Scuola Normale Superiore
Classe di Scienze Matematiche e Naturali
Corso di perfezionamento in Fisica

PhD thesis

Solving the $U(1)_A$ problem of QCD: new computational strategies and results

Marco Cè

27th February 2017

Advisor: Prof. Leonardo Giusti
Supervisor: Prof. Enrico Trincherini

Abstract

The low-energy spectrum and phenomenology of strongly-interacting particles is well described by Quantum Chromodynamics (QCD). Ab initio results have been obtained using Monte Carlo simulations of QCD regularized on the lattice, with a very good agreement with experiments. In particular, the mass pattern of an octet of light pseudoscalar mesons is well understood identifying them as pseudo Nambu–Goldstone bosons of spontaneous chiral symmetry breaking. A ninth pseudoscalar meson—the η'—evades this classification. It is associated to flavour-singlet axial U(1) transformations, but it is too heavy to be a pseudo Nambu–Goldstone boson. This is known as the $\mathrm{U}(1)_A$ problem.

A solution to this problem was proposed by Witten and Veneziano in 1979. Their proposal links the η' mass to the anomaly in the flavour-singlet axial Ward–Takahashi identities. They predict an anomalous contribution to the mass proportional to the topological susceptibility computed in Yang–Mills (YM) theory.

The verification of the Witten–Veneziano formula requires the computation of the YM topological susceptibility and the η' meson mass from first principles. This is possible only solving QCD at the non-perturbative level, as for instance in lattice simulations. This thesis is a step toward the implementation and verification of the Witten–Veneziano formula on the lattice.

In the first part of the thesis, we computed the topological susceptibility with percent-level accuracy in the SU(3) YM theory and for the first time in the $N_\mathrm{c} \to \infty$ limit. The computation was done on the lattice implementing a naïve discretization of the topological charge evolved with the YM gradient flow. We provided a field-theoretical proof that the cumulants of the topological charge distribution computed using this definition coincide, in the continuum limit, with those of the universal definition appearing in the anomalous Ward–Takahashi identities. We performed a range of high statistics Monte Carlo simulations with different lattice spacings and N_c values. The coverage of parameter space allowed us to extrapolate the topological susceptibility to the continuum and $N_\mathrm{c} \to \infty$ limits with confidence, keeping all systematic effects under control. As a by-product, we measured the non-Gaussianity of the topological charge distribution in SU(3) YM theory. This result is compatible with the expectations from the large-N_c expansion, while it rules out the θ behaviour of the vacuum energy predicted by the dilute instanton gas model.

In the last part of the thesis, we focused on the direct lattice determination of the mass of the η' meson. With state-of-the-art techniques, it is still not possible to compute this mass with an adequate accuracy. The root of the issue is the exponential worsening with distance of the signal-to-noise ratio of correlation functions. This is a very general unsolved problem, which is currently limiting the accuracy of a broad class of Monte Carlo computations. We proposed to address this problem generalising the multilevel Monte Carlo algorithm to theories with dynamical fermions. We devised the first step of this program, namely the factorization of hadronic two-point functions, focusing on the disconnected contribution which is relevant to the η' meson.

Acknowledgements

I would like to express my deep gratitude to my advisor Prof. Leonardo Giusti for his mentoring during the last four years. I thank him for introducing me to the topic of lattice QCD, for supporting me throughout the learning process and for his essential assistance in writing this thesis. I am indebted to him for the education that I received on how to approach physics problems, pursue research objectives and present the results.

Most of the original work that went into this thesis has been carried out in collaboration with Leonardo and other people, notably Stefan Schaefer and Miguel García Vera. It is a pleasure to share the credit for the results presented in this thesis with them.

I wish to thank the thesis referees Prof. Luigi Del Debbio and Prof. Uwe-Jens Wiese for sharing with me very helpful comments and corrections to the thesis, that I hope to have addressed in the final version.

I have been lucky to receive an education in high energy physics at Scuola Normale Superiore, following classes by Prof. Enrico Trincherini, Prof. Riccardo Barbieri, Dario Francia and many others. I am especially thankful to Enrico for allowing me to pursue a PhD program with Leonardo and other people not in Pisa.

