in computing capabilities, makes possible the direct computation of the disconnected contribution to single-flavor PDFs within the lattice-QCD framework. In Chap. 6, we show the results of the first exploratory study within the lattice QCD framework of the flavor decomposition of unpolarized, helicity, and transversity proton parton distribution functions.

Finally, in Chap. 7, we present the results of the first analysis of the discretization artifacts affecting the quasi-PDF operator and arising from the presence of a finite lattice spacing in lattice-QCD simulations.

Chapter 2

Theoretical framework of parton distribution functions

Parton distribution functions (PDFs) can be defined in the framework of the parton model formulated in 1969by Feynman [49]. It gained immediately great success since it was successful in modeling the electron-proton deep inelastic scattering [50]. Since then, a huge effort has been made to measure PDFs with ever increasing precision and to design experiments allowing to measure the most elusive distributions. In what follows, we will provide an overview of the parton model (Sec. 2.1) and the definition of PDFs and their moments (Sec. 2.2) and finally, in Sec. 2.3, a summary of the state-of-the-art phenomenological determination of the unpolarized, helicity and transversity nucleon parton distribution functions.

2.1 Parton model

The main idea behind the *parton model* is that hadrons are composed of subparticles, i.e. the partons, possessing decoherent wave functions. This means that one can treat the partons inside the hadron essentially as free particles, carrying a fraction of the parton's momentum. The model was originally introduced by Feynman [49] to study the high-energy hadron collisions and its effectiveness was immediately apparent when computing the $e^- p^+$ deep inelastic cross section. Indeed, within the parton model assumption, the process reduces to an elastic e-parton scattering, whose diagram is shown in Fig.2.1.

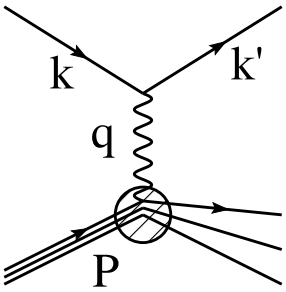


Figure 2.1: Diagram for elastic e-parton scattering. The circle represents the proton with momentum P^μ, and the three lines represent the partons within the hadron, only one of which participates in the interaction with the electron. q is the momentum transfer and $Q = \sqrt{-q^2}$.

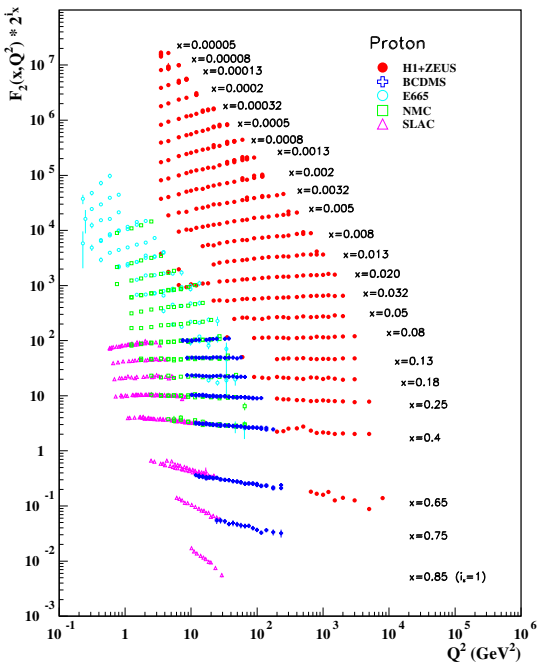


Figure 2.2: Proton structure function $F_2(x, Q^2)$ as a function of the momentum transfer Q^2. Figure taken from Ref. [51].

Assuming that the parton incoming momentum p_i^μ corresponds to a fraction ξ of the proton momentum, $p_i^\mu = \xi P^\mu$, then from momentum conservation, it follows that the Lorentz invariant

$$x = \frac{\xi Q^2}{2p_i \cdot q} = \xi \tag{2.1}$$

is a constant. It can be shown that, assuming the validity of the parton model, the cross section is approximately independent on Q^2 for fixed x. This result is known as **Bjorken scaling** and it is beautifully confirmed by the data shown in Fig. 2.2. The key ingredient of the parton model is the classical probability $f(\xi)d\xi$ to find a parton (quark or gluon) possessing a fraction ξ of the proton momentum. These $f_i(\xi)$ are known as **parton distribution functions (PDFs)**. From the parton model assumption, it follows that the cross section for the Deep Inelastic Scattering (DIS) process $e^- p_i \to e^- X$ reduces to

$$\sigma(e^- P^+ \to e^- X) = \sum_i \int_0^1 d\xi f_i(\xi) \hat{\sigma}\left(e^- p_i \to e^- X\right).$$

The previous formula coincides with the statement of the factorization theorem that has been proven only in a couple of examples. The physical justification is that the typical interaction time between partons is $\sim \Lambda_{QCD}^{-1}$, that is much slower than the time scales $\sim Q^{-1}$ that the photon probes.

2.2 Parton Distribution Functions

The parton distribution function of a hadron N with momentum $P = (P_0, 0, 0, P_3)$ is determined by the matrix element

$$h(\xi^- \lambda \cdot P) \equiv \frac{1}{2\lambda \cdot P} \langle N | \, \overline{\psi}(\xi^- \lambda) \not{\lambda} \Gamma W \left(\xi^- \lambda, 0\right) \psi(0) \, | N \rangle \,, \tag{2.2}$$

where ψ is the quark bilinear, $\lambda = (1, 0, 0, -1)/\sqrt{2}$ and $\xi^- = (\xi^0 - \xi^3)/\sqrt{2}$ is the minus-component of the nucleon position four-vector ξ expressed in the light-cone coordinate system. Due to the structure of λ, we can rewrite $\lambda \cdot P = P^+ = (P^0 + P^3)/\sqrt{2}$ and $\not{\lambda} = \lambda \cdot \gamma = \gamma^+$, with γ corresponding to the Dirac gamma matrices four-vector. The Wilson line $W(\zeta\lambda, \eta\lambda)$ reads

$$W\left(\zeta, \eta\right) \equiv \exp\left(-ig \int_{\eta}^{\zeta} d\rho \lambda \cdot A(\rho\lambda)\right), \tag{2.3}$$

with A being the gauge field and g the strong coupling constant. The matrix Γ acts in Dirac space, and it is defined as the product of Dirac matrices, depending on the kind of PDF:

- **unpolarized distribution** $q(x, \mu)$, $\Gamma = \mathbf{1}$, probability density to find a parton carrying a fraction x of the hadron momentum P;
- **helicity distribution** $\Delta q(x, \mu)$, $\Gamma = \gamma_5$, probability density to find a longitudinally polarized parton in a longitudinally polarized hadron carrying a fraction x of the hadron momentum P;
- **transversity distribution** $\delta q(x, \mu)$, $\Gamma = \gamma_j$, with the j-index purely spatial and orthogonal to the direction of the hadron momentum. This gives the probability density to find a transversely polarized parton in a transversely polarized hadron carrying a fraction x of the hadron momentum P.

The physical PDF of the hadron $|N\rangle$ is then defined as the Fourier transform [52]

$$\begin{aligned} q(x, P, \mu) &\equiv \int_{-\infty}^{\infty} \frac{d\xi^- \, \lambda \cdot P}{4\pi} \, e^{+ix\xi^- \lambda \cdot P} h(\xi^- \lambda \cdot P) \\ &= \int_{-\infty}^{\infty} \frac{d\xi^- P^+}{4\pi} \, e^{+ix\xi^- \cdot P^+} h(\xi^- P^+), \end{aligned} \tag{2.4}$$

where μ is the renormalization scale. The integration domain lies on the light-cone, defined as the locus specified by the equation $\xi^2 = \xi_0^2 - \vec{\xi} \cdot \vec{\xi} = 0$. A fundamental property of Eq. (2.4) is that it is invariant under Lorentz boosts along the z-direction. Indeed, by

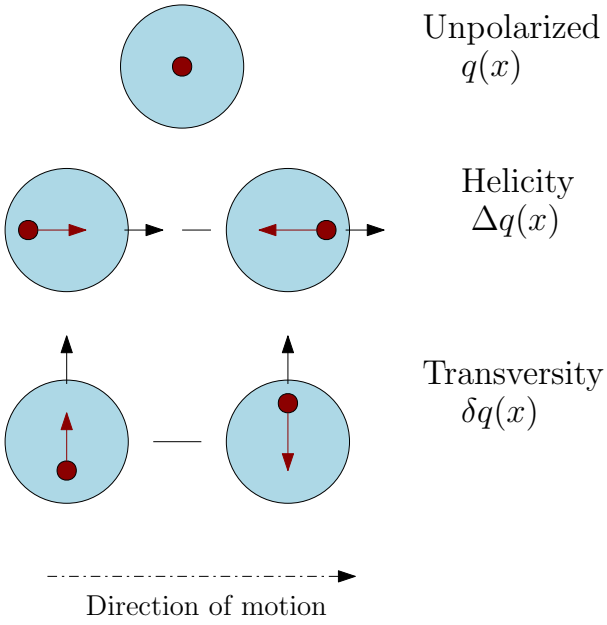


Figure 2.3: Schematic representation of the three PDFs. The red and black arrows correspond to the spin of the parton and the one of the hadron respectively.

applying such a transformation to the vector λ, we get

$$\lambda = \frac{1}{\sqrt{2}}(1,0,0,-1) \to \lambda' = \frac{\gamma(1+\beta)}{\sqrt{2}}(1,0,0,-1), \tag{2.5}$$

and changing integration variable $\xi'^{-} = \gamma(1+\beta)\xi^{-}$ the integral remains unchanged. In particular, this definition is valid in the rest frame where $P_3 = 0$.
Finally, a set of important identities, known as *crossing relations* [53], allows to relate the distribution of the antiquark $\bar{f}$ to the distribution of the quark f

$$\bar{q}_f(x) = -q_f(-x), \quad \Delta\bar{q}_f(x) = \Delta q_f(-x), \quad \delta\bar{q}_f(x) = -\delta q_f(-x)\,. \tag{2.6}$$

2.2.1 Moments of PDFs

Parton distribution functions have to satisfy some theoretical constraints, that can be formulated applying the *Mellin transform*. In particular, the n-th *moment* of the distribution $q_f(x)$ of the flavor f is defined as

$$\langle x^n \rangle_f \equiv \int_0^1 q_f(x) x^n. \tag{2.7}$$

In the previous definition we employed the unpolarized PDF q, however the quantity in Eq. (2.7) can be computed for any kind of distribution. As a consequence, the zeroth moment of the difference between quark-antiquark unpolarized distribution can be written as

$$\int_0^1 dx \left[q_f(x) - q_{\bar{f}}(x)\right] = \int_{-\infty}^{\infty} dx\, q_f(x), \tag{2.8}$$

where we use the crossing relations in Eq. (2.6) and we restrict the integration to the real axis, since the distributions have support in the interval $0 < x < 1$. Recalling the definition of the unpolarized PDF in Eq. (2.4), the zeroth moment in the previous equation can be written as

$$\begin{aligned}\int_0^1 dx \left[q_f(x) - q_{\bar{f}}(x)\right] &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} \frac{d\xi^- P^+}{4\pi} e^{+ix\xi^-\cdot P^+} h(\xi^- P^+) \\ &= \frac{1}{4\pi}\int_{-\infty}^{\infty} d\xi^- \delta(\xi^- P^+) h(\xi^- P^+) \\ &= \frac{1}{2P^+}\langle N|\bar{\psi}_f(0)\gamma^+\psi(0)_f|N\rangle = n_f,\end{aligned} \tag{2.9}$$

with n_f being the number of quarks with flavor f in the nucleon. Therefore, from the previous relation, it follows that, for the proton,

$$\begin{aligned}&\int_0^1 [q_u(x) - q_{\bar{u}}(x)] = 2, \\ &\int_0^1 [q_d(x) - q_{\bar{d}}(x)] = 1, \\ &\int_0^1 [q_f(x) - q_{\bar{f}}(x)] = 0, \text{ with } f = s, c, b, t.\end{aligned} \tag{2.10}$$

Moreover, the sum of the momentum fraction carried by the individual quarks, antiquarks and gluon has to be one . This statement corresponds with the so-called *momentum sum rule*, which can be expressed as

$$\int_0^1 \sum_f x q_f(x) dx = 1, \tag{2.11}$$

where the sum is performed over all the quark-antiquark flavors.
The zeroth moments of the helicity and transversity distributions provide information on the axial and tensor charges, g_A and g_T. These two quantities are defined as

$$\begin{aligned}g_A^f &= \int_0^1 \left[\Delta q_f + \Delta q_{\bar{f}}\right], \\ g_T^f &= \int_0^1 \left[\delta q_f - \delta q_{\bar{f}}\right].\end{aligned} \tag{2.12}$$

The axial charge can be measured with high precision from experiments involving beta decay, $n \to pe^-\nu_e$. For this reason, it can be seen as a benchmark quantity for lattice QCD calculations. The latest value provided by the PDG is $g_A = 1.2724(23)$ [51]. Conversely, the tensor charge cannot be easily measured experimentally. However, it constitutes a valuable input to reduce the uncertainty on the phenomenological determination of the transversity distribution [54], as we will discuss in Sec. 2.3.3.

2.3 Phenomenological determinations of PDFs

The parton distribution functions of the nucleons constitute key quantities for the ongoing experimental program of major facilities [55–57]. Unpolarized and helicity PDFs

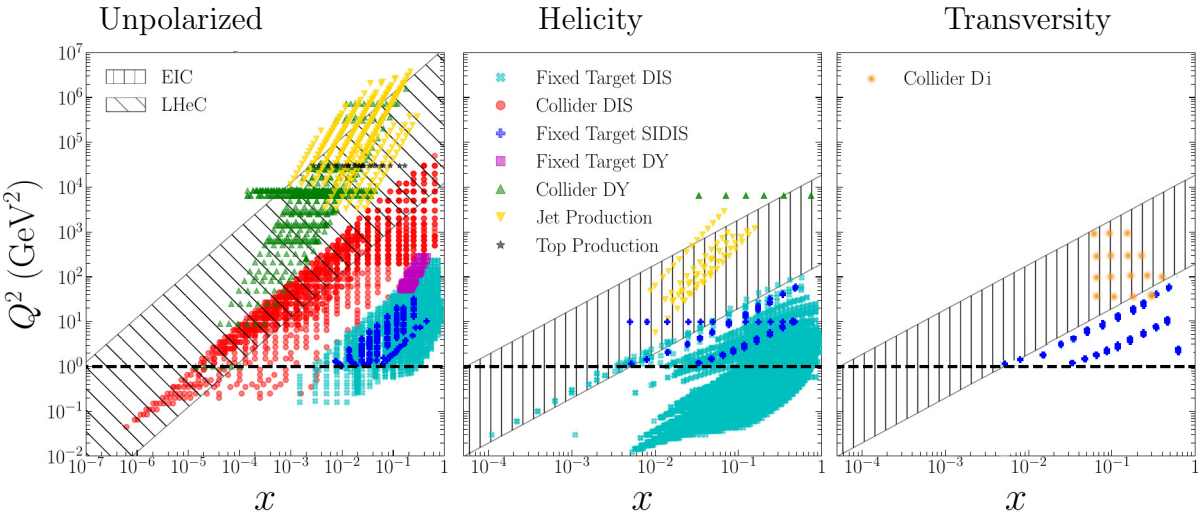


Figure 2.4: Kinematic coverage of the (x, Q^2)-plane of the cross-section data available to phenomenological analyses to reconstruct the three types of PDFs. The regions that will be accessible with the LHeC [58] and EIC future investigations [35, 59] are marked with the black lines. Sources: Refs. [60, 61].

can be accessed experimentally in inclusive and semi-inclusive deep-inelastic scattering (DIS and SIDIS), Drell-Yan (DY) and proton-proton (pp) scattering processes. On the other hand, transversity distributions can only be studied in processes with chiral-odd final states, such as di-hadron production in SIDIS or in pp collisions (Di) [61]. The phenomenologically determined PDFs are currently obtained through global analyses of the available experimental data [54, 62–68]. To reconstruct the x-dependence of PDFs through global fit determinations, it is necessary to sample the kinematic (x, Q^2)-plane, where Q is the characteristic energy scale of the data. While experiments provided thousands of dataset for the unpolarized PDFs, covering a large portion of the kinematic plane, for the helicity and transversity distributions, global analyses are restricted to hundreds and tens of datasets, respectively, as shown in Fig. 2.4. Due to the different amount of datasets available, the unpolarized PDFs can be estimated with significantly lower relative uncertainty compared to helicity and transversity. In what follows, we will provide a brief description of the state-of-the-art phenomenological determinations of the unpolarized, helicity and transversity distributions, describing the results that will be compared with the those of lattice QCD calculations. For a review of the state-of-the-art calculations of PDFs we refer to Ref. [69].

2.3.1 Unpolarized distributions

The phenomenological determination of the unpolarized PDFs is now in the precision era, since it can benefit from the large amount of experimental data from various ex-

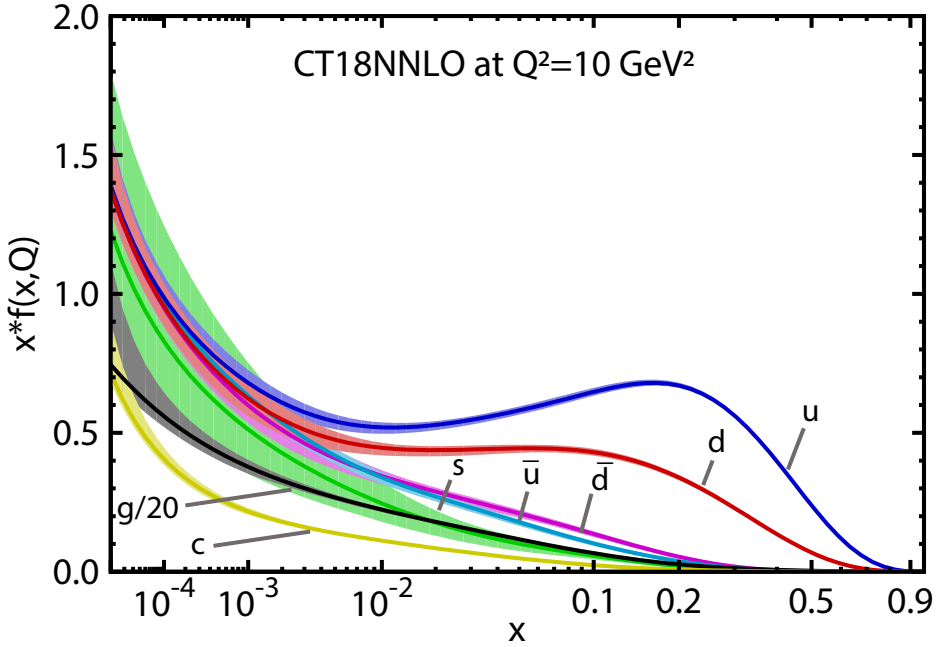


Figure 2.5: CT18 determination of the unpolarized PDFs at $\mu^2 = 10$ GeV^2. We show distribution $x \cdot q$, with $q = u, d, s, g, \bar{u}, \bar{d}, \bar{s}, \bar{g}$. The figure is from Ref. [71].

periments, such as Large Hadron Collider at CERN [70]. In Fig. 2.5 we show the CT18 phenomenological determination of the proton unpolarized PDFs. The distributions are obtained at the scale $\mu^2 = 10$ GeV^2. The gluon distribution is scaled by a factor of 1/20. In the small-x region, at $\mu^2 = 10\,\text{GeV}^2$, the gluon PDF dominates by far. This phenomenon is due to high energy collisions, which trigger processes initiated by gluons and sea quarks. While for $x > 0.1$ the sea quark and the gluon PDFs rapidly decay to zero, with the valence quark PDFs reaching a maximum for $x \approx 0.2$.

2.3.2 Helicity distributions

In what follows, we will indicate the quark (antiquark) helicity distribution as $\Delta q\,(\Delta\overline{q})$, while the distribution Δq^+ is defined as $\Delta q^+ \equiv (\Delta q + \Delta\overline{q})$. The phenomenological determination of the helicity PDFs with the currently available experimental data are carried out by the collaborations NNPDF$_{\text{POL}}$1.1 [66, 73], DSSV14 [68] and JAM17 [67]. In these analyses, the up and down contributions, $\Delta u(x)$, $\Delta d(x)$ are better constrained in the valence sector, with $\Delta u(x)$ being more precise. On the other hand, constraining $\Delta s(x)$ is not successful, as the kinematic regions of some of the data sets (e.g., the W-boson production data) are not sensitive to the strangeness [66]. The situation somewhat improves with the inclusion of kaon production SIDIS data, but it is still unsatisfactory, and influenced by theoretical assumptions, such as SU(3) symmetry. The helicity PDFs from the JAM17 parton set [67] at $Q^2 = 1\,\text{GeV}^2$ have been reported in Fig. 2.6. The authors used inclusive and semi-inclusive data, and find, for both sets of data, the strange polarization to be very small and consistent with zero. More details on the global analyses can be found in recent reports [61, 74].

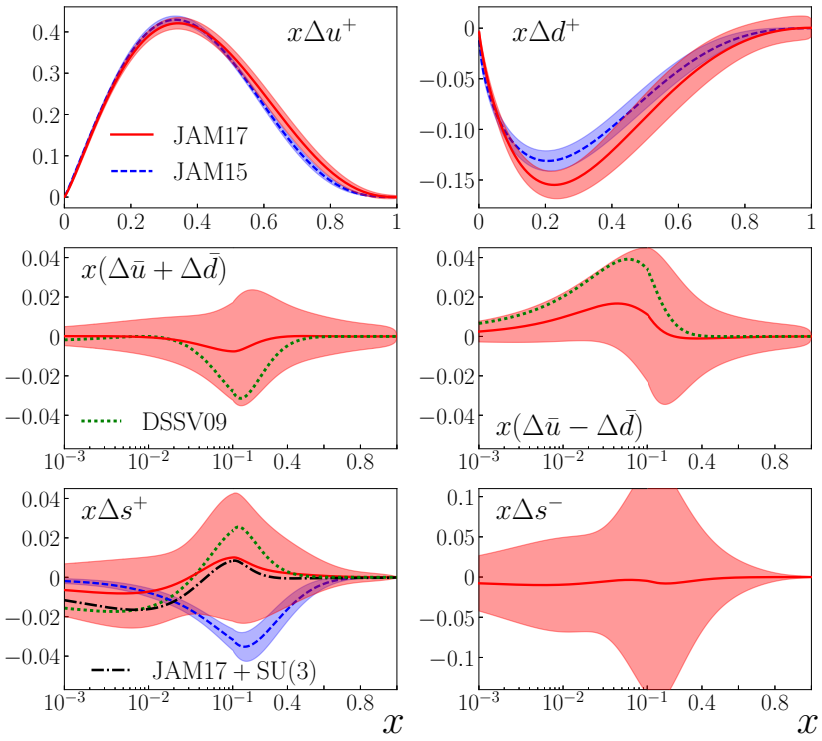


Figure 2.6: JAM17 [67] phenomenological determinations (red bands) of the polarized PDFs at an input scale $Q^2 = 1\,\mathrm{GeV}^2$. The distributions are compared with the JAM15 Δq^+ PDFs [72] (blue bands) and with the DSSV09 fit [65] for sea quark PDFs (green curves). Source: Ref. [67].

2.3.3 Transversity distributions

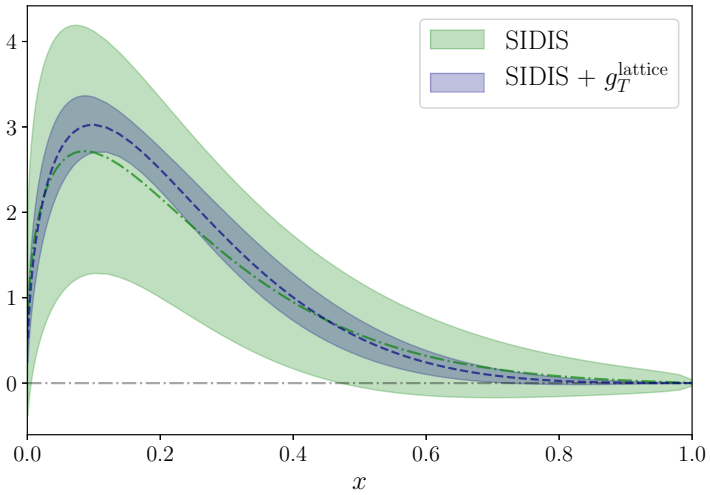


Figure 2.7: Quark transversity distribution of the isovector combination $u - d$ at $\mu^2 = 2\,\mathrm{GeV}^2$ obtained from SIDIS data (green) and SIDIS data constrained by the value of tensor charge g_T computed in lattice QCD [54] (blue).

Transversity $h_1(x)$ is the least known kind of PDF to date. As mentioned in Sec. 2.3, transversity distributions can be measured from processes having chiral-odd final states [75, 76], such as di-hadron production in SIDIS or in pp collisions (Di) [61]. The first determination of the up and down quark transversity distributions [77, 78] employed

SIDIS data available from Belle [79], COMPASS [80] and HERMES [81] experiments.

In Ref. [54], a Monte Carlo analysis of global data showed that the measurement of the x-dependence of PDFs is loosely constrained by SIDIS data (see Fig. 2.7). However, when including in the analysis the value of the tensor charge $g_T = \int_0^1 (h_1^u(x) - j_1^d(x))dx$ determined via lattice QCD calculations, the uncertainty on the $x-$dependence of the isovector $u - d$ transversity distribution is drastically reduced. For a thorough review of the status of the phenomenological determinations of the transversity distributions, including ongoing experiments aiming to reduce the experimental uncertainty, we refer to Refs. [69, 82].

Chapter 3

Lattice QCD formulation

3.1 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is a non-abelian gauge theory with gauge group $G = SU(N_C)$ and $N_C = 3$, the number of *colors*. The fundamental degrees of freedom of the theory are the six *quarks* of spin 1/2 described by Dirac 4-spinors and the *gluons*, the gauge bosons. Both quarks and gluons carry color degrees of freedom and they belong respectively to the fundamental and adjoint representation of the gauge group $SU(3)$. Defining ψ to be a vector composed by $N_f = 6$ quark fields we can write

$$\psi \equiv y = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_{N_f} \end{bmatrix}. \tag{3.1}$$

The six quarks *flavors* are called *up, down, strange, charm, top* and *bottom*. Each spinor $\psi^f_{\alpha,a}$ carries a Dirac index, here denoted by α and a color index, denoted by a, which can respectively take four and three different values. The degrees of freedom corresponding to the force carriers are represented by the $N_C^2 - 1 = 8$ gauge fields $A^a_\mu(x)$. A compact notation for the gauge field is given by

$$A_\mu = A^a_\mu T^a, \tag{3.2}$$

with T_a, $a = 1, \ldots, 8$ being the group generators $T_a = \lambda_a/2$, and λ_a the Gell-Mann matrices. The generators obey the relation $[T_a, T_b] = i f_{abc} T_c$, with f_{abc} the structure constants of the group, and normalization $\mathrm{Tr}\,[T_a T_b] = \delta_{ab}/2$.
The QCD lagrangian can be naturally split in two terms

$$\mathcal{L}^{QCD} = \mathcal{L}_F + \mathcal{L}_G, \tag{3.3}$$

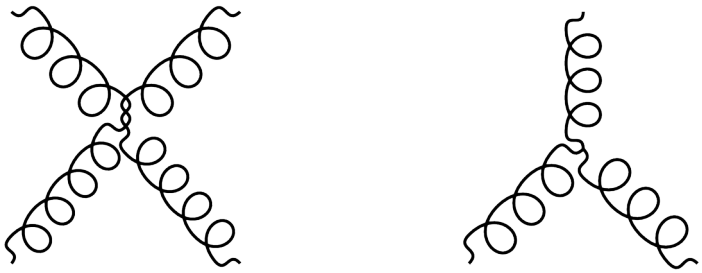


Figure 3.1: Gluon self interaction terms included in the Yang-Mills lagrangian: cubic vertex (right) and quartic vertex (left).

with $\mathcal{L}_F$ and $\mathcal{L}_G$ being, respectively, the fermion and gluon parts, respectively. The fermion part can be built following the so-called *minimal prescription*

$$\mathcal{L}_F = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi \mapsto \bar{\psi}(i\not{D} - m)\psi, \tag{3.4}$$

consisting of replacing in the Dirac lagrangian the standard derivative operator with the covariant derivative defined as

$$D_\mu \psi(x) = \partial_\mu \psi(x) + igA_\mu(x)\psi(x), \tag{3.5}$$

so that L_F includes a coupling between quarks and gluons.

The Yang-Mills lagrangian is given in terms of the field strength tensor $F_{\mu\nu}$, defined by

$$[D_\mu, D_\nu] = ig(\partial_\mu A_\nu(x) - \partial_\nu A_\mu(x) + ig[A_\mu, A_\nu] \equiv igF_{\mu\nu}. \tag{3.6}$$

$F_{\mu\nu}$ is a compact notation for $F_{\mu\nu}(x) = F^a_{\mu\nu}(x)T^a$. The gauge invariant term describing the dynamics of the $SU(3)$ gauge field is given by

$$\mathcal{L}_G = \frac{1}{2}\mathrm{Tr}\left[F_{\mu\nu}F^{\mu\nu}\right]. \tag{3.7}$$

The QCD lagrangian, thus, reads

$$\begin{aligned} \mathcal{L}^{QCD} &= \frac{1}{2}\mathrm{Tr}\left[F_{\mu\nu}F^{\mu\nu}\right] + \bar{\psi}(i\not{D} - m)\psi \\ &= -\frac{1}{4}F^a_{\mu\nu}F^{\mu\nu}_a + \sum_{f=1}^{N_f} \bar{\psi}_f(i\not{D} - m)\psi_f. \end{aligned} \tag{3.8}$$

The Yang-Mills lagrangian includes self interaction terms, such as the cubic and quartic vertices shown in Fig. 3.1. Under the gauge transformations $\mathcal{G} \in SU(3)$, the gluon field transforms as

$$A_\mu(x) \mapsto A'_\mu(x) = \mathcal{G}(x)A_\mu\mathcal{G}^\dagger(x) + \frac{i}{g}(\partial_\mu \mathcal{G}(x))\mathcal{G}^\dagger(x), \tag{3.9}$$

while the field strength tensor

$$F_{\mu\nu} \mapsto F'_{\mu\nu} = \frac{1}{ig}[D'_\mu, D'_\nu] = \mathcal{G}(x)F_{\mu\nu}(x)\mathcal{G}^{-1}(x). \tag{3.10}$$

With these gauge transformation of the gluon fields, the gauge invariance of L_G can be easily verified. The gauge transformation of the quark spinor is

$$\begin{aligned} \psi &\mapsto \psi' = \mathcal{G}(x)\psi \\ \bar{\psi} &\mapsto \bar{\psi}' = \bar{\psi}\mathcal{G}^\dagger(x) \end{aligned} \tag{3.11}$$

making L_F gauge invariant.

3.1.1 Strong coupling constant

The QCD lagrangian

$$\mathcal{L}^{QCD} = \frac{1}{2}\mathrm{Tr}\left[F_{\mu\nu}F^{\mu\nu}\right] + \bar{\psi}(i\not{D} - m)\psi \tag{3.12}$$

depends on the dimensionless coupling g. In what follows, we will review some of the main results provided by perturbation theory in dimensional regularization. The renormalized coupling g_R is related to the bare quantity g by

$$g = Z_g g_R, \tag{3.13}$$

where the renormalization constant reads

$$Z_g = 1 - \frac{g_R^2}{(4\pi)^2}\frac{N_C}{6}(11 - 2N_f)\frac{1}{\varepsilon} + \mathcal{O}(g_R^4), \tag{3.14}$$

with $\varepsilon = (4 - D)/2$ and D, and $D = 4$ is the number of space-time dimensions. The behavior of the renormalized coupling as a function of the energy scale can be studied from the Gell-Mann and Low beta function

$$\beta \equiv \mu\frac{d\tilde{g}_R(\mu)}{d\mu}, \tag{3.15}$$

with $\tilde{g}_R \equiv g_R\mu^\varepsilon$. Exploiting the independence of the bare coupling constant on the renormalization scale, one can easily show that

$$\beta = -\frac{\varepsilon Z_g g_R}{\frac{\partial}{\partial g_R}(Z_g g_R)}. \tag{3.16}$$

Replacing the explicit expression for Z_g reported in Eq. (3.14) we obtain

$$\beta(g_R) = -\frac{g_R^3}{(4\pi)^2}\left(\frac{11N_C - 2N_f}{3}\right) + \mathcal{O}(g_R^5) = -g_R^3\beta_0 + \mathcal{O}(g_R^5), \tag{3.17}$$

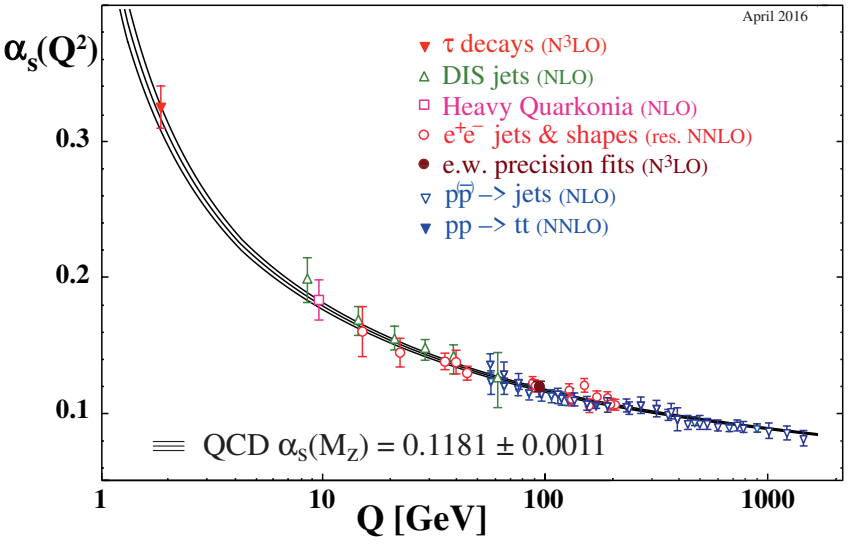


Figure 3.2: Experimental determination of the running of the strong coupling constant $\alpha_s = g_S^2/(4\pi^2)$. *Source: [51].*

with $\beta_0 = 21/(48\pi^2)$ for $N_C = 3$ and $N_f = 6$. This result provides very useful information on the dynamics of the theory in the perturbative regime, i.e. $g_R \ll 1$, when the expansion in powers of the renormalized coupling is justified. In particular, since $\beta_0 > 0$, Eq. (3.17) implies that

$$\frac{\partial g_R(\mu)}{\partial \mu} < 0, \tag{3.18}$$

which means that for $\mu \to \infty$ then $g_R \to 0$. This result represents one of the most important aspects of QCD, known as *asymptotic freedom*, namely at high energies the theory becomes free. This fundamental result has been independently obtained by David Gross and Frank Wilczek [83] and David Politzer [84], and earned them the Nobel prize in physics in 2004 [85].

The solution of the differential equation defining the beta function provides information about the behavior of the coupling constant as a function of the energy scale. In particular, from Eq. (3.17) we obtain that

$$\int_{\mu_1}^{\mu_2} \frac{d\mu}{\mu} = \int_{g_R(\mu_1)}^{g_R(\mu_2)} \frac{dg_R}{\beta(g_R)} \to \mu_2 = \mu_1 \exp\left(\int_{g_R(\mu_1)}^{g_R(\mu_2)} \frac{dg_R}{\beta(g_R)} \right). \tag{3.19}$$

Replacing the explicit expression for the β function at 1-loop in the previous equation we obtain

$$\mu e^{-\frac{1}{2\beta_0 g_R^2(\mu)}} \equiv \Lambda_{\rm QCD}. \tag{3.20}$$

The quantity $\Lambda_{\rm QCD}$ is a scheme-dependent constant with the dimension of an energy. The appearance of a physical dimensional parameter is sometimes referred to as *dimensional*

transmutation. Therefore, from the solution reported in Eq. (3.20) we are left with an equation describing the dependence of the coupling constant on μ

$$g_R^2(\mu) = \frac{1}{2\beta_0 \ln \mu/\Lambda_{\rm QCD})}, \tag{3.21}$$

whose validity is in the regime $\mu \gg \Lambda_{\rm QCD}$, which corresponds with the perturbative regime of QCD. Conversely, in case $\mu \sim \Lambda_{\rm QCD}$ then the coupling $g_R(\mu) \sim 1$ and the theory is strongly coupled. In this regime, QCD exhibits the *color confinement* phenomenon, stating that only color singlet states (i.e. $SU(3)_C$ singlets) are present in the hadronic spectrum. Although there is no analytic proof of this property, there is strong evidence of for this phenomenon both from experiments and from numerical simulations.
For further details about the renormalization of QCD refer to Ref. [86].

3.2 Expectation value of operators

We are interested in the expectation value of operators that are functionals of the field variables, both at finite temperature T and in the limit $T \to 0$. The main approach, which will be analyzed in what follows, consists of writing the theory in the path integral formalism [87]. In general, the expectation value $\langle O \rangle$ can be evaluated from the partition function

$$\mathcal{Z} = \mathrm{Tr}\left[e^{-\beta H}\right], \tag{3.22}$$

where H is the Hamiltonian operator of the system under investigation. The $\beta = 1/T$ factor is related with the temperature and the trace is a sum over all the possible configurations of the system. In the path integral formulation, an expectation value reads

$$\langle O \rangle_T \equiv \frac{1}{\mathcal{Z}} \mathrm{Tr}\left[O e^{-\beta H}\right]. \tag{3.23}$$

For a scalar field $\phi(x)$, the partition function reads

$$\mathcal{Z} = \mathcal{N} \int \mathcal{D}\phi(x) \exp\left(\frac{i}{\hbar}\int_0^t dt' L(\phi, \partial_\mu \phi)\right), \tag{3.24}$$

with $t = -i\beta\hbar$, $L(\phi, \partial_\mu\phi)$ the Lagrangian of the theory and $\mathcal{N}$ an appropriate normalization factor. Moreover, the bosonic field ψ satisfies periodical boundary condition in time $\phi(\mathbf{x}, t) = \phi(\mathbf{x}, 0)$. In the path integral formulation, the integration is carried out over the $\mathcal{D}\phi(x) \sim \prod_{\mathbf{x},t} d\phi(\mathbf{x}, t)$. To evaluate the integral using methods of statistical mechanics such as Monte Carlo techniques, it is necessary to reformulate the quantity in Eq. (3.23) as the expectation value of the operator O over a probability density. This can be achieved by performing a Wick rotation $\tau \equiv it$ and restricting the interval of

integration to $\tau \in [0, \hbar\beta]$, thus obtaining

$$\mathcal{Z} = \int \mathcal{D}\phi(x) \exp\left(-\frac{1}{\hbar}\int_0^{\hbar\beta} dt' L(\phi, \partial_\mu \phi)\right), \tag{3.25}$$

with $\phi(\mathbf{x}, \hbar\beta) = \phi(\mathbf{x}, 0)$. Consequently, the expectation value of Eq. (3.23) reads

$$\langle O \rangle_T = \int \mathcal{D}\phi P[\phi] O[\phi] \text{ with } P[\phi] = \frac{e^{-\mathcal{S}_E[\phi]/\hbar}}{\int \mathcal{D}\phi e^{-\mathcal{S}_E[\phi]/\hbar}}, \tag{3.26}$$

with $\mathcal{S}_E[\phi]$ being the Euclidean action of the theory. For a theory that includes fermionic fields, the expression for the expectation value at temperature T is equivalent to the one reported in Eq. (3.26), with the spinors replaced by Grassmann fields and satisfying antiperiodic boundary conditions in time i.e.

$$\begin{aligned} \overline{\psi}(\mathbf{x}, 0) &= -\overline{\psi}(\mathbf{x}, \hbar\beta) \\ \psi(\mathbf{x}, 0) &= -\psi(\mathbf{x}, \hbar\beta). \end{aligned} \tag{3.27}$$

Given a suitable discretization of the theory on a four-dimensional lattice with size $N_s^3 \times N_T$ with lattice spacing a, the expectation value $\langle O \rangle_T$ can be evaluated numerically as

$$\langle O \rangle_T = \frac{1}{\mathcal{Z}} \int \left(\prod_{i \in \text{sites}} d\phi\right) O[\phi_i] e^{-\mathcal{S}_E(\{\phi_i\})}, \tag{3.28}$$

where ϕ_i is the field ϕ evaluated on the i-th lattice site. The integral of Eq. (3.28) can be evaluated with Monte Carlo techniques.

3.3 QCD on the lattice

The discretization of the QCD lagrangian constitutes a challenging task since we need to preserve gauge invariance. The solution to that was provided by K. Wilson [88]. In what follows we will analyze the problems arising with the discretization of the fermionic sector of the QCD lagrangian, describing the most common discretization schemes, including the twisted mass action which is the one employed in our studies.
Let us consider the QCD lagrangian in Minkowski space

$$\mathcal{L}^{\text{QCD}} = -\frac{1}{2}\text{Tr}\left[F_{\mu\nu}F^{\mu\nu}\right] + \overline{\psi}(i\not{D} - m)\psi \tag{3.29}$$

where $\not{D} = \gamma^\mu(\partial_\mu + igA_\mu)$. As mentioned in Sec. 3.2, in order to evaluate the expectation value of relevant operators on a four-dimensional lattice, we need to perform a Wick rotation and express the QCD lagrangian in Euclidean space. This results in replacing

the position x and momentum k four-vectors with

$$\begin{aligned} x = (x_0, \mathbf{x}) \to \tilde{x}_0 = (\tilde{x}_0, \mathbf{x}) = (-ix_{E4}, \mathbf{x}_E) \\ k = (k_0, \mathbf{k}) \to \tilde{k}_0 = (\tilde{k}_0, \mathbf{k}) = (ik_{E4}, -\mathbf{k}_E), \end{aligned} \tag{3.30}$$

so that

$$\begin{aligned} \tilde{k} \cdot \tilde{x} = \tilde{k}_0 \tilde{x}_0 - \tilde{\mathbf{k}} \cdot \tilde{\mathbf{x}} = k_{E4} x_{E4} + \mathbf{k}_E \mathbf{x}_E \\ \tilde{x}^2 = \tilde{x}_0^2 - \mathbf{x}^2 = -x_E^2. \end{aligned} \tag{3.31}$$

In the aforementioned equations, all the four-vectors referred to with the subscript E are Euclidean vectors, e.g. the scalar product between $x_E = (x_{E4}, \mathbf{x}_E)$ and $k_E = (k_{E4}, \mathbf{k}_E)$ is the standard Euclidean scalar product. From Eq. (3.30), it follows that the derivative operator and the gauge field A_μ transform as

$$\begin{aligned} \partial_\mu = (\partial_0, \nabla) \to (\tilde{\partial}_0, \tilde{\nabla}) = (i\partial_{E4}, \nabla_E) \\ A_\mu(x) = (A_0(x), \mathbf{A}(x)) \to (A_0(\tilde{x}), \mathbf{A}(\tilde{x})) = (iA_{E4}(x_E), \mathbf{A}_E(x_E)), \end{aligned} \tag{3.32}$$

with the last equation following from the requirement to preserve the covariant transformation prescription. Given these definitions, the QCD lagrangian in Euclidean space reads

$$\mathcal{L}_E^{\text{QCD}} = \frac{1}{2}\text{Tr}\left[F_{E,\mu\nu}(x_E) F_{E,\mu\nu}(x_E)\right] + \overline{\psi}_E(x_E)\left(\gamma_{E,\mu} D_{E,\mu} + m\right)\psi_E(x_E), \tag{3.33}$$

with $\gamma_{E,4} = \gamma^0$, $\gamma_E^i = -i\gamma^i$. The Lorentz invariance $\text{SO}(3,1)$ of the QCD lagrangian in Minkowski space is now replaced with the invariance under $\text{O}(4)$ transformations of $\mathcal{L}_E^{\text{QCD}}$. The discretization of the Yang Mills action will be addressed in Sec. 3.3.6. In what follows we describe the discretization of the fermion fields.

3.3.1 Naivë discretization of the fermions

From now on, unless otherwise stated, all the vectors/tensors are in Euclidean space and the subscript E will be dropped. For convenience we rewrite the fermionic action $\mathcal{S}_F$ in terms of dimensionless lattice variables as

$$\begin{aligned} M \to \frac{1}{a}\hat{M} \\ \psi_\alpha(x) \to \frac{1}{a^{(3/2)}}\hat{\psi}_\alpha(n) \\ \bar{\psi}_\alpha(x) \to \frac{1}{a^{(3/2)}}\hat{\bar{\psi}}_\alpha(n) \\ \partial_\mu \psi_\alpha(x) \to \frac{1}{a^{(5/2)}}\hat{\partial}_\mu \hat{\psi}_\alpha(n), \end{aligned} \tag{3.34}$$

where we replace the continuous variables x with discrete ones, on sites na, $n \in [0, V]$, with a the lattice spacing and V the lattice volume. The variables with the hat symbol, such as $\hat{M}$, are now dimensionless and defined on a discrete set of lattice sites. The derivative operator acting on $\hat{\psi}$ is given by

$$\hat{\partial}\hat{\psi}_\alpha(n) \equiv \frac{1}{2}\left(\hat{\psi}_\alpha(n+\hat{\mu} - \hat{\psi}_\alpha(n-\hat{\mu}\right), \tag{3.35}$$

and it is defined to be symmetric. The discretized fermion action for a free theory reads

$$\begin{aligned}\hat{\mathcal{S}}_F &= \sum_{n,m}\left[\overline{\hat{\psi}}_\alpha(n)\left(\sum_\mu \gamma_\mu^{\alpha\beta}\frac{1}{2}(\delta_{m,n+\hat{\mu}} - \delta_{m,n-\hat{\mu}}) + \delta_{\alpha\beta}\delta m, n\hat{M}\right)\hat{\psi}_\beta(m)\right] \\ &= \sum_{m,n}\overline{\hat{\psi}}_\alpha(n)K^{\alpha\beta}_{mn}\hat{\psi}_\beta(m).\end{aligned} \tag{3.36}$$

The two-point function can thus be computed from

$$\langle\hat{\psi}_\alpha(n)\overline{\hat{\psi}}_\beta(m)\rangle = \frac{\int \mathcal{D}\overline{\hat{\psi}}\mathcal{D}\overline{\psi}\hat{\psi}_\alpha(n)\overline{\hat{\psi}}_\beta(m)e^{-\hat{\mathcal{S}}_F[\hat{\psi},\overline{\hat{\psi}}]}}{\int \mathcal{D}\overline{\hat{\psi}}\mathcal{D}\overline{\psi}e^{-\hat{\mathcal{S}}_F[\hat{\psi},\overline{\hat{\psi}}]}} = \left(K^{\alpha\beta}_{mn}\right)^{-1}. \tag{3.37}$$

This result, together with the computation of the inverse of the matrix $K^{\alpha\beta}_{mn}$ can be found in Ref. [89]. In particular, in the continuum limit the two point function reads

$$\langle\psi_\alpha(x)\bar{\psi}_\beta(y)\rangle = \lim_{a\to 0}\int_{-\pi/4}^{\pi/4}\frac{d^4k}{(2\pi)^4}\frac{\left[M - i\sum_\mu \gamma^\mu \frac{1}{a}\sin\left(k_\mu a\right)\right]_{\alpha\beta}}{\left[M^2 + \sum_\mu \frac{1}{a^2}\sin^2\left(k_\mu a\right)\right]}e^{ik(x-y)}, \tag{3.38}$$

which has to be equivalent to propagator of a free fermion field, given by

$$\langle\psi_\alpha(x)\bar{\psi}_\beta(y)\rangle = \int\frac{d^4k}{(2\pi)^4}\frac{M - i\not{k}}{M^2 + \not{k}^2}e^{ik(x-y)}. \tag{3.39}$$

This is indeed the case for $\frac{1}{a}\sin\left(k_\mu a\right) \sim k_\mu$. However, in the continuum limit one can also have $k_\mu = \pi/a$. As a consequence, in the continuum limit, Eq. (3.38) takes contribution from sixteen (i.e. 2^d with $d = 4$) fermion-like excitations, with just one of them corresponding to the continuum analog. The remaining fifteen contributions are lattice artifacts, having no physical meaning. This known as the *fermion doubling* problem, and the solution to it will be discussed in the next section.

3.3.2 Nielsen and Ninomiya theorem

The fermion doubling problem can be easily solved by taking the (asymmetrical) forward discretization of the derivative in Eq. (3.34). However, this would make the discretized action $\hat{\mathcal{S}}_F$ non-hermitian and, as a consequence, the corresponding Hamiltonian would be non-unitary. Different and more complicated solutions to the fermion doubling have

been proposed since the early days of lattice QCD, however, any one of them introduces some drawbacks in the formulation of the theory on the lattice. This is the consequence of the famous no-go theorem formulated by Nielsen and Ninomiya in 1981 [90]. It states that it is not possible to formulate a lattice fermionic action that concurrently satisfies the following four requirements:

1. is local;
2. lacks of doublers;
3. satisfies chiral symmetry in the massless limit;
4. in the limit $a \to 0$ matches the continuum theory.

In what follows we will analyze the Wilson discretization scheme and the twisted mass fermion formulation, which has been employed in this study. There are several fermion discretization procedures, which will be not discussed in this thesis. Ther are analyzed in detail in Ref. [89].

3.3.3 Wilson Fermions

The Wilson formulation introduces an additional term to the discretized fermion QCD action in Eq. (3.36), which vanishes in the continuum limit, $a \to 0$. The action, introduced in Ref. [88], reads

$$\mathcal{S}_F^W = \hat{\mathcal{S}}_F - \underbrace{\frac{r}{2}\sum_\mu \overline{\hat{\psi}}(n)\hat{\Box}\hat{\psi}(n)}_{=\Delta\hat{\mathcal{S}}^W}, \tag{3.40}$$

where $\hat{\Box}$ is the discretized d'Alembert operator $\Box = \partial_\mu\partial^\mu$, whose action on the fermion field can we written as

$$\hat{\Box}\hat{\psi}(n) = \sum_\mu \left[\hat{\psi}(n+\hat{\mu}) + \hat{\psi}(n-\hat{\mu}) - 2\hat{\psi}(n)\right] \tag{3.41}$$

. As a result, the term introduced by Wilson to get rid of the doublers is

$$\Delta\hat{\mathcal{S}}^W = -\frac{r}{2}\sum_{\mu,m,n} \overline{\hat{\psi}}(n)\left[\delta_{m,n+\mu} + \delta_{m,n-\mu} - 2\delta n, m\right]\hat{\psi}(m), \tag{3.42}$$

and the quadratic form K describing the fermion action now reads

$$K^W_{\alpha\beta}(n,m) = (\hat{M}+4r)\delta_{m,n} - \frac{1}{2}\sum_\mu \left[(r-\gamma_\mu)_{\alpha\beta}\delta_{m,n+\hat{\mu}} + (r+\gamma_\mu)_{\alpha\beta}\delta_{m,n-\hat{\mu}}\right]. \tag{3.43}$$

The inverse of K in the continuum limit takes contribution from the physical excitation only, suppressing the contribution coming from the doublers to the two-point function. The proof of this result, together with a more detailed description of the Wilson fermions

can be found in Ref. [89]. However, as a consequence of the Nielsen-Ninomiya theorem, the lattice formulation introduced by Wilson does not preserve chiral symmetry in the massless limit at finite lattice spacing.

Properties of the Wilson fermions

As mentioned in the previous section, the Wilson discretization does not preserve the chiral symmetry at finite lattice spacing. Besides, it has many interesting properties that is worth mentioning before moving to the description of the discretization scheme employed in this work. In particular, the complete Wilson-Dirac operator D_W^M (i.e. including the mass term), satisfies γ^5-Hermiticity, which reads

$$D_W^M(n,m)^\dagger = \gamma^5 D_W^M(n,m)\gamma^5. \tag{3.44}$$

Moreover, the Wilson action is invariant under parity, charge and time reversal transformations. Furthermore, the operator is not protected by zero modes, which can make its inverse ill-defined. The configurations corresponding to the zero modes are called *exceptional configurations*. This problem will be addressed by the twisted mass fermion formulation, introduced in the following section.

3.3.4 Twisted mass fermion formulation

In this section the twisted mass action [91], which is a variant of the Wilson formulation, will be introduced. In particular, we will discuss the discretization of the theory in the presence of two light degenerate quarks, i.e. $N_f = 2$. The generalization to $N_f = 2{+}1{+}1$, which can be found in Refs. [92, 93], will be discussed in the following chapters.

Let us consider a doublet field in flavor space χ representing two degenerate light quarks with bare mass m. Then the twisted mass action reads

$$\mathcal{S}_{\rm tm} = a^4 \sum_{n,m} \bar{\chi}(n) \left[D_W(n,m) + m\mathbf{1}_f + i\mu\gamma^5\tau^3 \right] \chi(m), \tag{3.45}$$

where $\mathbf{1}_f$ and $\tau^3 = \mathrm{diag}(1,-1)$ are, respectively, the identity and the third Pauli matrix, both acting in flavor space. D_W is the massless Wilson-Dirac operator with $r = 1$ and μ is the *twisted mass parameter*. It plays a key role in protecting the discretized Dirac operator against exceptional configurations linked with zero modes [94]. This is evident from the result of the computation of the determinant of the twisted mass operator

$$\begin{aligned}
\det\left[D_W + m\mathbf{1}_f + i\mu\gamma^5\tau^3\right] &= \det\left[D_W + m + i\mu\gamma^5\right] \det\left[D_W + m - i\mu\gamma^5\right] \\
&= \det\left[D_W + m + i\mu\gamma^5\right] \det\left[\gamma^5(D_W + m - i\mu\gamma^5)\gamma^5\right] \\
&= \det\left[(D_W + m + i\mu\gamma^5)(D_W^\dagger + m - i\mu\gamma^5)\right] \\
&= \det\left[(D_W + m)(D_W^\dagger + m) + \mu^2\right],
\end{aligned} \tag{3.46}$$

which is always larger than zero for $\mu \neq 0$ and where we use the γ^5-hermiticity property of the Wilson operator (see Eq. (3.44)). Therefore, the twisted mass term acts as an infrared regulator which, given an appropriate transformation, can be absorbed in the mass term. Let us consider the following chiral rotation

$$\begin{aligned} \psi &= e^{i\alpha\gamma_5\tau^3/2}\chi = \left(\cos(\alpha/2) + \gamma_5\tau^3 \sin(\alpha/2)\right)\chi \\ \bar{\psi} &= \bar{\chi}e^{i\alpha\gamma_5\tau^3/2} = \bar{\chi}\left(\cos(\alpha/2) + \gamma_5\tau^3 \sin(\alpha/2)\right), \end{aligned} \tag{3.47}$$

where α is the *twist angle*, and *psi* is the transformed quark doublet expressed in the so-called *physical basis*. Only the Wilson term of the Dirac-Wilson operator is affected by such a transformation, while the mass term reads

$$\bar{\chi}(m + i\mu\gamma^5\tau^3) \mapsto \bar{\psi}(e^{-i\alpha\gamma^5\tau^3}m + i\mu\gamma^5\tau^3)\psi, \tag{3.48}$$

with the quantity

$$Me^{i\alpha\gamma^5\tau^3}, \quad m = M\cos(\alpha), \ \mu = M\sin(\alpha) \tag{3.49}$$

referred to as *polar mass*. Therefore, in the physical basis, the twisted mass action reads

$$\mathcal{S}_{\rm tm} = a^4 \sum_{n,m} \bar{\psi}(n)\left[D_W^{\rm tm} + M\mathbf{1}_f\right]\psi(m), \tag{3.50}$$

with $M = \sqrt{m^2 + \mu^2}$ and $D_W^{\rm tm}$ the Dirac-Wilson operator obtained from the transformation of Eq. (3.47). In the continuum limit, the twisted mass action reproduces standard QCD, provided that the twist angle in Eq. (3.47) satisfies Eq. (3.49).

3.3.5 Properties of the twisted mass action

Because of the new term introduced in the twisted mass fermion action, the Dirac operator is not γ_5 hermitian. However, recalling the definition of the twisted mass fermion action of (3.45), it is easy to verify that the property that holds becomes:

$$\gamma_5\mathcal{D}_u^{\dagger\,\rm tm}\gamma_5 = \gamma_5\mathcal{D}_W^{\dagger}\gamma_5 - i\mu\gamma_5 = \mathcal{D}_W^{\dagger} - i\mu\gamma_5 = \mathcal{D}_d^{\rm tm}. \tag{3.51}$$

Since $(\mathcal{D}^{-1})^\dagger = \left(\mathcal{D}^\dagger\right)^{(-1)}$, we obtain a useful identity valid for twisted mass propagators

$$\mathcal{G}_u^{\rm tm} = \gamma_5\mathcal{G}_d^{\dagger\,\rm tm}\gamma_5 \tag{3.52}$$

3.3.6 Discretization of the gauge action

The QCD lagrangian in Minkowski space reported in Eq. (3.29) is written in terms of the gauge field $A_\mu(x)$ and the fermionic field $\psi(x)$. Under the transformation $\mathcal{G}(x) \in SU(3)_C$,

the fermion field transforms as

$$\begin{aligned}\psi(x) &\to \psi'(x) = \mathcal{G}(x)\psi(x)\\ \overline{\psi}(x) &\to \overline{\psi}'(x) = \overline{\psi}(x)\mathcal{G}^{-1}(x),\end{aligned} \tag{3.53}$$

while the gauge field transforms as

$$A_\mu(x) \to \mathcal{G}(x)A_\mu(x)\mathcal{G}^{-1}(x) - \frac{i}{g}\mathcal{G}(x)\left(\partial_\mu \mathcal{G}^{-1}(x)\right). \tag{3.54}$$

The lattice QCD Wilson fermion action, as well as the twisted mass fermion formulation, include bilinear terms like $\bar{\psi}(n)\psi(n+\hat{\mu}$. We want to show now how to make these quantities gauge invariant for an interacting theory and how to define the gauge action on the lattice. In the continuum limit, under gauge transformations the bilinear $\bar{\psi}(x)\psi(y)$ transforms as

$$\bar{\psi}(x)\psi(y) \mapsto \bar{\psi}(x)\mathcal{G}^{-1}(x)\mathcal{G}(y)\psi(y). \tag{3.55}$$

At this stage, we introduce the parallel transport

$$U(x,y) = \mathcal{P}e^{ig\int_x^y dz_\mu A^\mu(z)}, \tag{3.56}$$

where $\mathcal{P}$ is the path ordering operator, $A^\mu(z)$ the gauge field and g the strong coupling constant. Under gauge transformations the parallel transport transforms as

$$U(x,y) \mapsto \mathcal{G}(x)U(x,y)\mathcal{G}^{-1}(y), \tag{3.57}$$

and, as a result, the following quantity is gauge invariant

$$\bar{psi}(x)U(x,y)\psi(y) = \bar{\psi}(x)\mathcal{P}e^{ig\int_x^y dz_\mu A^\mu(z)}\psi(y). \tag{3.58}$$

On the lattice, we define a link variable in the μ -direction that connects site n with site $n + a\hat{\mu}$

$$U_\mu(n) \approx \mathbf{1} + iagA_\mu. \tag{3.59}$$

We thus have $U_\mu^\dagger(n) = U_{-\mu}(n+\hat{\mu})$, i.e. the dagger of the parallel transport going from n to $n + a\hat{\mu}$ yields the parallel transport going from $n + a\hat{\mu}$ to n. Therefore, to make the derivative term of the discretized Dirac operator gauge invariant, we perform the following replacement

$$\begin{aligned}\bar{\hat{\psi}}(n)\gamma_\mu\hat{\psi}(n+\hat{\mu}) &\mapsto \bar{\hat{\psi}}(n)\gamma_\mu U_\mu(n)\hat{\psi}(n+\hat{\mu});\\ \bar{\hat{\psi}}(n+\hat{\mu})\gamma_\mu\hat{\psi}(n) &\mapsto \bar{\hat{\psi}}(n+\hat{\mu})\gamma_\mu U_\mu^\dagger(n)\hat{\psi}(n).\end{aligned} \tag{3.60}$$

The Yang-Mills action on the lattice is written in terms of elementary parallel transports. The simplest non-trivial gauge invariant term that can be built starting from these build-

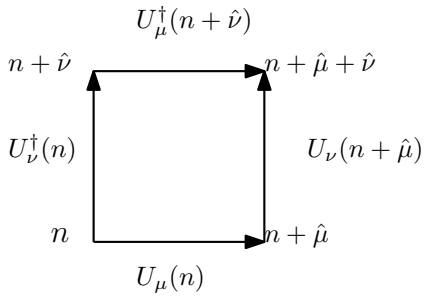


Figure 3.3: Representation of the elementary plaquette*, i.e. the shortest non-trivial closed path composed by parallel transports on the lattice.*

ing blocks is the trace of the elementary *plaquette* $U_{\mu\nu}$ among neighboring sites, defined as

$$U_{\mu\nu} \equiv U_\mu(n)U_\nu(n+\hat{\mu})U_\mu^\dagger(n+\hat{\nu})U_\nu^\dagger(n), \tag{3.61}$$

and represented in Fig. 3.3. The Wilson action [88], which forms the basis of the most widely used discretizations of the Yang-Mills theory, is given by a sum of the trace of plaquettes

$$\mathcal{S}^W \equiv \sum_{n,\mu<\nu} \beta \left[1 - \frac{1}{2N_C}\mathrm{Tr}\left[U_{\mu\nu}(n) + U_{\mu\nu}^\dagger(n)\right]\right], \tag{3.62}$$

where $N_C = 3$ is the number of colors and β is the coupling constant. In what follows, we will show that in the continuum limit, $a \to 0$, the Wilson action reduces to the standard Yang-Mills action. Recalling the Baker-Campbell-Hausdorff formula

$$e^A e^B = e^{A+B+\frac{1}{2}[A,B]\ldots}, \tag{3.63}$$

the plaquette can be written as

$$\begin{aligned} U_{\mu\nu}(n) = \exp(&igaA_\mu(n) + igaA_\mu(n+\hat{\mu}) - \frac{g^2a^2}{2}[A_\mu(n), A_\nu(n+\hat{\mu})] \\ &- igaA_\mu(n+\hat{\nu}) - iagA_\nu(n) - \frac{g^2a^2}{2}[A_\mu(n+\hat{\nu}), A_\nu(n)] \\ &+ \frac{g^2a^2}{2}[A_\nu(n+\hat{\mu}), A_\mu(n+\hat{\nu})] + \frac{g^2a^2}{2}[A_\mu(n), A_\nu(n)] \\ &+ \frac{g^2a^2}{2}[A_\mu(n), A_\mu(n+\hat{\nu})] + \frac{g^2a^2}{2}[A_\nu(n+\hat{\mu}), A_\nu(n)] + \mathcal{O}(a^3)). \end{aligned} \tag{3.64}$$

To perform the continuum limit, we take the Taylor expansion of the gauge field

$$A_\nu(n+\hat{\mu}) = A_\nu(n) + a\partial_\mu A_\nu(n) + \mathcal{O}(a^2), \tag{3.65}$$

and insert it in the expression for the plaquette in Eq. (3.64)

$$\begin{aligned} U_{\mu\nu}(n) &= \exp\left(iga^2\left(\partial_\mu A_\nu(n) - \partial_\nu A_\mu(n) + ig[A_\mu(n), A_\nu(n)]\right) + \mathcal{O}(a^3)\right) \\ &= \exp(ig^2a^2F_{\mu\nu}(n) + \mathcal{O}(a^3)). \end{aligned} \tag{3.66}$$

Substituting this expression for the plaquette in the Wilson action, we obtain

$$\begin{aligned}\mathcal{S}^W &= \sum_{n,\mu<\nu} \beta \left[1 - \frac{1}{2N_C}\mathrm{tr}\left(2\mathbb{1} - g^2a^4F_{\mu\nu}(n)F_{\mu\nu}(n) + \mathcal{O}(a^5)\right)\right] \\ &= \sum_{n,\mu<\nu} \beta\frac{g^2a^4}{2N_C}\mathrm{tr}\left[F_{\mu\nu}(n)F_{\mu\nu}(n)\right] + \mathcal{O}(a^5) \\ &\xrightarrow{a\to 0} \beta\int d^4x \frac{g^2}{2N_C}\mathrm{tr}\left[F_{\mu\nu}(n)F_{\mu\nu}(n)\right] + \mathcal{O}(a^2),\end{aligned} \tag{3.67}$$

which is equivalent to the pure gauge QCD action in the Euclidean space in case $\beta = (2N_C/g^2) = (6/g^2)$.

3.4 Quasi-PDFs approach

The definition in Eq. (2.4) of the PDFs cannot be employed to compute PDFs on a Euclidean lattice. Indeed, under the Wick rotation $\xi_0 \to i\xi_0$ the light cone shrinks to a point $\xi^2 = -\xi_0^2 - \xi^2 = 0$, thus making the integral singular. During the past few decades, different approaches have been developed in order to overcome this limitation [9, 52]. In this study, we mainly focus on the computation of nucleon PDFs employing the quasi-PDFs methodology [52]. This approach is based on the equivalence between two Lorentz frames, namely the Infinite Momentum Frame (IMF) [95], and the rest frame where $P_3 = 0$. Recalling the definition of PDF given in Sec. 2.2

$$q(x, P, \mu) \equiv \int_{-\infty}^{\infty} \frac{d\xi^-\,\lambda\cdot P}{4\pi}\, e^{+ix\xi^-\lambda\cdot P} h(\xi^-\lambda\cdot P), \tag{3.68}$$

and keeping the proton momentum finite, we move the vector λ away from the light-cone

$$\lambda = (1, 0, 0, -1)/\sqrt{2} \to \lambda' = (0, 0, 0, -1). \tag{3.69}$$

After performing this rotation, Eq. (3.68) reduces to

$$\tilde{q}(x, P_3, \mu) \equiv \int_{-\infty}^{\infty} \frac{dz}{4\pi}\, e^{+ixzP_3} \left\langle N\right| \overline{\psi}(z)\gamma^3\Gamma W\left(z, 0\right)\psi(0)\left|N\right\rangle. \tag{3.70}$$

The quantity $\tilde{q}(x, P_3, \mu)$ is called quasi-PDF and can be computed on in lattice QCD. The procedure that allows to relate the quasi-PDF to the light-cone PDF can be summarized in a few steps:

- the quasi-PDF is defined at finite-P_3 and $\lambda = (0, 0, 0, -1)$. The quark bilinear in the definition of the matrix elements in Eq. (2.2) is now at equal-time, and the matrix elements are not boost invariant;

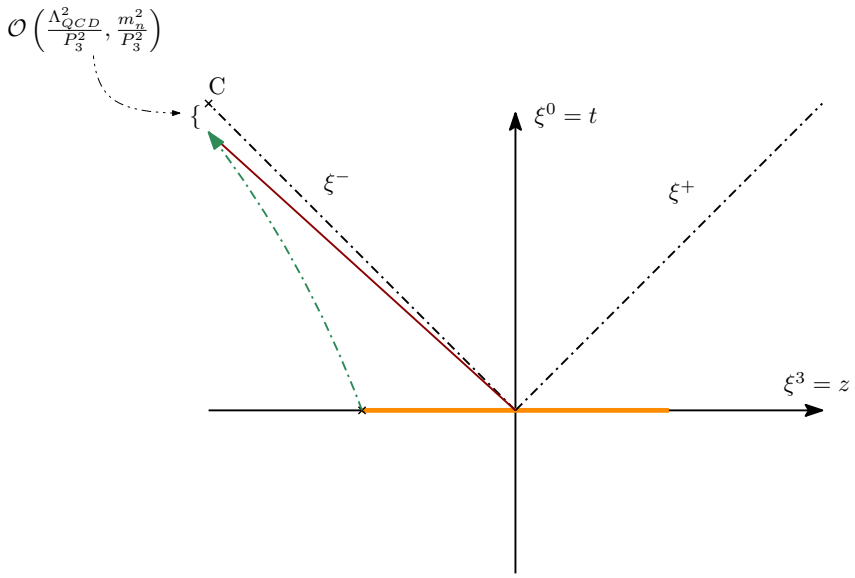


Figure 3.4: Schematic representation of the quasi-PDF approach. The dashed lines correspond to the light-cone directions and the green arrow represents the nucleon boost. The diagonal red solid line represents the estimate of the physical PDF from the one obtained within the quasi-PDF approach.

- we perform a Lorentz boost along z to the rest frame, where the spatial component of the proton momentum is $P_3 = 0$ and $\lambda = \gamma(\beta, 0, 0, -1)$, with $\beta = P_3/P_0$. The higher the quasi-PDF momentum is the closer will be λ to the light-cone. Since any boost can change the invariant length, the resulting vector cannot be exactly on the light-cone. However, if the momentum P_3 is large enough, the difference between the light-cone PDF and the quasi-PDF can be parametrized in terms of asymptotically vanishing power corrections in Λ_{QCD}/P_3 and M_n/P_3. The nucleon mass correction can be analytically computed to all orders [19];

- the finite momentum quasi-PDF can be related to the light-cone PDF or in the infinite momentum frame by a matching kernel computed in the large momentum effective theory (LaMET) [96]. Thus, the light-cone PDF $q(x, \mu)$ is given by

$$q(x,\mu) = \int_{\infty}^{\infty} \frac{dy}{|y|} C\left(y, \frac{\mu}{P_3}\right) \tilde{q}\left(\frac{x}{y}, \mu, P_3\right) + \mathcal{O}\left(\frac{\Lambda^2_{QCD}}{P_3^2}, \frac{m_N^2}{P_3^2}\right). \tag{3.71}$$

  The matching kernel C depends on the kind of PDF and can be computed in perturbation theory. The kernels employed in this work are computed to one-loop [97–101] and are discussed in Sec. 3.4.2.

3.4.1 Renormalization

The renormalization of matrix elements computed in lattice QCD constitutes a crucial step in the lattice computations. In what follows we will explain the main aspects of the procedure followed in order to remove divergences from the bare matrix elements of extended operators computed in lattice QCD.

The bare matrix elements of quasi-PDFs have a logarithmic divergence with respect to the regulator μ and a power divergence dependent on the Wilson line operator [102]. It assumes the form $\exp\{\delta m|z|/a + c|z|\}$, where z is the length of the Wilson line, $\delta m > 0$ is a dimensionless quantity parametrizing the strength of the power divergence and c is an arbitrary scale. To obtain a proper determination of δm, we employed a non-perturbative renormalization, introduced in Ref. [103], that we will briefly summarize in what follows. The parameter c can be fixed by such a renormalization prescription. The outcome of this procedure consists of a z-dependent function, the renormalization function Z

$$Z_\Gamma(z,\mu) = \hat{Z}_\Gamma e^{-\delta m|z| - c|z|}, \text{ with } \delta m > 0, \tag{3.72}$$

where $\hat{Z}_\Gamma$ is the multiplicative Z function that includes the contribution coming from the logarithmic divergence and Γ is the specific gamma structure characterizing the operator for the different kinds of PDFs. The procedure allowing to evaluate the Z-functions in the $\overline{\text{MS}}$ scheme consists of two steps:

- evaluation of the Z functions in the intermediate RI′-scheme at the scale μ_0. They can be extracted from appropriate *vertex functions*, evaluated via lattice computations;
- conversion of the renormalization function from the RI′-scheme $Z_\Gamma^{\text{RI}'}$ to the M$\overline{\text{MS}}$-scheme $Z_\Gamma^{\text{M}\overline{\text{MS}}}$, making use of perturbation theory. In this thesis, we refer to the Z-function in the M$\overline{\text{MS}}$-scheme simply as Z_Γ.

The definition of the Z-function suggests that the power divergence in the nucleon matrix elements can be removed by multiplying with Z_Γ

$$h_\Gamma^{\text{M}\overline{\text{MS}}}(z,P_3,\mu) = h_\Gamma^{\text{bare}}(z,P_3)\cdot Z_\Gamma^{\text{M}\overline{\text{MS}}}(z,\mu). \tag{3.73}$$

In the absence of a Wilson line ($z = 0$), the Z-function does not show any power divergence and Eq. (3.72) reduces to the renormalization function for local currents $\hat{Z}_\Gamma$.

The Z-function is a complex function of z, and thus the real and imaginary part of $h_\Gamma^{\text{M}\overline{\text{MS}}}$ are sums of two terms

$$\begin{aligned}
\text{Re}[h_\Gamma^{\text{M}\overline{\text{MS}}}] &= \text{Re}[Z_\Gamma^{\text{M}\overline{\text{MS}}}]\text{Re}[h_\Gamma^{\text{bare}}] - \text{Im}[Z_\Gamma^{\text{M}\overline{\text{MS}}}]\text{Im}[h_\Gamma^{\text{bare}}] \\
\text{Im}[h_\Gamma^{\text{M}\overline{\text{MS}}}] &= \text{Re}[Z_\Gamma^{\text{M}\overline{\text{MS}}}]\text{Im}[h_\Gamma^{\text{bare}}] + \text{Im}[Z_\Gamma^{\text{M}\overline{\text{MS}}}]\text{Re}[h_\Gamma^{\text{bare}}].
\end{aligned} \tag{3.74}$$

Further details regarding the renormalization of the specific matrix elements presented in this thesis will be provided in Chapters 7 and 6.

3.4.2 Matching procedure

Having obtained the renormalized quasi-PDFs in the M$\overline{\text{MS}}$ scheme, a matching procedure is necessary in order to make contact with the physical PDFs. It consists of a convolution

of the quasi-distribution with a kernel $C(\xi,\eta)$ evaluated in continuum perturbation theory within LaMET [96]. In this work we employ the matching formula

$$\begin{aligned} q(x,\mu) &= \int_{-\infty}^{\infty} \frac{d\xi}{|\xi|} C\left(\xi, \frac{\xi\mu}{xP_3}\right) \tilde{q}\left(\frac{x}{\xi}, \mu, P_3\right) \\ &= \int_{-\infty}^{\infty} \frac{dy}{|y|} C\left(\frac{x}{y}, \frac{\mu}{yP_3}\right) \tilde{q}\left(y, \mu, P_3\right), \end{aligned} \tag{3.75}$$

where the matching kernel $C\left(\xi,\eta\right)$ is computed in the $\overline{\text{MS}}$ scheme to one-loop in perturbation theory [101]. At leading order, the quasi-PDF is equivalent to the light-cone PDF, while the next-to-leading order (NLO) term $\mathcal{O}(\alpha_s)$ is nontrivial. Up to NLO the matching is given by

$$C\left(\xi,\eta\right) = \delta\left(1-\xi\right) + C^{\text{NLO}}(\xi,\eta)_{+(1)}. \tag{3.76}$$

The subscript $+(1)$ denotes the plus-prescription at $\xi = 1$, which implies

$$\begin{aligned} q(x,\mu) =& \tilde{q}(x,\mu,P_3) + \int_{-\infty}^{\infty} \frac{d\xi}{|\xi|} C^{\text{NLO}}\left(\xi, \frac{\mu\xi}{xP_3}\right) \tilde{q}\left(\frac{x}{\xi}, \mu, P_3\right) + \\ &- \tilde{q}\left(x,\mu,P_3\right) \int_{-\infty}^{\infty} \frac{d\xi}{|\xi|} C^{\text{NLO}}\left(\xi, \frac{\mu}{xP_3}\right). \end{aligned} \tag{3.77}$$

The next-to-leading order matching kernel for the unpolarized and helicity cases reads

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

with $\iota = 0$ for the unpolarized and $\iota = 1$ for the helicity. The matching kernel for the transversity PDFs reads

$$\begin{aligned} &C^{\text{NLO}}\left(\xi,\eta\right)_{+(1)} = \\ &+ \frac{\alpha_s C_F}{2\pi} \begin{cases} \left(\dfrac{2\xi}{1-\xi} \ln\left(\dfrac{\xi}{\xi-1}\right) + \dfrac{2}{\xi}\right)_{+(1)}, & \xi > 1, \\ \left(\dfrac{2\xi}{1-\xi} \ln\left[\dfrac{1}{\eta^2}\big(4\xi(1-\xi)\big)\right] - \dfrac{2\xi}{1-\xi}\right)_{+(1)}, & 0 < \xi < 1, \\ \left(-\dfrac{2\xi}{1-\xi} \ln\left(\dfrac{\xi}{\xi-1}\right) + \dfrac{2}{(1-\xi)}\right)_{+(1)}, & \xi < 0. \end{cases} \end{aligned}$$

Finally, to get rid of the residual dependence on the ratio between the nucleon mass to its momentum, we further apply the *nucleon mass corrections* (NMCs). In contrast to the corrections $\Lambda^2_{\text{QCD}}/P_3^2$ applied perturbatively through the matching procedure, the

NMCs can be computed analytically to all orders. In this work we employ the expressions derived in Ref. [19] for the unpolarized, helicity and transversity parton distributions.

Inverse matching procedure

The *inverse* matching procedure allows one to obtain the matrix elements starting from the phenomenological PDFs. Given the light-cone PDF, the quasi-PDF operator can be computed using

$$\tilde{q}\left(y,\mu,P_3\right) = \int_{-1}^{1} \frac{dy}{|y|} \tilde{C}\left(\frac{x}{y}, \frac{\mu}{yP_3}\right) q\left(y,\mu\right). \tag{3.78}$$

The two kernels $C(\xi,\eta)$ (direct matching kernel) and $\tilde{C}(\xi,\eta)$ (inverse matching kernel) only differ in the overall sign of the next-to-leading order term, which is minus for the inverse matching. For a detailed description of the inverse matching procedure and the technical aspects of the computation of the integral in Eq. (3.78) we refer to Ref. [104].

Chapter 4

Nucleon correlation functions in lattice QCD

4.1 Two-point functions

In this section we will briefly analyze the two-point correlator, showing how the energy levels of a hadron can be computed from it. The definition of the two-point correlator is

$$C^{2\mathrm{pt}}_{\alpha\beta}(x', x) \equiv \langle \Omega | \mathcal{J}^p_\alpha(x') \bar{\mathcal{J}}^p_\beta(x) | \Omega \rangle \,, \tag{4.1}$$

where $\mathcal{J}_p(x)$ is the proton *interpolating field* employed in this study, whose definition is

$$\mathcal{J}^p(x) \equiv \epsilon^{abc} u^a_\alpha(x) \left(d^{b^T}(x) C \gamma_5 u^c(x) \right) , \tag{4.2}$$

with $C = i\gamma_2\gamma_4$ the charge conjugation matrix. Besides reproducing the valence quark content of the proton, the interpolating field has the proton quantum numbers. Moreover, in order to fix the parity of the spinor $\mathcal{J}^p(x)$ one needs to project the correlation functions built from the interpolating field in the previous equation using a suitable parity projector. Given the plus and minus parity projectors $\Gamma = (1 \pm \gamma^4)/2$, the projected two-point correlator in position space can be written as

$$C^{2\mathrm{pt}}(x', x) \equiv \Gamma_{\alpha\beta} \langle \Omega | \mathcal{J}^p_\alpha(x') \overline{\mathcal{J}}^p_\beta(x) | \Omega \rangle \,. \tag{4.3}$$

This quantity can be interpreted as the probability amplitude for a proton propagating from x to x'. We call the space-time points x and x', respectively, *source position*, where a state with the nucleon quantum numbers is created, and *sink position*, where the state is annihilated. As a consequence of the translational invariance of the two-point correlator, this quantity depends just on the difference $x' - x$, and can be therefore assumed that the source position is set at zero, without loss of generality.

In what follows, we will study in detail the structure of the two-point correlators making explicit the dependence of this quantity on the energy states of the proton. In momentum space, the two-point correlator defined in Eq. (4.1) reads

$$C^{2\mathrm{pt}}_{\alpha\beta}(t; p) = \sum_{\vec{x}} e^{-i\vec{p}\cdot\vec{x}} C^{2\mathrm{pt}}_{\alpha\beta}(x), \tag{4.4}$$

where $x = (t, \vec{x})$ is now the only space-time index, corresponding to the sink position. In the Heisenberg picture, the interpolating field computed at x can be obtained as the operator computed in the origin transformed through the shift operator

$$C^{2\mathrm{pt}}_{\alpha\beta}(t;p) = \sum_{\vec{x}} e^{-i\vec{p}\cdot\vec{x}} \langle\Omega| e^{-i\hat{\vec{p}}\hat{\vec{x}}} e^{\hat{H}t} \mathcal{J}^p_\alpha(0) e^{-\hat{H}t} e^{i\hat{\vec{p}}\hat{\vec{x}}} \overline{\mathcal{J}}^p_\beta(0)|\Omega\rangle \,, \tag{4.5}$$

where $\hat{\vec{p}}$ is the momentum operator and $\hat{H}$ the QCD Hamiltonian operator. we insert a complete set of momentum and energy states

$$\mathbf{1} = \sum_{\vec{k},n} |n,\vec{k}\rangle \langle n,\vec{k}| \,, \tag{4.6}$$

where n counts the QCD Hamiltonian eigen states and $\vec{k}$ is the corresponding momentum. The resulting expression reads

$$C^{2\mathrm{pt}}_{\alpha\beta}(t;p) = \sum_{n} e^{-E(n;\vec{p})t} \langle\Omega|\mathcal{J}^p_\alpha(0)|n,\vec{k}\rangle \langle n,\vec{k}|\overline{\mathcal{J}}^p_\beta(0)|\Omega\rangle \,, \tag{4.7}$$

which can be written as

$$C^{2\mathrm{pt}}_{\alpha\beta}(t;p) = Z_0 e^{-E_0(\vec{p})t} \left(1 + Z_1 e^{-t\Delta E_1(\vec{p})} + Z_2 e^{-t\Delta E_2(\vec{p})} + \ldots\right), \tag{4.8}$$

with $Z_n \equiv |\langle\Omega|\mathcal{J}^p(0)|n,\vec{p}\rangle|$ and $\Delta E_n(\vec{p}) \equiv E_n(\vec{p}) - E_0(\vec{p})$, where $E_n(\vec{p}) \equiv E(n;\vec{p})$ is the $n-$th energy level of the QCD Hamiltonian in the boosted frame. Eq. (4.8) shows that we can explicitly write the two-point function as a sum over the energy levels of the QCD Hamiltonian each of which includes an exponentially decaying factor, with decay rate the energy of the state. In the limit where the source-sink time $t \to \infty$, the contribution of the excited states is exponentially suppressed and the two-point function reduces to the ground state energy

$$C^{2\mathrm{pt}}_{\alpha\beta}(t;p) = \sum_{n} e^{-E_n(\vec{p})t} |\langle\Omega|\mathcal{J}^p_\alpha(0)|n,\vec{k}\rangle|^2 \xrightarrow{t\to\infty} |\langle\Omega|\mathcal{J}^p(0)|n=0,\vec{p}\rangle|^2 e^{-E_0(\vec{p})t}. \tag{4.9}$$

Therefore, the energy of the ground state can be extracted from two point correlators in the infinite Euclidean time limit. The most widespread procedure for computing $E_0(\vec{p})$ consists of performing a constant fit in the large t region of the *effective energy* defined by

$$E_{\mathrm{eff}}(t;\vec{p}) \equiv \log\left(\frac{C^{2\mathrm{pt}}(t;p)}{C^{2\mathrm{pt}}(t+;p)}\right) = E_0(\vec{p}) + \log\left(\frac{1 + Z_1 e^{-\Delta E_1(\vec{p})t} + \ldots}{1 + Z_1 e^{-\Delta E_1(\vec{p})(t+1)} + \ldots}\right) \xrightarrow{t\to\infty} E_0(\vec{p}). \tag{4.10}$$

In Fig. 4.1 we show the nucleon effective energy and the *plateau fit* results, i.e. the results of a fit to a constant function. The two-point functions shown are produced using the twisted mass fermion ensemble cA211.30.32. The ensemble has $N_f = 2 + 1 +$

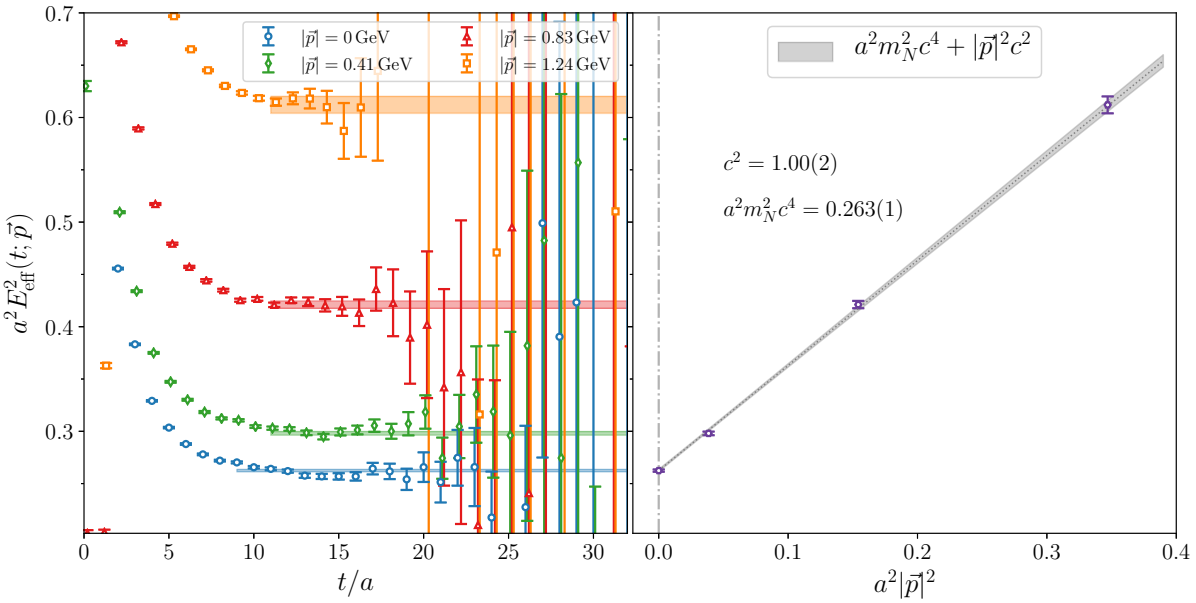


Figure 4.1: Left: effective energy computed using the ensemble cA211.30.32 at four different values of the nucleon boost $|\vec{p}| = 0, 0.41, 0.83, 1.24\,\mathrm{GeV}$. The values of the plateau fit are given in the legend. Right: Dispersion relation obtained using lattice data (blue points) and linear fit using Eq. (4.11).

1 dynamical twisted mass fermions, lattice volume $V = 32^2 \times 64$ and lattice spacing $a \approx 0.0938\,\mathrm{fm}$. The complete list of parameters regarding the cA211.30.32 ensemble is given in Table 6.2. The figure shows the effective energy obtained at four different values of the nucleon boost, i.e. $|\vec{p}| = \{0, 1, 2, 3\}\pi/L$ corresponding in physical units to $|\vec{p}| = \{0, 0.41, 0.83, 1.24\}\,\mathrm{GeV}$. We also show a plot of the results of the plateau fit as a function of the nucleon boost: the data for $E^2(\vec{p})$ follow a linear trend with $|\vec{p}|^2$, which satisfy the continuum dispersion relation

$$E^2(\vec{p}) = a^2 m_N^2 c^4 + |\vec{p}|^2 c^2, \tag{4.11}$$

m_N being the nucleon mass. A linear fit performed on the data shown in Fig. 4.1 with parameters $a^2 m_N^2 c^4$ and c^2 gives $c^2 = 1.00(2)$ and $a^2 m_N^2 c^4 = 0.263(1)$, where c is the speed of light confirming the fulfillment of the continuous dispersion relation.

4.1.1 Computation of two-point functions

Substituting in Eq. (4.3) the proton interpolator given in Eq. (4.2), and performing Wick contractions of the quark field we find

$$\begin{aligned} C^{2\mathrm{pt}}(x,0) =& \varepsilon^{abc}\varepsilon^{a'b'c'} \left(C\gamma_5\right)_{\alpha\beta} \left(C\gamma_5\right)_{\alpha'\beta'} \Gamma_{\gamma\gamma'} \\ & \langle\Omega| u^{a'}_{\gamma'}(x) d^{b'}_{\alpha'}(x) u^{c'}_{\beta'}(x) \overline{d}^{a}_{\alpha}(0) \overline{u}^{b}_{\beta}(0) \overline{u}^{c}_{\gamma}(0) u^{\dagger}|\Omega\rangle \\ =& \varepsilon^{abc}\varepsilon^{a'b'c'} \left(C\gamma_5\right)_{\alpha\beta} \left(C\gamma_5\right)_{\alpha'\beta'} \Gamma_{\gamma\gamma'} \mathcal{G}_d(x,0)^{b'a}_{\alpha'\alpha} \\ & \left[\mathcal{G}_u(x,0)^{a'c}_{\gamma'\gamma} \mathcal{G}_u(x,0)^{c'b}_{\beta'\beta} - \mathcal{G}_u(x,0)^{a'b}_{\gamma'\beta} \mathcal{G}_u(x,0)^{c'c}_{\beta'\gamma}\right], \end{aligned} \tag{4.12}$$

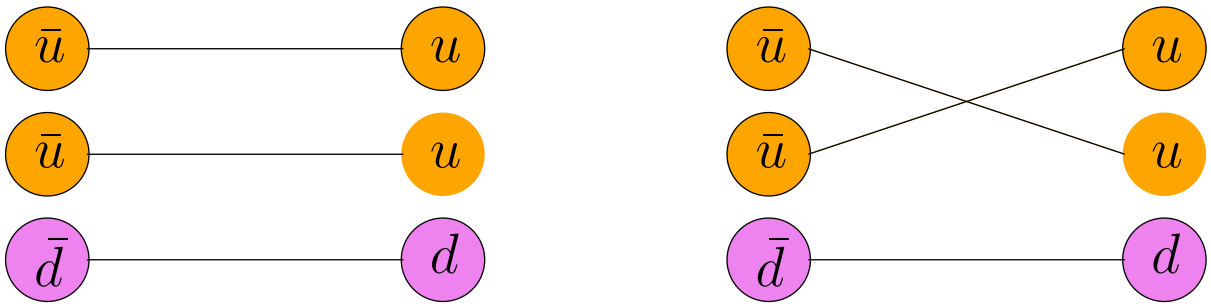


Figure 4.2: Schematic representation of the two terms contributing to the proton two-point function in Eq. (4.12). Each solid line correspond to a quark propagator, connecting the space-time points where the quark is created and annihilated.

where $\mathcal{G}_f(x, 0)$ is the quark propagator of flavor f, with source position 0 and sink position x. In Fig. 4.2 we show a schematic representation of the two terms contributing to the proton two-point function.

4.2 Three-point functions

In Sec. 3.4 we showed that the computation of quasi-PDFs requires the evaluation of three-point functions, that can be generically written as

$$C^{3\text{pt}}_{\alpha\beta}(x; y; x') = \langle \Omega | \mathcal{J}_\alpha(x) O(y) \overline{\mathcal{J}}_\beta(x') | \Omega \rangle \,, \tag{4.13}$$

where x and x' are respectively the sink and source positions and y is the operator insertion site. Analogously to the two-point functions, we take $x' = 0$ since the three-point function only depends on the separation between the source-sink and source-insertion positions, due to translational invariance and thus it can be written as

$$C^{3\text{pt}}_{\alpha\beta}(x; y) \equiv C^{3\text{pt}}_{\alpha\beta}(x; y; 0), \tag{4.14}$$

with $x = (t_s, \vec{x})$, $y = (t_{\text{ins}}, \vec{y})$ and $t_s < t_{\text{ins}} < 0$. In what follows we show the results for the spectral decomposition of the nucleon three-point functions, following a procedure analogous to the one employed in Sec. 4.1 for the two-point function. The proton three-point function reads

$$C_{3pt}(\vec{P}, \vec{q}; t_{\text{ins}}; t_s) = \mathcal{P}_{\alpha\beta} \sum_{\vec{x},\vec{y}} e^{-i\vec{P}\cdot\vec{x}} \, \langle \Omega | \mathcal{J}_\alpha(x) O(y) \overline{\mathcal{J}}_\beta(x') | \Omega \rangle \, e^{i\vec{q}\cdot\vec{y}}, \tag{4.15}$$

with $\vec{P}$ the final momentum and $\vec{q}$ the momentum transfer. In the Heisenberg picture, we can write

$$\begin{aligned} C_{3pt}(\vec{P},\vec{q};t_{\rm ins};t_s) = \mathcal{P}_{\alpha\beta}\sum_{\vec{x},\vec{y}} e^{-i\vec{P}\cdot\vec{x}}e^{i\vec{q}\cdot\vec{y}}\times \\ \times\langle\Omega|\mathcal{J}_\alpha(0)e^{-\hat{H}t_s}e^{i\vec{P}\cdot\vec{x}}e^{-i\vec{P}'\cdot\vec{y}}e^{\hat{H}t_{\rm ins}}O(0)e^{-\hat{H}t_{\rm ins}}e^{i\vec{P}'\cdot\vec{y}}\,\overline{\mathcal{J}}_\beta(x')|\Omega\rangle\,. \end{aligned} \tag{4.16}$$

Following the procedure employed with the two-point functions, we now insert two sets of complete eigen states defined in Eq. (4.6), obtaining

$$\begin{aligned} C_{3pt}(\vec{P},\vec{q};t_{\rm ins};t_s) = \mathcal{P}_{\alpha\beta}\sum_{\vec{x},\vec{y};\vec{k},\vec{k}',n,n'} \langle\Omega|\mathcal{J}_\alpha|n',k'\rangle\,\langle n',k'|O|n,k\rangle\,\langle n,k|\overline{\mathcal{J}}_\beta|\Omega\rangle\times \\ \times\, e^{-i\vec{x}\cdot(\vec{P}-\vec{k}')}e^{-i\vec{y}(\vec{k}'-\vec{k}-\vec{q})}e^{-E_{n'}(\vec{k}')(t_s-t_{\rm ins})}e^{-E_n(\vec{k})t_{\rm ins}}. \end{aligned} \tag{4.17}$$

The sum over the sink and insertion points provides useful identities involving the momenta vectors $\vec{k},\vec{k}',\vec{P}$ and $\vec{q}$. In particular, we obtain

$$\vec{k}'=\vec{P},\quad \vec{k}=\vec{P}-\vec{q}. \tag{4.18}$$

Therefore, performing the sum over the momenta and positions we finally obtain

$$\begin{aligned} C_{3pt}(\vec{P},\vec{q};t_{\rm ins};t_s) =& \mathcal{P}_{\alpha\beta}\sum_{n,n'}\langle\Omega|\mathcal{J}_\alpha|n',P\rangle\,\langle n',P|O|n,P-q\rangle\,\langle n,P-q|\overline{\mathcal{J}}_\beta|\Omega\rangle\times \\ &\times\, e^{-E_{n'}(\vec{P})(t_s-t_{\rm ins})}e^{-E_n(\vec{P}-\vec{q})t_{\rm ins}}. \end{aligned} \tag{4.19}$$

In the limit $t_s-t_{\rm ins}\to\infty$ and $t_{\rm ins}\to\infty$ (aka *forward limit*), the contribution associated with the ground state dominates the sum. In this limit we obtain

$$\begin{aligned} C_{3pt}(\vec{P},\vec{P}';t_{\rm ins};t_s) \xrightarrow{t_s-t_{\rm ins}\to\infty,\,t_{\rm ins}\to\infty} e^{-E_0(\vec{P})(t_s-t_{\rm ins})}e^{E_0(\vec{P}')t_{\rm ins}} \\ \times\langle\Omega|\mathcal{J}_\alpha|\mathcal{H}(P)\rangle\,\langle\mathcal{H}(P)|O|\mathcal{H}(P')\rangle\,\langle\mathcal{H}(P')|\overline{\mathcal{J}}_\beta|\Omega\rangle\,, \end{aligned} \tag{4.20}$$

$\mathcal{H}(\vec{p})$ being the proton ground state and $P'=P-q$. The large time limits allowing to isolate the ground-state contribution is not easy to obtain since the signal-to-noise ratio decreases exponentially with t_s, and thus the amount of statistics required to isolate the ground-state contribution with a sufficient degree of accuracy grow exponentially. Defining the quantity

$$Z_{\mathcal{H}}(P)=\sqrt{\frac{M}{E_0(P)}}\,\langle\Omega|\mathcal{J}_\alpha|\mathcal{H}(P)\rangle\,, \tag{4.21}$$

with M the hadron mass, we can write

$$C_{3pt}(\vec{P},\vec{P}';t_{\rm ins};t_s) \xrightarrow{t_s-t_{\rm ins}\to\infty,\, t_{\rm ins}\to\infty} Z(P)Z^*(P')\sqrt{\frac{M^2}{E_0(P)E_0(P')}}\,\langle\mathcal{H}(P)|O|\mathcal{H}(P')\rangle \times e^{-E_0(\vec{P})(t_s-t_{\rm ins})}e^{E_0(\vec{P}')t_{\rm ins}}. \tag{4.22}$$

The matrix element $\langle\mathcal{H}(P)|O|\mathcal{H}(P')\rangle$ can be obtained by taking the ratio

$$R(P,P';t_{\rm ins},t_s) = \frac{C_{\rm 3pt}(\vec{P},\vec{P}';t_{\rm ins};t_s)}{C_{\rm 2pt}(t;P')}\times \sqrt{\frac{C_{\rm 2pt}(t_s-t_{\rm ins};P)C_{\rm 2pt}(t_{\rm ins};P')C_{\rm 2pt}(t_s;P')}{C_{\rm 2pt}(t_s-t_{\rm ins};P')C_{\rm 2pt}(t_{\rm ins};P)C_{\rm 2pt}(t_s;P)}}, \tag{4.23}$$

since it cancels the factors $Z(P)$ and $Z(P')$, as well as the exponential terms of Eq. (4.20). The square root term in Eq. (4.23) reduces to the identity in the case of zero momentum transfer. All the computations presented in this thesis satisfy this condition, therefore to isolate the matrix element we will simply take the ratio between the three- and two-point function with zero momentum transfer.