Thanks to Alberto, Fabrizio, Luca and Stefano, together with many other people in Pisa and Milano, for any interesting discussion that contributed to form me as a physicist and a researcher.

Contents

List of publications

For an updated list see https://inspirehep.net/author/profile/M.Ce.1.

Articles

[1] M. Cè, C. Consonni, G. P. Engel and L. Giusti, 'Non-Gaussianities in the topological charge distribution of the SU(3) Yang-Mills theory', Phys. Rev. D **92**, 074502 (2015), arXiv:1506.06052 [hep-lat].

[2] M. Cè, L. Giusti and S. Schaefer, 'Domain decomposition, multilevel integration, and exponential noise reduction in lattice QCD', Phys. Rev. D **93**, 094507 (2016), arXiv:1601.04587 [hep-lat].

[3] M. Cè, M. García Vera, L. Giusti and S. Schaefer, 'The topological susceptibility in the large-N limit of SU(N) Yang-Mills theory', Phys. Lett. B **762**, 232–236 (2016), arXiv:1607.05939 [hep-lat].

[4] M. Cè, L. Giusti and S. Schaefer, 'A local factorization of the fermion determinant in lattice QCD', (2016), arXiv:1609.02419 [hep-lat].

Proceedings

[5] M. Cè, C. Consonni, G. P. Engel and L. Giusti, 'Testing the Witten-Veneziano mechanism with the Yang-Mills gradient flow on the lattice', PoS **LATTICE2014**, 353 (2014), arXiv:1410.8358 [hep-lat].

[6] M. Cè, 'Non-Gaussianity of the topological charge distribution in SU(3) Yang-Mills theory', PoS **LATTICE 2015**, 318 (2015), arXiv:1510.08826 [hep-lat].

[7] M. Cè, M. García Vera, L. Giusti and S. Schaefer, 'The large-N limit of the topological susceptibility of Yang-Mills gauge theory', PoS **LATTICE2016**, 350 (2016), arXiv:1610.08797 [hep-lat].

[8] M. Cè, L. Giusti and S. Schaefer, 'Domain decomposition and multilevel integration for fermions', PoS **LATTICE2016**, 263 (2016), arXiv:1612.06424 [hep-lat].

Introduction

The quantum theory of fields (QFT) is the theoretical framework that, since the second half of the twentieth century, has been successfully used to construct modern theories of high energy physics. It is the only known way to combine quantum dynamics with the theory of special relativity. The Standard Model (SM) of particle physics [9–11] is a self-consistent QFT which represents the current reference in the understanding of subatomic particles and their interactions. It describes three of the four known fundamental forces of nature—strong, weak and electromagnetic interactions—by means of a quantum gauge theory. That is to say, a QFT in which the Lagrangian is invariant under a continuous group of local transformations, forming the Lie group $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)$. The non-Abelian SU(3) subgroup describes strong interactions mediated by gauge bosons called *gluons* and the corresponding gauge theory is called Quantum Chromodynamics (QCD) [12, 13]. Of the SM elementary fermions, only those called *quarks* carry the *colour* non-Abelian charge and thus 'feel' the strong interactions, in contrast with the leptons, which are colour-neutral.

In the functional integral formulation of QFT, the path integral of QCD is, in a formal and very compact notation,

$$\mathcal{Z} = \int \mathcal{D}[\bar{\psi}, \psi, A] \exp\{\mathrm{i}S[\bar{\psi}, \psi, A]\}, \tag{0.1}$$

where ψ ($\bar{\psi}$) denotes the quark field (and its Dirac adjoint) and A the gluon field. $\mathcal{D}[\bar{\psi}, \psi, A]$ denotes the functional integration measure and $S[\bar{\psi}, \psi, A]$ is the action of QCD, a functional of the fields that can be split into the fermion action

$$S_{\mathrm{f}}[\bar{\psi}, \psi, A] = \int \mathrm{d}^4x\, \bar{\psi}(x)\big[\mathrm{i}\not{D}[A] - M\big]\psi(x) \tag{0.2}$$

and the gluon action

$$S_{\mathrm{g}}[A] = -\frac{1}{4g^2} \int \mathrm{d}^4x\, F^a_{\mu\nu}(x) F^{a\mu\nu}(x). \tag{0.3}$$

QCD is the subject studied in the thesis, and from now on we focus on it.