4.2.1 Computation of three-point functions

We consider the isoscalar operator

$$O^\rho(x) = \overline{u}(x)\gamma^\rho u(x) + \overline{d}\gamma^\rho d(x). \tag{4.24}$$

Performing the Wick contractions we obtain the three point function

$$\begin{aligned} C^{\rm 3pt,\rho}_{\alpha\beta}(x,y) = |\Omega\rangle\, \varepsilon^{abc}u^a_\mu(x)(C\gamma_5)_{\mu\nu}d^b_\nu(x)u^c_\alpha(x)\times \\ \left[u^f_\kappa(y)(\gamma_\rho)_{\kappa\lambda}u^f_\lambda(x_1) - \bar{d}^{f'}_{\kappa'}(y)(\gamma_\rho)_{\kappa'\lambda'}d^{f'}_{\lambda'}(y)\right]\times \\ \varepsilon^{a'b'c'}\overline{u}^{c'}_\beta(0)(C\gamma_5)_{\nu'\mu'}\overline{d}^{b'}_{\nu'}(0)\overline{u}^{a'}_{\mu'}(0)\,\langle 0|\,. \end{aligned} \tag{4.25}$$

Three distinct terms contribute to the three-point function in Eq. (4.25), that can be written as

$$C^{\rm 3pt,\rho}_{\alpha\beta} = \mathcal{U}^\rho_{\alpha\beta} - \mathcal{D}^\rho_{\alpha\beta} + D^\rho_{\alpha\beta}, \tag{4.26}$$

where $\mathcal{U}_{\alpha\beta}$ and $\mathcal{D}_{\alpha\beta}$ are respectively the contributions coming from the coupling of the current with the up and down quark fields, while $D_{\alpha\beta}$ is the disconnected term, resulting by the contraction of the quark fields within the operator. It is easy to show that the

three contributions can be written as

$$\begin{aligned}\mathcal{U}^{\rho}_{\alpha\beta}(x,y) =&\varepsilon^{abc}\varepsilon a'b'c' \left[C\gamma_5\mathcal{G}_d(x;0)C\gamma_5\right]^{aa'}_{\mu\nu} \times \\ &\Big\{ \left[\mathcal{G}_u(x,y)\gamma_\rho\mathcal{G}_u(y,0)\right]^{bb'}_{\alpha\beta} \mathcal{G}_u(x_2,0)^{cc'}_{\mu\nu} + \\ &\left[\mathcal{G}_u(x,y)\gamma_\rho\mathcal{G}_u(y,0)\right]^{bb'}_{\alpha\nu} \mathcal{G}_u(x_2,0)^{cc'}_{\mu\beta} + \\ &\left[\mathcal{G}_u(x,y)\gamma_\rho\mathcal{G}_u(y,0)\right]^{bb'}_{\mu\nu} \mathcal{G}_u(x_2,0)^{cc'}_{\alpha\beta} + \\ &\left[\mathcal{G}_u(x,y)\gamma_\rho\mathcal{G}_u(y,0)\right]^{bb'}_{\mu\beta} \mathcal{G}_u(x_2,0)^{cc'}_{\alpha\nu} \Big\},\end{aligned} \tag{4.27}$$

$$\begin{aligned}\mathcal{D}_{\alpha\beta}(x,y) =&\varepsilon^{abc}\varepsilon^{a'b'c'} \left[C\gamma_5\mathcal{G}_d(x,y)\gamma_\rho\mathcal{G}_d(y,0)C\gamma_5\right]^{aa'}_{\mu\nu'} \times \\ &\left\{\mathcal{G}_u(x,0)^{bb'}_{\alpha\beta}\mathcal{G}_u(x,0)^{cc'}_{\mu\nu} + \mathcal{G}_u(x,0)^{bb'}_{\mu\beta}\mathcal{G}_u(x_2,0)^{cc'}_{\alpha\nu}\right\},\end{aligned} \tag{4.28}$$

$$\begin{aligned}D_{\alpha\beta}(x,y) = -\mathrm{Tr}\left[\left(\mathcal{G}_u(y,y) + \mathcal{G}_d(y,y)\right)\gamma_\rho\right] \varepsilon^{abc}\varepsilon^{a'b'c'} \left[C\gamma_5\mathcal{G}_d(x_2,0)C\gamma_5\right]^{bb'}_{\mu\mu'} \times \\ \times \left(\mathcal{G}_u(x,0)^{cc'}_{\alpha\beta}\mathcal{G}_u(x,0)^{aa'}_{\mu\mu'} + \mathcal{G}_u(x,0)^{ca'}_{\alpha\mu'}\mathcal{G}^{ac'}_{\mu\beta}\right).\end{aligned} \tag{4.29}$$

Both the terms $\mathcal{U}_{\alpha\beta}$ and $\mathcal{D}_{\alpha\beta}$ do not include disconnected quark loops, i.e. quark propagators starting from a given point in space time and ending at the same point, the so-called all-to-all quark propagators. Indeed, we gathered such terms in the disconnected term $D_{\alpha\beta}$: their signal-to-noise ratio is much smaller as compared to the connected contributions and their evaluation requires the use of appropriate stochastic and gauge noise reduction techniques that will be discussed in the following section.

4.3 Stochastic techniques

The evaluation of *all-to-all* propagators is one of the most computationally demanding aspects of lattice QCD computations. In particular, the exact evaluation of this quantity would require a number of inversions of the Dirac operator which is the same as the size of the lattice volume, i.e. $\mathcal{O}(10^7)$ inversions. In contrast, stochastic techniques allow to carry out the computation of the all-to-all propagators with a feasible but yet high computational cost. Such techniques are based on the definition of noise sources $\zeta_r(x)$, which obey the following properties:

$$\begin{aligned}\frac{1}{N_r}\sum_r \xi_r(x) &= 0 + \mathcal{O}\left(\frac{1}{\sqrt{N_r}}\right) \\ \frac{1}{N_r}\sum_r \xi_r(x)\otimes\xi_r^*(y) &= \delta(x,y)\delta_{\alpha\beta}\delta_{ab} + \mathcal{O}\left(\frac{1}{\sqrt{N_r}}\right),\end{aligned} \tag{4.30}$$

where N_r is the total number of noise vectors and $\otimes$ denotes the tensor product in volume, spin and color subspaces. One possible choice for the noise vectors consists in randomly filling the entries with elements of $\mathcal{Z}_2 = \{1,-1\}$ or $\mathcal{Z}_4 = \{1,-1,i,-i\}$. Given a set of vectors $\zeta_r(x)$, the all-to-all propagator can be constructed by solving for $\phi_r(x)$ the

following equation

$$M(x,y)\phi_r(y) = \zeta_r(x), \tag{4.31}$$

where $M(x,y)$ is the Dirac twisted mass operator. Having the set of solutions $\phi_r(x)$, the all-to-all propagator can be estimated via

$$\mathcal{G}(x;y) = \frac{1}{N_r}\sum_r \phi_r(x)\,\zeta_r^\dagger(y) + \mathcal{O}\left(\frac{1}{\sqrt{N_r}}\right). \tag{4.32}$$

In order to show that the stochastic error becomes comparable to the gauge error, the number of stochastic vectors required is much smaller than calculating the all-to-all propagator exactly. The stochastic technique allows us to compute out the trace of the quark propagator required in the computation of the disconnected contribution.

Further improvements can be made by studying the properties of the Dirac matrix under examination, allowing one to design better strategies for the selection of the vectors $\zeta_r(x)$. In particular, given the sparsity pattern of the matrix, the following theorem can be applied

Theorem 4.3.1. If the i-th row of a matrix $Z \in \mathbb{C}^{V\times N_r}$ given by the N_r stochastic vectors, is orthogonal to the rows of Z for which $M(i,j) \neq 0$, then the stochastic estimator yields the exact trace [105].

In the following section, we will analyze in more detail advanced techniques allowing to reduce the contamination to the trace coming from off-diagonal terms when employing stochastic techniques.

4.3.1 Dilution

With stochastic techniques we can compute the all-to-all propagator with a precision related to the number of noise vector employed. Indeed, the truncation of the sum introduces stochastic noise to the all-to-all propagator, which is due to the mixing with off-diagonal contributions. In order to reduce this noise, one can perform different types of *dilutions*, that consist in reducing the density of the noise vectors. Time dilution is one of the simplest schemes that consists in defining a noise vector that is different from zero only for a specific time slice

$$\zeta_r^\tau = \zeta_r \delta(t-\tau) \tag{4.33}$$

Dilution may also be applied on Dirac and color indices. In this case, given the color index a and Dirac index α, the noise vector reads

$$\tilde{\zeta}_r(x)^a_\alpha = \zeta_r(x)^b_\beta \delta_{ba}\delta^{\beta\alpha}. \tag{4.34}$$

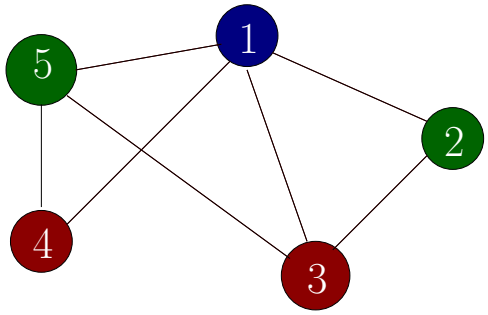


Figure 4.3: Graph $G = (V, E)$ with $V = \{1, 2, 3, 4, 5\}$ and $E = \{\{1, 3\}, \{1, 2\}, \{1, 4\}, \{1, 5\}, \{3, 5\}, \{4, 5\}, \{2, 3\}\}$. The chromatic number of the graph is $\chi_{k=1}(G) = 3$, and the corresponding optimal $3-$coloring map is shown.

Finally, there are also dilution schemes acting on the spatial volume. For instance one may divide the lattice into even and odd points, thus ensuring that closest neighbors do not pollute the diagonal entries of the all-to-all propagator. A very efficient dilution scheme, which acts on the volume is the *hierarchical probing* algorithm described in the next subsection.

Dilution techniques based on probing

Algorithms employing *probing* are among the most well performing methods for the stochastic evaluation of the trace of sparse matrices. The idea is to recover the non-zero entries of a sparse matrix by multiplying it with a small (compared to the size of the matrix) well-chosen set of vectors. In particular, given the sparsity pattern of the matrix B, special algorithms relying on *graph coloring* can be employed to identify a special set of probing vectors [106]. Before describing the *hierarchical probing* algorithm, which is the main topic of this section, the definition of some fundamental concepts will be provided. In what follows, we will assume that the matrix B is symmetric and non-singular.

Definition 4.3.1 (Undirected graph [107])**.** A graph is a pair $G = (V, E)$, where V are the vertices of the graph G and and E are its edges. In particular, each edge is associated with a couple of vertices, i.e. $E \subseteq [V]^2$ and, to avoid ambiguities, $V \cap E = \emptyset$. The usual graphical representation of a graph consists of drawing a dot for each vertex, while a line joining two of them represents an edge.

Definition 4.3.2 (Distance$-k$ neighbors [108])**.** Two distinct vertices u, v of the graph G are said to be distance$-k$ neighbors if the shortest path connecting them has at most length k.

Definition 4.3.3 (Distance$-k$ Graph coloring [107, 108])**.** Given a graph $G = (V, E)$, a distance$-k$ vertex coloring is a map $\mathcal{C} : V \to S$, where S is the set of *colors*, such that $\mathcal{C}(u) \neq \mathcal{C}(v)$ whenever u, v are distance$-k$ neighbors. The $k-$*chromatic number* $\chi_k(G)$ of a graph G, is the smallest k for which G is distance$-k$ colorable.

A simple example of a graph with 5 vertices and 7 edges is shown in Fig. 4.3. The chromatic number of this simple graph is $\chi_{k=1}(G) = 3$, and the corresponding optimal

3−coloring map is shown in the figure. Given the matrix B, for which we want to compute the trace of its inverse, it is possible to unambiguously define a graph of B as the intersection graph of the columns.

Definition 4.3.4 (Intersection graph of the matrix B [107])**.** Given the family $\mathcal{F}$ of sets, then the intersection graph G of $\mathcal{F}$ is defined by $XY \in E(G) \iff X \cap Y \neq \emptyset$. In the case of the matrix B, the sets are the rows of the matrix. Two rows, identified with sets whose elements are their entries, have non-empty intersection if their element-wise product returns a non-zero vector. Given a symmetric matrix B, the resulting intersection graph is undirected.

For instance, one of the possible matrices generating the graph of Fig. 4.3 is the following

$$B = \begin{bmatrix} b_{11} & 0 & b_{13} & 0 & b_{15} \\ 0 & b_{22} & b_{23} & 0 & 0 \\ b_{31} & b_{32} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & b_{45} \\ b_{51} & 0 & 0 & b_{54} & b_{55} \end{bmatrix} . \tag{4.35}$$

Given the intersection graph $G(V, E)$ with chromatic number $\chi_k(G)$ of the generic matrix B, with $n(c)$ we refer to the number of vertices of color c, with $c = 0, \ldots, \chi_k(G)$. We can then define the probing vectors

$$z_c(i) = \begin{cases} 1 & \text{if } \mathcal{C}(i) = c \\ 0 & \text{otherwise} \end{cases} , \tag{4.36}$$

which satisfy the requirements of theorem 4.3.1, allowing to compute exactly the trace with $\mathrm{Tr}\,[B] = \sum_{c=1}^{\chi_k(G)} z_c^T B z_c$. Then, if the chromatic number of the intersection graph is much smaller than the size of the matrix, provided a suitable coloring, this method allows exploiting the sparsity pattern to drastically reduce the computational cost of the trace.

In the case of the computation of the all-to-all propagator, we are interested in the trace of the inverse of the Dirac operator D. Therefore, given the sparsity pattern of the matrix D, it is crucial to know the pattern of the inverse. In general, the inverse of a sparse matrix is dense. However, in some cases, some elements of D^{-1} rapidly decay away from the non-zero entries of the sparsity pattern of D, and can be neglected. In this sense the inverse D^{-1} of the sparse matrix D is itself sparse and thus colorable. If it is the case, computing the distance-1 coloring of the inverse while dropping elements of D^{-1}_{ij} whose vertices are more than k links apart in the graph of D is equivalent of evaluating the distance-1 coloring of D^k [106]. In light of these considerations, the problem of computing the trace of the inverse of a large sparse symmetric matrix D can be now viewed in terms of how to perform probing on D^k, for large enough k. The *hierarchical probing* algorithm [109] allows to find a distance-1 probing basis of B^k by making certain

compromises, and can be summarized in two steps:

1. employ the hierarchical coloring algorithm [109], that recursively colors the submatrices obtained the previous iteration of the greedy coloring algorithm [110];

2. find a set of Hadamard vectors that spans exactly the probing basis.

A detailed description of the hierarchical coloring algorithm can be found in Ref. [109].

4.3.2 One-end trick

To compute the disconnected quark loop of the disconnected contribution to the three-point function, we also use the so-called *one-end trick*. It allows to exploit the properties of the noise vectors in order to transform a sum over the volume to a sum over the r index. We defined $\phi_r(x)$ as

$$\phi_r(\vec{x},t) = \mathcal{G}(x;x_0)\zeta_r(x_0), \tag{4.37}$$

where $\mathcal{G}(x;x_0)$ is the quark propagator. Then, we consider the expression

$$\begin{aligned}\frac{1}{N_r}\sum_r\sum_{\vec{x}}\phi_r^\dagger(\vec{x},t)\phi_r(\vec{x},t) &= \sum_{\vec{x},\vec{x}_0,\vec{x}_1}\mathcal{G}(x,x_0)^{*ab}_{\mu\nu}\mathcal{G}(x,x_1)^{ac}_{\mu\kappa}\frac{1}{N_r}\sum_r\zeta^{*b}_\nu(x_0)_r\zeta^c_\kappa(x_1)\\ &= \sum_{\vec{x},\vec{x}_0,\vec{x}_1}\mathcal{G}(x,x_0)^{*ab}_{\mu\nu}\mathcal{G}(x,x_1)^{ac}_{\mu\kappa}\delta^{bc}\delta_{\nu\kappa}\delta(x_0-x_1)\\ &= \sum_{\vec{x},\vec{x}_0}\mathrm{Tr}\Big[\,|\mathcal{G}(x,x_0)|^2\,\Big]\end{aligned} \tag{4.38}$$

This shows that one can transform a double sum over the space-time indices of the all-to-all propagator to a single sum over the volume and the noisy vectors.

The standard one-end trick

The standard one-end trick exploits properties of the twisted-mass discretization. In particular, the *standard one-end trick* can be employed in the case that τ^3 appears in the insertion operator Γ. In what follows, unless otherwise stated, operators and propagators are expressed in the twisted mass fermion formulation at maximal twist. For twisted mass fermions we know that

$$D_u - D_d = 2i\mu\gamma_5, \tag{4.39}$$

and, given the fact that $\mathcal{G}_u D_u = \mathbf{I}$, then

$$\mathcal{G}_u - \mathcal{G}_d = \mathcal{G}_u(D_d - D_u)\mathcal{G}_d. \tag{4.40}$$

using the previous two equations we find

$$\mathcal{G}_u - \mathcal{G}_d = -2i\mu\mathcal{G}_u\gamma_5.\mathcal{G}_d \tag{4.41}$$

Because of (4.41) we can write

$$L(t;\Gamma) = \sum_{\vec{x}} \left(\mathrm{Tr}\left[\mathcal{G}_u(x,x)\Gamma\right] - \mathrm{Tr}\left[\mathcal{G}_d(x,x)\Gamma\right]\right) = -2i\mu \sum_{\vec{x},y} \mathrm{Tr}\left[\mathcal{G}_u(x,y)\gamma_5\mathcal{G}_d(y,x)\Gamma\right]. \tag{4.42}$$

Thus, we can now employ (4.38) to write

$$\begin{aligned} L(t;\Gamma) &= -\frac{2i\mu}{N_r}\sum_r \sum_{\vec{x}} \mathrm{Tr}\left[\phi_r^\dagger(x)\gamma_5\Gamma\phi_r(x)\right] \\ &= \sum_{\vec{x}} \Big(\mathrm{Tr}\left[\mathcal{G}_u(x,x)\Gamma\right] - \mathrm{Tr}\left[\mathcal{G}_d(x,x)\Gamma\right]\Big) + \mathcal{O}\left(\frac{1}{\sqrt{N_r}}\right). \end{aligned} \tag{4.43}$$

To check this identity we need to use the property geven in (3.52)

$$\begin{aligned} \sum_r \sum_{\vec{x}} \mathrm{Tr}\left[\phi_r^\dagger(x)\gamma_5\Gamma\phi_r(x)\right] &= \sum_r \sum_{\vec{x},x_0,x_1} \mathrm{Tr}\left[\zeta_r^\dagger(x_0)\mathcal{G}_u^\dagger(x,x_0)\gamma_5\Gamma\mathcal{G}_u(x,x_1)\zeta_r(x_1)\right] \\ &= \sum_r \sum_{\vec{x},x_0,x_1} \mathrm{Tr}\left[\zeta_r^\dagger(x_0)\gamma_5\mathcal{G}_d(x,x_0)\Gamma\mathcal{G}_u(x,x_1)\zeta_r(x_1)\right] \\ &= \sum_{\vec{x},x_0} \mathrm{Tr}\left[\gamma_5\mathcal{G}_d(x,x_0)\Gamma\mathcal{G}_u(x,x_0)\right] + \mathcal{O}\left(\frac{1}{\sqrt{N_r}}\right) \\ &= \sum_{\vec{x},x_0} \mathrm{Tr}\left[\mathcal{G}_u(x,x_0)\gamma_5\mathcal{G}_d(x,x_0)\Gamma\right] + \mathcal{O}\left(\frac{1}{\sqrt{N_r}}\right). \end{aligned} \tag{4.44}$$

The generalized one-end trick

The standard one-end trick can be employed only in case the the operator includes τ^3. If this is not the case, the so-called *generalized one-end trick* comes in handy. In particular, we can write

$$D_u + D_d = 2D_W, \tag{4.45}$$

with D_W being the massless Wilson-Dirac operator. By exploiting the definition $\mathcal{G}_u D_u = \mathbf{1}$, we get

$$\mathcal{G}_u + \mathcal{G}_d = \mathcal{G}_u(D_u + D_d)\mathcal{G}_d, \tag{4.46}$$

which, using the identity in Eq. (4.45), can be written as

$$\mathcal{G}_u + \mathcal{G}_d = 2\mathcal{G}_u D_W \mathcal{G}_d. \tag{4.47}$$

From the definition of D_u it follows that

$$D_u\gamma^5 D_W = D_W\gamma^5 D_u, \tag{4.48}$$

which, multiplying on the left and right my $\mathcal{G}_u$, reads

$$\gamma^5 D_W \mathcal{G}_u = \mathcal{G}_u D_W \gamma^5. \tag{4.49}$$

We are interested in evaluating the loop

$$\mathcal{L}^{u+d}(t) = 2\mathrm{Tr}\left[\mathcal{G}_u D_W \mathcal{G}_d \Gamma\right]. \tag{4.50}$$

By exploiting the identities of Eq. (3.52) and Eq. (4.48), the previous equation can be written as

$$\mathcal{L}^{u+d}(t) = 2\sum_{\vec{x}}\sum_{y,z}\left[(\gamma^5)_{\alpha\beta} D_W(x;y)^{ab}_{\beta\gamma}\mathcal{G}_u(y,z)^{bc}_{\gamma\delta}\mathcal{G}^*_u(x,z)^{ac}_{\varepsilon\delta}(\gamma^5\Gamma)_{\varepsilon\alpha}\right]. \tag{4.51}$$

We insert now the unity in position, color and Dirac spaces in between the two propagators and introduce a set of stochastic vectors ζ_r satisfying the properties of Eq. (4.30)

$$\begin{aligned}\mathcal{L}^{u+d}(t) = 2\sum_{\vec{x}}\sum_{y,z}\left[(\gamma^5)_{\alpha\beta} D_W(x;y)^{ab}_{\beta\gamma}\mathcal{G}_u(y,z)^{bc'}_{\gamma\delta'}\frac{1}{N_r}\sum_r \zeta_r(z')^{c'}_{\delta'}\zeta^*_r(z)^{c}_{\delta}\mathcal{G}^*_u(x,z)^{ac}_{\varepsilon\delta}(\gamma^5\Gamma)_{\varepsilon\alpha}\right] + \\ + \mathcal{O}\left(\frac{1}{\sqrt{N_r}}\right).\end{aligned} \tag{4.52}$$

Using the property in Eq. (4.37), the equation above becomes

$$\mathcal{L}^{u+d}(t) = 2\frac{1}{N_r}\sum_{\vec{x}}\sum_{y}\sum_{r}\left[(\gamma^5)_{\alpha\beta} D_W(x;y)^{ab}_{\beta\gamma}\phi^u_r(y)^b_\gamma\phi^{u*}_r(x)^a_\varepsilon(\gamma^5\Gamma)_{\varepsilon\alpha}\right] + \mathcal{O}\left(\frac{1}{\sqrt{N_r}}\right). \tag{4.53}$$

In summary, we have

$$\begin{aligned}\mathcal{L}^{u+d}(t) &= 2\frac{1}{N_r}\sum_{\vec{x}}\sum_r \mathrm{Tr}\left[\phi^\dagger_r(x)\gamma^5\Gamma\gamma^5 D_W\phi_r(x)\right] \\ &= \sum_{\vec{x}}\left(\mathrm{Tr}\left[\mathcal{G}_u(x,x)\Gamma\right] + \mathrm{Tr}\left[\mathcal{G}_d(x,x)\Gamma\right]\right) + \mathcal{O}\left(\frac{1}{\sqrt{N_r}}\right),\end{aligned} \tag{4.54}$$

which is known as the *generalized one-end trick*.

4.4 Smearing techniques

The gauge field is subject to fluctuations which deteriorate the signal-to-noise ratio. Furthermore, one needs to optimize the overlap of the interpolating fields with the ground state when computing hadronic quantities. Various *smearing* methods are used to reduce gauge noise and to improve ground state dominance. In particular, Gaussian or Wuppertal smearing, APE smearing and momentum smearing are employed in our studies to reduce the contamination by excited states and gauge noise. Stout smearing is employed on gauge links to smooth the power divergences present in non-local operators with Wilson line.

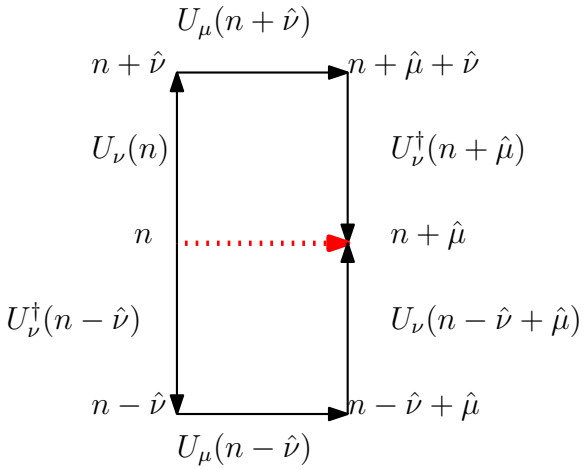


Figure 4.4: Staples used to APE smear the gauge link $U_\mu(n)$.

4.4.1 APE smearing

APE smearing [111], as well as stout smearing act on the gauge links. In both procedures each link variable $U_\mu(n)$ is replaced with a new one obtained from the sum over appropriately defined *staples*. A simple staple can be obtained by multiplying three elementary gauge links in a plane. In the case of APE smearing we have

$$U_\mu^{\text{APE}} = \text{Proj}_{\text{SU(3)}}[(1-\alpha_{\text{APE}})U_\mu(n) + \frac{\alpha_{\text{APE}}}{6}\sum_{\mu\neq\nu} C_{\mu\nu}(n)], \tag{4.55}$$

with $C_{\mu\nu}(x)$ the sum of simple staples around $U_\mu(x)$ given by

$$C_{\mu\nu}(n) = U_\nu(n)U_\mu(n+\hat{\nu})U_\nu^\dagger(n+\hat{\mu}) + U_\nu^\dagger(n-\hat{\nu})U_\mu(n-\hat{\nu})U_\nu(n-\hat{\nu}+\hat{\mu}). \tag{4.56}$$

The sum over the staples can be on the four lattice dimensions or on the spacial dimensions only. APE-smearing is a recursive method, since the replacement of the gauge links described in Eq. (4.55) is performed n_{APE} times. Since the sum of the staples is not an element of the $SU(3)$ group, an $SU(3)$ projection [112,113] is needed at each iteration to ensure that the algorithm remains effective. A schematic 2D representation of the staples employed to APE smear the gauge link $U_\mu(n)$ is shown in Fig. 4.4. Besides improving the overlap of the interpolating fields with the ground state of the hadron of interest, APE smearing also helps in limiting the presence of *exceptional configurations* (see Sec. 3.3.3), whose eigenvalues of the Dirac operator are close to zero. However, such smearing technique cannot be employed when performing Hybrid Monte Carlo [114] simulations, which require the gauge action to be differentiable.

4.4.2 Stout smearing

Stout smearing [115] is an iterative smearing technique obtained by replacing each link variable by

$$U_\mu(n) \to U'_\mu(n) = e^{iQ_\mu(n)} U_\mu(n), \tag{4.57}$$

where $Q_\mu(n)$ is a traceless hermitian matrix defined as

$$\begin{aligned} Q_\mu(n) &\equiv \frac{i}{2}\left(\Omega^\dagger_\mu(n) - \Omega_\mu(n) - \frac{\mathbf{1}}{3}\mathrm{Tr}\left[\Omega^\dagger_\mu(n) - \Omega_\mu(n)\right]\right); \\ \Omega_\mu(n) &\equiv \left(\sum_{\mu\neq\nu} \rho_{\mu\nu} C_{\mu\nu}(n)\right) U^\dagger_\mu(n), \end{aligned} \tag{4.58}$$

and $C_{\mu\nu}(n)$ is the sum of staples defined in Eq. (4.56). The smearing can be tuned by setting an appropriate value of the parameters $\rho_{\mu\nu}$. In contrast to the APE smearing procedure, the new link variables are in $SU(3)$, eliminating the need for any projection. As a consequence, the updated link variables are analytic in the finite complex plane and the stout smearing can be employed in combination with Monte Carlo methods. In the present study we will refer to a single stout smearing parameter ρ, defined as

$$\rho_{ij} = \rho, \quad rho_{4\mu} = \rho_{\mu 4} = 0, \tag{4.59}$$

which is often referred to as the isotropic three-dimensional smearing. In some cases, it might be useful to smear the gauge links in the temporal direction as well, thus employing the four-dimensional isotropic setup

$$\rho_{\mu\nu} = \rho. \tag{4.60}$$

4.5 Gaussian Smearing

To improve the overlap of the interpolating field with the ground state, thus suppressing the excited states contamination, we employ the so-called Gaussian smearing, also known as Wuppertal smearing [116]. Given a point source $\eta(\vec{y}, t)$, the smeared source reads

$$\eta^{\mathrm{sm}}(\vec{x}, t) = \sum_{\vec{y}} F(\vec{x}, \vec{y}; U(t))\eta(\vec{y}, t), \tag{4.61}$$

where the operator F is defined as

$$F(\vec{x}, \vec{y}; U(t)) = [1 + \alpha H(\vec{x}, \vec{y}; U(t))]^n \tag{4.62}$$

and

$$H(\vec{x}, \vec{y}; U(t)) = \sum_{k=1}^{3} \left[U_k(\vec{x}, t)\delta_{\vec{x},\vec{y}-a\hat{k}} + U_k(\vec{x} - a\hat{k}, t)\delta_{\vec{x},\vec{y}+a\hat{k}}\right]. \tag{4.63}$$

The parameter α and the number of smearing iteration n are tuned to yield the approximate root mean sqaure radius of the hadron. Finally, since the operator F depends on the gauge links, smearing techniques such as APE can be employed to reduce gauge noise.

Chapter 5

Bayes-Gauss-Fourier transform

On the lattice, renormalized matrix element is only available for a limited range of z-values, where z is the length of the Wilson line entering the non-local operators. Since the theory is defined on a discrete set of lattice sites, the matrix element is known only for discrete values of z. Thus, the integral defining the quasi-PDF becomes a finite sum

$$2P_3 \int_{-\infty}^{\infty} \frac{dz}{4\pi} e^{ixP_3 z} \to 2P_3 \sum_{z=-z_{\max}}^{z_{\max}} \frac{a}{4\pi} e^{ixP_3 z}, \tag{5.1}$$

where a is the lattice spacing. In particular, the transformation given in Eq. (5.1) means that, instead of computing the Fourier transform (FT), we compute an analytic continuation of the discrete Fourier transform (DFT) defined for continuous x values. The DFT frequencies are

$$\omega_k = \frac{2\pi}{N} k, \quad k \in [-k_{\max}, k_{\max}], \tag{5.2}$$

with $N = (2k_{\max}/a + 1)$ and $ak_{\max} = z_{\max}/a$. However, the FT frequencies relevant for the computation of quasi-PDFs are

$$\omega_x = xP_3 = 2\pi \frac{P}{L} x, \tag{5.3}$$

where the momentum $P_3 = 2\pi P/L$, L is the spatial extent of the lattice, $P \in \mathbb{N}$ and x assumes continuous values in the interval $[-1, 1]$.
The discretization procedure described in Eq. (5.1) introduces a systematic bias in the quasi-PDF [117]:

1. knowing the matrix element only at discrete z-values limits the high frequency components to $|x| < \frac{\pi}{aP_3}$. Therefore, higher frequency components cannot be resolved and they are wrongly measured as lower frequency components below the threshold. This phenomenon is also known as *aliasing*.

2. The finite spatial extent of the lattice introduces a cut-off on z that is limited by $z_{\max} \leq \frac{L}{2}$. A limitation in the number of points reduces the frequency resolution of the discretized Fourier transform defined in Eq. (5.1). When frequency components of the signal do not correspond to the discrete frequencies, the discretized Fourier

transform suffers from a distortive effect known as frequency *leakage*. Cut-off effects become significant for $aP_3 \sim 1$.

Different advanced reconstruction techniques attempting to overcome the limitations of the discrete FT have been proposed [48, 117]. In the next section, we will introduce the Bayes-Gauss-Fourier transform, an algorithm based on Gaussian process regression we introduced in Ref. [48].

5.1 Bayes-Gauss-Fourier Transform

These two problems could be solved if it were possible to reconstruct a continuous form of the renormalized lattice matrix element, before evaluating the FT. In Ref. [48], we proposed a new method to overcome the limitations of the discretized FT in Eq.(5.1). The method, called *Bayes-Gauss-Fourier Transform* (BGFT) is based on the formalism of Gaussian Process Regression (GPR) [118]. The choice of GPR for the continuous reconstruction is based on the following reasons:

- the interpolation is non-parametric, so it has the flexibility to adapt to any dataset without being restricted to a specific parametrized function;
- since GPR is based on Bayesian inference, the information about the behavior of the function towards infinity (extrapolation) can be incorporated into the prior distribution, and taken into account for the interpolation;
- the uncertainties of the measurements are incorporated into the interpolation through Bayes theorem;
- it is possible to impose a chosen level of smoothness to the interpolating function;
- the result of the interpolation is continuous, defined over whole domain of interest and its Fourier transform is computable in closed form.

In the following sections we give a brief introduction to the Gaussian Process Regression (GPR) and the way we employed it in order to obtain a robust estimate of the quasi-PDF through the Bayes-Gauss-Fourier transform.

5.2 Gaussian Process Regression

Gaussian process regression is a *nonparametric* regression technique based on Bayesian inference. Before moving to the description of GPR, it is crucial to define the meaning of *nonparametric* in contrast to parametric regression models. When performing a linear regression (parametric model) the number of parameters is fixed throughout the fit procedure. Nonparametric models instead, still are defined in terms of parameters, but their number increases with the number of instances populating the training set. The

more data we include in the regression, the larger will be the number of parameters the nonparametric regression model has and, as a consequence, the complexity of the model increases as well. This aspect will be evident once the definition of the Gaussian Process regression result will be provided.
The Bayes' rule reads

$$\begin{aligned} \text{Posterior} &= \frac{\text{Likelihood} \times \text{Prior}}{\text{Marginal likelihood}} \\ P(A|B) &= \frac{P(B|A)P(A)}{P(B)}, \end{aligned} \tag{5.4}$$

where the second line is equivalent to the first written in terms of probabilities. In our case A will be the unknown model describing the observed data, given by B. The *prior* $P(A)$ is a key quantity when performing Gaussian Process Regression, since it is not present in the *frequentist* approach to regression and express our beliefs about the target model before we look at the observation. This term can encode theoretical constraints on the regression curve. The *likelihood* express the probability to obtain the data given the model describing them: the better the model describes the training set, the higher will be the likelihood. Finally, the *marginal likelihood* in Eq. (5.4) is a normalization factor, that does not depend on the model A. The explicit expression for the quantities reported in Eq. (5.4) will be provided in what follows.

Definition 5.2.1 (Gaussian Process [118])**.** A Gaussian process (GP) is a collection of random variables, any finite number of which have a multivariate Gaussian distribution

$$f(z) \sim \mathcal{GP}\left(m(z), k(z, z')\right), \tag{5.5}$$

where $\mu(z)$ is the mean function and $k(z, z')$ the covariance function. This means that, given any number n_z of domain points $z_1, \ldots, z_n$, the random variables $f_1, \ldots, f_n$, where $f_i \equiv f(z_i)$, are distributed according to:

$$p(f_1, ..., f_n) = \det(2\pi K)^{-1/2} e^{-\frac{1}{2}\sum_{ij}(f_i-\mu_i)K^{-1}_{ij}(f_j-\mu_j)}, \tag{5.6}$$

where $\mu_i \equiv \mu(z_i)$ and K^{-1}_{ij} is the inverse matrix of $K_{ij} \equiv k(z_i, z_j)$.
In summary, the Gaussian Process is completely defined by two functions

$$\begin{aligned} \mu(z) &= \mathbf{E}[f(z)] \\ k(z, z') &= \mathbf{E}[(f(z) - \mu(z))(f(z') - \mu(z'))], \end{aligned} \tag{5.7}$$

corresponding to the mean and the covariance of the multivariate Gaussian prior distribution.

If we assume the target model $f(z)$ is a Gaussian Process

$$f(z) \sim \mathcal{GP}\left(m(z), k(z, z')\right), \tag{5.8}$$

then the likelihood in Eq. (5.4) can be written as a multivariate Gaussian distribution

$$\mathcal{L}(h|z,f) = \mathcal{N}(f,\mathbf{\Sigma}), \tag{5.9}$$

being f the mean and $\mathbf{\Sigma}$ the covariance matrix of the distribution. The latter is defined as a diagonal matrix of size n_z (number of observations), whose entries are the uncertainties on the data points $h(z)$ included in the training set

$$\mathbf{\Sigma} = \begin{pmatrix} \Delta h(z_1)^2 & 0 & \dots & 0 \\ 0 & \Delta h(z_2)^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Delta h(z_n)^2 \end{pmatrix}. \tag{5.10}$$

In our application, they correspond to the total error given to the matrix element for each value of the independent variable z. In addition, from the definition of the Gaussian process follows that the prior distribution is a Gaussian multivariate distribution, defined in Eq. (5.6). Finally, the marginal likelihood ensuring the normalization of the posterior distribution is defined as the integral over the Gaussian process f of the product of the likelihood times the prior

$$p(h|Z) = \int p(h|f,Z)\, p(f)\, df, \tag{5.11}$$

where Z is the set of values of the independent variable corresponding to the observations $h(z)$.

Once all the quantities appearing in the Bayes's rule in Eq. (5.4) are defined, the posterior can be evaluated and its mean correspond to the regression output. Denoting by $z_1, ..., z_n$ the z values in which the matrix element measures $h_1, ..., h_n$ are available with uncertainties $\Delta h_1, ..., \Delta h_n$ and assuming the errors to be Gaussian, it is possible to analytically compute the mean function $\mu(z)$ of the posterior Gaussian process $f(z)$ in terms of the prior GP [118]:

$$\mu(z) = \mu_P(z) + \sum_{ij} k_P(z,z_i)\widetilde{K}^{-1}_{ij}(h_j - \mu_P(z_j)) \tag{5.12}$$

where $\widetilde{K}_{ij} = k_P(z_i,z_j) + \Delta h_i^2 \delta_{ij}$.

The GPR described until now is defined on real-valued data. There are many possible ways to extend the procedure to complex data. The approach that we choose consists in performing two independent fits: a GPR for the absolute values and a minimum χ^2 linear regression for the complex argument.

5.2.1 Choice of the prior

It is possible to verify from Eq. (5.12) that, if the errors Δh_i are zero, then $\mu(z)$ passes through all measured points. This property, together with smoothness conditions that can be imposed on GP, as it will be detailed below, ensures that the dependence of the mean posterior on the mean prior becomes weaker as we get closer to measured points. Therefore, the choice of GP prior does not affect the domain areas with high density of measured points. On the other hand, for z values that are far from the measured values, the choice of the prior plays a decisive role.

With the choice of the prior GP, it is possible to impose different levels of smoothness to a GP [119, 120]. In particular, choosing $\mu_P(z), k_P(z, z') \in C^\infty$ guarantees that both the GP prior and the GP posterior are infinitely mean-square differentiable, which we expect to be the most appropriate description for our problem at hand.

In order to not compromise the performance of the method and to reduce over-fitting, the prior covariance function is commonly chosen to be stationary and symmetric so that $k_P(z, z') = k_P(|z - z'|, 0)$. Another common choice is to use a covariance function monotonically decreasing with distance: $k_P(a, 0) > k_P(b, 0)\ \forall a, b$ such that $0 < a < b$. This last property states that the correlation between predictions at different z_1 and z_2 decreases with their distance $|z_1 - z_2|$, which means that the value of $\mu(z)$ will be mostly determined by the value of the neighboring measured points.

We opted for the squared exponential covariance function, which is the standard choice for a C^∞ function that satisfy all the properties listed above. We thus consider

$$k_P(z, z') = \sigma^2 e^{-\frac{(z-z')^2}{2\ell^2}}, \tag{5.13}$$

where the real values σ and ℓ, also called hyper-parameters, are fixed using the maximum likelihood estimation of type II described in Ref. [118].

Since the behavior of the mean function tends to be independent of the measured values at long distances from them, the asymptotic behavior depends only on the choice of the prior mean function. Since the renormalized matrix element should tend to zero in the limit of $|z/a| \to \infty$, it is then possible to guarantee this limit for the posterior mean function by choosing a prior mean function that tends to zero at infinity.

5.2.2 Properties of the lattice matrix elements

We apply the BGFT approach to the $\overline{\text{MS}}$ renormalized matrix element of the unpolarized PDF. We use the results of Ref. [101] that were computed using a gauge ensemble with two dynamical mass degenerate light twisted mass quarks ($N_f = 2$) generated by the Extended Twisted Mass Collaboration (ETMC) [121]. The lattice spacing is $a = 0.0938(2)(3)$, the lattice size $48^3 \times 96$ and the source-sink time separation is fixed at $t_s = 12a \approx 1.1\,\text{fm}$. All the relevant parameters regarding the employed gauge ensemble are reported in Table 5.1. This ensemble is referred to as the *cA2.09.48* ensemble.

β=2.10,	$c_{\rm SW}$=1.57751,	a=0.0938(3)(2) fm
$48^3 \times 96$	$a\mu = 0.0009$	$m_N = 0.932(4)$ GeV
$L = 4.5$ fm	$m_\pi = 0.1304(4)$ GeV	$m_\pi L = 2.98(1)$

Table 5.1: Simulation parameters of the cA2.09.48 ensemble used to extract the unpolarized quasi-PDF. The nucleon mass (m_N), the pion mass (m_π) and the lattice spacing (a) were determined in Ref. [122].

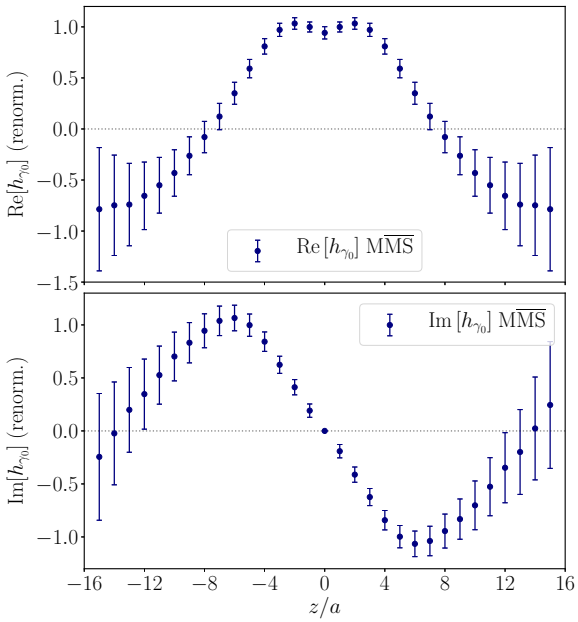


Figure 5.1: Real and imaginary part of the renormalized matrix element computed at $P_3 = 10\pi/L \simeq 1.38\,\text{GeV}$.

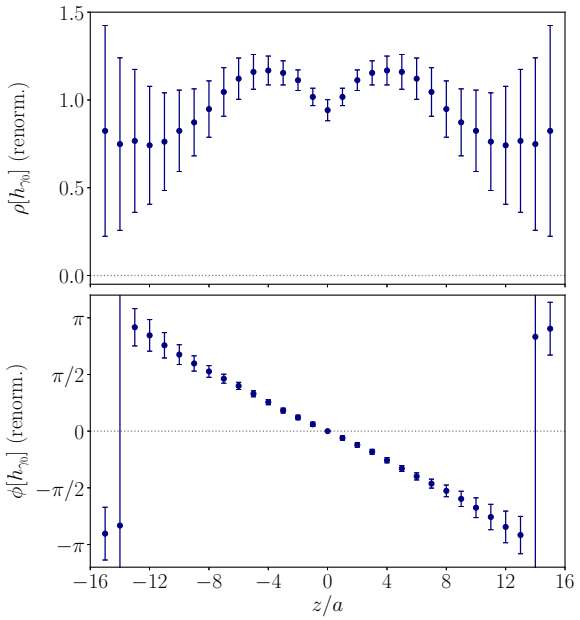


Figure 5.2: Modulus $\rho(z)$ and argument $\phi(z)$ of the renormalized matrix element $h(z)$.

We show the results on the renormalized nucleon matrix element of the unpolarized operator in Fig. 5.1 taken from ref. [101]. The momentum is $P_3 = 10\pi/L \simeq 1.38\,\text{GeV}$. A total of $N_{conf} = 811$ gauge configurations were analyzed, with a total number of measurements $N_{measures} = 73000$. The errors were obtained using the jackknife resampling method.

Before presenting the methods used to determine the quasi-PDFs, it is necessary to analyze some important aspects regarding the properties of these lattice matrix element. In a recent paper [101], based on the data from the *cA2.09.48* ensemble, the authors identified possible sources of systematic uncertainties in the reconstruction of PDFs from lattice QCD simulations, including the dependence on the cutoff $z_{\rm max}$.

The results showed that the PDFs extracted are contaminated by a larger noise and stronger oscillations as $z_{\rm max}$ increases above the value of $z_{\rm max}/a \gtrsim 10$. In particular, the higher the value of $z_{\rm max}$, the bigger the errors prohibiting the PDFs from reaching a zero value in the large positive x region. However, all the light-cone PDFs extracted using different values of the cut-off $z_{\rm max}/a$ are compatible within errors [101]. We chose to compare the results of the proposed non-parametric regression procedure with the DFT

quasi-PDF obtained with $z_{\max}/a = 10$.
Let us examine first some important properties exhibited by the renormalized matrix element that will be useful to better design our algorithm. Firstly, they are hermitian functions of z, namely

$$h(z, P_3, \mu) = h(-z, P_3, \mu)^\dagger, \tag{5.14}$$

which means that the real part is even and the imaginary part is odd in z. This property will be exploited by our BGFT method in order to make the algorithm more stable, as described in Sec. 5.2.4.

Secondly, both the real and imaginary parts are expected to decay to zero, with a rate that increases with the nucleon boost. This crucial aspect will be reflected in the choice of the prior mean of the GPR, which encodes all the relevant information about the renormalized matrix element, as illustrated in Sec. 5.2.1. Thirdly, the real and imaginary parts of the matrix element can be written in polar coordinates, i.e. at fixed z, the complex function $h(z) = \Re h(z) + i\Im h(z)$ can be written as

$$h(z) = \rho(z) e^{i\phi(z)},$$

with

$$\begin{aligned} \rho(z) &= \sqrt{\Re h(z)^2 + \Im h(z)^2} \\ \phi(z) &= \arg(h(z)) \\ &= \operatorname{arctan2}\left(\Im h(z), \Re h(z)\right). \end{aligned}$$

The function $\operatorname{arctan2}(y, x)$ reads

$$\operatorname{arctan2}(y, x) = \\ = \begin{cases} 2 \arctan \frac{y}{\sqrt{x^2+y^2}+x} & \text{if } x > 0 \text{ or } y \neq 0 \\ \pi & \text{if } x < 0 \text{ and } y = 0 \\ \text{undefined} & \text{if } x = 0 \text{ and } y = 0 \end{cases}$$

In Fig. 5.2 we show the functions $\rho(z)$ and $\phi(z)$ for the renormalized nucleon matrix element of the unpolarized operator. The modulus $\rho(z)$ of the complex function $h(z)$ is an even function that decays to zero with a rate that depends on the nucleon boost. However, due to the increasing errors in the renormalization function, the function $\rho(z)$ deviates from zero as z increases. On the other hand, the argument $\phi(z)$ shows a linear behavior for all relevant values of z. Such a property means that the matrix element can be written as

$$\begin{aligned} \Re h(z) &= \rho(z) \cos(\theta z) \\ \Im h(z) &= \rho(z) \sin(\theta z), \end{aligned} \tag{5.15}$$

or, equivalently

$$h(z) = \rho(z)e^{i\phi(z)} = \rho(z)e^{i\theta z}, \tag{5.16}$$

This property will be used in the BGFT to improve the results as described in Sec. 5.2.4. The linear dependence of $\phi(z)$ on z can also be observed directly from the phenomenological PDFs, as discussed in the following section.

5.2.3 Analysis of phenomenological data

To illustrate the relation between the real and imaginary parts of the underlying matrix element, we consider the NNPDF3.1 phenomenological determination of the unpolarized PDF [63]. This data set is shown in Fig. 5.3 and we apply the *inverse* matching procedure to derive the corresponding matrix element. The inverse matching can be interpreted as the inverse of the operation given in Eq. (3.71), which allows us to obtain the quasi-PDF from the light cone PDF by

$$\tilde{q}\,(x,\mu,P_3) = \int_{-1}^{1} \frac{dy}{|y|}\,\tilde{C}\left(\frac{x}{y},\frac{\mu}{yP_3}\right) q\,(y,\mu)\,. \tag{5.17}$$

The matching kernel $\tilde{C}(\xi,\eta)$ is reported in Sec. 3.4.2, while the quasi-PDF for $P_3 = 1.38\,\text{GeV}$ is shown in Fig. 5.4. Having the quasi-PDF, the lattice matrix element can be computed through the inverse FT

$$h(z,\mu,P_3) = \int_{-\infty}^{\infty} dx\, e^{-izxP_3}\tilde{q}\,(x,\mu,P_3)\,. \tag{5.18}$$

As can be seen in Fig. 5.5, the matrix elements extracted from the NNPDF3.1 phenomenological unpolarized PDF show qualitatively the same behavior as that seen in the lattice QCD data. In particular, they decay to zero at sufficient large z/a with a trend that depends on the nucleon boost P_3. It is interesting to observe that one needs to boost to $P_3 = 2.75$ GeV so that the matrix element decay to zero for z/a-values larger than 10. Moreover, the z-dependent phase $\phi(z)$ behaves linearly within the range of z/a-values for which the real matrix elements is non-zero, as shown in Fig. 5.6. In order to measure the deviation from the linear behavior of the function $\phi(z)$, we present in Fig. 5.7 the difference between the argument of the matrix element computed at $P_3 = 0.83\,\text{GeV}$ and a linear fit performed in the interval $z/a \in [-11, 11]$. The obtained curve is compatible with zero, therefore we have elements to presume that the linearity of $\phi(z)$ is indeed an intrinsic property of the matrix element that can be exploited to design an efficient regression algorithm. We will provide more details on how we exploit this property in the next section.