Asymptotic freedom

Even though QCD is superficially similar to Quantum Electrodynamics (QED), the Abelian gauge theory of electromagnetic interactions, the phenomenology of the two theories differs substantially. This is mainly due to the fact that non-Abelian gauge theories like QCD, in which the force-carrying gauge bosons are themselves charged under the gauge group, are *asymptotically free* theories [14, 15]. The strength of interactions, determined by the coupling g, becomes asymptotically weaker as energy increases, in contrast with QED in which the vacuum

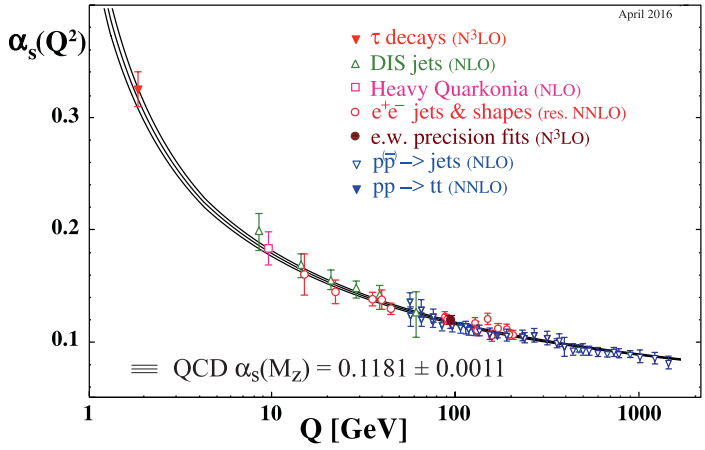


Figure 0.1.: Summary of experimental measurements of $\alpha_s = g^2/4\pi$ as a function of the energy scale Q. The respective order of QCD perturbation theory used in the extraction of α_s is indicated in brackets. The fit is performed including experimental determinations based on (at least) full NNLO QCD prediction and non-perturbative lattice determinations. The fit line shows the running of α_s using the $\overline{\text{MS}}$-scheme four-loop β-function and three-loop threshold matching at charm-, bottom- and top-quark pole masses.

polarization screens the electric charge at long distance. In mathematical terms, the change of the coupling g with energy is encoded in the so-called β-function

$$\beta(g) = Q\frac{\partial g}{\partial Q} = -\left(b_0 g^3 + b_1 g^5 + \cdots\right), \qquad b_0 = \frac{1}{16\pi^2}\left(\frac{11}{3}N_\text{c} - \frac{2}{3}N_\text{f}\right), \tag{0.4}$$

for a theory with SU(N_c) gauge group and N_f flavours of quarks in the fundamental representation. Asymptotic freedom is realized with a negative β-function for $g \to 0$, e.g. in QCD ($N_\text{c} = 3$ case) if the number of flavours is $N_\text{f} \leq 16$, which obviously include the physical case.

It is worth noting that even the theory without fermions ($N_\text{f} = 0$), described only by the gluon action in Eq. (0.3), is an asymptotically free theory. This purely-bosonic quantum gauge theory is know as Yang–Mills (YM) theory, as it was firstly proposed by Yang and Mills [16] for the SU(2) gauge group. Unlike the Abelian counterpart, which describes a free photon, YM theory is naturally interacting, with interactions completely determined just by the non-Abelian structure of gauge symmetry.

Perturbative methods based on Feynman diagrams treat interactions as a small perturbation of otherwise free particles. Asymptotic freedom allows perturbation theory to be used to describe the dynamics of quarks and gluons at extremely high energies. The perturbative expansion parameter $\alpha_s = g^2/4\pi$ is small in this regime and, to a given order in perturbation theory, the running given by the β-function in Eq. (0.4) can be integrated. For instance, at leading order

$$\alpha_s(Q) = \frac{g^2(Q)}{4\pi} \simeq \frac{1}{b_0 \ln \frac{Q^2}{\Lambda^2_\text{QCD}}}, \tag{0.5}$$

from which is clear that the strength of interactions changes *logarithmically* with the energy. Here, $\Lambda_{\rm QCD} \approx 300\,\text{MeV}$ is an integration constant whose precise value depends on the renormalization scheme. It sets a scale at which the perturbative α_s formally diverges: this means that perturbation theory is not reliable at energies $Q \approx \Lambda_{\rm QCD}$, and non-perturbative methods are needed.

Figure 0.1 shows how very different experimental measurements at different scales Q, matched with different order of perturbation theory, give values of $\alpha_s(Q)$ that reproduce the correct perturbative running. These measurements concur together with lattice results to give the average value of α_s, in the $\overline{\rm MS}$ scheme, at the energy scale $M_Z \approx 91\,\text{GeV}$ [17]

$$\alpha_s^{\overline{\rm MS}}\left(M_Z^2\right) = 0.1181(11). \tag{0.6}$$

An introduction to QCD, including renormalization, is given in Chapter 1.

Hadrons

Asymptotic freedom indicates that at lower energies the interactions become so strong that a perturbative approach is not applicable, since the dynamics is not approximated at all by free quarks and gluons. Indeed, the existence of quarks and gluons has been indirectly confirmed identifying them as the constituents, the *partons*, of composite particles called hadrons—such as the proton and the neutron—when scattered at high energies. However, neither free quarks nor free gluons have been observed directly in high energy physics experiment. This phenomenon is called *colour confinement* and asserts that all the asymptotic states present in the spectrum must be colour singlet.

A simple state which transforms trivially under SU(3) colour transformations can be obtained in two ways. A *meson* is a bound state obtained combining a quark and an antiquark which transform respectively under the fundamental $\mathbf{3}$ and antifundamental $\bar{\mathbf{3}}$ representation

$$\text{colour:} \quad \mathbf{3} \otimes \bar{\mathbf{3}} \supset \mathbf{1} \text{ of mesons.} \tag{0.7}$$

A *baryon* (*antibaryon*) is a bound state obtained combining three (anti)quarks in the (anti)fundamental representation

$$\text{colour:} \quad \mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} \supset \mathbf{1} \text{ of baryons.} \tag{0.8}$$

More complicated colour-singlet states, such as tetraquarks and pentaquarks, may exist as bound states. Even in the absence of quarks, in the YM theory case, it is still possible to obtain a *glueball* colour-singlet state made only of gluons. Glueballs form the spectrum of all excited states in YM theory. Despite being colour-neutral, hadrons and glueballs feel residual strong interactions in the form of nuclear interactions. These are in analogy with the van der Waals forces between electrically-neutral molecules.

Remarkably, the dynamics behind confinement generates a mass gap in the theory. Even if the elementary fields in the Lagrangian are massless, there are no massless particles in the spectrum, with the only exception of the NG bosons of chiral symmetry breaking if two or more flavours of massless quarks are present. The analytical proof of the fact that YM theory has a mass gap is an unsolved Millennium Prize problem [18]. In QCD, the strong dynamics

accounts for most of the mass of hadrons made of light quarks, which is typically $\approx 1\,\text{GeV}$. With the only exception of an octet of pseudoscalar mesons which are significantly lighter, these masses are proportional to the characteristic energy scale Λ_{QCD} with an $\mathcal{O}(1)$ coefficient. It is clear from Figure 0.1 that a perturbative expansion in α_s fails at these energy scales, thus a complete understanding of the hadron spectrum and colour confinement must come from a non-perturbative approach to the theory.

Lattice

Until 1974, no methods were available to make quantitative predictions for the low-energy regime. The introduction by Wilson [19] of a regularization of QCD on a spacetime lattice first opened the way to fully non-perturbative predictions. The use of Euclidean spacetime allows the direct application to QFT of Monte Carlo methods used in statistical mechanics and the study of non-perturbative phenomena through numerical simulations. Moreover, on the theoretical side, the lattice-regularized path integral is mathematically well-defined. In particular, gauge theories can be discretized in a way that gauge invariance is exactly preserved and gauge fields are angular variable, which makes gauge fixing unnecessary. In the thesis, we use the lattice regularization to study non-perturbative phenomena in QCD and YM theory. This formalism is introduced in Chapter 2.