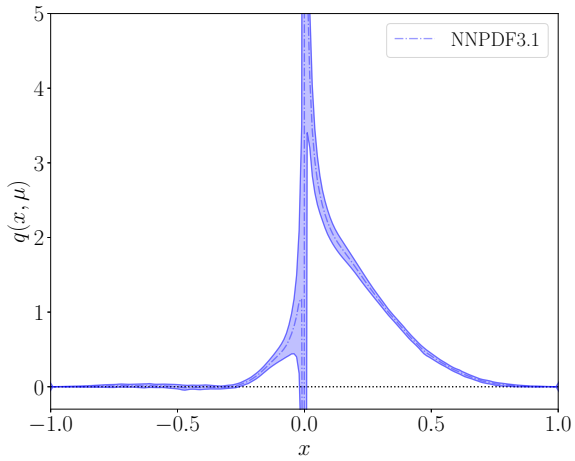


Figure 5.3: NNPDF3.1 *phenomenological determination of the unpolarized PDF.*

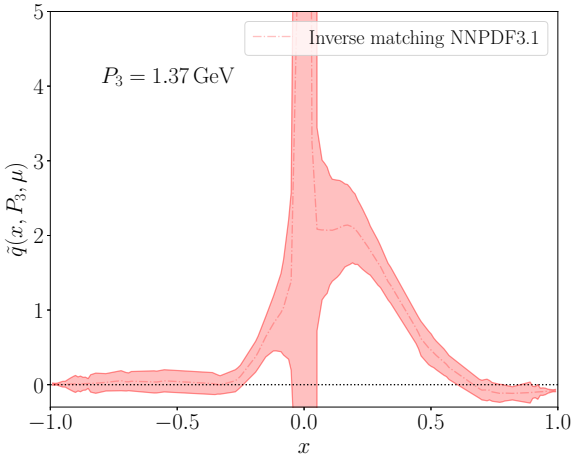


Figure 5.4: Quasi-PDF for $P_3 = 1.37\,\mathrm{GeV}$ *obtained through the inverse matching procedure of the* NNPDF3.1 *data.*

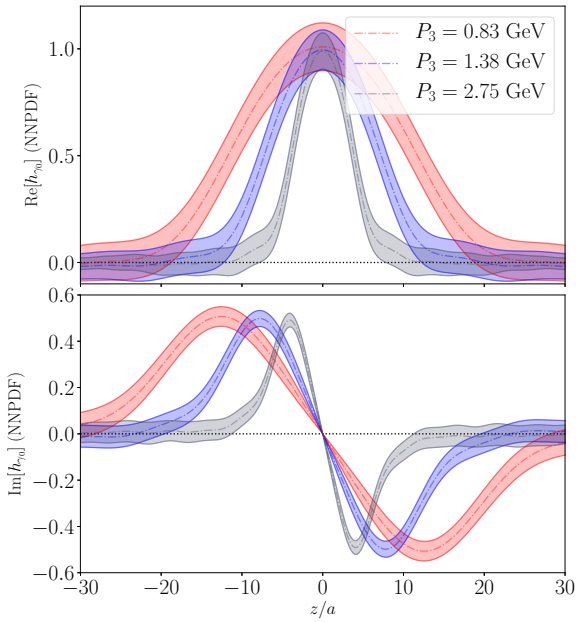


Figure 5.5: Matrix elements corresponding to the NNPDF3.1 phenomenological determination of the unpolarized PDF computed for three different values of the momentum P_3.

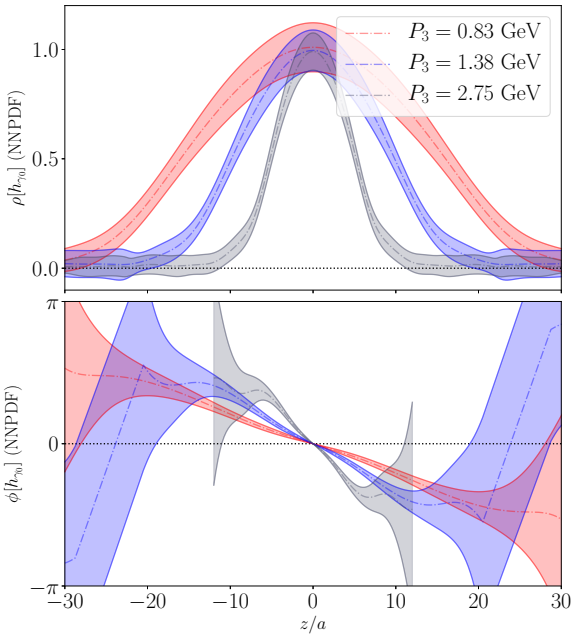


Figure 5.6: Modulus and argument of the matrix elements corresponding to the NNPDF3.1 phenomenological determination of the unpolarized PDF computed for three different values of the momentum P_3.

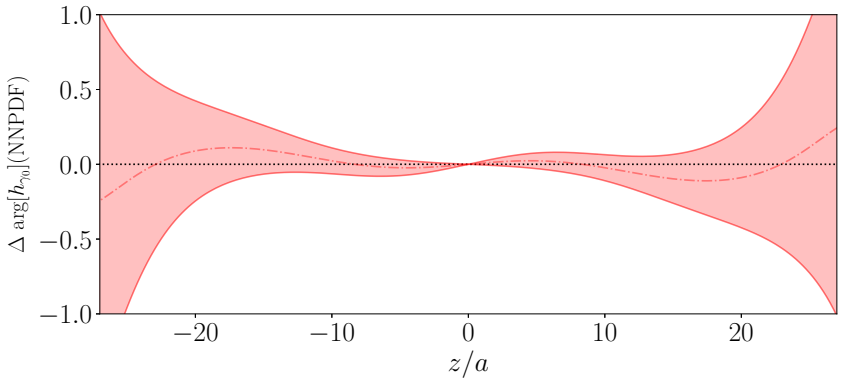


Figure 5.7: Deviation from the linear behavior of the argument of the matrix element corresponding to the NNPDF3.1 phenomenological determination of the unpolarized PDF computed for $P_3 = 0.83\,\mathrm{GeV}$.

5.2.4 Strategy for complex Hermitian data

The GPR described until now is defined on real-valued data. There are many possible ways to extend the procedure to complex data. The approach that we choose consists in performing two independent fits: a GPR for the absolute values and a minimum χ^2 linear regression for the complex argument.

Since our target functions are Hermitian, we can restrict the fit procedure to the positive semi axis and then use the Hermitian symmetry to obtain the results on the negative semi axis. With this strategy there is a reduction by half of the number of points used for the non-parametric regression, which improves the stability and the performance of the algorithm. Since the complex argument is an odd function of z, the linear regression is performed with the intercept fixed to zero.

Denoting by $\mu(z)$ the result of the GPR for the absolute value and θ the coefficient resulting from the linear regression, the surrogate model obtained for the renormalized matrix element in the positive semi axis reads:

$$h^{\mathrm{fit}}_{\mathrm{pos}}(z) = \mu(z)e^{i\theta z} \qquad \text{for } z \geq 0$$

The corresponding Hermitian function defined over the full real domain is then:

$$h^{\mathrm{fit}}(z) = \mu(|z|)e^{i\theta z}. \tag{5.19}$$

A side effect of this procedure is the loss of continuity of the derivative at $z/a = 0$. However this is not an issue for our procedure because no subsequent passages require this property to hold. A nice property of this parametrization consists of the possibility to evaluate analytically the Fourier transform of Eq. (5.19).

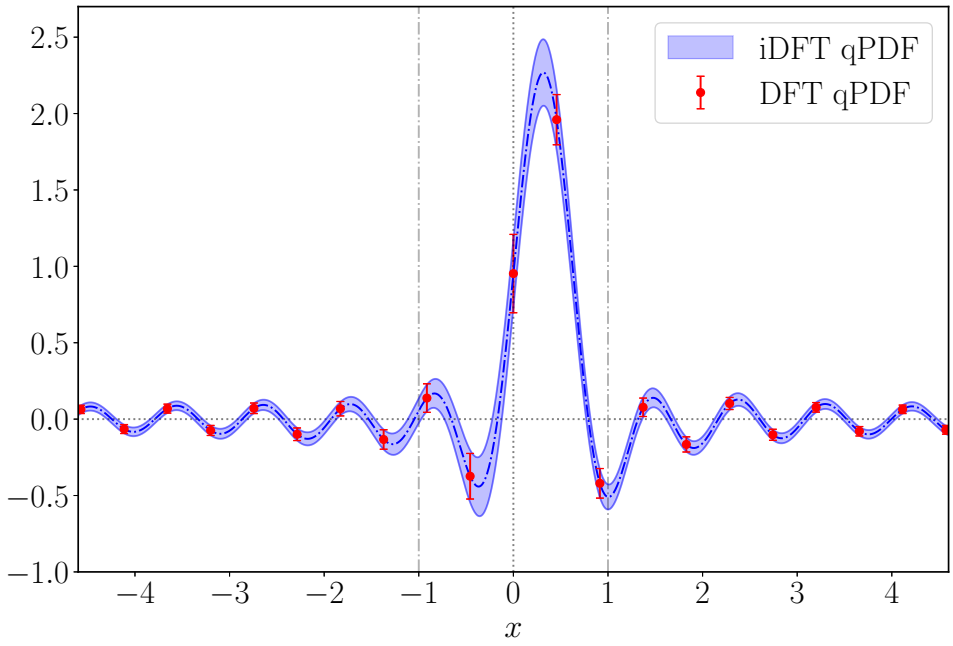


Figure 5.8: Comparison between DFT and iDFT corresponding to the matrix element computed on the cA2.09.48 ensemble, the parameters of which are given in Table 5.1.

5.3 Discrete Fourier Transform

The quasi-PDF given in Eq. (3.70) is defined as the Fourier transform of the renormalized lattice matrix element of the unpolarized operator. As mentioned in Sec. 3.4, the physical PDF can be obtained from the quasi-PDF by applying a perturbative matching procedure. In particular, at one-loop order, the integral of Eq. (3.71) consists of a convolution of the quasi-PDF with a function possessing a singularity at $x/\xi = 1$. Therefore, to compute the light-cone PDF it is crucial to obtain a trustworthy reconstruction of the quasi-PDF for continuous x-values and, in particular, in the region $x \in [-1, 1]$. In what follows we give more details on the issues that arise in the reconstruction of a continuous quasi-PDF originating from the renormalized matrix element. In particular, we show the difficulties in accessing the small-x region having available a limited amount of lattice data, corresponding to $\mathcal{O}(10)$ values of the Wilson line length z.

The problem of reconstructing a continuous momentum-space function starting from a discrete and finite set of momentum-space points is mathematically ill-posed. Indeed, a well-defined transformation is the one that maps $N-$points discrete position-space sequence $h(z, P_3, \mu)$ defined in the interval $z \in [-z_{\max}, z_{\max}]$ into a N-points discrete momentum-space sequence $\tilde{q}(\omega_k, \mu)$. Such a transformation is the Discrete Fourier Transform (DFT), that can be defined as follows

$$\tilde{q}(\omega_k, \mu) = 2P_3 \sum_{z=-z_{\max}}^{z_{\max}} \frac{a}{4\pi} e^{i\omega_k z} h(z, P_3, \mu), \tag{5.20}$$

with

$$\omega_k = \frac{2\pi}{N}k, \;\; k \in [-k_{\max}, k_{\max}]\,, \tag{5.21}$$

where $N = (2z_{\max}/a + 1)$ and $ak_{\max} = z_{\max}/a$. However, the DFT assumes periodicity of the matrix element $h(z, P_3)$ as a function of z

$$h(z, P_3) = h(z + T, P_3), \tag{5.22}$$

with a period T that is a function of z_{max} and the DFT frequencies given in Eq. (5.21) have a meaning only under this property. However, the matrix element is not periodic and Eq. (5.22) does not hold. In particular, this implies that we cannot attribute to the DFT frequencies of Eq. (5.21) any special meaning. However, as it will become clear later on, it is still interesting to examine the DFT of the renormalized matrix element. In Fig. (5.8) we show the DFT corresponding to the matrix element of the unpolarized operator (introduced in Sec. 5.2.2) with a boost $P_3 = 10\pi/L$ and $z_{\max}/a = 10$.
As already pointed out, to compute the physical PDF a continuous momentum-space function $q(x, P_3)$ is required. A possible solution is to compute the sum in the right-hand side of Eq. (5.1) for continuous x values. The resulting transformation can be written as follows[1]

$$\tilde{q}(\omega, P_3) = 2P_3 \sum_{z=-\infty}^{\infty} \frac{a}{4\pi} e^{ixP_3 z} h(z, P_3), \tag{5.23}$$

with $x \in \mathbb{R}$. Such a transformation maps an N-point discrete position-space sequence $h(z, P_3)$ defined on the interval $z \in [-z_{\max}, z_{\max}]$ into a continuous momentum-space function $\tilde{q}(x, P_3)$. However, since the matrix element is available up to $z_{\max}$, the definition in Eq. (5.23) cannot be applied. What we compute instead is the convolution of the matrix element with a function $\chi_I(z)$ defined by

$$\chi_I(z) = \begin{cases} 1 & \text{if } |z| < z_{\max} \\ 0 & \text{if } |z| > z_{\max}, \end{cases} \tag{5.24}$$

that restricts z in the interval $I = [-z_{\max}, z_{\max}]$. This procedure is equivalent to computing the sum in Eq. (5.23) within the interval $[-z_{\max}, z_{\max}]$, where the matrix element is known. In what follows, we will show that the transform obtained through this procedure is an interpolation of the DFT and, for this reason, will be referred to as interpolated-DFT (iDFT). The results of applying iDFT to the matrix element as compared to the DFT are shown in Fig. (5.8). The iDFT is the transformation that is being employed to compute the quasi-PDF starting from the renormalized matrix element and, for this reason, it is interesting to analyze in depth the features characterizing this quantity.

[1] In signal processing, this kind of transformation is referred to as the Discrete-Time Fourier Transform (DTFT).

The matrix element can be expressed as the inverse DFT

$$h(z, P_3) = \frac{4\pi}{2aP_3} \frac{1}{2N+1} \sum_{k=-k_{\max}}^{k_{\max}} e^{-i\omega_k z} \tilde{q}(\omega_k, P_3), \tag{5.25}$$

where $\tilde{q}(\omega_k, P_3)$ is the k-th DFT coefficient, and $k_{\max} = z_{\max}/a$. Substituting Eq. (5.25) into Eq. (5.23), and computing the sum over z we get

$$\tilde{q}(\omega, P_3) = \sum_{k=-k_{\max}}^{k_{\max}} \tilde{q}(\omega_k, P_3) D_n(\Delta\omega_k), \tag{5.26}$$

where $\Delta\omega_k = \omega - \omega_k$ and

$$D_n(x) = \frac{1}{n} \frac{\sin(x\, n/2)}{\sin(x/2)} \tag{5.27}$$

is the so-called Dirichlet kernel. In particular, in the limit $x \to 0$, the Dirichlet kernel converges to 1, thus the iDFT is equivalent to the DFT for $\omega = \omega_k,\ k \in [-k_{\max}, k_{\max}]$. In summary, given an $N-$point discrete position space sequence $h(z, P_3, \mu)$, it is not possible to define an appropriate discrete transformation returning an N-point discrete momentum-space sequence $\tilde{q}(\omega_k, \mu)$. The most straightforward solution is to use analytical continuation for real values of the variable x of the DFT. In particular, the iDFT is the convolution of a step function $\chi_I(z)$ defined in the interval $[-z_{\max}, z_{\max}]$ with the matrix element. This is equivalent to setting the value of the matrix element to zero for $|z| > z_{\max}$ and, as it will be shown in Sec. 5.3.1, this procedure introduces nonphysical oscillations in the quasi-PDF.

5.3.1 Testing BGFT on a mock dataset

To gain an insight on the artifacts that may be introduced by the discrete Fourier transform and on the effectiveness of the proposed method, we produced a mock dataset that mimics the behavior of the matrix element shown in Fig. 5.1. Given the rescaled Gaussian

$$g(x; \mu, \sigma, c) = c\, \exp -(x-\mu)^2/2\sigma^2 \tag{5.28}$$

its inverse quasi-PDF transform of Eq. (A.2b) reads

$$\mathcal{T}^{-1}[g(x)](z) = c\sqrt{2\pi}\sigma e^{-\sigma^2 P_3^2 z^2/2 + i\mu P_3 z} \tag{5.29}$$

To be consistent with the results reported in Sec. 5.4, we choose $P = 5$ and $L = 48$. The complex function of Eq. (5.29) is then sampled in the interval $z \in [-25, 25]$, $z \in \mathbb{N}$, with $c = 2.22$, $\mu = 0.315$, $\sigma = 0.230$. The employed coefficients μ, σ and c correspond to the best fit performed with the function of Eq. (5.28) on the data obtained by evaluating the iDFT using the discrete set of grid points. The resulting fit is shown in Fig. 5.9. In order to mimic the behavior of lattice data, we generated $N = 100$ numbers at each fixed

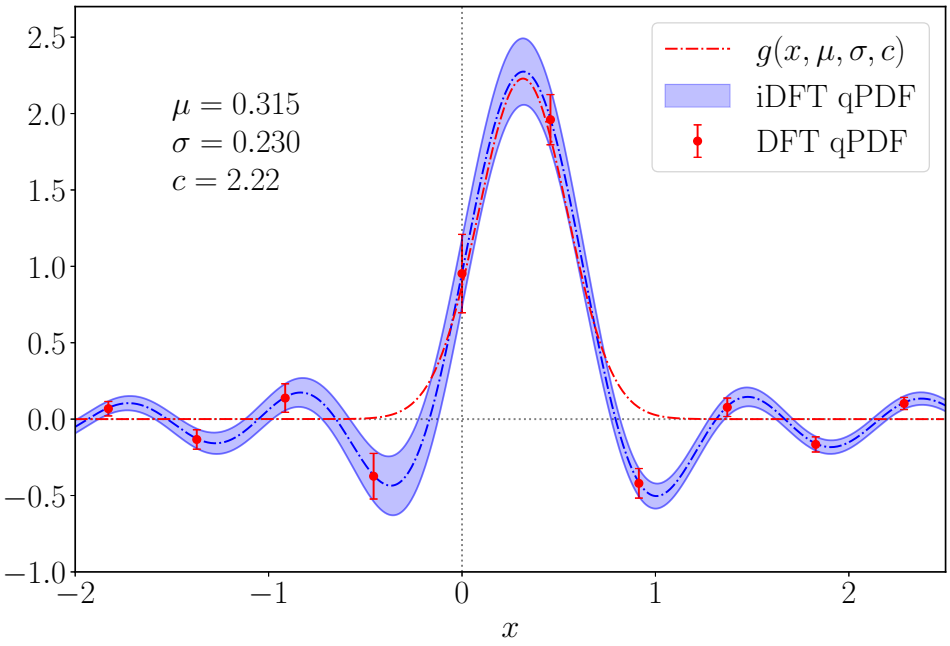


Figure 5.9: The function $g(x; \mu, \sigma, c)$ (line dashed line) fitted on the iDFT data extracted using the grid of points (red circles).

integer z from a Gaussian distribution centered in $\mathcal{T}^{-1}[g(x)](z)$ with variance increasing linearly with z^2, obtaining a sample of N mock matrix element. The average and the Jackknife standard deviation of this sample are shown in Fig. 5.10.

The dependence of the discrete Fourier transform on the cutoff $z_{\max}$ is investigated. In Fig. 5.12 we show the DFT computed with four different values of $z_{\max}$, together with the shifted Gaussian $g(x, \mu, \sigma, c)$ from which we generate the mock dataset. In particular, setting $z_{\max} = 5$, a huge bias is introduced in the DFT and big oscillations afflict the final result. Moreover, the bias becomes negligible only with $z_{\max} = 18$, where the iDFT coincides with the analytical FT within the error. The observed behavior of the iDFT is due to the fact that, if the z cutoff is too small, then the frequency resolution of the DFT is not fine enough to capture the behavior of the analytical FT in the small x region. Considering the case of $z_{\max} = 7$, we apply the regression described in Sec. 5.1 to the mock data set, using a constant zero function as the prior mean function.

The results are illustrated in Fig. 5.11. The conclusion is that, given the mock matrix element up to $z_{\max} = 7$ and their asymptotic behavior specified by the prior as a zero constant function, the nonparametric regression is able to reproduce the data outside the fit range. As a consequence, the results, after applying the Bayes-Gauss-Fourier transform shown in Fig. 5.13, are compatible with the analytical transform of the mock matrix-element within error.

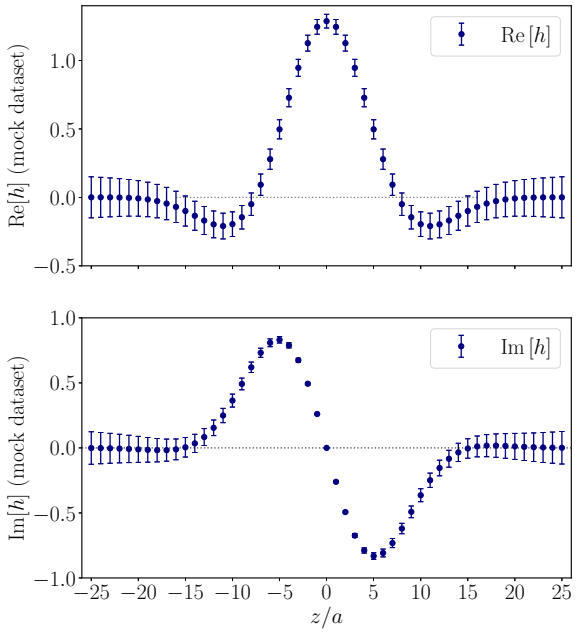


Figure 5.10: Mock matrix element generated according to Eq. (5.29).

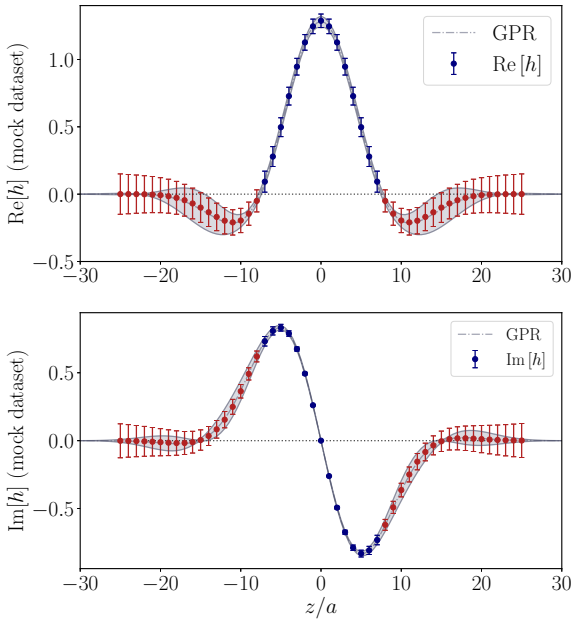


Figure 5.11: nonparametric regression performed on the mock data set. Only data up to $z_{\max} = 7$ (blue circles) are included in the fit.

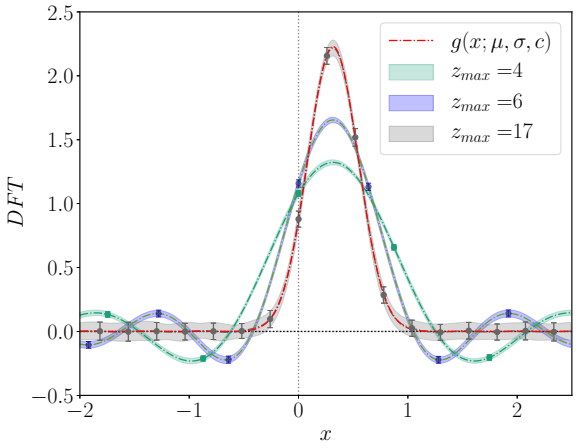


Figure 5.12: Dependence of the DFT on the cutoff compared to the shifted Gaussian $g(x, \mu, \sigma, c)$ from which we generated the mock dataset.

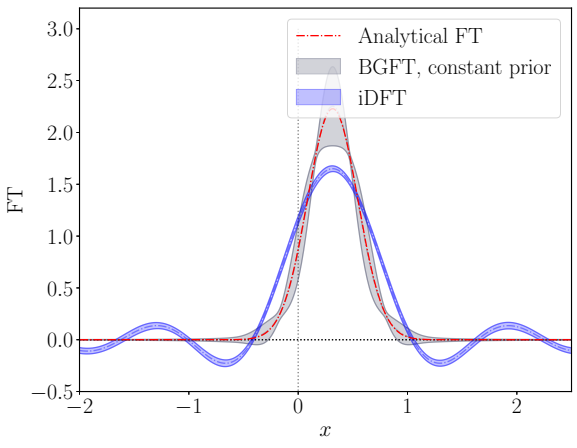


Figure 5.13: BGF transform performed on the mock data set and comparison with the analytical FT (red dashed line) and the DFT (blue band).

5.4 Application of BGFT to the renormalized lattice matrix element

We apply the method detailed in Sec. 5.1 on the case of interest, namely the renormalized nucleon matrix element of the unpolarized operator, for which the data are given in Sec. 5.2.2. As pointed out in Sec. 5.2.1, in the region with high density of data-points, the posterior mean is strongly dependent on the renormalized matrix element rather than on the prior mean.

However, as discussed in Sec. 5.2.2, the matrix elements at $z_{\max}/a \gtrsim 7$ carry large statistical errors, and, for this reason, they are excluded from the regression. As a consequence of this, the outcome of the posterior mean become increasingly more dependent on the prior mean function starting from $|z/a| > 7$. The asymptotic behavior is almost entirely determined by the choice of the prior mean function. Therefore, any theoretical consideration on the asymptotic behavior should be incorporated into the prior mean function.

We choose to test our method with two different prior mean functions, namely the uniformly zero function and the function of Eq. (5.29) obtained through applying the inverse transform to the Gaussian fit of the DFT. Both of the chosen priors satisfy the requirement of being asymptotically zero. Using different prior distributions is a method to cross-check that the final conclusions are independent of the prior choice.

Let us here summarize the key steps of the procedure:

1. At a fixed z-value we rewrite the complex number $h(z) = \Re h(z) + i\Im h(z)$ in the polar complex plane as

   $$h(z) = \rho(z)e^{i\phi},$$

   with

   $$\begin{aligned}\rho(z) &= \sqrt{\Re h(z)^2 + \Im h(z)^2} \\ \phi(z) &= \arg(h(z)) \\ &= \arctan2\left(\Im h(z), \Re h(z)\right).\end{aligned}$$

   The function $\arctan2(y, x)$ is defined in Sec. 5.2.2

2. As pointed out in Sec. 5.2.2, the function $\rho(z)$ is asymptotically zero, while $\phi(z)$ can be taken as a linear function of z. After choosing a prior mean function, we perform a non parametric regression of the function $\rho(z)$, while a linear fit is sufficient to reproduce $\phi(z)$, as shown in Fig. 5.14;

3. In order to check the result of the fit to the renormalized matrix element, we can go back to the Cartesian coordinates, as shown in Fig. 5.15;

4. Employ the formula in Eq. (A.4) to compute the quasi-PDF.

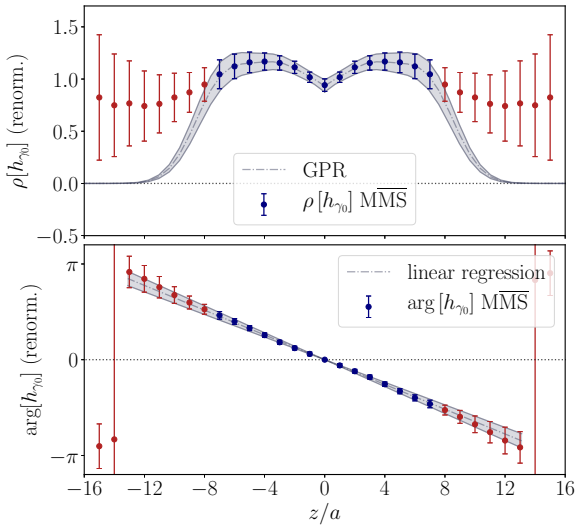


Figure 5.14: Lattice data for the matrix element rotated in the polar complex plane as described in Sec. 5.4. Only the points shown with the blue circles are considered in the fit.

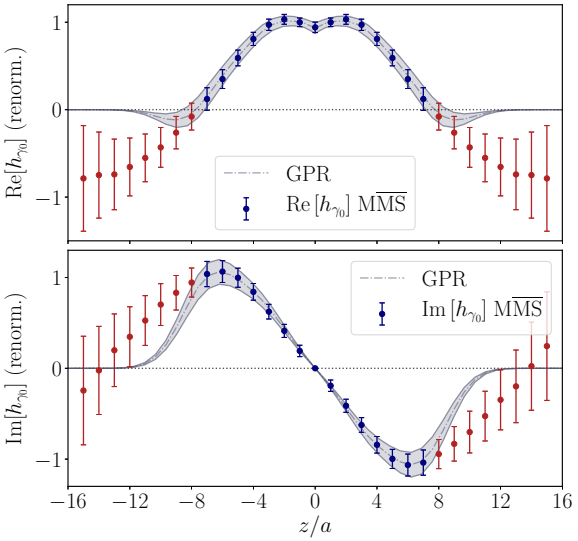


Figure 5.15: Lattice data for the matrix element together with the non parametric regression. Only the points shown with the blue circles are considered in the fit.

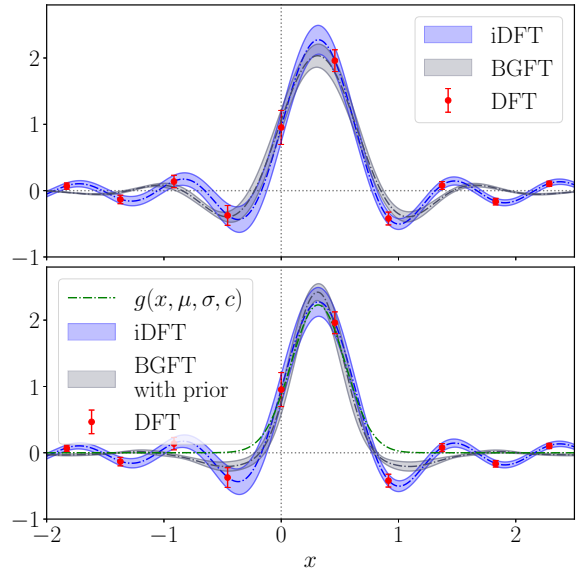


Figure 5.16: Unpolarized quasi-PDF computed with discrete Fourier transform (red circles), iDFT (blue band) and Gauss-Bayes-Fourier transform (grey band) with zero mean prior (upper panel) and with mean prior $g(x, \mu, \sigma, c)$ (lower panel).

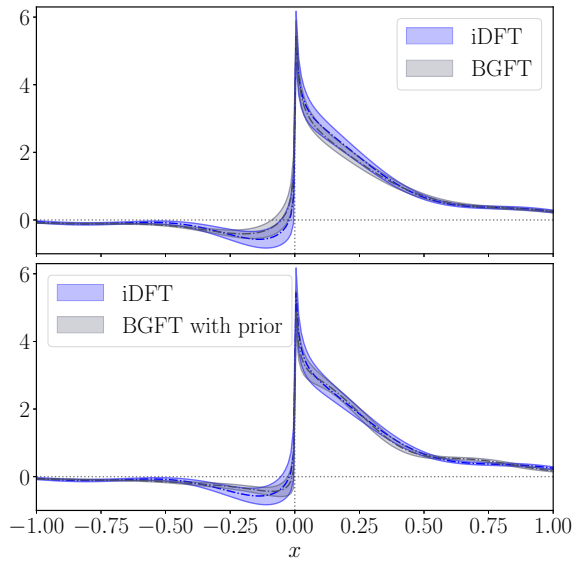


Figure 5.17: Comparison of light-cone PDFs obtained via iDFT reconstruction and BGFT with zero mean prior (upper panel) and with mean prior $g(x, \mu, \sigma, c)$ (lower panel).

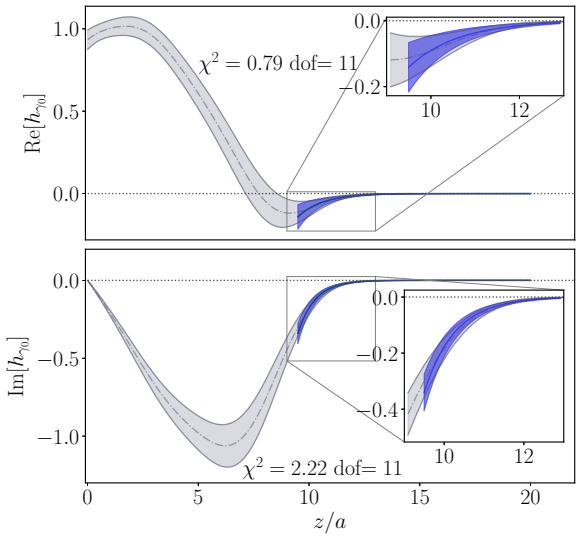


Figure 5.18: Fit (blue bands) with an exponential model of the tails of the nonparametric regression curves (grey bands) corresponding to the real (upper panel) and imaginary (lower panel) part of the matrix elements.

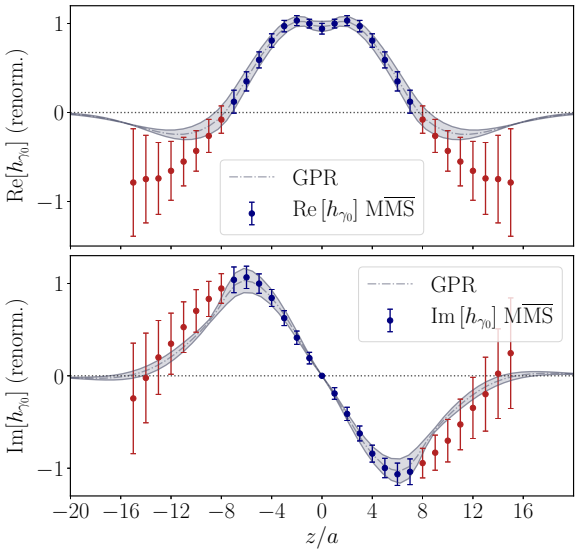


Figure 5.19: Lattice data for the matrix element of the unpolarized operator together with the non parametric regression performed with nonzero prior. Only points shown with blue circles are in the fit.

In the upper panel of Fig. 5.16 we compare the iDFT quasi-PDF to the BGFT quasi-PDF. While in the physical region $x \in [-1, 1]$ the two results are compatible, for larger $|x|$ nonphysical oscillations due to the periodicity of the discrete FT are strongly suppressed. However, the physical meaning of quasi-PDF can be made explicit only after having performed the matching procedure. In Fig. 5.17 we display the light-cone PDF reconstructions obtained via iDFT and BGFT. As can be seen, although the the nonphysical oscillations in the quasi-PDF are suppressed, in the physical PDF the effect is small. The nonphysical negative PDF in the antiquark region $x \sim -0.1$ remains, as well as a mild oscillatory behavior in the large $|x|$ region. This means that this behavior does not appear to be caused by the cutoff in z and the discrete FT.

Finally, it is interesting to investigate how the nonparametric regression curves for the real and imaginary parts of the matrix element go to zero. In particular, both the tail of the real and imaginary parts of the nonparametric regression curves can be modeled by the function

$$s(z) = -a^2 \exp -zb. \tag{5.30}$$

The parameters a and b have been computed by minimizing χ^2. We obtained $a = 25(16)$, $b = 9.4(9)$ fm^{-1} for the real part and $a = 270(80)$, $b = 13.8(7)$ fm^{-1} for the imaginary part. In Fig. 5.18 , we show the result of the fit procedure performed with the exponential model of Eq.(5.30).

As previously stated, we use as an alternative prior mean the function of Eq. (5.29) with $\mu = 0.315$, $\sigma = 0.23$ and $c = 2.22$ in order to cross-check our results. The outcome of the nonparametric regression with nonzero mean prior is shown in Fig. 5.19, while the lower panel of Fig. 5.16 shows the FT of this function together with the resulting BGFT.

As pointed out in Sec. 5.2.1, the choice of the prior mean function modifies the result of the GPR in the region where there is a low density of data points. In this specific case, it slightly modifies the decay rate in the large-z region, further reducing the amplitude of the remnant oscillations present in the BGFT. However, the effect of the prior mean is not observable in the light-cone PDF that is still compatible with the reconstruction obtained with iDFT, as shown in the lower panel of Fig. 5.17.

The choice of the prior within the BGFT approach as well as the choice of the cutoff z_{max} introduce systematic uncertainties. In this work, we show that the differences between the light-cone PDFs obtained with two different priors $\mu_P(z)$ are negligible, suggesting that the systematic effect due to the choice of the model for the matrix elements outside the fit region does not lead to large effects. This finding suggests that the presence of the nonphysical negative values in the light-cone PDF for $x < 0$ cannot be ascribed to the discrete Fourier transform, or at least, this cannot be the sole cause of this behavior.

Chapter 6

Flavor decomposition of the nucleon unpolarized, helicity and transversity parton distribution functions

In this chapter, we present the results of the study of Ref. [47] on the quark unpolarized, helicity and transversity parton distributions functions of the nucleon. We use the quasi-parton distribution approach within the lattice QCD framework and perform the computation using an ensemble of twisted mass fermions with the strange and charm quark masses tuned to approximately their physical values and light quark masses giving pion mass of 260 MeV. Parton distribution functions can be obtained from a number of scattering processes and have a wide kinematical coverage (see, e.g., Ref. [60,66]) and the individual-flavor contributions are also studied in phenomenological fits for the collinear PDFs.

The present calculation is motivated by the fact that not all PDFs are well-constrained from global analyses. The number of available experimental data sets in the case of the transversity is less by $\mathcal{O}(10)$ compared to the helicity, and $\mathcal{O}(100)$ compared to the unpolarized PDFs. In addition, isolating the strange-quark PDF from the down-quark PDFs can be challenging, as most of the high-energy processes cannot differentiate between the two flavors. For example, there is a disagreement on the sign of the strange-quark helicity PDF, $\Delta s(x) + \Delta\bar{s}(x)$, from analysis of polarized inclusive deep inelastic scattering [123, 124] and global analyses of inclusive and semi-inclusive deep inelastic scattering data sets [65, 125–127]. The large uncertainties in the strange PDFs have an effect on other quantities, such as the W-boson mass and the determination of the CKM matrix element V_{cs} [128, 129]. Therefore, calculations of the individual-quark PDFs from lattice QCD can, eventually, be used as input in analysis requiring knowledge of PDFs.

The flavor decomposition of proton charges and form factors had been under investigation in the last few years with calculations of disconnected diagrams using ensembles at or near the physical values for the quark masses(*physical point*) [130–143]. Such a success in lattice calculations of hadron structure is partly due to available computational resources, but also due to novel methodologies to extract the individual quark Mellin moments and form factors. A notable example is the hierarchical probing [109], which improves the signal significantly. In this work, we extend hierarchical probing to non-local operators. Such an approach was shown to be successful in our first calculation on

the helicity PDF [46].

Matrix elements of non-local operators are of great interest in recent years. These can be related to light-cone PDFs through a factorization and matching procedure. Methods to access the x dependence of PDFs, such as the quasi-PDFs [8, 96], pseudo-ITDs [144], and current-current correlators [145, 146] are now well established. Progress in terms of the renormalizability, renormalization prescription, and factorization of light-cone PDFs has been made, alleviating major sources of systematic uncertainties. For their application in lattice QCD see Refs. [10,11,14,15,17–21,24,26,31,45,46,101,103,104,117,147–168] and the recent reviews of Ref. [169–171]. Results from lattice QCD simulations on the x-dependence of PDFs are very promising, and therefore, their flavor decomposition is the extension of these investigations. The matching for the singlet case, as well as the mixing with the gluon PDFs have been recently addressed [153, 172]. In Ref. [46] we presented the first calculation for the helicity PDFs including disconnected contributions and the flavor decomposition for the up-, down- and strange-quark PDFs. Here we extend the calculation to the three types of collinear PDFs, that is the unpolarized, helicity and transversity PDFs. While such calculations are becoming feasible, there are a number of computational challenges before taming the statistical uncertainties. To date, calculations of disconnected contributions for matrix elements of non-local operators at the physical point do not exist.

In Sec. 6.1 we will provide an overview about the operators we evaluated on the lattice, including connected and disconnected diagrams. Next, in Secs. 6.1.1-6.1.2 a description of the perturbative renormalization and matching procedures employed in this study is provided. The lattice setup and the numerical methods allowing to evaluate the disconnected diagrams have been reported in Secs. 6.2.1. In Secs. 6.2.2-6.2.1 we provide the results of the excited states contamination analysis respectively on the two- and disconnected three-point functions. For the latter, we show as well the momentum dependence of the unpolarized, helicity and transversity operators. An analogous analysis for the connected diagram is reported in Sec. 6.4. In Sec. 6.5 we provide the results of the computation of the nucleon charges and, finally, in Sec. 6.6 the parton distribution functions resulting from our analysis.

6.1 Nucleon bare matrix elements

The main component of this study is the calculation of the nucleon matrix elements of non-local operators, that is

$$h_{\Gamma}^{\Upsilon}(z, P_3) = \langle N(P_3)|\overline{\psi}(z)\Upsilon\,\Gamma\,W(z)\psi(0)|N(P_3)\rangle\,, \tag{6.1}$$

where $|N(P_3)\rangle$ is the nucleon state with momentum boost along the z-direction, i.e. $\vec{P} = (0, 0, P_3)$. The fermionic field $\psi(x) \equiv \psi(\vec{x}, t)$ can be either the light quark doublet $\psi \equiv (u, d)^T$ or the strange quark field $\psi(x) \equiv s(x)$. The Wilson line $W(z)$ is constructed

in the direction parallel to the nucleon boost $\vec{P}$ and extends from zero length to up to half of the lattice, $L/2$, in both positive and negative directions. The Dirac structure of the operator, Γ, acts in spin space and depends on the type of the collinear PDF under study. Without loss of generality, one can take that the momentum boost is in the z-direction (k).

The matrix elements are computed from the ratio of three- and two-point functions defined as

$$\begin{aligned} C_{2pt}(\vec{P}; t_s, 0) &= \mathcal{P}_{\alpha\beta} \sum_{\vec{x}} e^{-i\vec{P}\cdot\vec{x}} \langle \Omega | N_\alpha(\vec{x}, t_s) \overline{N}_\beta(\vec{0}, 0) | \Omega \rangle \\ C_{3pt}(\vec{P}; t_{\text{ins}}; t_s, 0) &= \tilde{\mathcal{P}}_{\alpha\beta} \sum_{\vec{x},\vec{y}} e^{-i\vec{P}\cdot\vec{x}} \langle \Omega | N_\alpha(\vec{x}, t_s) \mathcal{O}(\vec{y}, t_{\text{ins}}; z) \overline{N}_\beta(\vec{0}, 0) | \Omega \rangle \,, \end{aligned} \tag{6.2}$$

where t_s is the source-sink separation, and t_{ins} the insertion time of the three-point function. We use the proton interpolating field $N_\alpha = \varepsilon^{abc} u^a_\alpha(x) \left(d^{bT}(x) \mathcal{C}\gamma^5 u^c(x) \right)$ with $\mathcal{C} = \gamma^0\gamma^2$. The three-point function projector depends on the operator under study and can be found in Table 6.1 and for the two-point function we use $\mathcal{P} = (1 \pm \gamma_0)/2$. To increase the number of measurements for the disconnected diagrams, we average the three- and two- point functions over plus and minus parity projectors.

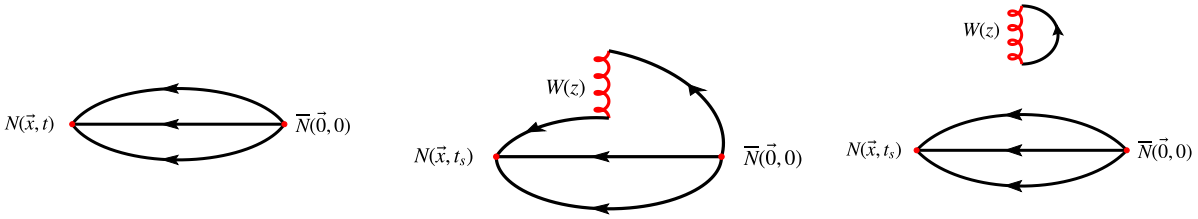
PDF	Γ	$\tilde{\mathcal{P}}_{\alpha\beta}$
Unpolarized	γ^0	$\frac{\mathbf{1}\pm\gamma^0}{2}$
Helicity	$\gamma^5\gamma^3$	$i\gamma^3\gamma^5 \left(\frac{\mathbf{1}\pm\gamma^0}{2}\right)$
Transversity	σ_{3j}	$i\gamma^5\gamma^i \left(\frac{\mathbf{1}\pm\gamma^0}{2}\right),\ i \neq j$

Table 6.1: List of parity projectors and insertions for each collinear PDFs. The nucleon momentum boost is assumed to be in the z-direction, $P = (0, 0, P_3)$.

The operator $\mathcal{O}$ is defined as

$$\mathcal{O}(\vec{y}, t_{\text{ins}}; z) = \bar{\psi}(\vec{y} + z\hat{z}, t_{\text{ins}}) \Upsilon \, \Gamma \, W(\vec{y} + z\hat{z}, \vec{y}) \psi(\vec{y}, \tau) \,, \tag{6.3}$$

and is inserted in the three-point function of Eq. (6.2). The Wick contractions lead to two topologically different diagrams, as shown in Fig. 6.1b and 6.1c. In Fig. 6.1a, we also show pictorially the two-point function of Eq. (6.2). For the case that the fermionic field in Eq. (6.3) is $\psi = (u, d)^T$ and $\Upsilon = \tau^3$, we obtain the matrix elements for the isovector distribution $u - d$, which receive contribution from the connected diagram only (Fig. 6.1b). However, in the case where $\Upsilon = \mathbb{1}$, the three-point function takes contributions from both connected and disconnected diagrams. For the nucleon, the strange-quark contribution comes exclusively from the disconnected diagram. We emphasize that, disconnected contributions have a considerably smaller signal-to-noise ratio compared to the connected ones and their evaluation requires the use of stochastic and gauge noise



(a) Nucleon two-point function. *(b) Nucleon connected three-point function.* *(c) Nucleon disconnected three-point function.*

Figure 6.1: Schematic representation of the two- and three-point functions. The time $t_s\,(t)$ indicates the source-sink separation for the three(two)-point function. The solid lines correspond to quark propagators, while the curly lines represent the Wilson lines of length z.

reduction techniques described in detail in Sec. 6.2.1.

The ratio of three- over two-point functions becomes,

$$\frac{\langle C_{3pt}(P_3;t_{\text{ins}};t_s,0\rangle}{\langle C_{2pt}(P_3;t_s,0)\rangle} \overset{0\ll t_{\text{ins}}\ll t_s}{=} \left(h_\Gamma^\Upsilon\right)^{\text{bare}}(z,P_3)\ , \tag{6.4}$$

and is used to obtain the matrix elements of Eq. (6.1). This relation is meaningful for the ground-state contribution. To isolate the latter, we apply constant (plateau) fits in a range where the operator insertion time is large enough, and away from the source and sink. Besides the plateau fit method, we employ different techniques allowing the extraction of the nucleon matrix element, as described in Sec. 6.3.1.

6.1.1 Non-perturbative renormalization

The bare matrix elements of Eq. ((6.1)) must be renormalized in coordinate space prior to obtaining the quasi-PDFs, which are defined in the momentum space (see Sec. 3.4). The renormalization of both the non-singlet and singlet quantities is multiplicative. In this work, we use the non-singlet renormalization function, as the difference with the singlet is expected to be small [173], which was demonstrated numerically for local operators [137, 139]. The mixing between the unpolarized and helicity singlet-quark PDFs with the gluon PDFs arises at the matching level because there is no additional non-local ultraviolet divergence in the quasi-PDF [153, 172, 174].

To renormalize the matrix elements we apply the regularization independent (RI′) scheme, and use the momentum source method [175] that offers high statistical accuracy. More details on the setup can be found in Refs. [101, 176]. In Refs. [102, 103], we proposed an extension of the renormalization prescription to include non-local operators, which we also follow in this work. The conditions for the renormalization functions of the non-local

operator, Z_Γ, and the quark field, Z_q, are

$$\left(Z_q^{\rm impr}(\mu_0)\right)^{-1}\, Z_\Gamma(z,\mu_0)\,{\rm Tr}\left[\mathcal{V}_\Gamma(p,z)\,\not{p}\right]\Big|_{p^2=\mu_0^2} = {\rm Tr}\left[\mathcal{V}_\Gamma^{\rm Born}(p,z)\,\not{p}\right]\Big|_{p^2=\mu_0^2}\,, \tag{6.5}$$

$$Z_q^{\rm impr}(\mu_0) = \left(\frac{1}{12}{\rm Tr}\left[(S(p))^{-1}\,S^{\rm Born}(p)\right] - dZ_q^{\infty}(p)\right)\Big|_{p^2=\mu_0^2}\,. \tag{6.6}$$

$\mathcal{V}(p,z)$ ($S(p)$) is the amputated vertex function of the operator (fermion propagator) and $S^{\rm Born}(p)$ is the tree-level of the propagator. These conditions are applied at each value of z separately. We improve Z_q, and consequently Z_Γ, by subtracting lattice artifacts calculated to one-loop level in perturbation theory and to all orders in the lattice spacing, $dZ_q^{\infty}(p)$. The details of the calculation can be found in Ref. [176]. The RI-type schemes are mass-independent, and therefore Z_Γ is calculated at several ensembles with different values for the quark masses. Eventually, a chiral extrapolation is applied to remove residual artifacts related to the quark mass. Here we use five ensembles that are generated at different pion masses in the range of 350 MeV - 520 MeV with a lattice volume of $24^3 \times 48$. For the chiral extrapolation to be meaningful, all quark masses should be degenerate. Therefore, we use $N_f = 4$ ensembles generated by the Extended Twisted Mass collaboration (ETMC) that are dedicated to the renormalization program. These ensembles have the same lattice spacing and action parameters as the $N_f = 2+1+1$ ensemble used for the production of the nucleon matrix elements.

Z_Γ depends on the RI renormalization scale μ_0, and it will be converted to $\overline{\rm MS}$ and evolved at a scale of choice. To reliably perform this procedure we use several values of μ_0, and the conversion and evolution formulas is applied on Z_Γ obtained at each scale. We choose the initial scale μ_0 such that discretization effects are small [176]. In particular, the 4-vector momentum p, which is set equal to μ_0, has the same spatial components: $p = (p_0, p_1, p_1, p_1)$. The values of p_0 and p_1 are chosen such that the ratio $\frac{p^4}{(p^2)^2}$ is less than 0.28, as suggested in Ref. [177]. The values of $a\,\mu_0$ cover the range $[1, 5]$. For each μ_0 value, we apply a chiral extrapolation using the fit

$$Z_\Gamma^{\rm RI}(z,\mu_0,m_\pi) = Z_{\Gamma,0}^{\rm RI}(z,\mu_0) + m_\pi^2\, Z_{\Gamma,1}^{\rm RI}(z,\mu_0)\,, \tag{6.7}$$

to extract the mass-independent $Z_{\Gamma,0}^{\rm RI}(z,\mu_0)$ at each value of the initial scale. $Z_{\Gamma,0}^{\rm RI}(z,\mu_0)$ is converted to the $\overline{\rm MS}$ scheme and evolved to μ=2 GeV (μ=$\sqrt{2}$ GeV) using the results of Ref. [102] for the unpolarized and helicity (transversity) PDFs. The conversion and evolution depends on both the initial scale μ_0 and the final scale in the $\overline{\rm MS}$. The appropriate expressions have been obtained to one-loop perturbation theory in dimensional regularization. Therefore, there is residual dependence on the initial scale μ_0, which is eliminated by taking the limit $(a\,\mu_0)^2 \to 0$ using a linear fit on the data in the region $(a\,\mu_0)^2 \in [1 - 2.6]$. The last step of the renormalization program is the conversion to a

modified $\overline{\text{MS}}$ scheme (M$\overline{\text{MS}}$), developed in Ref. [101], and is given by

$$\mathcal{Z}_{\Gamma,0}^{\text{M}\overline{\text{MS}}}(z,\bar{\mu}) = \mathcal{Z}_{\Gamma,0}^{\overline{\text{MS}}}(z,\bar{\mu})\,\mathcal{C}^{\overline{\text{MS}},\text{M}\overline{\text{MS}}}\,, \tag{6.8}$$

where

$$\begin{aligned}\mathcal{C}_{\Gamma}^{\overline{\text{MS}},\text{M}\overline{\text{MS}}}=1+\frac{C_F g^2}{16\pi^2}\Bigg[&e_\Gamma^{(1)}+e_\Gamma^{(2)}\ln\left(\frac{\bar{\mu}^2}{4\mu_F^2}\right)+e_\Gamma^{(3)}\left(\frac{i\pi\,|\mu_F z|}{2\mu_F z}-\ln(|\mu_F z|)+\right.\\ &\left.-\text{Ci}(\mu_F z)-i\text{Si}(\mu_F z)+\ln(\mu_F z)\right)\\ &+e_\Gamma^{(4)}\left(e^{i\mu_F z}(2\text{Ei}(-i\mu_F z)+i\pi\text{sgn}(\mu_F z)-\ln(-i\mu_F z)+\ln(i\mu_F z))\right)\Bigg]\,.\end{aligned} \tag{6.9}$$

This scheme was introduced to satisfy particle number conservation. In Ref. [101] we showed that the difference between $\overline{\text{MS}}$ and M$\overline{\text{MS}}$ is numerically very small, but brings the PDFs closer to the phenomenological ones. In Eq. ((6.9)) μ_F is the factorization scale set equal to the $\overline{\text{MS}}$ scale. Ci, Si, Ei and sgn are the special functions cosine integral, sine integral, exponential integral, and sign function, respectively. The coefficients $e_\Gamma^{(i)}$ depend on the operator: $\{e_\Gamma^{(1)}, e_\Gamma^{(2)}, e_\Gamma^{(3)}, e_\Gamma^{(4)}\}$ is $\{-5, -3, +3, -3/2\}$, $\{-7, -3, +3, -3/2\}$, $\{-4, -4, +4, -4/2\}$, for the vector, axial and tensor operator, respectively. Due to the presence of the Wilson line, both the matrix elements and renormalization functions are complex functions. As a consequence, a complex multiplication is required to extract the renormalized matrix element, that is

$$\begin{aligned}h_\Gamma^\Upsilon = Z_\Gamma^{\text{M}\overline{\text{MS}}}\cdot\left(h_\Gamma^\Upsilon\right)^{\text{bare}} &= \left(\text{Re}[Z_\Gamma^{\text{M}\overline{\text{MS}}}]\,\text{Re}[\left(h_\Gamma^\Upsilon\right)^{\text{bare}}]-\text{Im}[Z_\Gamma^{\text{M}\overline{\text{MS}}}]\,\text{Im}[\left(h_\Gamma^\Upsilon\right)^{\text{bare}}]\right)\\ &+ i\left(\text{Re}[Z_\Gamma^{\text{M}\overline{\text{MS}}}]\,\text{Im}[\left(h_\Gamma^\Upsilon\right)^{\text{bare}}]+\text{Im}[Z_\Gamma^{\text{M}\overline{\text{MS}}}]\,\text{Re}[\left(h_\Gamma^\Upsilon\right)^{\text{bare}}]\right)\end{aligned} \tag{6.10}$$

For simplicity in the notation, the dependence on z, P_3, scheme and scale is implied. As can be seen, the real (imaginary) part of the renormalized matrix elements are not simple multiples of the real (imaginary) part of the bare matrix element. Therefore, controlling systematic uncertainties in the renormalization is an important aspect of the calculation.

6.1.2 Quasi-PDFs and matching to light-cone PDFs

Quasi-PDFs are defined as the Fourier transform of the renormalized nucleon matrix elements in Eq. (6.4) with respect to the Wilson line length z

$$\tilde{q}(x,P_3) = \int_{-\infty}^{\infty}\frac{dz}{4\pi}e^{-ixzP_3}\,h_\Gamma^\Upsilon(z,P_3)\,. \tag{6.11}$$

Note that the renormalized matrix elements, h_Γ^Υ, depend on the scheme and scale, which also propagates to $\tilde{q}(x, P_3)$. For simplicity in the notation, this dependence is implied. As mentioned in the previous paragraph, the matrix elements are renormalized in the M$\overline{\text{MS}}$ scheme and evolved to 2 GeV ($\sqrt{2}$ GeV) for the unpolarized and helicity (transversity) PDFs.

On the lattice, we can only evaluate the matrix elements for discrete and finite values of z. Therefore, the integral of Eq. (6.11) is replaced by a discrete sum over a finite number of Wilson line lengths

$$\tilde{q}(x, P_3) = \sum_{-z_{\max}}^{z_{\max}} \frac{dz}{4\pi} e^{-ixzP_3} \, h_\Gamma^\Upsilon(z, P_3) \,, \tag{6.12}$$

where $dz/a = 1$. For the summation in Eq. ((6.12)) to accurately reproduce Eq. ((6.11)), both the real and imaginary parts of the matrix element should be zero beyond $z_{\max}$. Practically, this is not always possible due to the finite momentum boost and limited volume in the lattice formulation. The choice of the cutoff $z_{\max}$, which is anyway limited up to $L/2$, requires an extensive study. We note that systematic effects related to the reconstruction of the PDFs is operator dependent, as each matrix element may have different large-z behavior. We will show the results of such analysis in Sec. 6.6.2.

From the finite-momentum quasi-PDF, it is possible to obtain the light-cone parton distribution function (infinite momentum) through the so-called matching procedure. This is accomplished through a convolution of the quasi distribution with a kernel evaluated in continuum perturbation theory within the large momentum effective theory (LaMET) [170, 178]. The matching formula reads

$$q(x, \mu) = \int_{-\infty}^{\infty} \frac{d\xi}{|\xi|} \, C\left(\xi, \frac{\mu}{xP_3}\right) \widetilde{q}\left(\frac{x}{\xi}, \mu, P_3\right) \,, \tag{6.13}$$

and the factorization scale μ is chosen to be the same as the renormalization scale. The matching kernel C contains information on P_3 which, in principle, is eliminated in $q(x, \mu)$. However, there is residual P_3 due to the limitations in accessing large values of momentum from lattice QCD (see, e.g., Ref. [101]) and the matching kernel being available to limited order in perturbation theory. Most of the calculations of the matching kernel have been performed to one-loop level (see, e.g., Refs. [97–100, 145, 179–181]). Recently, the computation of the kernel C was extended to two loops [181–184]. In this study, we employ the kernel in the M$\overline{\text{MS}}$-scheme which is known at one-loop level [101]. This matching kernel relates the quasi-PDFs defined in the M$\overline{\text{MS}}$ scheme at some scale, to the light-cone PDFs in the $\overline{\text{MS}}$ at the same scale. For the unpolarized and helicity we choose a scale of 2 GeV, while for the transversity we choose $\sqrt{2}$ GeV.

To calculate the anti-quark distributions from $q(x)$, we exploit the crossing rela-

tions [53], that is

$$\bar{q}^f(x) = -q^f(-x), \quad \Delta\bar{q}^f(x) = \Delta q^f(-x), \quad \delta\bar{q}^f(x) = -\delta q^f(-x)\,. \tag{6.14}$$

6.2 Lattice setup

The computation is performed using one gauge ensemble of $N_f = 2+1+1$ clover-improved twisted mass fermions and the Iwasaki improved gluonic action [185] generated by ETMC [186]. The fermionic action of the light quarks in the "twisted basis" takes the form

$$S^l_{\rm tm}(\chi_l, \overline{\chi}_l, U) = a^4 \sum_x \overline{\chi}_l(x) \left[D_W[U] + i\mu_l\gamma_5\tau_3 + m_l + \frac{i}{4} c_{\rm SW}\sigma^{\mu\nu}F^{\mu\nu}[U] \right] \chi_l(x)\,. \tag{6.15}$$

Here, $\chi_l^T(x) = (u, d)$ is the light quark doublet in the twisted basis, $D_W[U]$ is the massless Wilson-Dirac operator and $F_{\mu\nu}[U]$ is the field strength tensor. The last term is weighted by $c_{\rm SW}$, the Sheikoleslami-Wohlert [187] clover coefficient. The heavy quark twisted mass action is similar to the light quark action in Eq. (6.15). However, it contains an additional term, proportional to the parameter μ_δ, due to the non-degeneracy of the heavy quarks and reads

$$\begin{aligned} S^h_{\rm tm}(\chi_h, \overline{\chi}_h, U) = \\ = a^4 \sum_x \overline{\chi}_h(x) \left[D_W[U] + i\mu_\sigma\gamma_5\tau_3 + m_h - \mu_\delta\tau_1 + \frac{i}{4} c_{\rm SW}\sigma^{\mu\nu}F^{\mu\nu}[U] \right] \chi_h(x)\,, \end{aligned} \tag{6.16}$$

where $\chi_h^T(x) = (s, c)$. Moreover, μ_l and μ_σ are the twisted mass parameter. The mass terms m_l and m_h are the (untwisted) Wilson quark masses tuned to the critical value $m_{\rm crit}$ (i.e. at *maximal twist*), which ensures automatic $\mathcal{O}(a)$ improvement [188] for parity even operators. The equivalent discussion for non-local operators can be found in Refs. [45, 174, 189, 190]. Fields in the "physical basis" can be obtained from the twisted basis through the transformation

$$\overline{\psi}(x) \equiv \overline{\chi}(x) e^{i\frac{\alpha}{2}\gamma^5\tau^3}, \qquad \psi(x) \equiv e^{i\frac{\alpha}{2}\gamma^5\tau^3}\chi(x), \tag{6.17}$$

with $\alpha = \pi/2$ at maximal twist. From now on we will use fields in the physical basis.

The ensemble we use has lattice volume $V = 32^3 \times 64$, with a lattice spacing of $a = 0.0938$ fm. The pion mass is approximately equal to $m_\pi = 260$ MeV and $m_\pi L \approx 3$. In Table 6.2, one can find the summary of the main parameters characterizing the ensemble. For further details see Ref. [186].

$\beta = 1.726$, $c_{\rm SW} = 1.74$, $a = 0.0938(3)(2)$ fm	
$32^3 \times 64$, $L = 3.0\,\text{fm}$	$a\mu_l = 0.003$
	$m_\pi \approx 260\,\text{MeV}$
	$m_\pi L \approx 3$
	$m_N = 1.09(6)$ GeV

Table 6.2: Parameters of the ensemble used in this work. The nucleon mass (m_N), the pion mass (m_π) and the lattice spacing (a) are determined in Ref. [122].

6.2.1 Numerical methods

Connected diagrams

To improve the overlap between the states generated by the interpolating field $N_\alpha(x) = \varepsilon^{abc} u^a_\alpha(x) \left(d^{bT}(x)\mathcal{C}\gamma^5 u^c(x)\right)$ and the proton ground state we employ Gaussian smearing [116, 191]. In addition, we use APE smearing for the gauge links that enter the Gaussian smearing. The optimal parameters for the Gaussian and APE smearing techniques, determined in Ref. [192], are $(\alpha_{\rm G}, N_{\rm G}) = (4.0, 50)$ and $(\alpha_{\rm APE}, N_{\rm APE}) = (0.5, 50)$, respectively. Moreover, to further improve the overlap with the boosted proton ground state, we use momentum smearing [193], as it has been proven to drastically reduce the statistical noise in the matrix elements of boosted hadrons [20]. In Ref. [31] and in this work, the momentum smearing parameter has been tuned to $\xi = 0.6$, which minimizes the statistical errors. The momentum smearing operator $\mathcal{S}$ on a quark field $\psi(x)$ reads

$$\mathcal{S}\psi(x) = \frac{1}{1+6\alpha_G}\left(\psi(x) + \alpha_G \sum_{j=1}^{3} U_j(x) e^{i\xi \vec{P}\cdot\hat{j}} \psi(x+\hat{j})\right), \tag{6.18}$$

where ξ is the momentum smearing parameter and j runs over the spatial directions, with $U_j(x)$ being the link in the spatial j-direction.

To evaluate the connected contributions to the three-point functions, we employ the sequential method [194] through the sink. Moreover, to further increase the number of measurements, we compute the three-point functions with $N_{\rm src}$ different source positions on each configuration and we boost the nucleon along all the spatial directions and orientations, i.e. $\pm x, \pm y, \pm z$. Indeed, in Ref. [101] it was found that the statistical uncertainty decreases as $1/\sqrt{N_{\rm src} N_{\rm dirs}}$ for all the operators under consideration, with $N_{\rm dirs} = 6$ being the number of directions of the nucleon boost. The number of source positions employed depends on the nucleon boost, and is $N_{\rm src} = 8$ for the two lowest values of the momentum and $N_{\rm src} = 14$ at $P_3 = 1.24$ GeV for the third. The source-sink separation is $t_s = 0.94$ fm for the lowest momentum value, and $t_s = 1.13$ fm for the two highest ones. The value employed for t_s at $P_3 = 1.24$ GeV is expected to be large enough to suppress excited-states contamination [101]. In Table 6.3 we report the number of

measurements for the connected contributions at each value of P_3.

P_3 [GeV]	$N_{\rm conf}$	$N_{\rm src}$	$N_{\rm meas}$	t_s [fm]
0.41	50	8	400	0.94
0.83	194	8	1552	1.13
1.24	709	14	9926	1.13

Table 6.3: Number of measurements, $N_{\rm meas}$, used for the connected diagrams. For each value of P_3 we report the number of configurations and source positions employed, as well as the source-sink separation, t_s, in physical units.

Disconnected contribution

The evaluation of the disconnected quark loops with a Wilson line in the boosted frame constitutes the most computationally demanding aspect of this work. The isoscalar three-point function of Eq. (6.2) ($\tau = \mathbf{1}$ and $\psi(x) = (u(x), d(x))^T$) reads

$$\begin{aligned} &C_{3pt}(\vec{P}; t_{\rm ins}; t_s, 0) = \\ &= \tilde{\mathcal{P}}_{\alpha\beta} \sum_{\vec{x},\vec{y}} e^{-i\vec{P}\cdot\vec{x}} \, \langle \Omega | \mathcal{J}_N(x) [\overline{u}(y+z)\Gamma W(z) u(y) + \overline{d}(y+z)\Gamma W(z) d(y)] \overline{\mathcal{J}}_N(0) | \Omega \rangle_{\beta\alpha} \, . \end{aligned} \tag{6.19}$$

The three-point function C_{3pt} contains a connected and disconnected part. The latter is given by

$$\begin{aligned} &C^{\rm disc}_{3pt}(\vec{P}; t_{\rm ins}; t_s, 0) \\ &= -\tilde{\mathcal{P}}_{\alpha\beta} \sum_{\vec{x},\vec{y}} e^{-i\vec{P}\cdot x} \, \langle (\mathcal{J}_N(x) \overline{\mathcal{J}}_N(0))_{\beta\alpha} {\rm Tr} \left[(\mathcal{G}_u(y; y+z) + \mathcal{G}_d(y; y+z)) \, \Gamma W(z) \right] \rangle \ , \end{aligned} \tag{6.20}$$

where $x = (t_s, \vec{x})$, $y = (t_{\rm ins}, \vec{y})$ and $z = (0, 0, 0, z)$. The quantity $\mathcal{G}_f(\vec{x}, t_x; \vec{y}, t_y)$ is the *all-to-all* propagator with quark flavor $f = u, d, s$, from each lattice point x to any point y. Eq. ((6.19)) is a correlation of two parts: the nucleon two-point function and the quark loop with a Wilson line. The latter can be written as

$$\begin{aligned} \mathcal{L}^{\rm u+d}(\tau; z; \Gamma) &= \sum_{\vec{y}} {\rm Tr} \left[(\mathcal{G}_u(y; y+z) + \mathcal{G}_d(y; y+z)) \, \Gamma W(z) \right] \\ &= \sum_{\vec{y}} {\rm Tr} \left[\overline{\psi}(y+z) \Gamma W(z) \psi(y) \right] \ , \end{aligned} \tag{6.21}$$

where the trace in the second line is intended over volume, spin and flavor indices. Due to the presence of the *all-to-all* propagator $\mathcal{G}_f(\vec{x}, t_x; \vec{y}, t_y)$, the exact evaluation of the disconnected contribution in Eq. (6.21) would require $\approx 10^7$ inversions of the Dirac operator per configuration for the lattice that we are considering. As explained in Sec. 4.3, stochastic techniques allow to overcome this limitation, and to compute the all-to-all

propagator with reduced computational cost. In addition, exploiting a property of the twisted mass operator, it is possible to design a stochastic algorithm that further reduce the computational cost. Recalling the transformation of Eq. (6.17), the insertion operator Γ in twisted basis reads

$$\Gamma^{\rm tm} \equiv e^{i\frac{\alpha}{2}\gamma^5\tau^3}\Gamma e^{i\frac{\alpha}{2}\gamma^5\tau^3}\,, \tag{6.22}$$

with $\alpha = \pi/2$ at maximal-twist. Depending on the operator Γ it is possible to exploit two properties of the twisted-mass operator to evaluate the loop of q. (6.21) with a stochastic technique [195, 196]:

1. If $[\Gamma^{\rm tm}, \gamma^5] = 0$ for a particular Γ, then τ^3 appears in the loop of Eq. (6.21) when expressed in the twisted basis

$$\begin{aligned}\mathcal{L}^{u+d}(\tau; z; \Gamma) &= \mathcal{L}^{u+d}_{\rm tm}(\tau; z; i\gamma^5\tau^3\Gamma) \\ &= \sum_{\vec{y}} \mathrm{Tr}\left[\left(\mathcal{G}^{\rm tm}_u(y; y+z) - \mathcal{G}^{\rm tm}_d(y; y+z)\right) i\gamma^5\Gamma W(z,0)\right]\,.\end{aligned} \tag{6.23}$$

   In this case, we apply the *standard one-end trick*, introduced in Sec. 4.3.2, that exploit the following property of the twisted mass operator

$$\mathcal{G}^{\rm tm}_u - \mathcal{G}^{\rm tm}_d = -2i\mu(\mathcal{M}^\dagger_u\mathcal{M}_u)^{-1}\gamma^5\,. \tag{6.24}$$

   The transversity operator $\Gamma = \sigma_{3j}$ belongs in this category.

2. If $\{\Gamma^{\rm tm}, \gamma^5\} = 0$, then the loop in twisted basis possess the same analytical form as in the physical basis

$$\begin{aligned}\mathcal{L}^{u+d}(\tau; z; \Gamma) &= \mathcal{L}^{u+d}_{\rm tm}(\tau; z; \Gamma) \\ &= \sum_{\vec{y}} \mathrm{Tr}\left[\left(\mathcal{G}^{\rm tm}_u(y; y+z) + \mathcal{G}^{\rm tm}_d(y; y+z)\right) \Gamma W(z,0)\right]\,.\end{aligned} \tag{6.25}$$

   The quantity of interest can be computed with the *generalized one-end trick*, introduced in Sec. 4.3.2, exploiting the following property

$$\mathcal{G}^{\rm tm}_u + \mathcal{G}^{\rm tm}_d = 2\gamma^5 D_W(\mathcal{M}^\dagger_u\mathcal{M}_u)^{-1}\gamma^5\,, \tag{6.26}$$

   where D_W is the massless Wilson-Dirac clover operator. Note that if the twisted mass parameter becomes very small (close to the physical point) this type of one-end trick is approaching the standard definition in Eq. (4.32). The helicity operator $\Gamma = \gamma^5\gamma^3$ and the unpolarized $\Gamma = \gamma^0$ belong to this category.

One of the technical aspects of the calculation is the evaluation of the traces in Eqs. (6.23) - (6.25). With small quark masses, the contribution to the loops coming from the low modes of the spectrum of the Dirac operator may be sizeable, and contributes significantly

to the stochastic noise [197]. Therefore, we compute the first $N_{ev} = 200$ eigen-pairs $\lambda_j, |v_j\rangle$ of the squared Dirac operator $\mathcal{M}_u\mathcal{M}_u^\dagger$, that allow to reconstruct exactly the low-mode contribution to the disconnected quark loops. At this stage, stochastic techniques can be employed with the deflated operator to evaluate the high-modes contribution to the traces. To reduce the stochastic noise, we use the *hierarchical probing* algorithm [109], that allows to reduce the contamination to the trace coming from off-diagonal terms up to a distance 2^k. This improvement is achieved by partitioning the lattice with $2^{d(k-1)+1}$ Hadamard vectors, where $d = 4$ is the number of dimensions of the lattice. Finally, to remove the contamination from off-diagonal terms in spin-color subspaces, we apply full dilution [198]. The algorithm employed in the present work has been successfully used in other studies involving the evaluation of disconnected contributions [136–139].

For each value of the proton boost $P_3 = 0.41, 0.83, 1.24$ GeV, we evaluated the two-point functions contributing to the disconnected diagram of Eq. (6.19) with $N_{srcs} = 200$ source positions (see Sec. 6.2.2). In addition, apart from averaging over all possible directions and orientations of the nucleon boost, we also average over forward and backward projections. In Table 6.4, we report the total statistics collected for the disconnected three point correlators. We note that, for the second largest value of the momentum used, namely $P_3 = 1.24$ GeV we use $\approx 10^6$ measurements. In addition, we also compute the matrix elements for $P_3 = 1.65$ GeV and all Γ using approximately the same statistics as the previous smaller boost. While this number of statistics is not sufficient to obtain the same statistical accuracy as lower momenta, it allows us to check whether convergence with P_3 is reached.

	Loops					Two-point functions		
P_3 [GeV]	N_{ev}	N_{conf}	N_{had}	N_{sc}	N_{inv}	N_{srcs}	N_{dir}	N_{meas}
0.41	200	330	512	12	6144	200	6	$396 \cdot 10^3$
0.83	200	349	512	12	6144	200	6	$418.8 \cdot 10^3$
1.24	200	1103	512	12	6144	200	6	$1.3236 \cdot 10^6$
1.65	200	1160	512	12	6144	200	6	$1.392 \cdot 10^6$

Table 6.4: Number of measurements (last column) for each momentum (first column) used for computing the disconnected contributions. N_{ev} is the number of eigen-modes (second column), N_{conf} the number of configurations (third column) and N_{had} the number of Hadamard vectors (fourth column). N_{inv} is the number of inversions per configuration (sixth column), computed as the product of the number of stochastic vectors multiplied by N_{sc} (fifth column) which takes into account the spin-color dilution. The number of source positions for the two-point functions N_{srcs} (seventh column) contributes to the total statistics N_{meas} of the disconnected diagrams, as well as the number of directions and orientations of the nucleon boost N_{dir} (eighth column).

6.2.2 Two-point functions

The two-point functions enter the calculation through the ratio of Eq. (6.4), but also contribute to the evaluation of the disconnected diagram, as shown in Eq. (6.19). For this reason, to obtain a significant amount of measurements for the disconnected contributions

we compute the two-point functions with a large number of source positions, $N_{\rm src}^{\rm 2pt} = 200$, and consider boosts of the nucleon along the different spatial directions and orientations. This procedure allowed us to considerably reduce the statistical error in the disconnected contributions at small computational cost because the same loops is combined with all 200 two-point functions on the same configurations. Given that the computational cost of the two-points function is considerably lower compared to the one required to evaluate the disconnected quark loops, using multiple source positions is highly beneficial. In Table 6.4 we report the number of measurements of the two-point function performed at each nucleon boost P_3.

The two-point function can be written as

$$C_{2pt}(\vec{P};t,0) = \sum_n |\,\langle\Omega|N(\vec{0},0)|n\rangle\,|^2 e^{-tE_n(P)}\,, \tag{6.27}$$

with $|n\rangle$ being the $n^{\rm th}$ energy state of the interpolator $N_\alpha(x)$ and $E_n(P)$ its energy. We performed the analysis by keeping up to two terms in the expansion of Eq. (6.27). In particular, the two-state fit function of the two-point correlator consists of

$$\begin{aligned} C_{2pt}(\vec{P};t,0) &= c_0 e^{-tE_0} + c_1 e^{-tE_1} \\ &= c_0 e^{-tE_0}\left(1 + \frac{c_1}{c_0} e^{-\Delta E t}\right)\,, \end{aligned} \tag{6.28}$$

while the effective energy reads

$$E_{\rm Eff}(\vec{P};t,0) \equiv \log\left(\frac{C_{2pt}(\vec{P};t,0)}{C_{2pt}(\vec{P};t+1,0)}\right) = E_0 + \log\left(\frac{1+Be^{-\Delta E t}}{1+Be^{-\Delta E(t+1)}}\right), \tag{6.29}$$

with $\Delta E = (E_1 - E_0)$ and $B = c_1/c_0$. In Fig. 6.2, we show the results of the two-state fits of the correlator and the effective energy for the parameters E_0, ΔE, c_0 and c_1/c_0, varying the low-end of the fit interval $t_{\rm min}$. The results show that the fits on the correlator or the effective energy lead to the same ground state energy. Furthermore, the plateau and two-state fits converge at $t_s/a = 9$ for momentum 1.24 GeV.

In Table 6.5 we report the parameters extracted using one- and two-state fits. The results for E_0 are obtained with the plateau fit of the effective energy, and they are compatible with the values extracted using two-state fit results. In Fig. 6.3 we show the effective energy for the second largest momentum $P_3 = 1.24$ GeV, together with the plateau fit and two-state fit results. By iterating the fit procedure described above over the data for the different nucleon boosts, we reconstructed the dispersion relation

$$a^2E^2 = a^2m_N^2c^4 + a^2\vec{P}^2c^2, \tag{6.30}$$

with m_N being the nucleon mass. In Fig. 6.3 we show the observed trend of the energy with the nucleon boost P_3 together with a linear fit performed with the function of

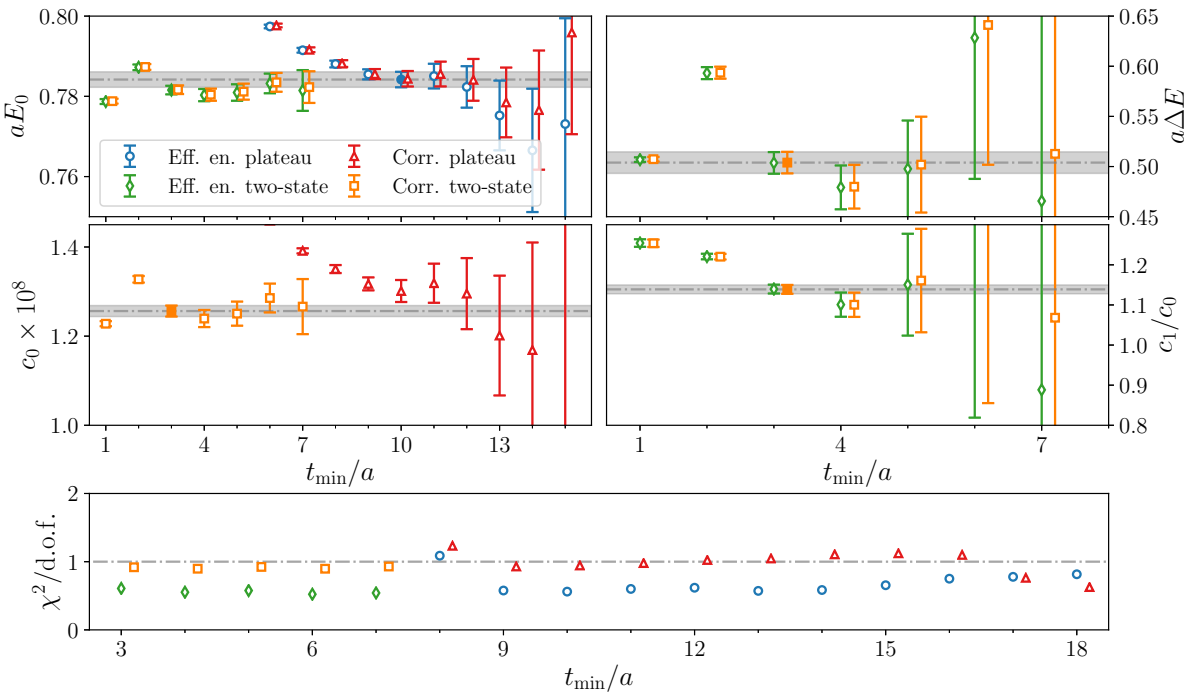


Figure 6.2: Results of the two-state and plateau fits performed on the two-point correlator and on the effective energy at $P_3 = 1.24$ GeV as a function of the lowest time, $t_{\min}$ used in the fit, using the expansion of Eqs. (6.28) - (6.29). In the lower panel we report the reduced $\bar{\chi}^2$ for each fitting procedure. The gray bands correspond with the selected values for E_0 and the remaining parameters ΔE, c_0, c_1/c_0, respectively obtained with the plateau fit of the effective energy and the two-state fit of the correlator. The numerical results for the parameters are reported in Tab. 6.5.

Eq. (6.30), giving $a^2 m_N^2 c^4 = 0.2678(8)$ and $c^2 = 1.003(7)$. As can be seen, the lattice data are fully compatible with the dispersion relation for all values of P_3.

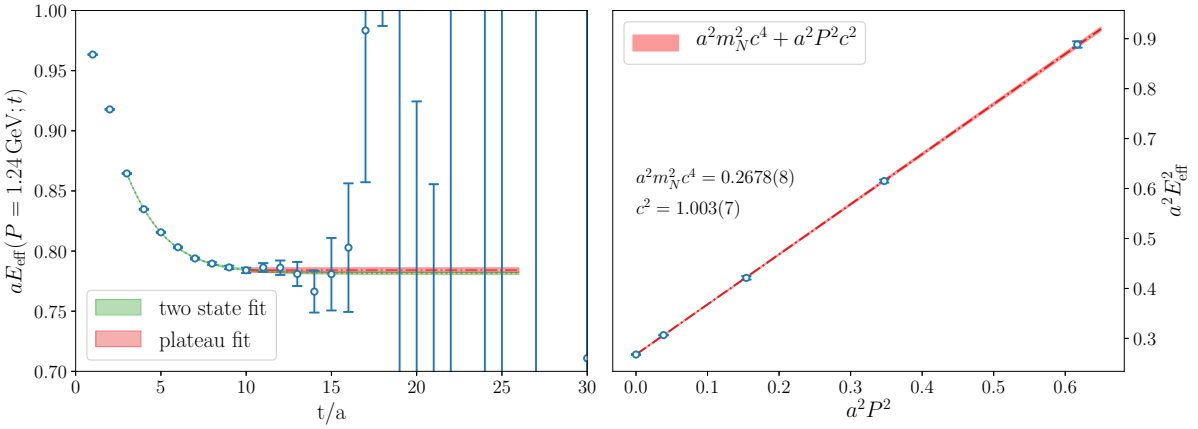


Figure 6.3: Left panel: effective energy computed for $P_3 = 1.24$ GeV, together with the two-state fit (red) and plateau-fit (green) results. Right panel: dispersion relation obtained using the plateau fit for $P_3 = 0$, 0.41, 0.83, 1.24 GeV (blue points). We also report the results for the linear fit using Eq. (6.30) (red line).

P_3 [GeV]	aE_0	$a\Delta E$	c_1/c_0	c_0
0	0.5139(9)	0.51(9)	0.80(2)	$8.99(9) \times 10^{-8}$
0.41	0.5504(9)	0.49(2)	0.82(2)	$6.74(7) \times 10^{-8}$
0.83	0.647(4)	0.48(3)	0.88(4)	$3.59(5) \times 10^{-8}$
1.24	0.784(2)	0.50(1)	1.14(1)	$1.26(1) \times 10^{-8}$
1.65	0.942(3)	0.53(1)	1.34(2)	$3.28(5) \times 10^{-9}$

Table 6.5: Results for the parameters E_0, ΔE, c_1/c_0 and c_0 for $P_3 = 0, 0.41, 0.83, 1.24, 1.65$ GeV. The remaining parameters are obtained with the two-state fit of the two-point correlator of Eq. (6.28).

6.3 Disconnected matrix elements

Obtaining the disconnected contributions to the up-, down- and strange-quark PDFs is the central goal, and most laborious aspect of this work. We use the techniques outlined in Section 6.2.1 to extract the matrix elements and study systematic uncertainties, such as excited-states contamination.

6.3.1 Excited-states contamination

To extract reliably the ground-state contribution to the matrix elements, we evaluate the ratio between the three- and two-point functions at seven source-sink separations, ranging from $t_s = 0.563\,\text{fm}$ to $t_s = 1.126\,\text{fm}$ in steps of $a = 0.0938$ fm. For disconnected contributions, the evaluation of different source-sink separations does not require new inversions. This allowed us to study the excited-states contamination to the matrix elements using several t_s values and three analysis methods: plateau fit, two-state fit and summation method. We briefly summarize these methods.

1. **Two-state fit**. In Sec. 6.2.2, the two-point correlator is expanded up to the first excited state. Likewise, we can expand the three-point correlator keeping terms up to the first excited state. This gives four terms, that is

$$\begin{aligned} C_{3pt}(\vec{P}; t_s, \tau) = &\mathcal{A}_{0,0}(\vec{P}) e^{-E_0(\vec{P})t_s} \\ &+ \mathcal{A}_{0,1}(\vec{P}) e^{-E_0(\vec{P})t_s} e^{-\Delta E(\vec{P})\tau} \\ &+ \mathcal{A}_{1,0}(\vec{P}) e^{-E_1(\vec{P})t_s} e^{\Delta E(\vec{P})\tau} \\ &+ \mathcal{A}_{1,1}(\vec{P}) e^{-E_1(\vec{P})t_s} \,. \end{aligned} \tag{6.31}$$

   Being interested in the forward kinematic limit allows one to reduce the number of independent parameters, since $\mathcal{A}_{0,1} = \mathcal{A}_{1,0}$. We performed a fit of the ratio of

Eq. (6.4) with the function

$$\frac{\langle C_{3pt}(\vec{P};t;t_s,0\rangle}{\langle C_{2pt}(\vec{P};t,0)\rangle} = \\ = \frac{\mathcal{A}_{0,0}}{c_0}\frac{\left[1+(\mathcal{A}_{0,1}/\mathcal{A}_{0,0})\,e^{-\Delta E\tau}+(\mathcal{A}_{0,1}/\mathcal{A}_{0,0})\,e^{-\Delta E(t_s-\tau)}+(\mathcal{A}_{1,1}/\mathcal{A}_{0,0})\,e^{-\Delta E t_s}\right]}{\left[1+\frac{c_1}{c_0}e^{-\Delta E t_s}\right]}\,, \tag{6.32}$$

where the parameters c_1/c_0 and ΔE are determined through the effective energy fit and the results are reported in Table 6.5. Thus, the parameters determined by fitting the ratio of three- and two-point functions are $\mathcal{A}_{0,0}/c_0$, $\mathcal{A}_{0,1}/\mathcal{A}_{0,0}$ and $\mathcal{A}_{1,1}/\mathcal{A}_{0,0}$. $\mathcal{A}_{0,0}/c_0$ corresponds to the matrix element we are interested in. Such fits are weighted by the statistical errors, and therefore the fit is driven by the most accurate data. Since the statistics for the disconnected contributions are independent of th t_s value, we repeat the two-state fits modifying, each time, the starting value of t_s entering the fit (t_s^{low}). The results from the two-state fit method, allows us to verify ground-state dominance by comparing the matrix elements from the individual t_s values.

2. **Plateau fit**. For $0 \ll \tau \ll t_s$ and $\Delta E\, t_s \gg 0$, the first term in the ratio of Eq. (6.32) dominates. Thus, the matrix elements can be extracted by performing a constant fit on the ratio of Eq. (6.4) in the region defined $0 \ll \tau \ll t_s$, with large enough source-sink separation. We exclude from the fit range three points from left and right, i.e. we evaluate the weighted average in the interval $\tau \in [3, t_s - 3]$. While the excited-states contamination decreases with t_s, at the same time the statistical uncertainty exponentially increases. For this reason, the determination of the ground state of the matrix elements with the plateau fit method is a challenging task, and the results need to be compared with other analysis techniques.

3. **Summation method**. Summing over the insertion time τ of the ratio of the three- and two-point functions we find [199, 200]

$$S(t_s) = \sum_{\tau=2}^{\tau=t_s-2} \frac{\langle C_{3pt}(\vec{P};t;t_s,0\rangle}{\langle C_{2pt}(\vec{P};t,0)\rangle} = t_s\frac{A_{0,0}}{c_0} + c + \mathcal{O}\left(e^{-\Delta E t_s}\right). \tag{6.33}$$

Thus, the matrix elements corresponds with the slope of the straight line $S(t_s)$, and can be measured by performing a linear regression.

Using the three aforementioned approaches we analyze the excited-states effects on the matrix elements for the unpolarized, helicity and transversity PDFs. In the next three subsections, we present the analysis of the real and imaginary parts of the matrix elements at $P_3 = 1.24$ GeV, as a representative example.

Unpolarized

We start by discussing the analysis of the unpolarized isoscalar $u+d$ disconnected matrix elements. In Fig. 6.4 we show the ratio of three- and two-point functions at $P_3 = 1.24$ GeV for the unpolarized operator for $z/a = 3$ and we compare the results obtained with the three analysis methods reported in Sec. 6.3.1. The real part of the matrix elements shows no substantial dependence on the source-sink separation, and the plateau fit results obtained at different t_s give all compatible results. The dependence of the two-state fit on the lowest source-sink separation t_s^{low}, included in the fit shows a constant trend, which is also compatible with the results obtained with the summation method. The reduced chi-square $\chi^2/\text{d.o.f} = 0.96$ suggests that the function of Eq. (6.32) provides a good description of the data. To extract the matrix elements, we compute the constant correlated fit of the plateau fit results starting from $t_s^{\text{low}}/a = 9$. In contrast to the real part, the imaginary part shows a large effect due to the excited-states contamination. However, the two-state fit is compatible with the plateau value using $t_s/a = 11$. Also, the results obtained at different t_s^{low} using the summation method are compatible with the other methods within uncertainties. As final results for the unpolarized matrix element, we report the ones from the plateau fit for $t_s/a = 11$, which is compatible with the results obtained with the two-state fit and summation method.

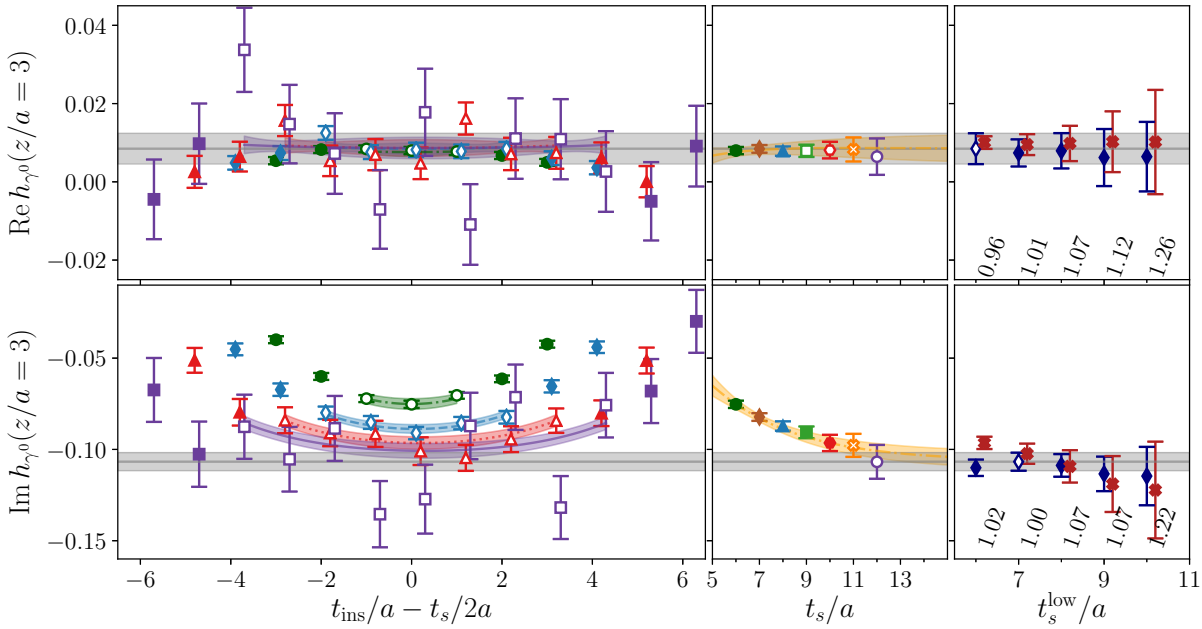


Figure 6.4: Left: Results on $C_{3pt}(t;t_s)/C_{2pt}(t)$ for the unpolarized PDFs for $P_3 = 1.24$ GeV, at $t_s/a = 6, 8, 10, 12$ for $z/a = 3$. The data for $t_s/a = 7, 9, 11$ are omitted to improve the readability. The two-state fit results (gray band), and the value of the two-state fit of Eq. (6.31) evaluated at the same t_s as the data-points are also shown. Only the data-points with open symbols are taken into account in the two-state fit procedure. Center: the plateau fit results as a function of t_s/a. Each source-sink separation is associated with a different color. The orange band is the predicted t_s dependence of the function in Eq. (6.31) at $t_{\rm ins} = t_s/2$. Our final value for the matrix elements is determined as the correlated constant fit of the plateau values shown with open symbols. Right: results of the two-state fit (navy blue) as a function of the lowest source-sink separation $t_s^{\rm low}$ included in the fit. The empty data-point is the selected two-state fit result, which corresponds to the gray band. For each $t_s^{\rm low}$ we report the reduced χ^2 of the two-state fit. The results obtained with the summation method are reported with the red open crosses as a function of $t_s^{\rm low}$.

Helicity

The disconnected contributions to the helicity isoscalar matrix elements is purely real and exhibits a non-negligible dependence on the source-sink separation. We observe a decreasing behavior as t_s increases, which is a behavior also observed in the axial charge [137]. In addition, the plateau fit for $t_s/a > 10$ are compatible with the two-state fit. Therefore, we use the plateau fit for $t_s/a = 10$ as our final results, so that statistical uncertainties are controlled.

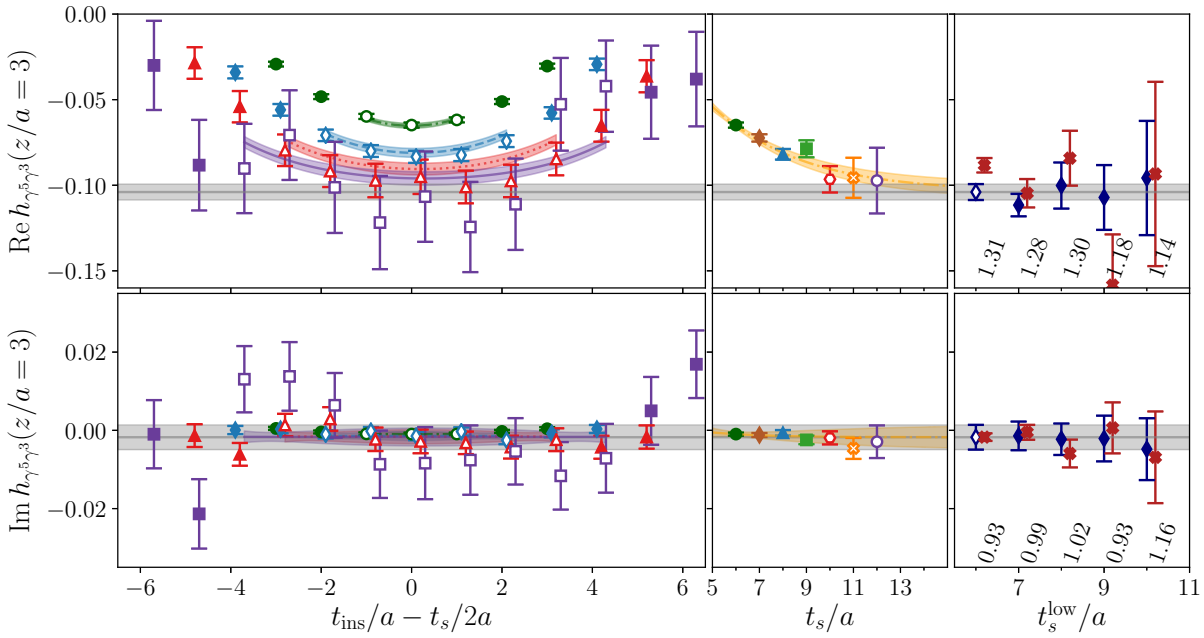


Figure 6.5: The same as Fig. 6.4 but for the isoscalar helicity matrix elements.

Transversity

The ratio of the three- and two-point functions for the disconnected contributions to the isoscalar transversity matrix elements does not shows dependence on the source-sink separation for both the real and imaginary parts (see Fig. 6.6). Thus, the matrix elements are computed from the plateau fit for $t_s/a = 9$, both for the real and the imaginary parts.

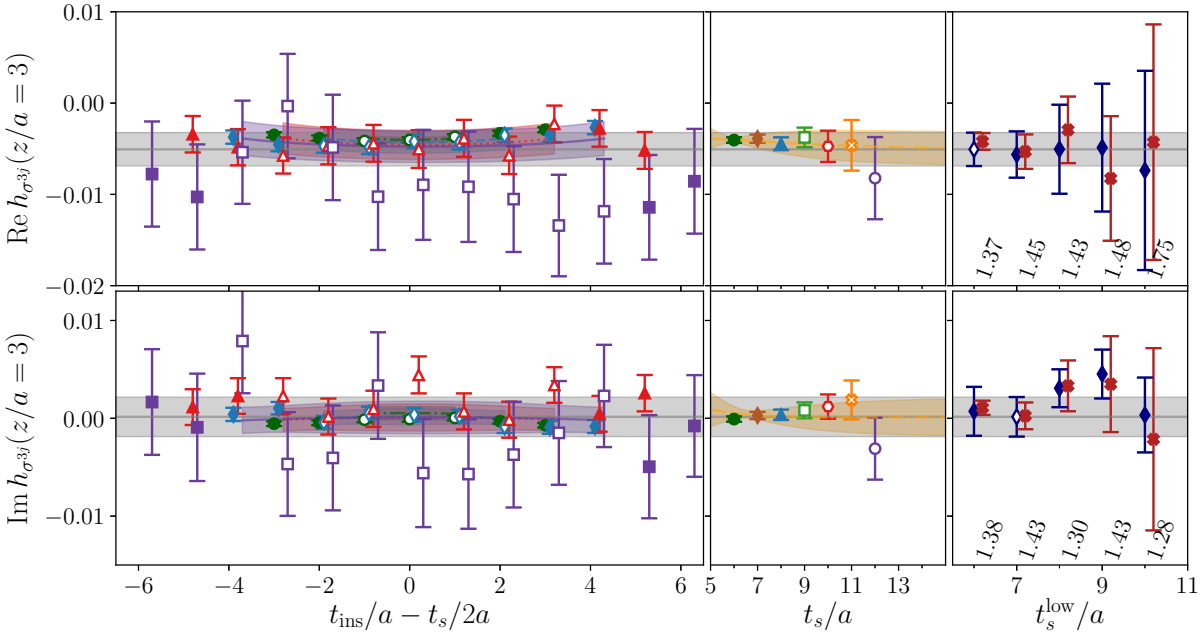


Figure 6.6: The same as Fig. 6.4 but for the isoscalar transversity matrix elements.