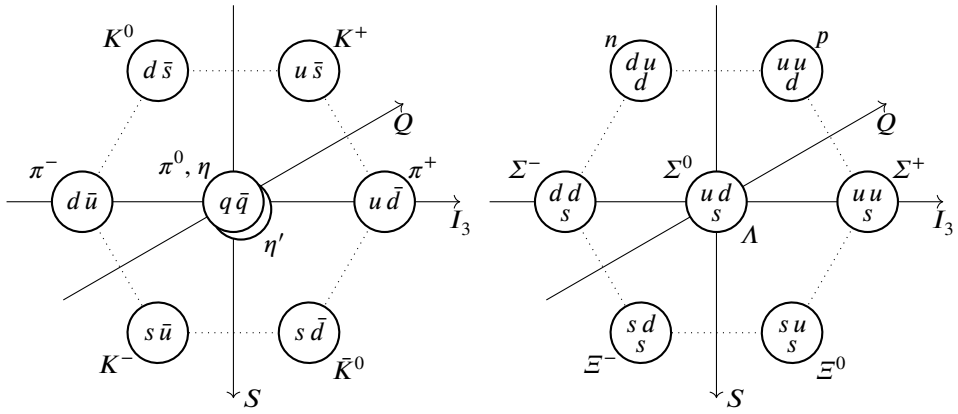
Flavour

Besides colour, quarks are labelled by another quantum number called *flavour* which denotes the presence of more than one species of them. From the point of view of QCD, flavour is a global quantum number and it is not associated to a force mediated by a gauge boson. Six different flavours exist and, neglecting for the moment the difference between quark masses, the theory is completely blind to flavour, which thus generates a global SU(6) symmetry. However, the electroweak sector of the Standard Model breaks flavour symmetry in two ways. First, the six quarks are embedded in pairs in three generations of weak doublets. The doublet components have different electric charge and weak interactions allow transitions between different flavours. Still, these effects provide only a small correction to the dynamics of coloured states: electromagnetic interactions have a coupling $\alpha \approx 1/137$ which is much smaller with respect to the strong coupling in Eq. (0.6); weak interactions at $\approx 1\,\text{GeV}$ are well described by irrelevant four-fermion operators proportional to the Fermi coupling G_{F}, suppressed by the electroweak symmetry breaking scale v

$$G_{\text{F}} \equiv \frac{1}{\sqrt{2}v^2}, \qquad v \approx 246\,\text{GeV}. \tag{0.9}$$

A second more important fact is that electroweak spontaneous symmetry breaking is responsible for the mass of elementary fermions, including quarks, and the range of masses is widespread[1]

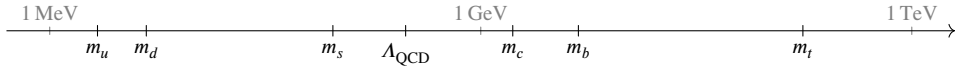
[1] The values quoted here are [17]: running masses $\bar{m}_q^{\overline{\text{MS}}}(2\,\text{GeV})$ for u, d and s; pole masses for c, b and t.



(a) Meson ($B = 0$) nonet. **(b)** Baryon ($B = 1$) octet.

Figure 0.2.: Eightfold Way particle multiplets. Particles along the same horizontal line share the same strangeness S, while those on the same vertical line share the same isospin projection I_3. Electric charge, given by the Gell-Mann–Nishijima formula $Q = I_3 + \frac{1}{2}(B+S)$, is represented on a diagonal axis.

$$\begin{aligned} m_u &\approx 2.15(15)\,\text{MeV}, \quad m_c \approx 1.67(7)\,\text{GeV}, \quad m_t \approx 174.2(14)\,\text{GeV}, \\ m_d &\approx 4.70(20)\,\text{MeV}, \quad m_s \approx 93.8(24)\,\text{MeV}, \quad m_b \approx 4.78(6)\,\text{GeV}. \end{aligned} \tag{0.10}$$



These numbers have to be taken with some care. Indeed, they are not true particle masses since quarks, because of confinement, do not exists as asymptotic states of the interacting QFT. They are parameters of the Lagrangian of the theory, which are not observable and can be precisely defined only after fixing some renormalization scheme. The mass of three of the six flavours of quarks—charm, bottom and top—is higher than the scale of strong interactions, thus when considering light hadrons it is a good approximation to integrate out and completely neglect them.