Using the criterion adopted for selecting the final results for each PDF case we compared the extracted matrix elements in Fig. 6.7 for all values of z. In summary, the plateau

fits are evaluated at $t_s/a = 9, 10, 9$ ($t_s/a = 11, 10, 9$) for the real (imaginary) part of the unpolarized, helicity and transversity, respectively. For the summation method, we employ all t_s values available, except for the imaginary part of the unpolarized and the real part of the helicity matrix elements, where $t_s/a = 6$ is excluded as explained above. The two-state fit is performed in the range $t_s/a \in [6, 12]$ in all cases, except for the imaginary part of the unpolarized and transversity matrix elements, where the $t_s/a = 6$ value, as explained, is not included in the regression.

Our conclusions for the isoscalar disconnected matrix elements apply also to the strange matrix elements, with the excited-states contamination showing similar effects.

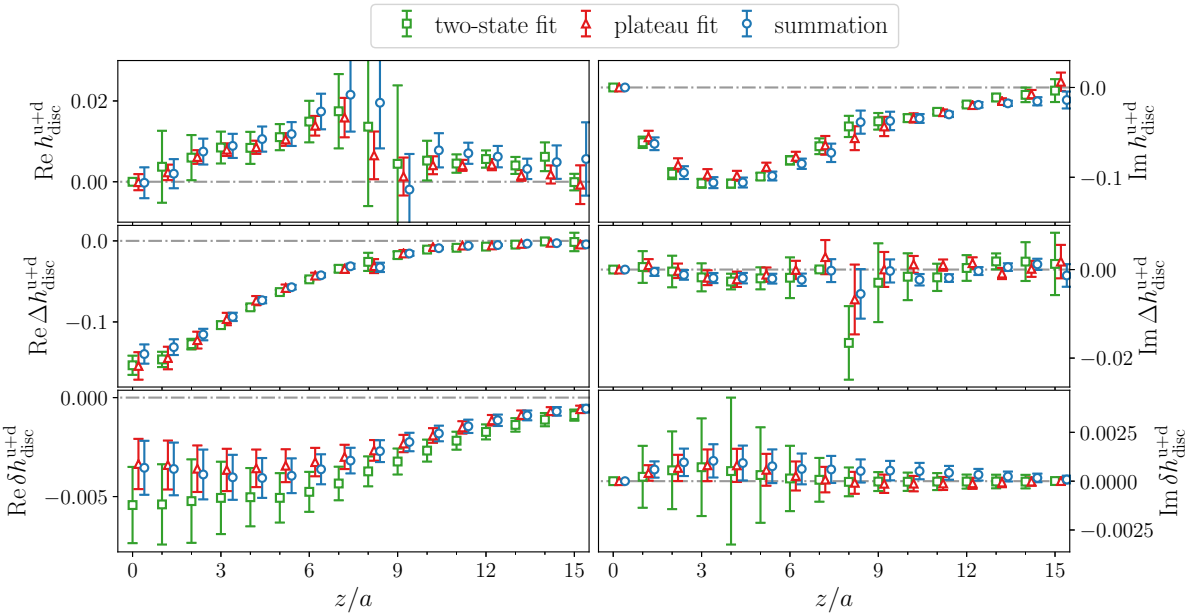


Figure 6.7: Comparison of the matrix elements obtained from the one- (red points) and two-state (green points) fits and the summation method (blue points) for the disconnected isoscalar matrix elements at $P_3 = 1.24$ GeV. From top to bottom we show the unpolarized, helicity and transversity PDFs. See text for more details.

6.3.2 Momentum dependence

As explained in Sec. 6.1, the matrix elements and the quasi-PDFs have a dependence on the nucleon boost P_3, which also enters the matching formula leading to the PDFs. An important aspect of the study is the investigation of the momentum dependence of the matrix elements, which affects the convergence to the light-cone PDFs. In Fig. 6.8, we present the results for the renormalized strange and isoscalar disconnected matrix elements as a function of the momentum boost. For the unpolarized case, the real part decreases in magnitude as the P_3 increases, and becomes compatible with zero. In contrast, its imaginary part is non-zero and shows convergence for the two largest values of P_3. We find that the isoscalar disconnected matrix elements share the same qualitative behavior as the strange-quark ones.

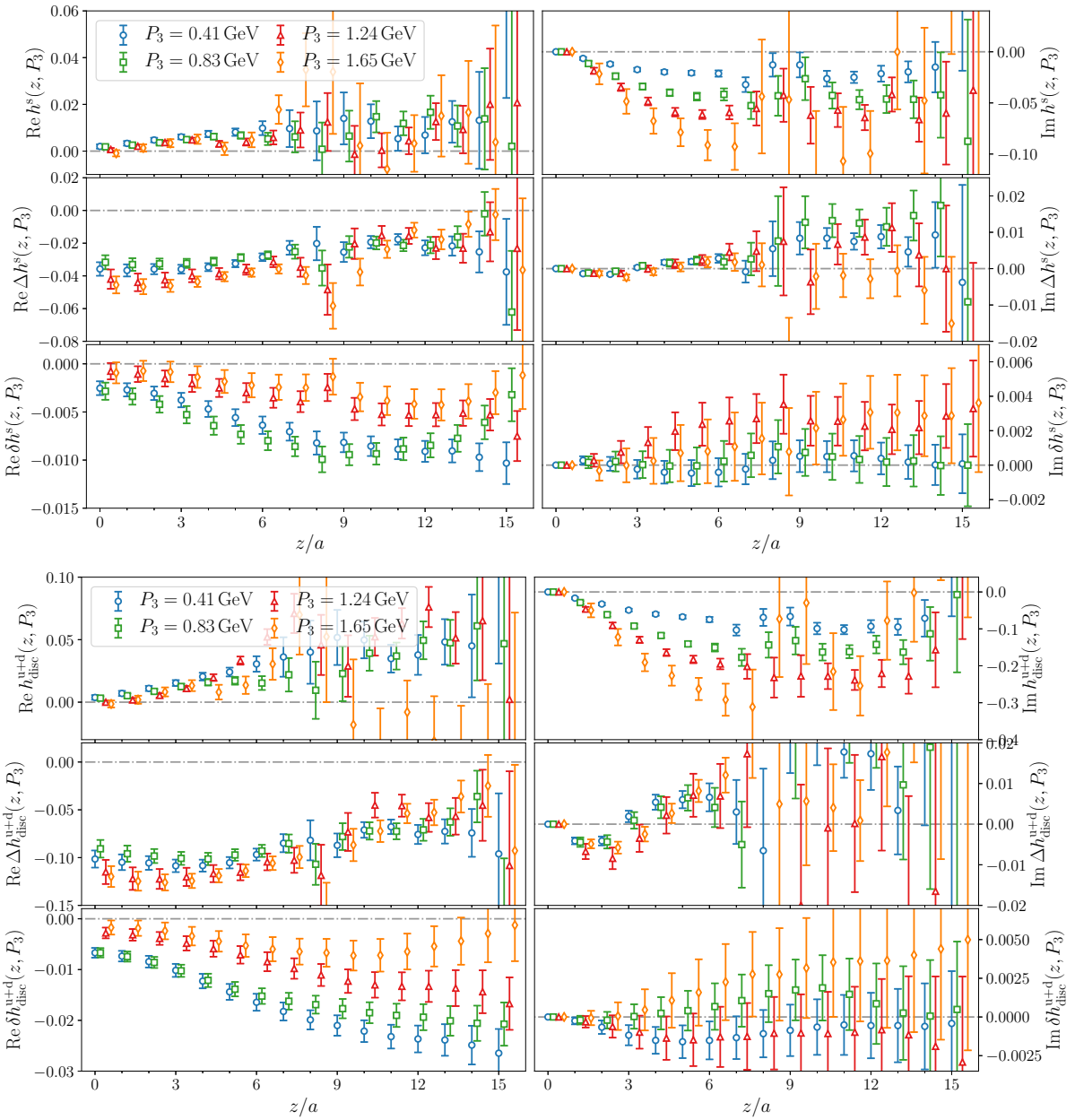


Figure 6.8: Momentum dependence of the renormalized matrix elements for the strange (upper figure) and isoscalar disconnected (lower figure) unpolarized (top panels), helicity (middle panels) and transversity (bottom panels) distributions. We show the matrix elements computed at $P_3 = 0.41$ GeV *(blue),* 0.83 GeV *(green),* 1.24 GeV *(red) and* 1.65 GeV *(yellow). Data points are slightly shifted to improve readability.*

First results on the helicity distribution appeared in Ref. [46]. Here we show results with increased statistics, and with the addition of $P_3 = 1.65$ GeV. The matrix elements show a mild residual dependence on the momentum. The imaginary part of the renormalized matrix elements arises entirely from the complex multiplication with the renormalization function and the bare matrix elements (see Eq. ((6.10))). Indeed, as mentioned already in Sec. 6.3.1, the disconnected contribution to the bare matrix element for the helicity distribution is purely real. The real part of the matrix elements for the transversity distribution exhibit a strong dependence on the nucleon boost, changing dramatically as we increase P_3 from 0.83 GeV to 1.24 GeV. However, results obtained

for $P_3 = 1.65$ GeV show agreement with those for $P_3 = 1.24$ GeV albeit the large uncertainty. From the current results it is still unclear if convergence is reached. However, in order to fully check this would require a larger momentum and much more measurements to reach the required accuracy. This is beyond the current study and will be tested in a followup work. We thus, construct the PDFs using the results for $P_3 = 1.24$ GeV. In Sec. 6.6.4, we comment on the region of x affected by the gap observed in the real part of the matrix elements as momentum increases. In contrast, the imaginary part is fully compatible with zero for the two lowest momenta, while it is slightly non-zero at large z at the highest momentum.

6.4 Connected matrix elements

The evaluation of the connected matrix elements contributing to the three types of PDFs has been studied in our previous works. In particular, we refer the reader to the study of Ref. [101], where several sources of systematic uncertainties were discussed in great detail. For completeness, we briefly discuss here the connected contributions which are needed for the flavor decomposition. In Fig. 6.9, we show the momentum dependence of the bare connected contributions to the isoscalar and isovector matrix elements for the three types of PDFs. In all cases, as the nucleon boost increases, the real part of the matrix elements decay to zero faster, and the magnitude for the imaginary part increases in the region $z/a \lesssim 9$. The unpolarized matrix elements show convergence with P_3 while the imaginary parts of the helicity and transversity distributions increase in magnitude. We note that in order to compute the connected contributions for a fourth larger boost would require new inversions and large number of measurements to reduce the errors sufficiently enough to check convergence. Thus for the current work we opt to use the results for $P_3 = 1.24$ GeV for the connected parts since the focus of this work is the evaluation of the disconnected contributions. The uncertainty on the unpolarized distribution is smaller as compared to the other two distributions. This behavior is due to the fact that both the three- and two-point functions share the same projector, $(1 + \gamma^0)/2$, which increases the correlation between the two quantities and, as a result, drastically decreases the noise-to-signal ratio. We note also that the transversity distribution reported here is the average over the two insertions σ_{3j} with $j = 1, 2$.

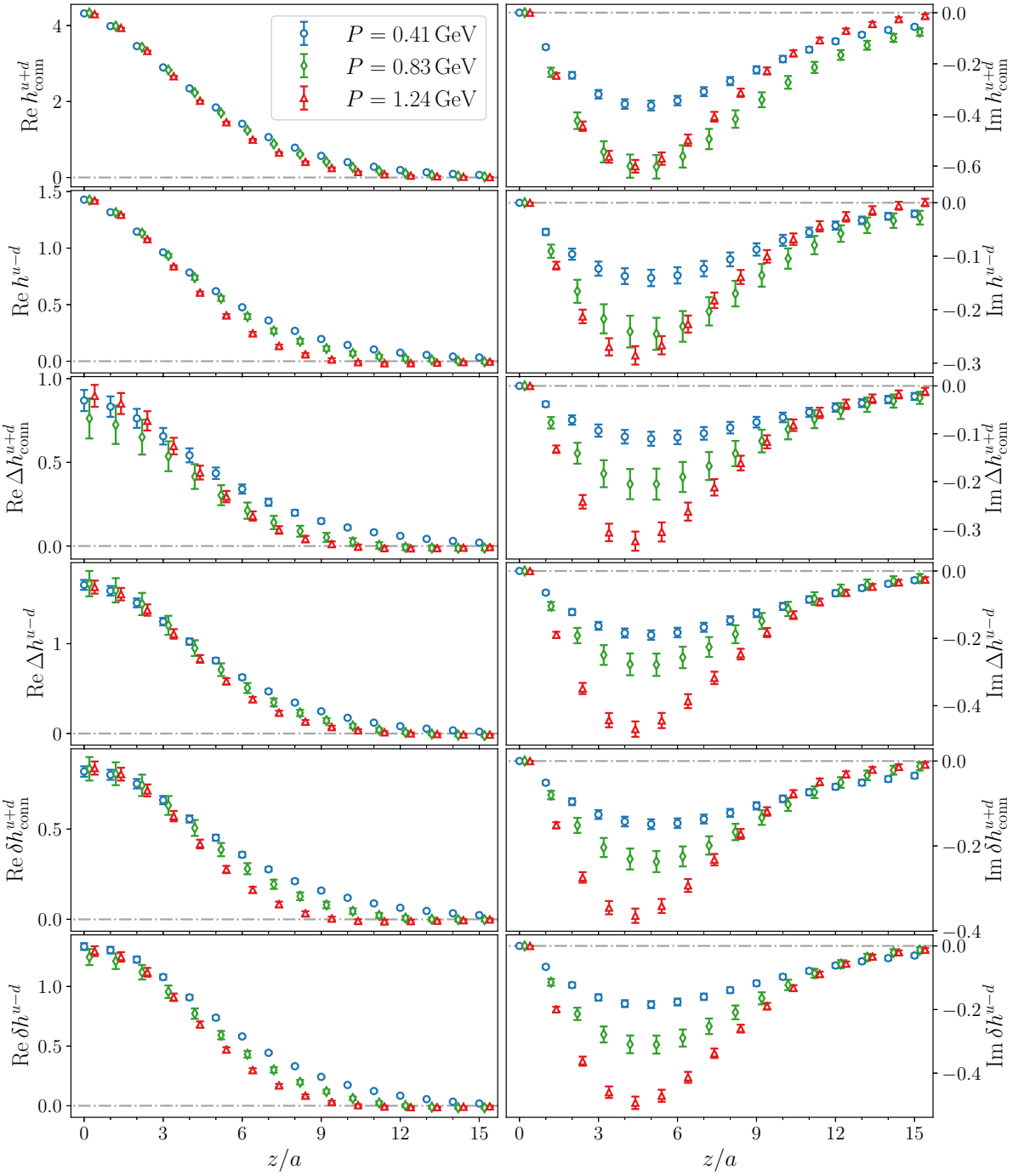


Figure 6.9: Momentum dependence of the bare connected contributions to the matrix elements. The first two rows show respectively the isoscalar connected contribution and the isovector unpolarized matrix elements. The left column shows the real part and the right the imaginary part. The same flavor combinations are reported respectively for the helicity and transversity distributions in the 3rd and 4th rows, and in the last two rows. We show the matrix elements at $P_3 = 0.41$ GeV (blue), 0.83 GeV (green) and 1.24 GeV (red). Data points are slightly shifted to to improve readability.

6.5 Nucleon charges

The nucleon charges are usually extracted from the nucleon matrix elements of local operators. This limit is obtained from the matrix elements of non-local operators at

$z = 0$. Since the charges are frame independent, any value of P_3 may be used. Indeed, in Sec. 6.4 we demonstrate that the $z = 0$ have little dependence on P_3 For the disconnected contributions, we have the matrix elements at $P_3 = 0$ (rest frame). For the connected contributions to the charges, we use the lowest momentum, so we control statistical uncertainties. In what follows, we will show the results obtained for the isovector $u - d$, isoscalar $u + d$ and strange-quark vector, axial and tensor charges, g_V, g_A and g_T.

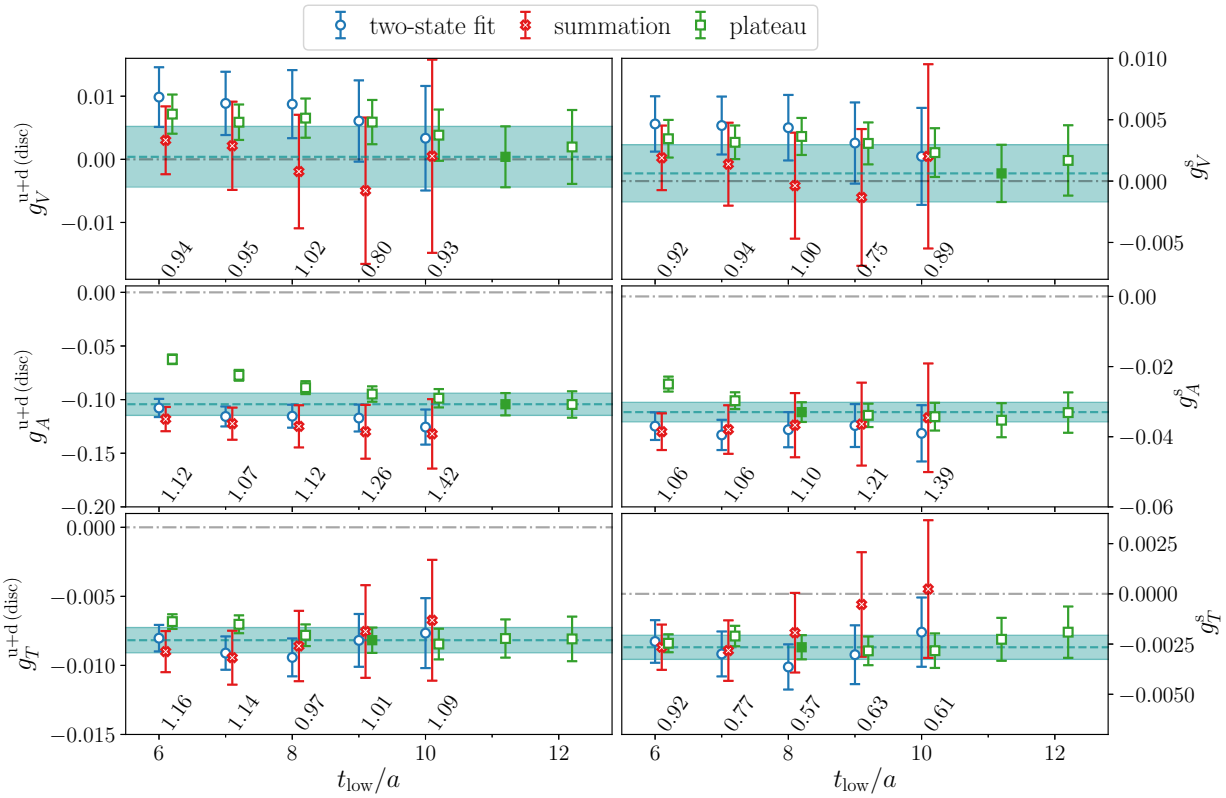


Figure 6.10: Isoscalar $u - d$ (left) and strange (right) disconnected contributions to the renormalized g_V (top panels), g_A (middle panels) and g_T (bottom panels). In each subplot we show the results obtained with the plateau fit (open green squares), two-state fit (open blue circles) and summation method (open red crosses) as a function of t_{low}. We also include χ^2/d.o.f. for the two-state fit. The horizontal band corresponds to the selected plateau fit result.

First, we describe our results for the disconnected contributions, whose total number of measurements in the rest frame is $N_{meas} = 66 \cdot 10^3$. In Fig. 6.10 we show our results for the renormalized disconnected vector, axial and tensor charges. The integral over the volume (i.e. Fourier transform at zero momentum transfer) of the trace of the vector current $\bar{\psi}(x)\gamma^0\psi(x)$ is zero because quark and antiquark loops contributions cancel each other. Thus, the disconnected contribution to the unpolarized isoscalar and strange matrix elements in the absence of the Wilson line and at $P_3 = 0$ expected to be zero is verified and this constitutes a consistency check of our computations. Due to excited-states contamination, the disconnected isoscalar and strange vector charges $g_V^{u+d\,(disc)}$ and g_V^s obtained with the two-state and plateau fits are not compatible with zero at small source-sink separation. In particular, the two-state fit results become compatible with zero when $t_{low} \geq 10$. The results obtained with the summation method have the largest

uncertainties and are compatible with zero.
The axial charge shows the larger contamination from excited states. In particular, the plateau fit results show a decreasing trend with the t_{low}, converging to a constant value for $t_{\text{low}}/a \geq 11$ for the isoscalar and $t_{\text{low}}/a \geq 8$ for the strange charges, which are selected as our final values. In contrast, both the results obtained with two-state fit and summation are constant and compatible with the selected plateau fit results. We find

$$g_A^{\text{u+d (disc)}} = -0.104(10), \quad g_A^s = -0.0320(28)\,. \tag{6.34}$$

The results on the disconnected contributions of the tensor charge show very mild excited states effects. We use the plateau value extracted by fitting the ratio to a constant for $t_s/a \geq 9$ for the isoscalar connected tensor charge $g_T^{\text{u+d(disc)}}$ and for $t_s/a \geq 8$ for g_T^s. Our final results for these two quantities are

$$g_T^{\text{u+d (disc)}} = -0.00818(91), \quad g_T^s = -0.00265(60)\,. \tag{6.35}$$

We stress that despite the agreement of the strange tensor charge with the value extracted using local operators [131, 201], a direct comparison is not meaningful since we are using gauge ensembles simulated with heavier than physical pion mass. It thus comes with no surprise that the disconnected isoscalar tensor charge differs from the value obtained at the physical pion mass.

The connected contributions to the nuclear charges are computed for the smallest momentum $P_3 = 0.41$ GeV. Using these results we can extract the values for each quark flavor for the vector, axial an tensor charges. The details on the computation of the connected isoscalar and isovector contributions are given in Sec. 6.4. The connected contributions used to extract the charges are obtained from plateau fits with $t_s/a = 12$. The nucleon axial and tensor charges are given in Table 6.6. We note that for the vector charge g_V we find results that are consistent with charge conservation.

	$u-d$	$u+d$ (conn.)	$u+d$ (disc.)	u	d	s
g_A	1.25(4)	0.66(7)	-0.104(10)	0.90(2)	-0.35(2)	-0.0320(28)
g_T	1.11(2)	0.68(2)	-0.00818(91)	0.89(1)	-0.22(1)	-0.00265(60)

Table 6.6: Results for the isovector (first column), isoscalar connected (second column) and disconnected (third column) and for the up (fourth column), down (fifth column) and strange (sixth column). We show our results on the axial (second row) and tensor (third row) nucleon charges.

P_3	$h^{u-d}(z=0)$	$h^{u+d}(z=0)$	$\Delta h^{u-d}(z=0)$	$\Delta h^{u+d}(z=0)$	$\delta h^{u-d}(z=0)$	$\delta h^{u+d}(z=0)$
0.41 GeV	1.005(4)	3.046(4)	1.25(4)	0.52(5)	1.11(2)	0.67(2)
0.83 GeV	1.004(8)	3.053(8)	1.26(11)	0.45(9)	1.04(6)	0.69(5)
1.24 GeV	1.000(4)	3.026(5)	1.23(5)	0.52(5)	1.08(3)	0.69(3)

Table 6.7: Momentum dependence of the unpolarized, helicity and transversity isovector (first column) and isoscalar (second column) matrix elements at $z = 0$.

6.6 Parton distribution functions

6.6.1 Isoscalar and isovector renormalized matrix elements

In Fig. 6.12 we show the momentum dependence of the total renormalized isoscalar and isovector matrix elements, including disconnected contributions. The renormalized matrix elements are reported as a function of zP_3. We renormalize and apply the matching procedure independently for the isoscalar and isovector distributions, allowing us to obtain the individual up and down quark PDFs.

The source-sink separation used is $t_s = 0.94$ fm for the lowest momentum and $t_s = 1.13$ fm for $P_3 = 0.83$ and 1.24 GeV. From previous studies of the isovector distributions (see, e.g. Ref. [101]) and nucleon charges [137], we expect that excited-states contamination is more significant for the nucleon three-point correlators of the axial and tensor currents as compared to the vector. However, the statistical uncertainty is larger for these quantities and within our current errors the source-sink separation employed at the highest momentum is sufficient to suppress excited states to this level of accuracy [101].

A summary of the results for the connected matrix elements in the absence of the Wilson line are reported in Table 6.7. The momentum dependence of all the matrix elements analyzed is negligible for $z = 0$ as expected For example, the isoscalar connected matrix elements for the unpolarized distribution, $h^{\mathrm{u+d}}(z = 0)$, for the largest momentum differs from the others by less then 1%. The isovector unpolarized matrix elements at $z = 0$ is independent of the momentum boost, and equal to 1, as expected from charge conservation. Regarding the isoscalar helicity case, we still find agreement for different P_3 within uncertainties, but with larger fluctuations of the mean values, as the disconnected contribution is about $\sim 17\%$ of the connected part. We note that the $\Delta h^{u-d}(z = 0)$ is compatible with the experimental value $g_A^{\mathrm{u-d}} = 1.27641(56)$ [202]. Insensitivity to the momentum boost is also observed in the transversity case.

6.6.2 Truncation of the Fourier transform

In order to construct the x-dependence of PDFs we need to take the Fourier transform. Since the matrix elements are determined for discrete finite number of z values, we study the dependence on the cutoff z_{max} to understand systematic effects related to the reconstruction. In particular, for all types of distributions we verify that a z_{max} exists

such that, addition of information for $z > z_{\max}$ in the Fourier transform leaves the PDF unchanged within statistical uncertainties. This value of z is selected as the maximum value $z_{\max}$ included in the Fourier transform and, typically, the matrix elements at this value has a vanishing real part. Note that the latter is just a qualitative criterion and in practice we always check by increasing z for convergence.

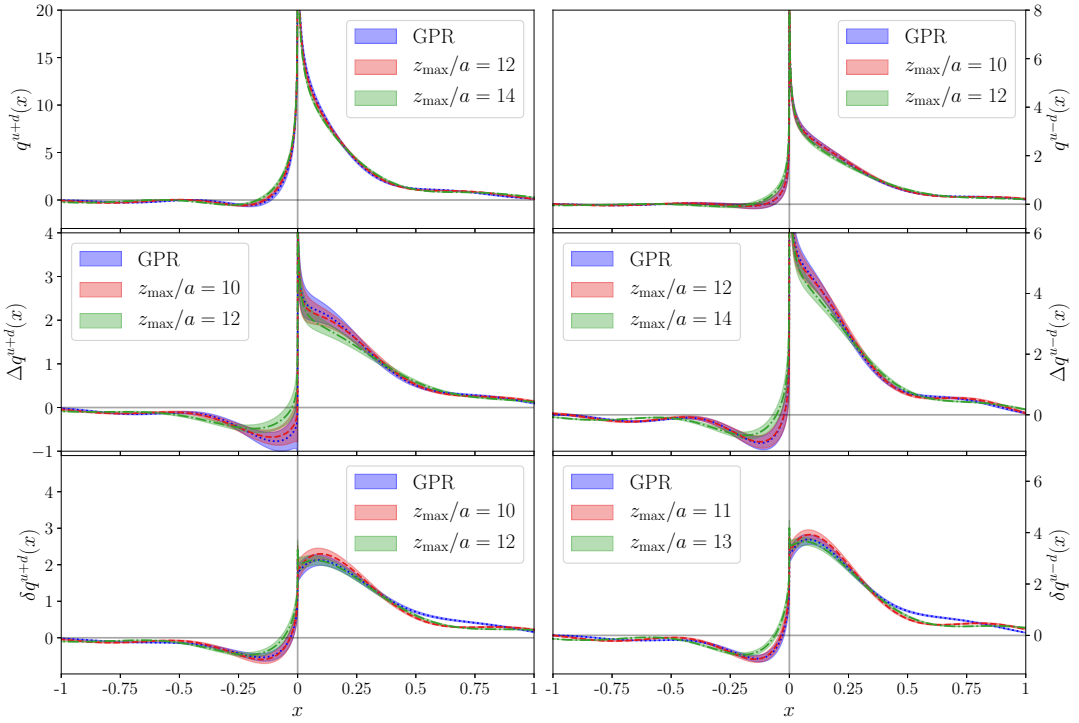


Figure 6.11: Cutoff dependence ($z_{\max}$) of the isoscalar (left) and isovector (right) unpolarized (upper panels), helicity (middle panels) and transversity (bottom panels) at $P_3 = 1.24\,\text{GeV}$. Results from BGFT are shown with a blue band. The distributions corresponding to the value of the cutoff reported in Tab. 6.8 are reported in red.

In Fig. 6.11 we show the dependence on the cutoff $z_{\max}$ for the isoscalar and isovector distributions at $P_3 = 1.24\,\text{GeV}$. We compare the results obtained with the discrete Fourier transform of Eq. ((6.12)) with the results from the Bayes-Gauss-Fourier transform (BGFT) [48]. The latter is an advanced reconstruction technique based on Gaussian process regression, which allows to obtain an improved estimate of quasi-PDF for continuous values of x, starting from a discrete set of data obtained with lattice QCD computations. The chosen values of $z_{\max}$ for each quark flavor and each operator are given in Table. 6.8. From Fig. 6.11 it is clear that increase of the cutoff beyond the reported values of Table 6.8, does not affect the results for the PDFs. We also find compatible of the discreet Fourier transform with the results using BGFT.

	Isoscalar	Isovector	Strange
Unpolarized	15,14,12	15,13,10	15,12,12
Helicity	15,10,10	15,12,12	14,14,14
Transversity	15,11,10	15,11,11	14,12,12

Table 6.8: Values of $z_{\max}/a$ used in the Fourier transform for each type of distribution. Each triplet of numbers corresponds to the cases for $P_3 = 0.41, 0.83$ and 1.24 GeV, respectively.

6.6.3 Isoscalar and isovector distributions

The isoscalar and isovector PDFs are extracted from the corresponding renormalized matrix elements shown in Fig. 6.12. For the isoscalar combination, we add both the connected and disconnected contributions. We plot the matrix elements against zP_3, which is the argument of the exponential in the Fourier transform.

In all cases, we find that the matrix elements for the lowest momentum $P_3 = 0.41$ GeV do not decayed to zero for large z, demonstrating, as expected, that the momentum is not large enough. By increasing the momentum to $P_3 = 1.24$ GeV, the matrix elements become consistent with zero within their uncertainties. While the imaginary parts show a residual momentum dependence, the convergence must be checked at the level of the reconstructed PDF. This is due to the fact that P_3 enters the matching kernel and affects the convergence. Therefore, to address the momentum convergence as we increase P_3, we show in Fig. 6.13 the momentum dependence of the isoscalar and isovector PDFs. We use the standard Fourier transform, with the values of $z_{\max}$ given in Table 6.8, as discussed in Sec. 6.6.2. As can be seen in Fig. 6.13, the overall dependence on the two largest values of the momentum is relatively small. Dependence on P_3 is observed in the unpolarized isoscalar PDF. In general, the PDFs for the smallest momentum, do not show convergence, and exhibit non-physical oscillations due to the presence of systematic effects in the reconstruction of the x-dependence. However, such oscillations are suppressed for the higher values of P_3. The isoscalar and isovector helicity distributions have a similar magnitude and exhibit milder dependence on the boost as compared to the unpolarized. In particular, both isoscalar and isovector helicity distributions are consistent for $P_3 = 0.83$ GeV and $P_3 = 1.24$ GeV. Finally, the isoscalar and isovector transversity distributions also show nice convergence with P_3 for the two largest values. These distributions will be used for the flavor decomposition presented in Sec. 6.6.4 together with comparison of our data with phenomenology.

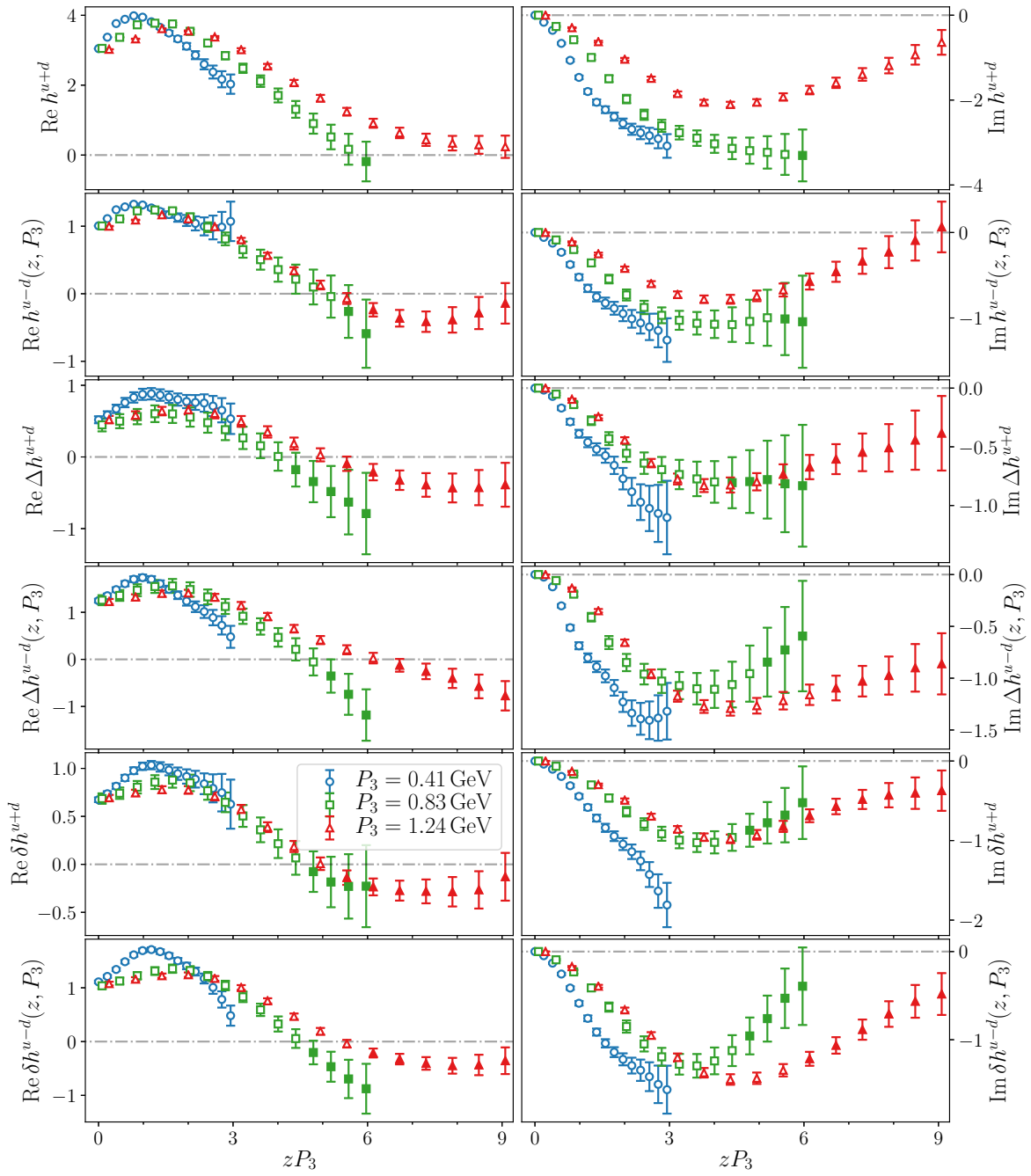


Figure 6.12: Real (left) and imaginary (right) parts of the renormalized matrix elements as a function of zP_3. From top to bottom and in rows of two we show the isoscalar and the isovector matrix elements for the unpolarized, helicity and transversity cases, respectively. The points included in the Fourier transform of Eq. (6.11) are shown with open symbols. Each sub-figure shows the momentum dependence of the corresponding matrix element, where the blue circles correspond to $P_3 = 0.41\,\mathrm{GeV}$, the green squares to $P_3 = 0.83\,\mathrm{GeV}$ and the red triangles to $P_3 = 1.24\,\mathrm{GeV}$.

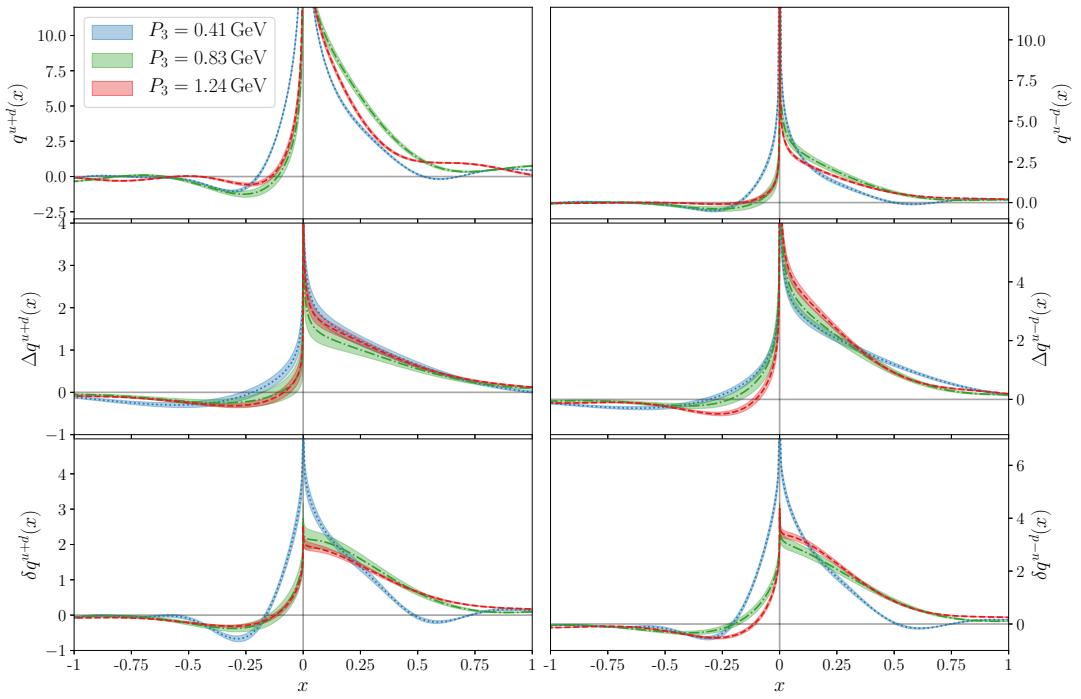


Figure 6.13: Results for the isoscalar (left) and isovector (right) unpolarized (first row), helicity (middle row) and transversity (bottom row) PDFs for different values of P_3. Each sub-figure shows the momentum dependence of the corresponding distribution, where the blue line corresponds to $P_3 = 0.41\,\mathrm{GeV}$, the green line to $P_3 = 0.83\,\mathrm{GeV}$, and the red one to $P_3 = 1.24\,\mathrm{GeV}$.

6.6.4 Flavor decomposition and comparison with phenomenology

Light quark distributions

Our results on the isoscalar and isovector distributions presented in Sec. 6.6.3 allow us to extract the up and down quark contributions for the unpolarized, helicity and transversity distributions. The disconnected contributions are taken into account in all cases. We stress that the comparison with phenomenology can only be qualitative for a number of reasons: i) We use an ensemble with larger than physical pion mass. We know from previous studies that there is a non-negligible pion mass dependence on the PDFs; ii) lattice systematics, such as cut-off effects, are not taken into account; iii) the renormalization ignores mixing present in the case of the unpolarized and helicity singlet PDFs; and iv) errors are still sizable and may hide systematics, such as convergence with the boost. However, it is still interesting to compare with phenomenology keeping these caveats in mind. The results for the unpolarized PDF at the largest momentum are compared with data by NNPDF3.1 [63], while the helicity distribution is compared with JAM17 [67] and $\mathrm{NNPDF}_{\mathrm{POL1.1}}$ [66]. Finally, the quark transversity distribution obtained in this study is compared against the SIDIS data [54] and SIDIS data constrained by the value of tensor charge g_T computed in lattice QCD [54]. For the anti-quark region for the NNPDF3.1 data, we include the crossing relations of Eq. (6.14), such that we show the antiquark distributions in the negative-x region. The light-quark contributions to

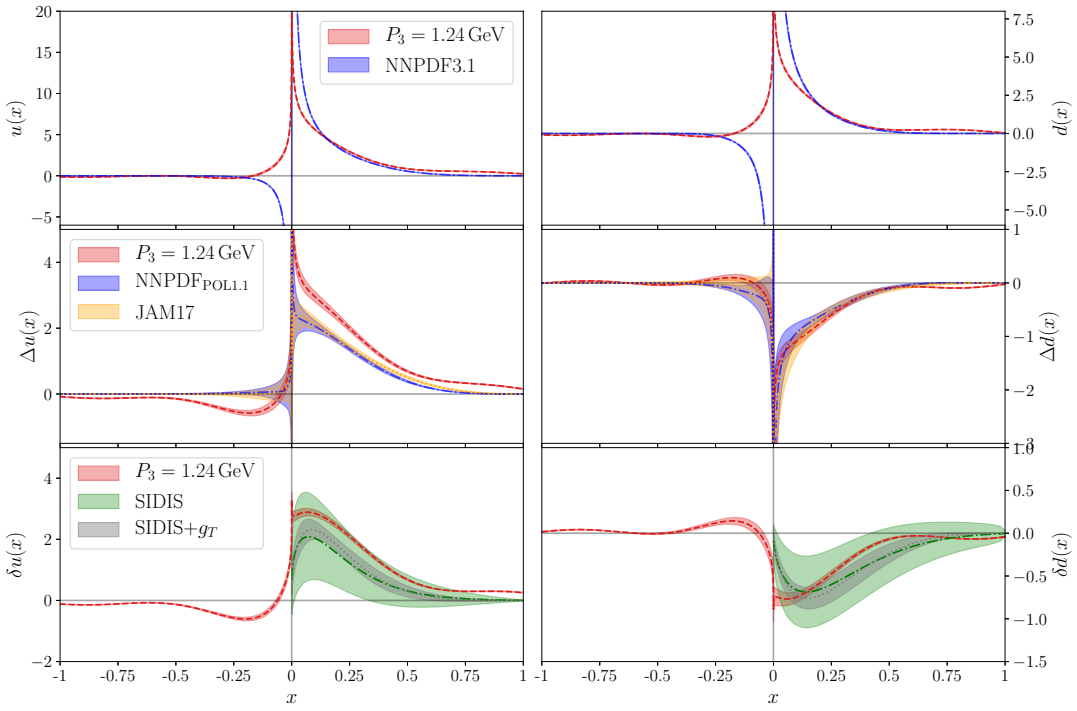


Figure 6.14: Up (left) and down (right) quark unpolarized (upper panels), helicity (middle panels) and transversity (bottom panels) distributions at $P_3 = 1.24$ GeV (red band). We also show the NNPDF *results [63, 66, 203] (blue band) and JAM17 [67] (orange band) phenomenological results. For the transversity PDF we compare against the SIDIS data [54] (green band) and SIDIS data constrained by the value of tensor charge g_T computed in lattice QCD [54] (gray band).*

the unpolarized PDF show good agreement with phenomenology in the region $x \gtrsim 0.2$. Also, the region $x \lesssim -0.2$ both estimates are compatible with zero. Note that lattice results for the small-x region ($|x| \lesssim 0.15$) suffer from uncontrolled uncertainties due to the reconstruction of the PDFs and the values of the lattice spacing used. The case of the helicity distributions is very interesting, as it has non-negligible contribution from the disconnected diagram. Our results for the up quark helicity show similar features as the NNPDF data, but are have higher values. The down quark distribution gives compatible results both with $\mathrm{NNPDF}_{\mathrm{POL1.1}}$ and JAM17 data for all x in the physical region $[-1, 1]$. The transversity distribution is the least known collinear PDF and it is not well-constrained by SIDIS data. As a result, global fits for the light quark $\delta q(x)$ carry large relative error of $\approx 50-100\%$ [54]. A more precise phenomenological estimate of the transversity PDFs can be obtained by constraining the distributions with the value of the tensor charge g_T computed within lattice QCD [54]. A comparison with the latter, reveals a similar agreement as for the helicity PDFs. We would like to stress that the overall qualitative agreement is very promising, as this computation is done using simulations with heavier than physical pions.

Strange quark distributions

The strange distributions presented here are computed using the renormalized matrix elements shown in Fig. 6.8. The values of $z_{\max}$ employed in the Fourier transform defining the quasi-PDF are reported in Table 6.8. The criterion adopted to select $z_{\max}$ is to analyze the dependence of the PDF as $z_{\max}$ is increased, as discussed in the previous section. In Fig. 6.15 we show the unpolarized, helicity and transversity PDFs. The anti-quark distribution reported here takes into account the crossing relations in Eq. (6.14), showing the anti-quark distributions in the negative x region. Although the unpolarized PDFs extracted from the matrix element using the two largest momenta tend towards the phenomenological result, there is still some residual dependence, which points to the need to increase the momentum boost to check the independence on P_3. Due to the simultaneous suppression of the real part of the matrix elements and the enhancement of the imaginary part, $\bar{s}(x)$ becomes symmetrical with respect to $x = 0$ as the momentum boost increases. This symmetry feature is exploited in the global fits. The results for the

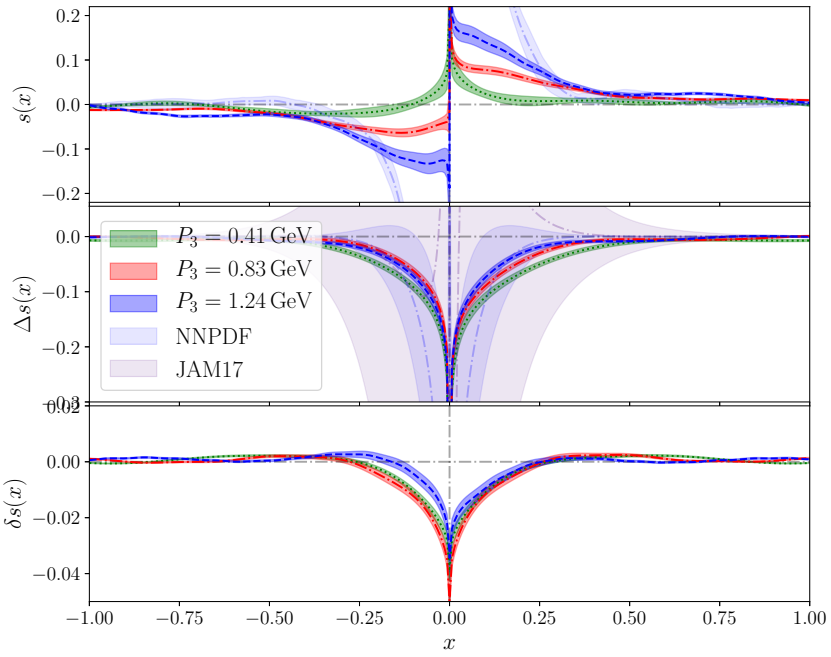


Figure 6.15: Results on the strange unpolarized (top panel), helicity (center panel) and transversity (bottom panel) distributions for three values of P_3. We compare with the $\mathrm{NNPDF}_{\mathrm{POL1.1}}$ *[66, 203] (light blue) and JAM17 [67] (light purple) phenomenological data. Lattice data for $P_3 = 0.41$, 0.83, 1.24 GeV are shown with green, red and dark blue bands, respectively.*

helicity distribution are approximately symmetric in the quark and antiquark regions, and are compatible with the results from the $\mathrm{NNPDF}_{\mathrm{POL1.1}}$ [66] and with JAM17 global fits analysis both of which have larger uncertainties. Our results, thus, provide valuable input for phenomenological studies. In fact, this is more evident for the strange transversity distribution where experimental results are lacking. We obtained results on the transversity PDF with small uncertainties that show no residual momentum dependence for the two largest momentum values.

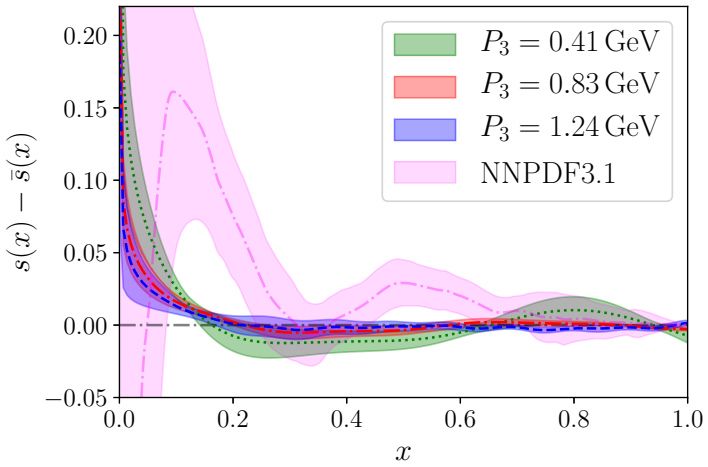


Figure 6.16: The strange-quark asymmetry for the unpolarized PDF for three values of P_3. We compare with NNPDF *[203] (pink) phenomenological data. Lattice data for $P_3 = 0.41$, 0.83, 1.24 GeV are shown with green, red and dark blue bands, respectively.*

Besides the individual $s(x)$ and $\bar{s}(x)$ distributions, there is also an interest on the strange-quark asymmetry. This is partly due to the fact that there is no symmetry to suggest that the two distributions have to be the same. The strange and anti-strange asymmetry has been discussed within chiral effective theory [204, 205], perturbative evolution of QCD [206], and a physical model for parton momenta [207]. Constraints on the $s(x) - \bar{s}(x)$ asymmetry have also been discussed in Ref. [208] using lattice results for the strange form factor with predictions based on a baryon-meson fluctuation model. Here, we study the asymmetry using our data for $P_3 = 0.41$, 0.83, 1.24 GeV, and the results are shown in Fig. 6.16. In contrast to the individual $s(x)$ and $\bar{s}(x)$ distributions, here we find that there is no momentum dependence in the strange-quark asymmetry. We also note that the difference between $s(x)$ and $\bar{s}(x)$ is a non-singlet combination and, thus, does not mix with the gluon PDFs. Focusing on the most accurate results at P_3, we find that the asymmetry vanishes at $x \gtrsim 0.2$ and is small but non negligible in the small-x region. This conclusion is, at present stage, qualitative, and an investigation of systematic effects is needed before drawing quantitative conclusions. We observe that our conclusions are similar to those of Ref. [208], that is, the $s(x) - \bar{s}(x)$ asymmetry is very small with its peak value less than 0.01.

6.6.5 Moments of nucleon PDFs

In this section, we calculate the moments $\langle x^n \rangle$ of the three PDFs considering $n = 0, ..., 3$. The $n-$th moment of the unpolarized, helicity and transversity distributions are defined

as

$$\begin{aligned}
\langle x^n \rangle_q &= \int_0^1 x^n \, \left[q(x) + (-1)^{n+1} \bar{q}(x) \right] \, dx = \int_{-1}^1 x^n \, q(x) \, dx, \\
\langle x^n \rangle_{\Delta q} &= \int_0^1 x^n \, \left[\Delta q(x) + (-1)^{n} \Delta \bar{q}(x) \right] \, dx = \int_{-1}^1 x^n \, \Delta q(x) \, dx, \\
\langle x^n \rangle_{\delta q} &= \int_0^1 x^n \, \left[\delta q(x) + (-1)^{n+1} \delta \bar{q}(x) \right] \, dx = \int_{-1}^1 x^n \, \delta q(x) \, dx,
\end{aligned} \tag{6.36}$$

where we employed the crossing relations of Eq. (6.14), to write the moments as a function of the quark distributions only. In Table 6.9 we report the results for the isovector, isoscalar and flavor diagonal moments. The zero-th moments are compatible with the nucleon charges reported in Tab 6.6. This is a non-trivial check, as the calculation of the charges follows a totally different procedure and undergoes a Fourier transform and matching.

PDF		$u-d$	$u+d$	u	d	s
Unpolarized	$\langle x \rangle_q$	0.28(1)	0.75(2)	0.51(2)	0.234(9)	0.030(2)
	$\langle x^2 \rangle_q$	0.118(5)	0.23(1)	0.176(8)	0.058(4)	-0.00054(46)
Helicity	$\langle 1 \rangle_{\Delta q}$	1.26(6)	0.50(6)	0.88(5)	-0.38(3)	-0.033(3)
	$\langle x \rangle_{\Delta q}$	0.49(2)	0.32(2)	0.40(2)	-0.087(9)	-0.00029(26)
	$\langle x^2 \rangle_{\Delta q}$	0.127(9)	0.067(7)	0.097(7)	-0.030(4)	-0.0019(4)
Transversity	$\langle 1 \rangle_{\delta q}$	1.06(4)	0.67(4)	0.86(3)	-0.20(2)	-0.0015(6)
	$\langle x \rangle_{\delta q}$	0.49(2)	0.33(2)	0.41(2)	-0.075(7)	-0.00038(37)
	$\langle x^2 \rangle_{\delta q}$	0.118(6)	0.086(5)	0.102(5)	-0.016(2)	0.00038(9)

Table 6.9: Moments of the unpolarized, helicity and transversity PDFs. We refer to the zero-th moment $\langle x^0 \rangle$ as $\langle 1 \rangle$.

Chapter 7

Continuum limit of nucleon quasi-PDFs

In this chapter, we present the results of the study [45] analyzing the approach to the continuum limit of isovector nucleon unpolarized and helicity parton distributions using three lattice ensembles, each having a different lattice spacing but with otherwise similar parameters. We use three lattice ensembles that differ primarily in their lattice spacings $a = 0.0644$, 0.0820, and 0.0934 fm. These have dynamical degenerate up and down quarks with pion mass approximately 370 MeV and dynamical strange and charm quarks with near-physical masses, i.e. $N_f = 2 + 1 + 1$. The gauge action is Iwasaki [185, 209] and the fermions use Wilson twisted mass tuned to maximal twist. These ensembles were generated by ETMC [210]; parameters for the three used in this work are given in Table 7.1. The ensemble with intermediate lattice spacing, B55, was previously used by some of us for studying quasi-PDFs in Refs. [18, 20, 103].

7.1 Excited-states contamination

In this study, we computed the ratio between the three- and two-point function at seven values of the source-sink separation for the ensemble A60. In Tab. 7.2 we reported the statistics required for each value of t_s. In Sec. 6.3.1 we introduced the analysis techniques that have been employed in this study to measure the excited states contamination to the quantities of interest. The results of the analysis for the unpolarized and helicity matrix elements have been reported respectively in Fig. 7.1 and 7.2. Although the helicity matrix elements show a much higher signal-to-noise ratio, especially in the real part, the two

Name	$a\mu_l$	size	a (fm)	m_π (MeV)	$p_z L/(2\pi)$	p_z (GeV)	t_s/a	t_s (fm)	N_{conf}	N_{samp}
A60	0.0060	$24^3 \times 48$	0.0934(13)(35)	365	3	1.66	10	0.934	1260	40320
B55	0.0055	$32^3 \times 64$	0.0820(10)(36)	373	4	1.89	12	0.984	1829	58528
D45	0.0045	$32^3 \times 64$	0.0644(07)(25)	371	3	1.80	15	0.966	1259	40288

Table 7.1: Parameters of the three $N_f = 2+1+1$ lattice ensembles: gauge coupling β, bare light quark mass $a\mu_l$, and size. The pion mass m_π and lattice spacing a (determined via the nucleon mass) are taken from Ref. [211]. Nucleon three-point functions are computed with momentum $\vec{p} = (0, 0, p_z)$ and source-sink time separation t_s. The total number of gauge configurations is given by N_{conf}; on each one, we use an evenly-spaced grid of 32 source positions, with a random overall displacement, yielding $N_{samp} = 32 N_{conf}$ samples.

t_s/a	N_{conf}	N_{src}	N_{samp}
$\{4,5,6,7\}$	315	4	1260
8	315	8	2520
9	315	16	5040
10	1260	32	40320

Table 7.2: Statistics used for excited-state study on ensemble A60.

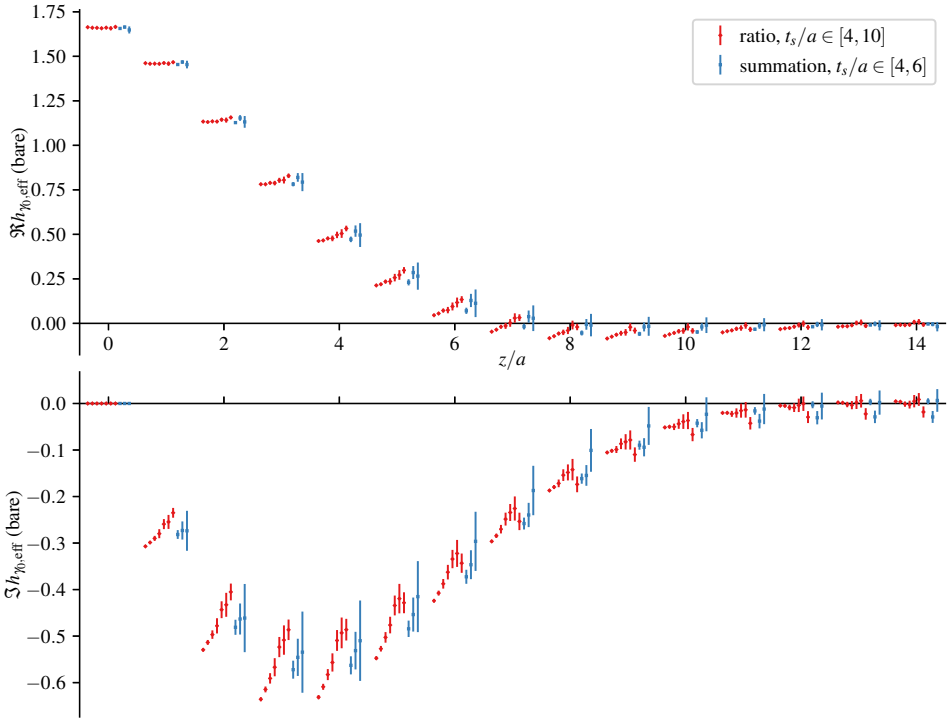


Figure 7.1: Unpolarized matrix elements for the A60 ensemble as a function of the Wilson line length z. The red points have been obtained using seven source-sink separations increasing from left to right. The blue points are obtained with the summation method, varying the left-end of the linear regression using the shortest three source-sink separations.

quantities show the same qualitative behavior with the source-sink separation. In most cases, the results obtained with the plateau fit method show a constant trend within uncertainties for $t_s \geq 8$, as well as compatibility with the summation method results. Therefore, it is well established that the excited-contamination is reasonably under control when considering the plateau fit results at the largest source-sink separation. To prevent the possibility that the excited states effect on the ensembles B55 and D45 may be larger than what observed with the A60, we chose slightly larger source-sink separations on the two ensemble with smaller lattice spacings, that lack an excited-states study.

7.2 Continuum limit

7.2.1 Renormalized matrix elements

Given the renormalized ratio between the three- and two-point function for the three ensembles employed in this study, we need to define how the continuum limit has to

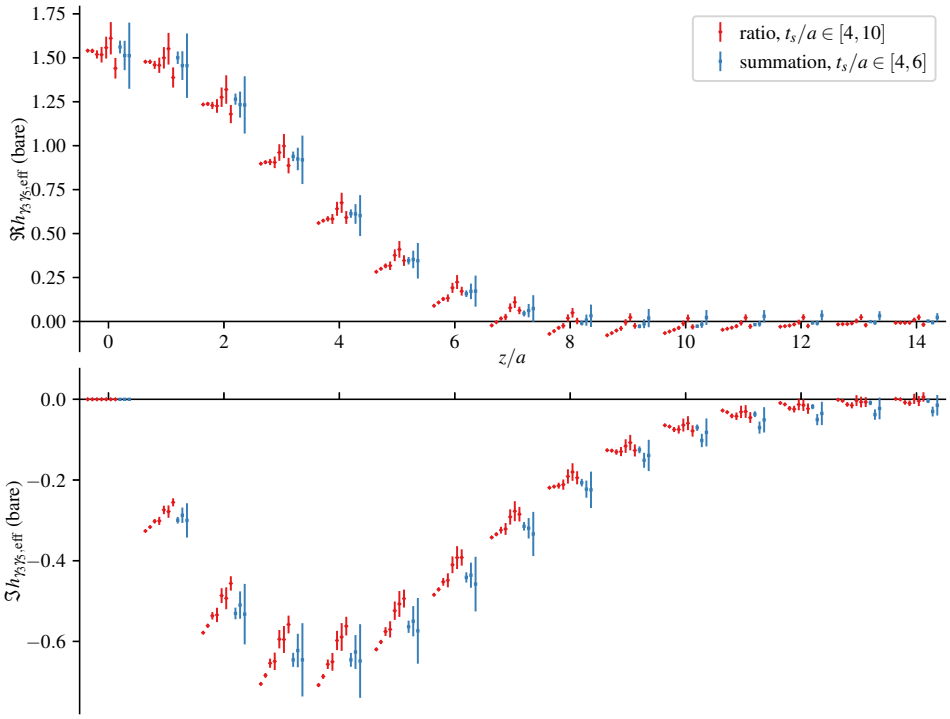


Figure 7.2: Helicity matrix elements for the A60 ensemble as a function of the Wilson line length z. The red points have been obtained using seven source-sink separations increasing from left to right. The blue points are obtained with the summation method, varying the left-end of the linear regression using the shortest three source-sink separations.

be taken. In the first place, the matrix elements corresponding to the three ensembles have slightly different boosts (see Tab. 7.1). To account at leading order for the small differences in p_z we interpolate the matrix elements at finite lattice spacing, extracting the lattice data at a common value of zp_z in physical units. In addition, different models for the continuum extrapolation can be employed. At zero Wilson line length, the matrix elements are defined from a local operator which, since we work at maximal twist, benefits from automatic $O(a)$ improvement [188, 212]. At $z \neq 0$, the operator is non-local and it can be contaminated by $O(a)$ contributions [190]. Therefore, at non-zero Wilson line length we extrapolate both with a and a^2 models.

In Fig. 7.3 we show the results of the continuum extrapolation for the unpolarized and helicity matrix elements. For most values of zp_z, there is a large dependence on the lattice spacing and the extrapolated values are far from those of the individual ensembles. The extrapolations tend to reduce the magnitude of the matrix element, except for the unpolarized case at large zp_z, where both the real and imaginary parts are positive and growing. For the unpolarized matrix element, the $O(a)$ and $O(a^2)$ extrapolations are not in good agreement, particularly in the real part at large zp_z and the imaginary part at medium zp_z. Details of these continuum extrapolations for selected values of zp_z are shown in Figs. 7.4 and 7.5.

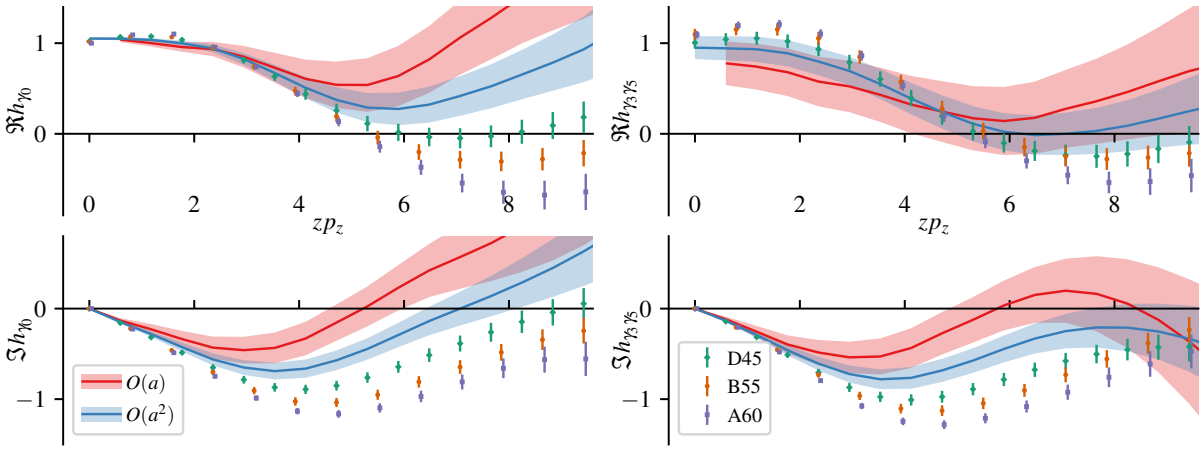


Figure 7.3: Real part (top) and imaginary part (bottom) of the unpolarized (left) and helicity (right) renormalized matrix elements, converted to the $\overline{MS}$ scheme at scale 2 GeV. The curves with error bands depict the results from the continuum extrapolations assuming leading artifacts linear in a (red) and quadratic in a (blue).

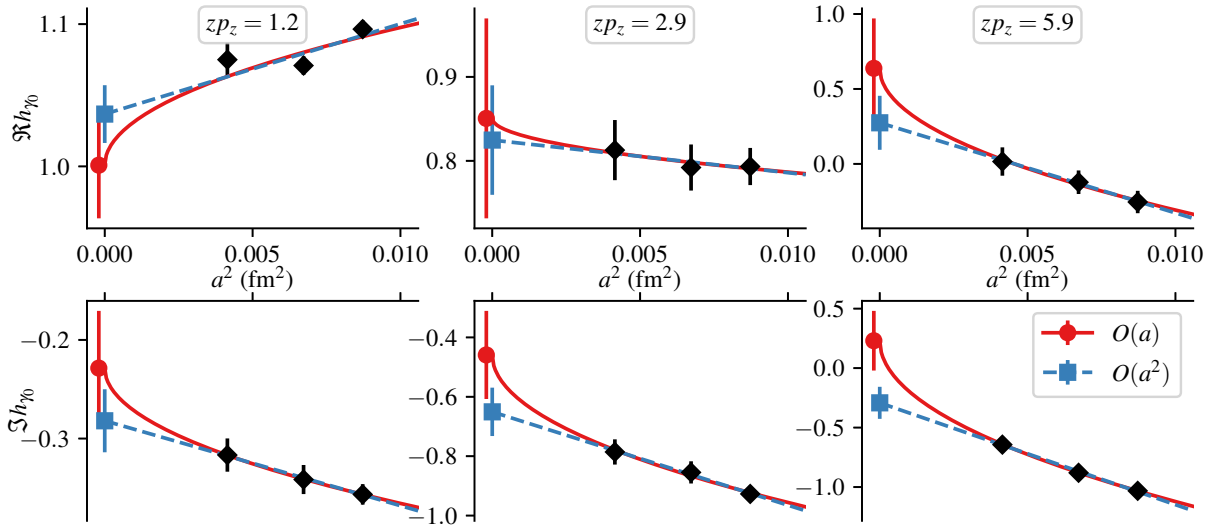


Figure 7.4: Continuum extrapolation of unpolarized matrix elements renormalized using the whole operator approach: real part (top) and imaginary part (bottom), versus a^2. Results are shown for $zp_z = 1.2$ (left), 2.9 (center), and 5.9 (right). Black diamonds show the data from ensemble D45 and interpolated data from ensembles B55 and A60. The solid red curve and red circle show the $O(a)$ extrapolation; the dashed blue line and blue square show the $O(a^2)$ extrapolation.

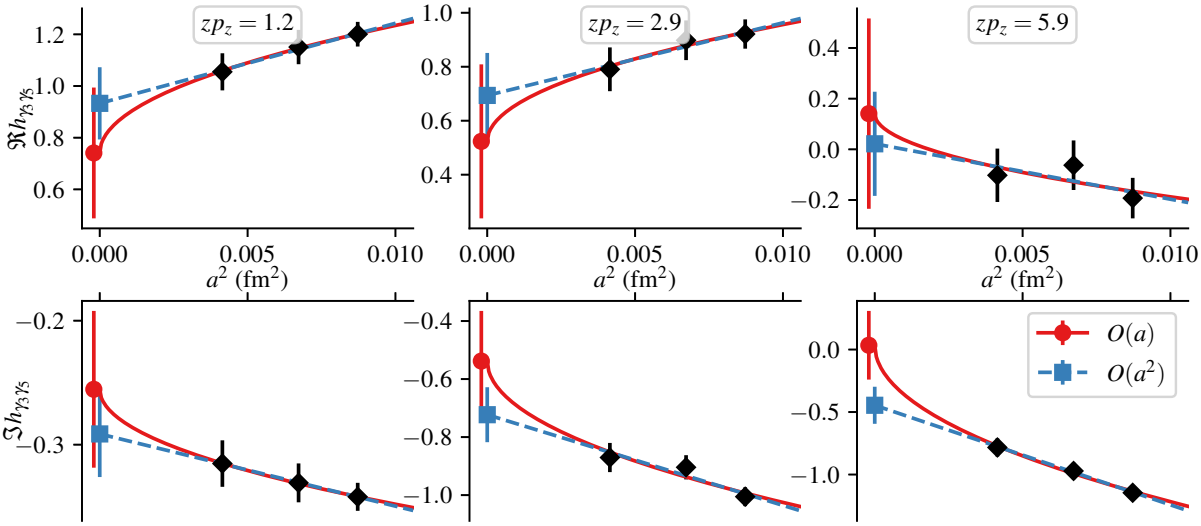


Figure 7.5: Continuum extrapolation of helicity matrix elements renormalized using the whole operator approach. See the caption of Fig. 7.4.

7.2.2 Comparison with phenomenology - Matrix elements

The ultimate goal of this analysis is to perform the matching of the continuum-extrapolated matrix elements obtaining the physical PDFs and comparing them with phenomenology. However, in this section, we show the results of the reverse exercise. Specifically, starting with phenomenological parton distributions determined by NNPDF [63,66], we invert the matching and the Fourier transform to determine the position-space matrix elements that yield those PDFs, up to higher-order corrections in the matching. Figure 7.6 compares this with the continuum-extrapolated lattice matrix elements. Full agreement cannot be expected, since the lattice calculation was done at a heavy pion mass and other systematics such as the dependence on p_z and finite-volume effects have not been included in this study.

The real part of the unpolarized matrix elements shows agreement with phenomenology up to $zp_z \leq 5$, while at larger Wilson line length the phenomenological curve goes to zero, while the continuum extrapolations diverge. The same qualitative behavior is observed in the imaginary part, with the NNPDF results being compatible with the $O(a)$ extrapolation in the small zp_z region and showing different asymptotic behavior. Both the $O(a)$ and $O(a^2)$ extrapolations of the real part of the helicity matrix element lay below the phenomenological results in the small zp_z region. This behavior can be partly understood by recalling that at heavy pion masses, the nucleon axial charge (i.e. the helicity matrix element at $z = 0$) lies below its physical value. The extrapolated imaginary parts of the unpolarized and helicity matrix elements show larger slopes close to $z = 0$ and a faster decay to zero compared to phenomenology. However, comparing with Fig. 7.3, it is evident that the continuum extrapolation produces results that lie closer to phenomenology.

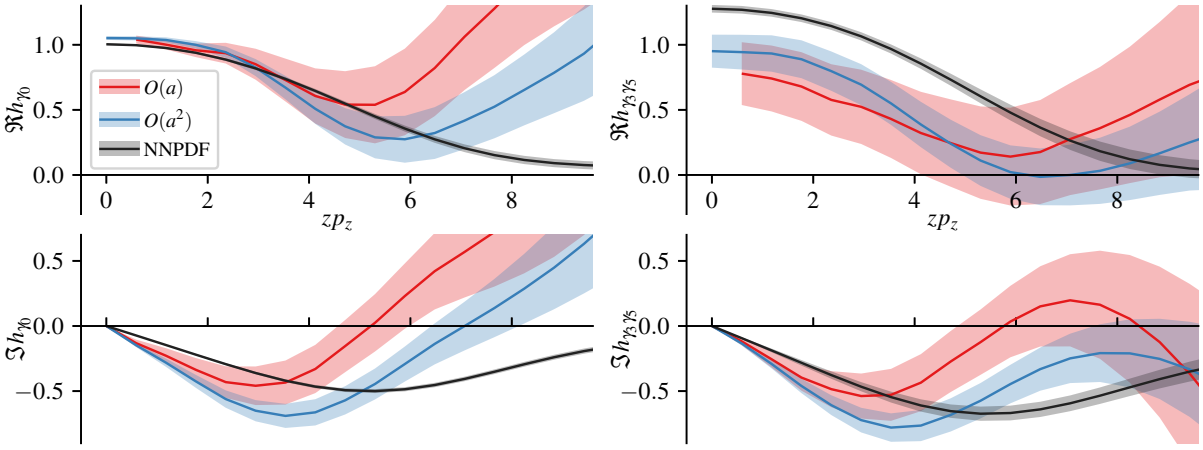


Figure 7.6: Unpolarized (left) and helicity (right) matrix elements from continuum extrapolation of lattice data renormalized using the whole operator approach via the RI′-MOM intermediate scheme (blue, red) and from the inverse Fourier transform of the quasi-PDFs obtained by applying inverse matching to phenomenological PDFs from NNPDF [63, 66] (dark gray). Note that in the lattice calculation, the pion mass is much larger than in nature, so that exact agreement should not be expected.

7.2.3 Parton distributions

In this section, we present the main results of this study, namely the effect of the continuum extrapolation on PDFs. However, we first discuss another source of systematic uncertainty: how to perform the Fourier transform in the definition of the quasi-PDF using a finite set of position-space data. We illustrate this using data on the finest ensemble, D45. Next, we perform the continuum extrapolation at fixed x, using the PDFs determined on each ensemble, and compare the result with the PDF determined from the continuum-limit matrix elements obtained in the previous section. Finally, we compare our continuum-limit PDFs with phenomenology.

In Chap. 5 we discussed how the Fourier transform introduces a systematic uncertainty. For this reason, we supplement the naïve truncated FT with more sophisticated reconstruction techniques [48, 117]. In these approaches, obtaining the Fourier transform from a finite number of data points is seen as an ill-defined inverse problem. Its solution is not unique and one approach is to use explicit models for the shape of the (quasi-)PDF. By contrast, we choose to use two approaches that do not contain an explicit model: the Bayes-Gauss-Fourier Transform (BGFT) [48] which was discussed in depth in Chap. 5 and the Backus-Gilbert method, first applied for PDFs calculations in [117]. The latter address the reconstruction problem as follows.

Backus-Gilbert (BG) The inverse problem is obtained by inverting Eq. (3.70) to write the real and imaginary parts of the unpolarized matrix element in terms of the

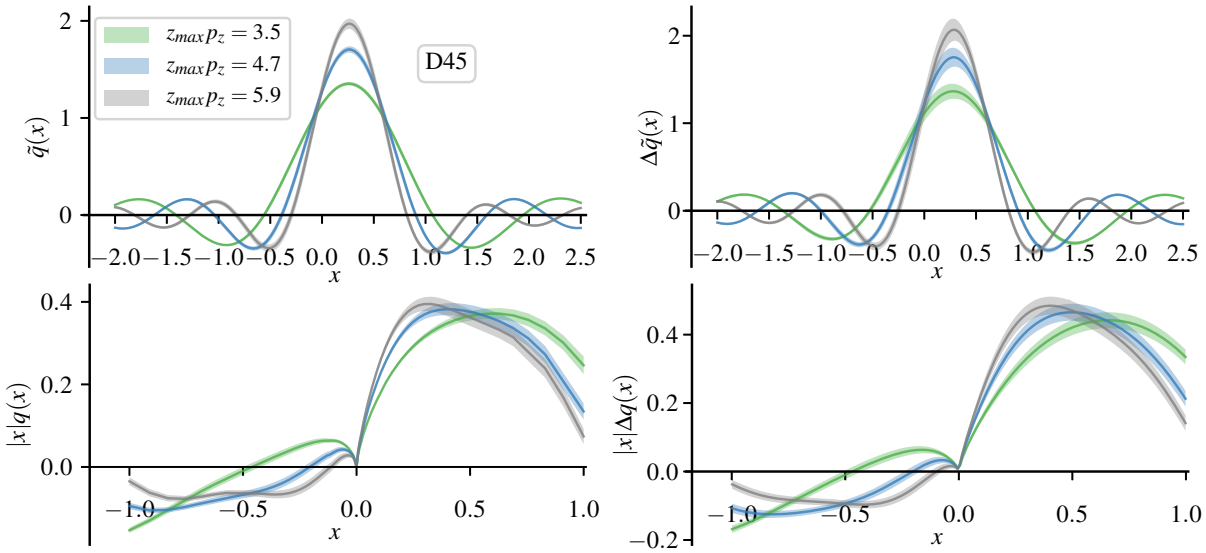


Figure 7.7: Unpolarized (left) and helicity (right) quasi-PDFs (top panels) and PDFs (bottom panels) of the D45 ensemble for different values of the cutoff $z_{\max}$. The curves with cutoff at $z_{\max}p_z = \{3.5, 4.7, 5.9\}$ are depicted in green, blue and gray.

quasi-PDF:

$$\begin{aligned} \Re h_{\gamma_0}(p_z, z; \mu) &= \int_0^\infty dx\, \cos(x p_z z)\, \tilde{q}_+(x, p_z; \mu), \\ \Im h_{\gamma_0}(p_z, z; \mu) &= \int_0^\infty dx\, \sin(x p_z z)\, \tilde{q}_-(x, p_z; \mu), \end{aligned} \tag{7.1}$$

where for $x \geq 0$, $\tilde{q}_\pm(x) = \tilde{q}(x) \pm \tilde{q}(-x)$, and likewise for the helicity case[1]. The reconstruction is applied independently to $\tilde{q}_+$ and $\tilde{q}_-$, so for brevity we describe the procedure applied to $\tilde{q}_+$. We also omit the labels p_z and μ. For each x, the solution is assumed to be a linear combination of the finite set of lattice data:

$$\tilde{q}_+^{\mathrm{BG}}(x) = \sum_{z/a=0}^{z_{\max}/a} \mathbf{a}_+(x, z) \Re h_{\gamma_0}(z), \tag{7.2}$$

where $\mathbf{a}_+$ can be understood as an approximation to the inverse of the Fourier transform in Eq. (7.1). The accuracy of this approximation is governed by the function

$$\Delta_+(x, x') = \sum_{z/a=0}^{z_{\max}/a} \mathbf{a}_+(x, z) \cos(x' p_z z) \tag{7.3}$$

that approximates $\delta(x - x')$. Specifically, the result is an integral over the quasi-PDF:

$$\tilde{q}_+^{\mathrm{BG}}(x) = \int_0^\infty dx'\, \Delta_+(x, x') \tilde{q}_+(x'). \tag{7.4}$$

The function $\mathbf{a}_+$ is determined by the Backus-Gilbert procedure [213], which minimizes the width of $\Delta_+(x, x')$. For more details, see Refs. [117, 161].

[1]For the unpolarized case, this is not the same as the convention commonly used for PDFs, where $q_\pm(x) \equiv q(x) \pm \bar{q}(x) = q(x) \mp q(-x)$. For helicity, $\Delta q_\pm(x) \equiv \Delta q(x) \pm \Delta\bar{q}(x) = \Delta q(x) \pm \Delta q(-x)$.

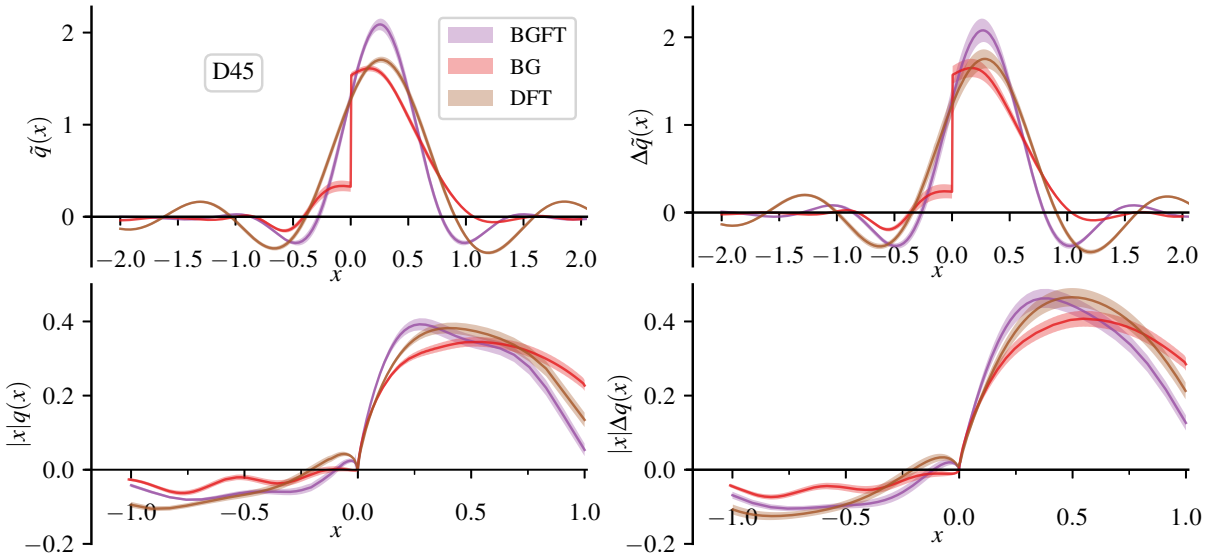


Figure 7.8: Comparison of quasi-PDFs (top panels) and PDFs (bottom panels) of the D45 ensemble obtained from Bayes-Gauss-Fourier Transform (BGFT), Backus-Gilbert (BG) and discrete FT (DFT) for the unpolarized (left) and helicity (right).

In Fig. 7.8, we compare results from the truncated discrete Fourier transform, Eq. (5.1), and the BG and BGFT reconstruction methods described above, again using ensemble D45 as our reference data set. For a fair comparison, in all cases we use $z_{\max}p_z = 4.7$. We begin by discussing the quasi-PDFs (upper two panels). The most striking difference is that the Backus-Gilbert result has a discontinuity at $x = 0$ that is not present in the other results. This is because $\tilde{q}^{\rm BG}_{-}(x)$ is not constrained to vanish at $x = 0$. Such a discontinuity could occur if $\Im h(z)$ has a slowly decaying tail $\sim 1/z$. For x between -0.5 and 1.0, the DFT and BGFT results are similar, although the BGFT distribution is slightly narrower. For larger values of $|x|$, the DFT produces stronger oscillations, which are suppressed by the BGFT. The BG result is the outlier, being considerably smaller at small negative x and also having a smaller dip below zero.

We next discuss the physically relevant parton distributions, obtained after matching and nucleon mass corrections (lower two panels). For most values of x, the DFT and BGFT method produce very similar results, although for BGFT the the dip below zero in the antiquark region occurs at smaller negative x and the magnitude is smaller at $x = -1$ and $+1$. Again, the BG result is somewhat different: in the antiquark region at small negative x, the small positive bump is gone and the result is either consistent with zero (unpolarized) or slightly negative (helicity). This discrepancy at small x may be associated with a lack of data for the matrix element at large $|z|$; better data or a more rigorous understanding of the large-$|z|$ behavior could help to improve this situation. In the quark region for x greater than about 0.5, the BG result has a much weaker downward trend than the other two methods. Given that the DFT produces a result not substantially different from BGFT, we exclude the DFT from further analyses presented in the next sections.

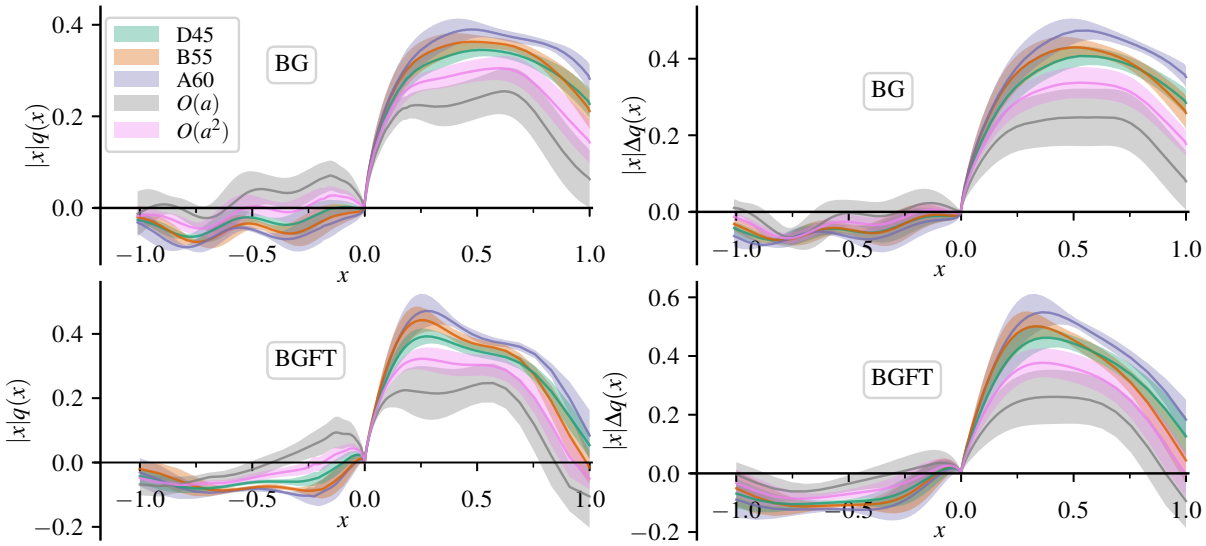


Figure 7.9: Matched unpolarized (left) and helicity (right) PDFs obtained using the gauge ensembles A60 (blue), B55 (orange), D45 (green), whose lattice spacings are reported in Table 7.1. The PDF in the continuum, after $O(a)$ extrapolation (gray) and $O(a^2)$ extrapolation (pink), is also shown.

7.2.4 Continuum extrapolation

In what follows, we compare the distributions at finite lattice spacings with continuum extrapolations. In the reconstruction of the quasi-PDFs we use the lattice data with $|zp_z| \leq z_{\max}p_z = 4.7$, at which point either the real part or the imaginary part of the continuum matrix element is compatible with zero, as shown in Fig. 7.3. Moreover, we estimate the systematic uncertainty from this choice of the cutoff by varying $z_{\max}$:

$$\varepsilon_{\text{cutoff}}(x) = \frac{|q_{z_{\max}p_z=5.9}(x) - q_{z_{\max}p_z=3.5}(x)|}{2}. \tag{7.5}$$

Finally, we estimate the combined uncertainty as the quadrature sum of $\varepsilon_{\text{cutoff}}(x)$ and the statistical uncertainty.

One approach for obtaining continuum-limit PDFs is to take the PDF determined on each ensemble and then perform an $O(a)$ or $O(a^2)$ extrapolation of the data at each x. This is shown in Fig. 7.9, for both unpolarized and helicity PDFs determined using the BG and BGFT methods. In the quark region with x between roughly 0 and 0.7, the PDFs decrease monotonically with the lattice spacing; at larger x, the D45 data (with the finest lattice spacing) move relatively upward to lie between those of the other two ensembles. For all $x > 0$, the $O(a^2)$ extrapolation lies below all of the individual lattice spacings and the $O(a)$ extrapolation is even lower. Using the BGFT approach, both of the extrapolations are consistent with the expected value of zero at $x = 1$, whereas for BG, this is true only of the $O(a)$ extrapolation. In the antiquark region, the extrapolated results lie above the PDFs determined at finite lattice spacing, except for the BGFT unpolarized distribution near $x = -1$. This produces a more prominent positive region at small negative x, particularly in the unpolarized case. At larger negative x, the

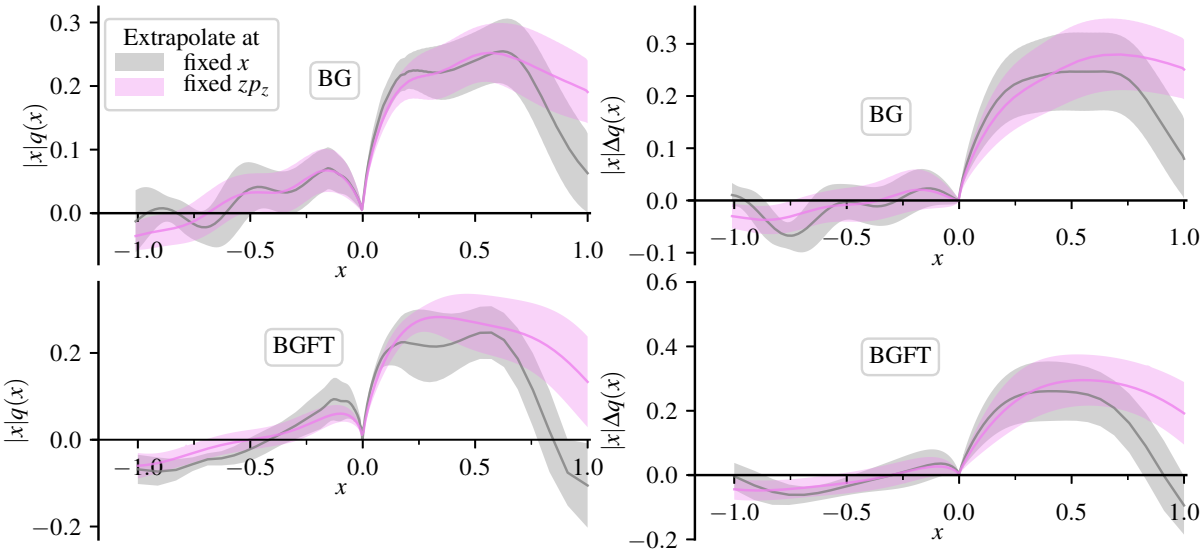


Figure 7.10: Comparison between the results for the unpolarized (left) and helicity (right) PDFs in the continuum limit obtained with the $O(a)$ extrapolation at fixed x (gray; see Fig. 7.9) and at fixed zp_z (pink, based on the continuum-limit data in Fig. 7.3). The distributions has been obtained using the BG (top panels) and BGFT (bottom panels) reconstruction techniques.

extrapolations are generally closer to zero.

Another approach is to obtain PDFs from the continuum limit of $h(z)$ as determined in Section 7.2.1 by extrapolating data at fixed zp_z. By changing the order in which the continuum limit and the combination of the Fourier transform and PDF matching are performed, we obtain results affected by different systematic effects. The comparison of the $O(a)$ extrapolations from both approaches is shown in Fig. 7.10. They are consistent within uncertainties, except near $x = 1$, where the fixed-x extrapolation is in all cases lower than the fixed-zp_z extrapolation and only the former is consistent with zero at $x = 1$.

For comparing with phenomenology in the next section, we take the fixed-x extrapolation as our central value and add an additional systematic uncertainty in quadrature, namely half the difference with the fixed-zp_z extrapolation.

7.2.5 Comparison with phenomenology - PDFs

In Fig. 7.11, we compare the distributions obtained using $O(a)$ and $O(a^2)$ extrapolations with those obtained from phenomenology by NNPDF [63, 66]. This comparison is intended to be qualitative, since our calculation was not done at the physical pion mass and does not include a study of other sources of systematic uncertainty such as finite-volume effects or the dependence on p_z.

In the antiquark region ($x < 0$), the NNPDF result is slightly positive for $x > -0.25$, particularly in the unpolarized case. Focusing on the latter case, both of the extrapolations using both BG and BGFT methods reproduce this feature, although the $O(a)$ extrapolation (which has a larger uncertainty) prefers a wider and larger positive region. This agreement with NNPDF is only present after the continuum extrapolation

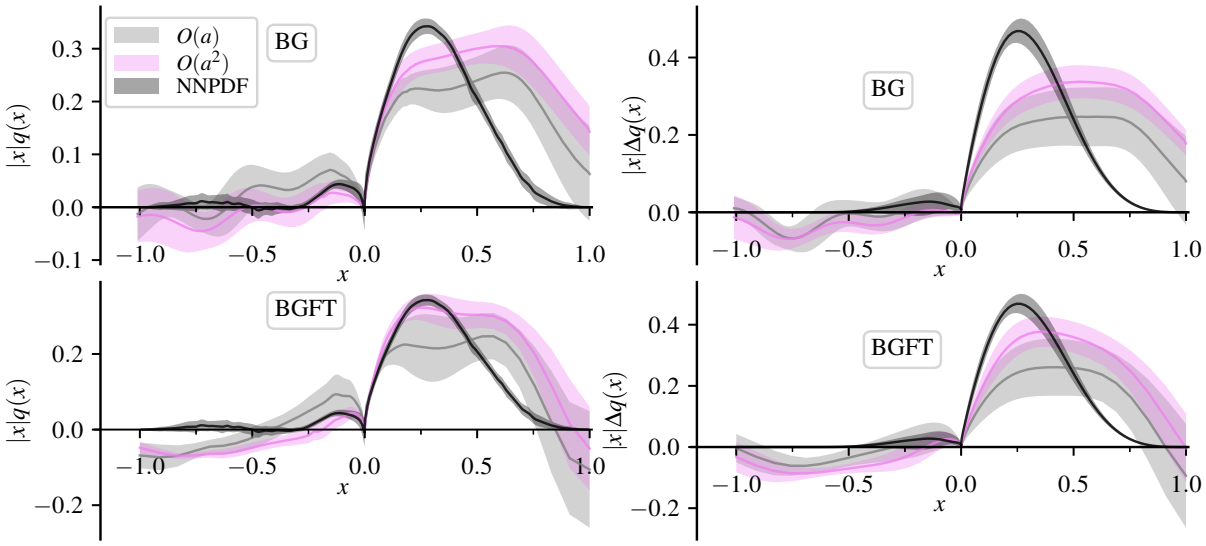


Figure 7.11: Unpolarized (left) and helicity (right) distributions in the continuum, using BG (top) and BGFT (bottom) methods. $O(a)$ and $O(a^2)$ extrapolations are shown in gray and pink, respectively. PDFs extracted through global fits from the releases NNPDF [63, 66] (dark gray) are included for qualitative comparison.

and does not appear in the analyses of any of the individual ensembles. For larger negative x, the NNPDF distributions are close to zero. However, the BGFT result is below zero, particularly when using an $O(a^2)$ extrapolation.

In the quark region ($x > 0$), the distributions obtained from our data tend to have smaller peaks at larger x than phenomenology and fall off more slowly at large x. All of the analyses are consistent with zero at $x = 1$, except for the $O(a^2)$-extrapolated BG data. For small x, the lattice unpolarized distributions are consistent with phenomenology, whereas the lattice helicity distributions have smaller slopes. In the unpolarized case, the agreement holds for a wider range of x when using the BGFT approach, and this approach also produces less disagreement in the helicity distribution.

Chapter 8

Conclusions

A number of different approaches are being developed to evaluate PDFs from first principles. Some of them, including quasi-PDFs, are now in an advanced status, where the methodology is consolidated and many exploratory studies have been performed.

A major outcome of this work is to demonstrate that one can compute the flavor decomposition of the proton unpolarized, helicity and transversity distributions. To accomplish this task we use algorithmic developments and access to large computational resources to perform the first computation of the disconnected diagrams with extended operators contributing to the quark isoscalar $u + d$ and strange matrix elements. We obtain results using a twisted mass fermion ensemble with a pion mass of $m_\pi \approx 260$ MeV. We find that the light-quark disconnected contributions are largest in the case of the helicity distribution. However, a clear non-zero signal is obtained also for the unpolarised and transversity PDFs. The strange-quark PDFs are also nonzero up to $x \sim 0.5$ for all all PDFs with largest contributions for the unpolarized and helicity PDFs. For the transversity they suppressed by an order of magnitude. These distributions are very challenging to extract from experimental data due to the lack of sensitivity to the strange quark. In a qualitative comparison of our results with phenomenologically extracted PDFs we find: i) our results on the unpolarized have a statistical precision which is similar to the NNPDF data; ii) the helicity strange-quark PDF is significantly more accurate than the JAM and NNPDF results and iii) our results for the strange-quark transversity PDF serve as a prediction. Although there are a number of systematic effects that still need to be addressed, this work clearly demonstrates the great potential in the extraction of the x-dependence of individual quark PDFs from lattice QCD.

A number of systematic effects are still to be studied in depth. Another aim of this work was to address some of the lattice artifacts. One such systematic error arises from the finite lattice spacing. How large cut-off effects are was not known at the beginning of this work. In our study we use three twisted mass fermion ensembles simulated with pion mass $m_\pi \approx 370\,\text{MeV}$ and different lattice spacings to investigate the the continuum limit of the unpolarized and helicity distributions. We observed significant discretization effects both in the position-space matrix elements and in the final parton distribution functions. In particular, in the antiquark region, the continuum limit brings better agreement with what is expected from phenomenology. This finding suggests that the

discretization effects are to a large extend responsible for the observed discrepancy between phenomenology and lattice QCD results at finite lattice spacing in the negative-x region.

Another issue arising from the discretization of space-time is connected to taking the Fourier transform using a discrete data set. The naive discrete Fourier transform (DFT), which is usually employed to overcome this limitation, may introduce large artifacts in the resulting quasi-PDF thus preventing in principle the light-cone distribution from reaching a good agreement with phenomenology. In this work, we implemented a new data-driven approach, the Bayes-Gauss Fourier transform, that is proven successful in reducing the unphysical oscillations in the quasi-PDF that arise due to the truncation of the Fourier transform. However, the final results on the light-cone distribution show little improvement compared to the standard DFT, suggesting that the truncation and discretization of the FT may not be the sole cause of disagreement with phenomenology. The BGFT technique is also employed in the study of the continuum limit and compared with the well-established Backus-Gilbert method [213], showing agreement within uncertainties.

In conclusion, the results presented in this thesis constitute a significant step forward for a more precise determination of PDFs through lattice QCD simulations.

Appendices

Chapter A

Analytic Fourier transform of the GPR prediction

A useful feature of the GPR is the possibility to perform analytically the FT of the posterior mean, obtaining an improved stability and performance compared to what is achievable with a numerical integration.

The FT definition that we adopt is the following:

$$\mathcal{F}[f(z)](x) \equiv \frac{1}{2\pi}\int_{-\infty}^{\infty} dz\, e^{ixz} f(z) \tag{A.1a}$$

$$\mathcal{F}^{-1}[F(x)](z) \equiv \int_{-\infty}^{\infty} dx\, e^{-ixz} F(x) \tag{A.1b}$$

If $\mathcal{T}$ is the integral transform defined by Eq. (3.70), it is possible to write $\mathcal{T}$ and $\mathcal{T}^{-1}$ in terms of the FT of Eq. (A.1a) as follows

$$\mathcal{T}[h(z)](x) = P_3 \mathcal{F}[h(z)](P_3 x) \tag{A.2a}$$

$$\mathcal{T}^{-1}[H(x)](z) = \mathcal{F}^{-1}[H(x)](P_3 z) \tag{A.2b}$$

Thus, after performing the fit, it is possible to estimate the quasi-PDF by computing the FT of the fit function of Eq. (5.19) using the convention of Eq. (A.1a), and then by evaluating it using Eq. (A.2a).

When computing the FT, the phase of Eq. 5.19 simply corresponds to a shift in the FT:

$$\mathcal{F}[h^{\text{fit}}(z)](x) = \mathcal{F}[\mu(|z|)](x+\theta)$$

In order to compute the FT of $\mu(|z|)$, it is useful to observe that Eq. (5.12) is just a linear combination of the covariance function reported in Eq. (5.13):

$$\mu(|z|) = \mu_P(|z|) + \sum_i w_i k_P(|z|, z_i),$$

where $w_i \equiv \sum_j \widetilde{K}^{-1}_{ij}(h_j - \mu_P(z_j))$ and $\mu_P(z)$ is a generic prior mean assigned to the absolute value of the renormalized matrix element.

The FT of $k_P(|z|, z_i)$ is available in closed form given by

$$\begin{aligned}\mathcal{F}[k_P(|z|, z_i)](x) =& \frac{e^{-ixz_i-\ell^2x^2/2}}{2\sqrt{2\pi}/\ell}\left(1 + e^{2ixz_i} + \text{erf}\left(\frac{z_i/\ell - i\ell x}{\sqrt{2}}\right) + \right. \\ &\left. + e^{2ixz_i}\text{erf}\left(\frac{z_i/\ell - i\ell x}{\sqrt{2}}\right)\right).\end{aligned} \tag{A.3}$$

Thus, the quasi-PDF transform of the fit is

$$\begin{aligned}\mathcal{T}[h^{\text{fit}}(z)](x) =& P_3\mathcal{F}[\rho_P(|z|)](P_3(x+\theta)) \\ &+ P_3\sum_i w_i\mathcal{F}[k_P(|z|, z_i)](P_3(x+\theta))\end{aligned} \tag{A.4}$$

If the chosen prior mean $\rho_P(|z|)$ has a known analytical FT, then the quasi-PDF transform of the fit is available in closed form.

Bibliography

[1] Jülich Supercomputing Centre, "JURECA: Modular supercomputer at Jülich Supercomputing Centre," *J. Large-Scale Res. Facil.* **4** (2018) A132.

[2] P. A. Boyle, G. Cossu, A. Yamaguchi, and A. Portelli, "Grid: A next generation data parallel C++ QCD library," *PoS* **LATTICE2015** (2016) 023, arXiv:1512.03487 [hep-lat].

[3] A. Frommer, K. Kahl, S. Krieg, B. Leder, and M. Rottmann, "Adaptive aggregation-based domain decomposition multigrid for the lattice Wilson–Dirac operator," *SIAM J. Sci. Comput.* **36** (2014) A1581–A1608, arXiv:1303.1377 [hep-lat].

[4] C. Alexandrou, S. Bacchio, J. Finkenrath, A. Frommer, K. Kahl, and M. Rottmann, "Adaptive aggregation-based domain decomposition multigrid for twisted mass fermions," *Phys. Rev. D* **94** (2016) 114509, arXiv:1610.02370 [hep-lat].

[5] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, *et al.*, "Array programming with NumPy," *Nature* **585** (2020) 357–362, arXiv:2006.10256 [cs.MS].

[6] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, J. Weckesser, W. Bright, S. J. van der Walt, *et al.*, "SciPy 1.0: fundamental algorithms for scientific computing in Python," *Nature Methods* **17** (2020) 261–272, arXiv:1907.10121 [cs.MS].

[7] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Comput. Sci. Eng.* **9** no. 3, (2007) 90–95.

[8] X. Ji, "Parton Physics on a Euclidean Lattice," *Phys. Rev. Lett.* **110** (2013) 262002, arXiv:1305.1539 [hep-ph].

[9] A. V. Radyushkin, "Quasi-parton distribution functions, momentum distributions, and pseudo-parton distribution functions," *Phys. Rev. D* **96** no. 3, (2017) 034025, arXiv:1705.01488 [hep-ph].

[10] K. Orginos, A. Radyushkin, J. Karpie, and S. Zafeiropoulos, "Lattice QCD exploration of parton pseudo-distribution functions," *Phys. Rev. D* **96** no. 9, (2017) 094503, arXiv:1706.05373 [hep-ph].

[11] **Lattice Parton** Collaboration, Y.-S. Liu *et al.*, "Unpolarized isovector quark distribution function from lattice QCD: A systematic analysis of renormalization and matching," *Phys. Rev. D* **101** no. 3, (2020) 034020, arXiv:1807.06566 [hep-lat].

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

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

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

[15] C. Alexandrou, K. Cichy, M. Constantinou, K. Jansen, A. Scapellato, and F. Steffens, "Transversity parton distribution functions from lattice QCD," *Phys. Rev. D* **98** no. 9, (2018) 091503, arXiv:1807.00232 [hep-lat].

[16] C. Alexandrou, K. Cichy, V. Drach, E. Garcia-Ramos, K. Hadjiyiannakou, K. Jansen, F. Steffens, and C. Wiese, "First results with twisted mass fermions towards the computation of parton distribution functions on the lattice," *PoS* **LATTICE2014** (2014) 135, arXiv:1411.0891 [hep-lat].

[17] H.-W. Lin, J.-W. Chen, S. D. Cohen, and X. Ji, "Flavor Structure of the Nucleon Sea from Lattice QCD," *Phys. Rev. D* **91** (2015) 054510, arXiv:1402.1462 [hep-ph].

[18] C. Alexandrou, K. Cichy, V. Drach, E. Garcia-Ramos, K. Hadjiyiannakou, K. Jansen, F. Steffens, and C. Wiese, "Lattice calculation of parton distributions," *Phys. Rev. D* **92** (2015) 014502, arXiv:1504.07455 [hep-lat].

[19] J.-W. Chen, S. D. Cohen, X. Ji, H.-W. Lin, and J.-H. Zhang, "Nucleon Helicity and Transversity Parton Distributions from Lattice QCD," *Nucl. Phys. B* **911** (2016) 246–273, arXiv:1603.06664 [hep-ph].

[20] C. Alexandrou, K. Cichy, M. Constantinou, K. Hadjiyiannakou, K. Jansen, F. Steffens, and C. Wiese, "Updated Lattice Results for Parton Distributions," *Phys. Rev. D* **96** no. 1, (2017) 014513, arXiv:1610.03689 [hep-lat].

[21] Y. Chai *et al.*, "Parton distribution functions of Δ^+ on the lattice," *Phys. Rev. D* **102** no. 1, (2020) 014508, arXiv:2002.12044 [hep-lat].

[22] R. S. Sufian, C. Egerer, J. Karpie, R. G. Edwards, B. Joó, Y.-Q. Ma, K. Orginos, J.-W. Qiu, and D. G. Richards, "Pion Valence Quark Distribution from Current-Current Correlation in Lattice QCD," *Phys. Rev. D* **102** no. 5, (2020) 054508, arXiv:2001.04960 [hep-lat].

[23] J.-W. Chen, H.-W. Lin, and J.-H. Zhang, "Pion generalized parton distribution from lattice QCD," *Nucl. Phys. B* **952** (2020) 114940, arXiv:1904.12376 [hep-lat].

[24] T. Izubuchi, L. Jin, C. Kallidonis, N. Karthik, S. Mukherjee, P. Petreczky, C. Shugert, and S. Syritsyn, "Valence parton distribution function of pion from fine lattice," *Phys. Rev. D* **100** no. 3, (2019) 034516, arXiv:1905.06349 [hep-lat].

[25] J.-H. Zhang, J.-W. Chen, L. Jin, H.-W. Lin, A. Schäfer, and Y. Zhao, "First direct lattice-QCD calculation of the x-dependence of the pion parton distribution function," *Phys. Rev. D* **100** no. 3, (2019) 034505, arXiv:1804.01483 [hep-lat].

[26] R. S. Sufian, J. Karpie, C. Egerer, K. Orginos, J.-W. Qiu, and D. G. Richards, "Pion Valence Quark Distribution from Matrix Element Calculated in Lattice QCD," *Phys. Rev. D* **99** no. 7, (2019) 074507, arXiv:1901.03921 [hep-lat].

[27] X.-D. Ji, "Gauge-Invariant Decomposition of Nucleon Spin," *Phys. Rev. Lett.* **78** (1997) 610–613, arXiv:hep-ph/9603249.

[28] A. V. Radyushkin, "Scaling limit of deeply virtual Compton scattering," *Phys. Lett. B* **380** (1996) 417–425, arXiv:hep-ph/9604317.

[29] M. Diehl, "Generalized parton distributions," *Phys. Rept.* **388** (2003) 41–277, arXiv:hep-ph/0307382.

[30] X. Ji, "Generalized parton distributions," *Ann. Rev. Nucl. Part. Sci.* **54** (2004) 413–450.

[31] C. Alexandrou, K. Cichy, M. Constantinou, K. Hadjiyiannakou, K. Jansen, A. Scapellato, and F. Steffens, "Unpolarized and helicity generalized parton distributions of the proton within lattice QCD," *Phys. Rev. Lett.* **125** no. 26, (2020) 262001, arXiv:2008.10573 [hep-lat].

[32] C. Alexandrou, K. Cichy, M. Constantinou, K. Hadjiyiannakou, K. Jansen, A. Scapellato, and F. Steffens, "Transversity GPDs of the proton from lattice QCD," arXiv:2108.10789 [hep-lat].

[33] J. C. Collins and D. E. Soper, "Back-To-Back Jets in QCD," *Nucl. Phys. B* **193** (1981) 381. [Erratum: Nucl.Phys.B 213, 545 (1983)].

[34] J. C. Collins and D. E. Soper, "Parton Distribution and Decay Functions," *Nucl. Phys. B* **194** (1982) 445–492.

[35] A. Accardi *et al.*, "Electron Ion Collider: The Next QCD Frontier: Understanding the glue that binds us all," *Eur. Phys. J. A* **52** no. 9, (2016) 268, arXiv:1212.1701 [nucl-ex].

[36] R. Angeles-Martinez *et al.*, "Transverse Momentum Dependent (TMD) parton distribution functions: status and prospects," *Acta Phys. Polon. B* **46** no. 12, (2015) 2501–2534, arXiv:1507.05267 [hep-ph].

[37] M. A. Ebert, I. W. Stewart, and Y. Zhao, "Determining the Nonperturbative Collins-Soper Kernel From Lattice QCD," *Phys. Rev. D* **99** no. 3, (2019) 034505, arXiv:1811.00026 [hep-ph].

[38] M. A. Ebert, I. W. Stewart, and Y. Zhao, "Towards Quasi-Transverse Momentum Dependent PDFs Computable on the Lattice," *JHEP* **09** (2019) 037, arXiv:1901.03685 [hep-ph].

[39] M. A. Ebert, I. W. Stewart, and Y. Zhao, "Renormalization and Matching for the Collins-Soper Kernel from Lattice QCD," *JHEP* **03** (2020) 099, arXiv:1910.08569 [hep-ph].

[40] M. A. Ebert, S. T. Schindler, I. W. Stewart, and Y. Zhao, "One-loop Matching for Spin-Dependent Quasi-TMDs," *JHEP* **09** (2020) 099, arXiv:2004.14831 [hep-ph].

[41] A. A. Vladimirov and A. Schäfer, "Transverse momentum dependent factorization for lattice observables," *Phys. Rev. D* **101** no. 7, (2020) 074517, arXiv:2002.07527 [hep-ph].

[42] X. Ji, L.-C. Jin, F. Yuan, J.-H. Zhang, and Y. Zhao, "Transverse momentum dependent parton quasidistributions," *Phys. Rev. D* **99** no. 11, (2019) 114006, arXiv:1801.05930 [hep-ph].

[43] X. Ji, Y. Liu, and Y.-S. Liu, "TMD soft function from large-momentum effective theory," *Nucl. Phys. B* **955** (2020) 115054, arXiv:1910.11415 [hep-ph].

[44] X. Ji, Y. Liu, and Y.-S. Liu, "Transverse-momentum-dependent parton distribution functions from large-momentum effective theory," *Phys. Lett. B* **811** (2020) 135946, arXiv:1911.03840 [hep-ph].

[45] C. Alexandrou, K. Cichy, M. Constantinou, J. R. Green, K. Hadjiyiannakou, K. Jansen, F. Manigrasso, A. Scapellato, and F. Steffens, "Lattice continuum-limit study of nucleon quasi-PDFs," *Phys. Rev. D* **103** (2021) 094512, arXiv:2011.00964 [hep-lat].

[46] C. Alexandrou, M. Constantinou, K. Hadjiyiannakou, K. Jansen, and F. Manigrasso, "Flavor decomposition for the proton helicity parton distribution functions," *Phys. Rev. Lett.* **126** no. 10, (2021) 102003, arXiv:2009.13061 [hep-lat].

[47] C. Alexandrou, M. Constantinou, K. Hadjiyiannakou, K. Jansen, and F. Manigrasso, "Flavor decomposition of the nucleon unpolarized, helicity and transversity parton distribution functions from lattice QCD simulations," arXiv:2106.16065 [hep-lat].

[48] **Extended Twisted Mass** Collaboration, C. Alexandrou, G. Iannelli, K. Jansen, and F. Manigrasso, "Parton distribution functions from lattice QCD using Bayes-Gauss-Fourier transforms," *Phys. Rev. D* **102** no. 9, (2020) 094508, arXiv:2007.13800 [hep-lat].

[49] R. P. Feynman, "The behavior of hadron collisions at extreme energies," *Conf. Proc. C* **690905** (1969) 237–258.

[50] J. D. Bjorken and E. A. Paschos, "Inelastic electron-proton and γ-proton scattering and the structure of the nucleon," *Phys. Rev.* **185** (Sep, 1969) 1975–1982. https://link.aps.org/doi/10.1103/PhysRev.185.1975.

[51] **Particle Data Group** Collaboration, M. Tanabashi and *et al.* , "Review of particle physics," *Phys. Rev. D* **98** (Aug, 2018) 030001. https://link.aps.org/doi/10.1103/PhysRevD.98.030001.

[52] X. Ji, "Parton physics on a euclidean lattice," *Phys. Rev. Lett.* **110** (Jun, 2013) 262002. https://link.aps.org/doi/10.1103/PhysRevLett.110.262002.

[53] J. Collins, *Foundations of perturbative QCD*, vol. 32. Cambridge University Press, 11, 2013.

[54] H.-W. Lin, W. Melnitchouk, A. Prokudin, N. Sato, and H. Shows, "First Monte Carlo Global Analysis of Nucleon Transversity with Lattice QCD Constraints," *Phys. Rev. Lett.* **120** no. 15, (2018) 152502, arXiv:1710.09858 [hep-ph].

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

[56] C. A. Aidala, S. D. Bass, D. Hasch, and G. K. Mallot, "The Spin Structure of the Nucleon," *Rev. Mod. Phys.* **85** (2013) 655–691, arXiv:1209.2803 [hep-ph].

[57] K. J. Eskola, P. Paakkinen, H. Paukkunen, and C. A. Salgado, "EPPS16: Nuclear parton distributions with LHC data," *Eur. Phys. J. C* **77** no. 3, (2017) 163, arXiv:1612.05741 [hep-ph].

[58] **LHeC Study Group** Collaboration, J. Abelleira Fernandez *et al.*, "A Large Hadron Electron Collider at CERN: Report on the Physics and Design Concepts for Machine and Detector," *J. Phys. G* **39** (2012) 075001, arXiv:1206.2913 [physics.acc-ph].

[59] D. Boer *et al.*, "Gluons and the quark sea at high energies: Distributions, polarization, tomography," arXiv:1108.1713 [nucl-th].

[60] J. J. Ethier and E. R. Nocera, "Parton Distributions in Nucleons and Nuclei," *Ann. Rev. Nucl. Part. Sci.* **70** (2020) 43–76, arXiv:2001.07722 [hep-ph].

[61] M. Constantinou *et al.*, "Parton distributions and lattice QCD calculations: toward 3D structure," arXiv:2006.08636 [hep-ph].

[62] S. Alekhin, J. Blümlein, S. Moch, and R. Placakyte, "Parton distribution functions, α_s, and heavy-quark masses for LHC Run II," *Phys. Rev. D* **96** no. 1, (2017) 014011, arXiv:1701.05838 [hep-ph].

[63] **NNPDF** Collaboration, R. D. Ball *et al.*, "Parton distributions from high-precision collider data," *Eur. Phys. J. C* **77** no. 10, (2017) 663, arXiv:1706.00428 [hep-ph].

[64] A. Accardi, L. Brady, W. Melnitchouk, J. Owens, and N. Sato, "Constraints on large-x parton distributions from new weak boson production and deep-inelastic scattering data," *Phys. Rev. D* **93** no. 11, (2016) 114017, arXiv:1602.03154 [hep-ph].

[65] D. de Florian, R. Sassot, M. Stratmann, and W. Vogelsang, "Extraction of Spin-Dependent Parton Densities and Their Uncertainties," *Phys. Rev. D* **80** (2009) 034030, arXiv:0904.3821 [hep-ph].

[66] **NNPDF** Collaboration, E. R. Nocera, R. D. Ball, S. Forte, G. Ridolfi, and J. Rojo, "A first unbiased global determination of polarized PDFs and their uncertainties," *Nucl. Phys. B* **887** (2014) 276–308, arXiv:1406.5539 [hep-ph].

[67] J. J. Ethier, N. Sato, and W. Melnitchouk, "First simultaneous extraction of spin-dependent parton distributions and fragmentation functions from a global QCD analysis," *Phys. Rev. Lett.* **119** no. 13, (2017) 132001, arXiv:1705.05889 [hep-ph].

[68] D. De Florian, G. A. Lucero, R. Sassot, M. Stratmann, and W. Vogelsang, "Monte Carlo sampling variant of the DSSV14 set of helicity parton densities," *Phys. Rev. D* **100** no. 11, (2019) 114027, arXiv:1902.10548 [hep-ph].

[69] M. Constantinou *et al.*, "Parton distributions and lattice QCD calculations: toward 3D structure," arXiv:2006.08636 [hep-ph].

[70] "LHC Machine," *JINST* **3** (2008) S08001.

[71] K. Kovařík, P. M. Nadolsky, and D. E. Soper, "Hadronic structure in high-energy collisions," *Rev. Mod. Phys.* **92** no. 4, (2020) 045003, arXiv:1905.06957 [hep-ph].

[72] **Jefferson Lab Angular Momentum** Collaboration, N. Sato, W. Melnitchouk, S. Kuhn, J. Ethier, and A. Accardi, "Iterative Monte Carlo analysis of spin-dependent parton distributions," *Phys. Rev. D* **93** no. 7, (2016) 074005, arXiv:1601.07782 [hep-ph].

[73] E. R. Nocera, "Impact of Recent RHIC Data on Helicity-Dependent Parton Distribution Functions," in *22nd International Symposium on Spin Physics*. 2, 2017. arXiv:1702.05077 [hep-ph].

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

[75] R. L. Jaffe and X. Ji, "Chiral-odd parton distributions and polarized drell-yan process," *Phys. Rev. Lett.* **67** (Jul, 1991) 552–555. https://link.aps.org/doi/10.1103/PhysRevLett.67.552.

[76] J. L. Cortes, B. Pire, and J. P. Ralston, "Measuring the transverse polarization of quarks in the proton," *Z. Phys. C* **55** (1992) 409–416.

[77] M. Anselmino, M. Boglione, U. D'Alesio, A. Kotzinian, F. Murgia, A. Prokudin, and C. Turk, "Transversity and Collins functions from SIDIS and e+ e- data," *Phys. Rev. D* **75** (2007) 054032, arXiv:hep-ph/0701006.

[78] M. Anselmino, M. Boglione, U. D'Alesio, A. Kotzinian, F. Murgia, A. Prokudin, and S. Melis, "Update on transversity and Collins functions from SIDIS and e+ e- data," *Nucl. Phys. B Proc. Suppl.* **191** (2009) 98–107, arXiv:0812.4366 [hep-ph].

[79] **Belle** Collaboration, K. Abe *et al.*, "Measurement of azimuthal asymmetries in inclusive production of hadron pairs in e+ e- annihilation at Belle," *Phys. Rev. Lett.* **96** (2006) 232002, arXiv:hep-ex/0507063.

[80] **COMPASS** Collaboration, E. S. Ageev *et al.*, "A New measurement of the Collins and Sivers asymmetries on a transversely polarised deuteron target," *Nucl. Phys. B* **765** (2007) 31–70, arXiv:hep-ex/0610068.

[81] **HERMES** Collaboration, A. Airapetian *et al.*, "Single-spin asymmetries in semi-inclusive deep-inelastic scattering on a transversely polarized hydrogen target," *Phys. Rev. Lett.* **94** (2005) 012002, arXiv:hep-ex/0408013.

[82] Z. Ye, N. Sato, K. Allada, T. Liu, J.-P. Chen, H. Gao, Z.-B. Kang, A. Prokudin, P. Sun, and F. Yuan, "Unveiling the nucleon tensor charge at Jefferson Lab: A study of the SoLID case," *Phys. Lett. B* **767** (2017) 91–98, arXiv:1609.02449 [hep-ph].

[83] D. J. Gross and F. Wilczek, "Ultraviolet Behavior of Nonabelian Gauge Theories," *Phys. Rev. Lett.* **30** (1973) 1343–1346.

[84] H. D. Politzer, "Reliable Perturbative Results for Strong Interactions?," *Phys. Rev. Lett.* **30** (1973) 1346–1349.

[85] F. Wilczek, "Asymptotic freedom: From paradox to paradigm," *Proc. Nat. Acad. Sci.* **102** (2005) 8403–8413, arXiv:hep-ph/0502113.

[86] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.

[87] R. P. Feynman, "Space-time approach to non-relativistic quantum mechanics," *Rev. Mod. Phys.* **20** (Apr, 1948) 367–387. https://link.aps.org/doi/10.1103/RevModPhys.20.367.

[88] K. G. Wilson, "Confinement of quarks," *Phys. Rev. D* **10** (Oct, 1974) 2445–2459. https://link.aps.org/doi/10.1103/PhysRevD.10.2445.

[89] H. J. Rothe, *Lattice gauge theories: An Introduction*, vol. 43. 1992.

[90] H. Nielsen and M. Ninomiya, "A no-go theorem for regularizing chiral fermions," *Physics Letters B* **105** no. 2, (1981) 219–223. https://www.sciencedirect.com/science/article/pii/0370269381910261.

[91] **Alpha** Collaboration, R. Frezzotti, P. A. Grassi, S. Sint, and P. Weisz, "Lattice QCD with a chirally twisted mass term," *JHEP* **08** (2001) 058, arXiv:hep-lat/0101001.

[92] R. Frezzotti and G. C. Rossi, "Twisted mass lattice QCD with mass nondegenerate quarks," *Nucl. Phys. B Proc. Suppl.* **128** (2004) 193–202, arXiv:hep-lat/0311008.

[93] C. Pena, S. Sint, and A. Vladikas, "Twisted mass QCD and lattice approaches to the Delta I = 1/2 rule," *JHEP* **09** (2004) 069, arXiv:hep-lat/0405028.

[94] R. Frezzotti, P. A. Grassi, S. Sint, and P. Weisz, "A Local formulation of lattice QCD without unphysical fermion zero modes," *Nucl. Phys. B Proc. Suppl.* **83** (2000) 941–946, arXiv:hep-lat/9909003.

[95] R. P. Feynman, "Very high-energy collisions of hadrons," *Phys. Rev. Lett.* **23** (1969) 1415–1417.

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

[97] X. Xiong, X. Ji, J.-H. Zhang, and Y. Zhao, "One-loop matching for parton distributions: Nonsinglet case," *Phys. Rev. D* **90** no. 1, (2014) 014051, arXiv:1310.7471 [hep-ph].

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

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

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

[101] C. Alexandrou, K. Cichy, M. Constantinou, K. Hadjiyiannakou, K. Jansen, A. Scapellato, and F. Steffens, "Systematic uncertainties in parton distribution functions from lattice QCD simulations at the physical point," *Phys. Rev. D* **99** no. 11, (2019) 114504, arXiv:1902.00587 [hep-lat].

[102] M. Constantinou and H. Panagopoulos, "Perturbative renormalization of quasi-parton distribution functions," *Phys. Rev. D* **96** no. 5, (2017) 054506, arXiv:1705.11193 [hep-lat].

[103] C. Alexandrou, K. Cichy, M. Constantinou, K. Hadjiyiannakou, K. Jansen, H. Panagopoulos, and F. Steffens, "A complete non-perturbative renormalization prescription for quasi-PDFs," *Nucl. Phys. B* **923** (2017) 394–415, arXiv:1706.00265 [hep-lat].

[104] K. Cichy, L. Del Debbio, and T. Giani, "Parton distributions from lattice data: the nonsinglet case," *JHEP* **10** (2019) 137, arXiv:1907.06037 [hep-ph].

[105] C. Bekas, E. Kokiopoulou, and Y. Saad, "An estimator for the diagonal of a matrix," *Applied Numerical Mathematics* **57** no. 11, (2007) 1214–1229. https://www.sciencedirect.com/science/article/pii/S0168927407000244. Numerical Algorithms, Parallelism and Applications (2).

[106] J. M. Tang and Y. Saad, "A probing method for computing the diagonal of a matrix inverse," *Numerical Linear Algebra with Applications* **19** no. 3, (2012) 485–501.

[107] R. Diestel, *Graph Theory*. Electronic library of mathematics. Springer, 2006. https://www.springer.com/gp/book/9783662536216.

[108] A. H. Gebremedhin, F. Manne, and A. Pothen, "What color is your jacobian? graph coloring for computing derivatives," *SIAM Review* **47** no. 4, (2005) 629–705. https://doi.org/10.1137/S0036144504444711.

[109] A. Stathopoulos, J. Laeuchli, and K. Orginos, "Hierarchical probing for estimating the trace of the matrix inverse on toroidal lattices," (2, 2013) , arXiv:1302.4018 [hep-lat].

[110] Y. Saad, *Iterative Methods for Sparse Linear Systems*. Society for Industrial and Applied Mathematics, second ed., 2003. https://epubs.siam.org/doi/abs/10.1137/1.9780898718003.

[111] **APE** Collaboration, M. Albanese *et al.*, "Glueball Masses and String Tension in Lattice QCD," *Phys. Lett. B* **192** (1987) 163–169.

[112] Y. Liang, K.-F. Liu, B.-A. Li, S. J. Dong, and K. Ishikawa, "Lattice calculation of glueball matrix elements," *Phys. Lett. B* **307** (1993) 375–382, arXiv:hep-lat/9304011.

[113] M. Della Morte, A. Shindler, and R. Sommer, "On lattice actions for static quarks," *JHEP* **08** (2005) 051, arXiv:hep-lat/0506008.

[114] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, "Hybrid Monte Carlo," *Phys. Lett. B* **195** (1987) 216–222.

[115] C. Morningstar and M. J. Peardon, "Analytic smearing of SU(3) link variables in lattice QCD," *Phys. Rev. D* **69** (2004) 054501, arXiv:hep-lat/0311018.

[116] S. Gusken, "A Study of smearing techniques for hadron correlation functions," *Nucl. Phys. B Proc. Suppl.* **17** (1990) 361–364.

[117] 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 [hep-lat].

[118] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*, vol. 2. MIT press Cambridge, MA, 2006.

[119] R. Adler, "The geometry of random fields," *The Geometry of Random Fields, Chichester: Wiley, 1981* (1981) .

[120] P. Abrahamsen, "A review of gaussian random fields and correlation functions," 1997.

[121] **ETM** Collaboration, A. Abdel-Rehim *et al.*, "First physics results at the physical pion mass from $N_f = 2$ Wilson twisted mass fermions at maximal twist," *Phys. Rev. D* **95** no. 9, (2017) 094515, arXiv:1507.05068 [hep-lat].

[122] C. Alexandrou and C. Kallidonis, "Low-lying baryon masses using $N_f = 2$ twisted mass clover-improved fermions directly at the physical pion mass," *Phys. Rev. D* **96** no. 3, (2017) 034511, arXiv:1704.02647 [hep-lat].

[123] E. Leader, A. V. Sidorov, and D. B. Stamenov, "Impact of CLAS and COMPASS data on Polarized Parton Densities and Higher Twist," *Phys. Rev. D* **75** (2007) 074027, arXiv:hep-ph/0612360.

[124] E. Leader, A. V. Sidorov, and D. B. Stamenov, "New analysis concerning the strange quark polarization puzzle," *Phys. Rev. D* **91** no. 5, (2015) 054017, arXiv:1410.1657 [hep-ph].

[125] E. Leader, A. V. Sidorov, and D. B. Stamenov, "Determination of Polarized PDFs from a QCD Analysis of Inclusive and Semi-inclusive Deep Inelastic Scattering Data," *Phys. Rev. D* **82** (2010) 114018, arXiv:1010.0574 [hep-ph].

[126] F. Arbabifar, A. N. Khorramian, and M. Soleymaninia, "QCD analysis of polarized DIS and the SIDIS asymmetry world data and light sea-quark decomposition," *Phys. Rev. D* **89** no. 3, (2014) 034006, arXiv:1311.1830 [hep-ph].

[127] E. Leader, A. V. Sidorov, and D. B. Stamenov, "A Possible Resolution of the Strange Quark Polarization Puzzle ?," *Phys. Rev. D* **84** (2011) 014002, arXiv:1103.5979 [hep-ph].

[128] **ATLAS** Collaboration, M. Aaboud *et al.*, "Measurement of the W-boson mass in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector," *Eur. Phys. J. C* **78** no. 2, (2018) 110, arXiv:1701.07240 [hep-ex]. [Erratum: Eur.Phys.J.C 78, 898 (2018)].

[129] S. Alekhin, J. Blümlein, and S. Moch, "Strange sea determination from collider data," *Phys. Lett. B* **777** (2018) 134–140, arXiv:1708.01067 [hep-ph].

[130] **ETM** Collaboration, A. Abdel-Rehim, C. Alexandrou, M. Constantinou, K. Hadjiyiannakou, K. Jansen, C. Kallidonis, G. Koutsou, and A. Vaquero Aviles-Casco, "Direct Evaluation of the Quark Content of Nucleons from Lattice QCD at the Physical Point," *Phys. Rev. Lett.* **116** no. 25, (2016) 252001, arXiv:1601.01624 [hep-lat].

[131] C. Alexandrou *et al.*, "Nucleon scalar and tensor charges using lattice QCD simulations at the physical value of the pion mass," *Phys. Rev. D* **95** no. 11, (2017) 114514, arXiv:1703.08788 [hep-lat]. [Erratum: Phys.Rev.D 96, 099906 (2017)].

[132] C. Alexandrou, M. Constantinou, K. Hadjiyiannakou, K. Jansen, C. Kallidonis, G. Koutsou, and A. Vaquero Aviles-Casco, "Nucleon axial form factors using $N_f = 2$ twisted mass fermions with a physical value of the pion mass," *Phys. Rev. D* **96** no. 5, (2017) 054507, arXiv:1705.03399 [hep-lat].

[133] C. Alexandrou, M. Constantinou, K. Hadjiyiannakou, K. Jansen, C. Kallidonis, G. Koutsou, and A. Vaquero Aviles-Casco, "Nucleon electromagnetic form factors using lattice simulations at the physical point," *Phys. Rev. D* **96** no. 3, (2017) 034503, arXiv:1706.00469 [hep-lat].

[134] C. Alexandrou, M. Constantinou, K. Hadjiyiannakou, K. Jansen, C. Kallidonis, G. Koutsou, A. Vaquero Avilés-Casco, and C. Wiese, "Nucleon Spin and Momentum Decomposition Using Lattice QCD Simulations," *Phys. Rev. Lett.* **119** no. 14, (2017) 142002, arXiv:1706.02973 [hep-lat].

[135] C. Alexandrou, M. Constantinou, K. Hadjiyiannakou, K. Jansen, C. Kallidonis, G. Koutsou, and A. Vaquero Avilés-Casco, "Strange nucleon electromagnetic form factors from lattice QCD," *Phys. Rev. D* **97** no. 9, (2018) 094504, arXiv:1801.09581 [hep-lat].

[136] C. Alexandrou, S. Bacchio, M. Constantinou, J. Finkenrath, K. Hadjiyiannakou, K. Jansen, G. Koutsou, and A. Vaquero Aviles-Casco, "Proton and neutron electromagnetic form factors from lattice QCD," *Phys. Rev. D* **100** no. 1, (2019) 014509, arXiv:1812.10311 [hep-lat].

[137] C. Alexandrou, S. Bacchio, M. Constantinou, J. Finkenrath, K. Hadjiyiannakou, K. Jansen, G. Koutsou, and A. Vaquero Aviles-Casco, "Nucleon axial, tensor, and scalar charges and σ-terms in lattice QCD," *Phys. Rev. D* **102** no. 5, (2020) 054517, arXiv:1909.00485 [hep-lat].

[138] C. Alexandrou, S. Bacchio, M. Constantinou, J. Finkenrath, K. Hadjiyiannakou, K. Jansen, and G. Koutsou, "Nucleon strange electromagnetic form factors," *Phys. Rev. D* **101** no. 3, (2020) 031501, arXiv:1909.10744 [hep-lat].

[139] C. Alexandrou, S. Bacchio, M. Constantinou, J. Finkenrath, K. Hadjiyiannakou, K. Jansen, G. Koutsou, H. Panagopoulos, and G. Spanoudes, "Complete flavor decomposition of the spin and momentum fraction of the proton using lattice QCD simulations at physical pion mass," *Phys. Rev. D* **101** no. 9, (2020) 094513, arXiv:2003.08486 [hep-lat].

[140] **PNDME** Collaboration, T. Bhattacharya, V. Cirigliano, S. Cohen, R. Gupta, A. Joseph, H.-W. Lin, and B. Yoon, "Iso-vector and Iso-scalar Tensor Charges of the Nucleon from Lattice QCD," *Phys. Rev. D* **92** no. 9, (2015) 094511, arXiv:1506.06411 [hep-lat].

[141] J. Liang, M. Sun, Y.-B. Yang, T. Draper, and K.-F. Liu, "Ratio of strange to u/d momentum fraction in disconnected insertions," *Phys. Rev. D* **102** no. 3, (2020) 034514, arXiv:1901.07526 [hep-ph].

[142] D. Djukanovic, K. Ottnad, J. Wilhelm, and H. Wittig, "Strange electromagnetic form factors of the nucleon with $N_f = 2 + 1$ $\mathcal{O}(a)$-improved Wilson fermions," *Phys. Rev. Lett.* **123** no. 21, (2019) 212001, arXiv:1903.12566 [hep-lat].

[143] J. Green, S. Meinel, M. Engelhardt, S. Krieg, J. Laeuchli, J. Negele, K. Orginos, A. Pochinsky, and S. Syritsyn, "High-precision calculation of the strange nucleon electromagnetic form factors," *Phys. Rev. D* **92** no. 3, (2015) 031501, arXiv:1505.01803 [hep-lat].

[144] A. Radyushkin, "Nonperturbative Evolution of Parton Quasi-Distributions," *Phys. Lett. B* **767** (2017) 314–320, arXiv:1612.05170 [hep-ph].

[145] Y.-Q. Ma and J.-W. Qiu, "Extracting Parton Distribution Functions from Lattice QCD Calculations," *Phys. Rev. D* **98** no. 7, (2018) 074021, arXiv:1404.6860 [hep-ph].

[146] Y.-Q. Ma and J.-W. Qiu, "QCD Factorization and PDFs from Lattice QCD Calculation," *Int. J. Mod. Phys. Conf. Ser.* **37** (2015) 1560041, arXiv:1412.2688 [hep-ph].

[147] A. J. Chambers, R. Horsley, Y. Nakamura, H. Perlt, P. E. L. Rakow, G. Schierholz, A. Schiller, K. Somfleth, R. D. Young, and J. M. Zanotti, "Nucleon Structure Functions from Operator Product Expansion on the Lattice," *Phys. Rev. Lett.* **118** no. 24, (2017) 242001, arXiv:1703.01153 [hep-lat].

[148] T. Ishikawa, Y.-Q. Ma, J.-W. Qiu, and S. Yoshida, "Renormalizability of quasiparton distribution functions," *Phys. Rev. D* **96** no. 9, (2017) 094019, arXiv:1707.03107 [hep-ph].

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

[150] A. Radyushkin, "One-loop evolution of parton pseudo-distribution functions on the lattice," *Phys. Rev. D* **98** no. 1, (2018) 014019, arXiv:1801.02427 [hep-ph].

[151] J.-H. Zhang, J.-W. Chen, L. Jin, H.-W. Lin, A. Schäfer, and Y. Zhao, "First direct lattice-QCD calculation of the x-dependence of the pion parton distribution function," *Phys. Rev. D* **100** no. 3, (2019) 034505, arXiv:1804.01483 [hep-lat].

[152] 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 [hep-lat].

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

[154] S. Bhattacharya, C. Cocuzza, and A. Metz, "Generalized quasi parton distributions in a diquark spectator model," *Phys. Lett. B* **788** (2019) 453–463, arXiv:1808.01437 [hep-ph].

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

[156] B. Joó, J. Karpie, K. Orginos, A. Radyushkin, D. Richards, and S. Zafeiropoulos, "Parton Distribution Functions from Ioffe time pseudo-distributions," *JHEP* **12** (2019) 081, arXiv:1908.09771 [hep-lat].

[157] A. V. Radyushkin, "Generalized parton distributions and pseudodistributions," *Phys. Rev. D* **100** no. 11, (2019) 116011, arXiv:1909.08474 [hep-ph].

[158] B. Joó, J. Karpie, K. Orginos, A. V. Radyushkin, D. G. Richards, R. S. Sufian, and S. Zafeiropoulos, "Pion valence structure from Ioffe-time parton pseudodistribution functions," *Phys. Rev. D* **100** no. 11, (2019) 114512, arXiv:1909.08517 [hep-lat].

[159] X. Ji, "Fundamental Properties of the Proton in Light-Front Zero Modes," *Nucl. Phys.* **B** (2020) 115181, arXiv:2003.04478 [hep-ph].

[160] V. M. Braun, K. G. Chetyrkin, and B. A. Kniehl, "Renormalization of parton quasi-distributions beyond the leading order: spacelike vs. timelike," *JHEP* **07** (2020) 161, arXiv:2004.01043 [hep-ph].

[161] M. Bhat, K. Cichy, M. Constantinou, and A. Scapellato, "Flavor nonsinglet parton distribution functions from lattice QCD at physical quark masses via the pseudodistribution approach," *Phys. Rev. D* **103** no. 3, (2021) 034510, arXiv:2005.02102 [hep-lat].

[162] J. Bringewatt, N. Sato, W. Melnitchouk, J.-W. Qiu, F. Steffens, and M. Constantinou, "Confronting lattice parton distributions with global QCD analysis," *Phys. Rev. D* **103** no. 1, (2021) 016003, arXiv:2010.00548 [hep-ph].

[163] W.-Y. Liu and J.-W. Chen, "Renormalon Effects in Quasi Parton Distributions," arXiv:2010.06623 [hep-ph].

[164] L. Del Debbio, T. Giani, J. Karpie, K. Orginos, A. Radyushkin, and S. Zafeiropoulos, "Neural-network analysis of Parton Distribution Functions from Ioffe-time pseudodistributions," *JHEP* **02** (2021) 138, arXiv:2010.03996 [hep-ph].

[165] W.-Y. Liu and J.-W. Chen, "Chiral Perturbation for Large Momentum Effective Field Theory," arXiv:2011.13536 [hep-lat].

[166] **Lattice Parton Collaboration (LPC)** Collaboration, Y.-K. Huo *et al.*, "Self-renormalization of quasi-light-front correlators on the lattice," *Nucl. Phys. B* **969** (2021) 115443, arXiv:2103.02965 [hep-lat].

[167] W. Detmold, A. V. Grebe, I. Kanamori, C. J. D. Lin, R. J. Perry, and Y. Zhao, "Parton Physics from a Heavy-Quark Operator Product Expansion: I. Formalism and Wilson Coefficients," arXiv:2103.09529 [hep-lat].

[168] J. Karpie, K. Orginos, A. Radyushkin, and S. Zafeiropoulos, "The Continuum and Leading Twist Limits of Parton Distribution Functions in Lattice QCD," arXiv:2105.13313 [hep-lat].

[169] 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 [hep-lat].

[170] X. Ji, Y. Liu, Y.-S. Liu, J.-H. Zhang, and Y. Zhao, "Large-Momentum Effective Theory," arXiv:2004.03543 [hep-ph].

[171] M. Constantinou, "The x-dependence of hadronic parton distributions: A review on the progress of lattice QCD," *Eur. Phys. J. A* **57** no. 2, (2021) 77, arXiv:2010.02445 [hep-lat].

[172] W. Wang, J.-H. Zhang, S. Zhao, and R. Zhu, "Complete matching for quasidistribution functions in large momentum effective theory," *Phys. Rev. D* **100** no. 7, (2019) 074509, arXiv:1904.00978 [hep-ph].

[173] M. Constantinou, M. Hadjiantonis, H. Panagopoulos, and G. Spanoudes, "Singlet versus nonsinglet perturbative renormalization of fermion bilinears," *Phys. Rev. D* **94** no. 11, (2016) 114513, arXiv:1610.06744 [hep-lat].

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

[175] M. Gockeler, R. Horsley, H. Oelrich, H. Perlt, D. Petters, P. E. L. Rakow, A. Schafer, G. Schierholz, and A. Schiller, "Nonperturbative renormalization of composite operators in lattice QCD," *Nucl. Phys. B* **544** (1999) 699–733, arXiv:hep-lat/9807044.

[176] **ETM** Collaboration, C. Alexandrou, M. Constantinou, and H. Panagopoulos, "Renormalization functions for Nf=2 and Nf=4 twisted mass fermions," *Phys. Rev. D* **95** no. 3, (2017) 034505, arXiv:1509.00213 [hep-lat].

[177] **ETM** Collaboration, M. Constantinou *et al.*, "Non-perturbative renormalization of quark bilinear operators with $N_f = 2$ (tmQCD) Wilson fermions and the tree-level improved gauge action," *JHEP* **08** (2010) 068, arXiv:1004.1115 [hep-lat].

[178] X. Ji, "Why is LaMET an effective field theory for partonic structure?," arXiv:2007.06613 [hep-ph].

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

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

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

[182] L.-B. Chen, W. Wang, and R. Zhu, "Quasi parton distribution functions at NNLO: flavor non-diagonal quark contributions," *Phys. Rev. D* **102** no. 1, (2020) 011503, arXiv:2005.13757 [hep-ph].

[183] L.-B. Chen, W. Wang, and R. Zhu, "Master integrals for two-loop QCD corrections to quark quasi PDFs," *JHEP* **10** (2020) 079, arXiv:2006.10917 [hep-ph].

[184] L.-B. Chen, W. Wang, and R. Zhu, "Next-to-next-to-leading order corrections to quark Quasi parton distribution functions," *Phys. Rev. Lett.* **126** no. 7, (2021) 072002, arXiv:2006.14825 [hep-ph].

[185] Y. Iwasaki, "Renormalization group analysis of lattice theories and improved lattice action: Two-dimensional nonlinear $O(N)$ sigma model," *Nucl. Phys. B* **258** (1985) 141–156.

[186] C. Alexandrou *et al.*, "Simulating twisted mass fermions at physical light, strange and charm quark masses," *Phys. Rev. D* **98** no. 5, (2018) 054518, arXiv:1807.00495 [hep-lat].

[187] B. Sheikholeslami and R. Wohlert, "Improved Continuum Limit Lattice Action for QCD with Wilson Fermions," *Nucl. Phys. B* **259** (1985) 572.

[188] R. Frezzotti and G. C. Rossi, "Chirally improving Wilson fermions. 1. O(a) improvement," *JHEP* **08** (2004) 007, arXiv:hep-lat/0306014.

[189] **LP3** Collaboration, J.-W. Chen, T. Ishikawa, L. Jin, H.-W. Lin, J.-H. Zhang, and Y. Zhao, "Symmetry properties of nonlocal quark bilinear operators on a Lattice," *Chin. Phys. C* **43** no. 10, (2019) 103101, arXiv:1710.01089 [hep-lat].

[190] J. R. Green, K. Jansen, and F. Steffens, "Improvement, generalization, and scheme conversion of Wilson-line operators on the lattice in the auxiliary field approach," *Phys. Rev. D* **101** no. 7, (2020) 074509, arXiv:2002.09408 [hep-lat].

[191] C. Alexandrou, S. Gusken, F. Jegerlehner, K. Schilling, and R. Sommer, "The Static approximation of heavy - light quark systems: A Systematic lattice study," *Nucl. Phys. B* **414** (1994) 815–855, arXiv:hep-lat/9211042.

[192] C. Alexandrou *et al.*, "Quark masses using twisted mass fermion gauge ensembles," (4, 2021) , arXiv:2104.13408 [hep-lat].

[193] 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** no. 9, (2016) 094515, arXiv:1602.05525 [hep-lat].

[194] G. Martinelli and C. T. Sachrajda, "A Lattice Study of Nucleon Structure," *Nucl. Phys. B* **316** (1989) 355–372.

[195] **ETM** Collaboration, P. Boucaud *et al.*, "Dynamical Twisted Mass Fermions with Light Quarks: Simulation and Analysis Details," *Comput. Phys. Commun.* **179** (2008) 695–715, arXiv:0803.0224 [hep-lat].

[196] **ETM** Collaboration, C. Michael and C. Urbach, "Neutral mesons and disconnected diagrams in Twisted Mass QCD," *PoS* **LATTICE2007** (2007) 122, arXiv:0709.4564 [hep-lat].

[197] A. Abdel-Rehim, C. Alexandrou, M. Constantinou, J. Finkenrath, K. Hadjiyiannakou, K. Jansen, C. Kallidonis, G. Koutsou, A. V. Avilés-Casco, and J. Volmer, "Disconnected diagrams with twisted-mass fermions," *PoS* **LATTICE2016** (2016) 155, arXiv:1611.03802 [hep-lat].

[198] W. Wilcox, "Noise methods for flavor singlet quantities," in *Interdisciplinary Workshop on Numerical Challenges in Lattice QCD*. 8, 1999. arXiv:hep-lat/9911013.

[199] L. Maiani, G. Martinelli, M. L. Paciello, and B. Taglienti, "Scalar Densities and Baryon Mass Differences in Lattice QCD With Wilson Fermions," *Nucl. Phys. B* **293** (1987) 420.

[200] S. Capitani, M. Della Morte, G. von Hippel, B. Jager, A. Juttner, B. Knippschild, H. B. Meyer, and H. Wittig, "The nucleon axial charge from lattice QCD with controlled errors," *Phys. Rev. D* **86** (2012) 074502, arXiv:1205.0180 [hep-lat].

[201] R. Gupta, B. Yoon, T. Bhattacharya, V. Cirigliano, Y.-C. Jang, and H.-W. Lin, "Flavor diagonal tensor charges of the nucleon from (2+1+1)-flavor lattice QCD," *Phys. Rev. D* **98** no. 9, (2018) 091501, arXiv:1808.07597 [hep-lat].

[202] B. Märkisch *et al.*, "Measurement of the Weak Axial-Vector Coupling Constant in the Decay of Free Neutrons Using a Pulsed Cold Neutron Beam," *Phys. Rev. Lett.* **122** no. 24, (2019) 242501, arXiv:1812.04666 [nucl-ex].

[203] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, "LHAPDF6: parton density access in the LHC precision era," *Eur. Phys. J. C* **75** (2015) 132, arXiv:1412.7420 [hep-ph].

[204] X. G. Wang, C.-R. Ji, W. Melnitchouk, Y. Salamu, A. W. Thomas, and P. Wang, "Strange quark asymmetry in the proton in chiral effective theory," *Phys. Rev. D* **94** no. 9, (2016) 094035, arXiv:1610.03333 [hep-ph].

[205] X. G. Wang, C.-R. Ji, W. Melnitchouk, Y. Salamu, A. W. Thomas, and P. Wang, "Constraints on $s - \bar{s}$ asymmetry of the proton in chiral effective theory," *Phys. Lett. B* **762** (2016) 52–56, arXiv:1602.06646 [nucl-th].

[206] S. Catani, D. de Florian, G. Rodrigo, and W. Vogelsang, "Perturbative generation of a strange-quark asymmetry in the nucleon," *Phys. Rev. Lett.* **93** (2004) 152003, arXiv:hep-ph/0404240.

[207] J. Alwall and G. Ingelman, "Strange quark asymmetry in the nucleon and the NuTeV anomaly," *Phys. Rev. D* **70** (2004) 111505, arXiv:hep-ph/0407364.

[208] R. S. Sufian, T. Liu, G. F. de Téramond, H. G. Dosch, S. J. Brodsky, A. Deur, M. T. Islam, and B.-Q. Ma, "Nonperturbative strange-quark sea from lattice QCD, light-front holography, and meson-baryon fluctuation models," *Phys. Rev. D* **98** no. 11, (2018) 114004, arXiv:1809.04975 [hep-ph].

[209] Y. Iwasaki, "Renormalization group analysis of lattice theories and improved lattice action. II. Four-dimensional non-abelian SU(N) gauge model." University of tsukuba preprint uthep-118, 1983.

[210] R. Baron *et al.*, "Light hadrons from lattice QCD with light (u, d), strange and charm dynamical quarks," *JHEP* **2010** no. 06, (2010) 111, arXiv:1004.5284 [hep-lat].

[211] C. Alexandrou, M. Constantinou, S. Dinter, V. Drach, K. Jansen, C. Kallidonis, and G. Koutsou, "Nucleon form factors and moments of generalized parton distributions using $N_f = 2 + 1 + 1$ twisted mass fermions," *Phys. Rev. D* **88** no. 1, (2013) 014509, arXiv:1303.5979 [hep-lat].

[212] A. Shindler, "Twisted mass lattice QCD," *Phys. Rept.* **461** (2008) 37–110, arXiv:0707.4093 [hep-lat].

[213] G. Backus and F. Gilbert, "The resolving power of gross earth data," *Geophys. J. Int.* **16** no. 2, (10, 1968) 169–205.