On the other hand, the mass of the up and down quarks is much smaller than the strong scale Λ_{QCD}, and so is their mass splitting. Therefore, SU(2) flavour rotations acting on the up and down quarks leaves the action almost invariant. This manifests itself as an approximate symmetry of the spectrum: mesons and baryons containing up and down quarks are organized in multiplets of representations of $\text{SU}(2)_I$ classified by the so-called *isospin* quantum number, by analogy with three-dimensional rotations classified by spin.

The mass of the strange quark is approximatively smaller but comparable with the strong scale. Nevertheless, it is worth studying the transformation properties of hadrons under the $\text{SU}(3)_F$ flavour group. This was noticed even before the introduction of QCD and gave rise to a group-theoretical classification of hadrons know as Eightfold Way [20, 21]. For instance, combining an up, down or strange quark with an antiquark we can obtain nine different pseudoscalar mesons. The lightest eight—three pions, four kaons and the η—are nicely organized in an octet

Table 0.1.: The nonet of mesons with pseudoscalar quantum numbers: $J^P = 0^-$. The columns show respectively the transformation properties under $SU(3)_F$ and $SU(2)_I$, the valence quark content and the mass from Ref. [17]. The η and η' mesons are orthogonal linear combinations of the octet meson $\eta_{\mathbf{8}}$ and the singlet meson $\eta_{\mathbf{1}}$ given by

$$\begin{aligned} \eta_{\mathbf{8}} &= (d\bar{d} + u\bar{u} - 2s\bar{s})\Big/\sqrt{6}, \\ \eta_{\mathbf{1}} &= (d\bar{d} + u\bar{u} + s\bar{s})\Big/\sqrt{3}, \end{aligned} \qquad \theta \approx -11.4°.$$

		Meson	Quark content	Mass [MeV]
8	**3**	π^+	$u\bar{d}$	139.570 18(35)
		π^0	$(d\bar{d} - u\bar{u})/\sqrt{2}$	134.9766(6)
		π^-	$d\bar{u}$	139.570 18(35)
	2	K^+	$u\bar{s}$	493.677(16)
		K^0	$d\bar{s}$	497.611(13)
	2	$\bar{K}^0$	$s\bar{d}$	497.611(13)
		K^-	$s\bar{u}$	493.677(16)
	1	η	$\eta_{\mathbf{8}} \cos\theta + \eta_{\mathbf{1}} \sin\theta$	547.862(17)
1	mixing **1**	η'	$-\eta_{\mathbf{8}} \sin\theta + \eta_{\mathbf{1}} \cos\theta$	957.78(6)

representations of flavour, while the ninth—the η'—transforms as a singlet

$$\text{flavour:} \quad \mathbf{3} \otimes \bar{\mathbf{3}} = \mathbf{8} \oplus \mathbf{1} \text{ of pseudoscalar mesons.} \tag{0.11}$$

The organization of this nonet of pseudoscalar mesons under representations of $SU(3)_F$ is pictured in Figure 0.2a, where it is also suggested how the nonet decomposes under the $SU(2)_I$ isospin subgroup. The same information is given in Table 0.1 along with the experimental values of the pseudoscalar meson masses.

In the baryon case, the combination of three up, down or strange (anti)quarks decomposes in

$$\text{flavour:} \quad \mathbf{3} \otimes \mathbf{3} \otimes \mathbf{3} = \mathbf{10} \oplus \mathbf{8} \oplus \mathbf{8} \oplus \mathbf{1} \text{ of baryons (antibaryons).} \tag{0.12}$$

Fermi–Dirac statistics allows only for a spin-3/2 decuplet and a spin-1/2 octet. The latter, which includes the nucleons, is pictured in Figure 0.2b.

Chiral symmetry

In the limit in which the mass of N_f flavours of quarks is sent to zero, the theory is symmetric under a larger group of transformations, acting on left- and right-handed chiral components of Dirac fermions independently. This is called *chiral symmetry* and is described by the group $SU(N_\mathrm{f})_L \times SU(N_\mathrm{f})_R$. $SU(N_\mathrm{f})_F$ flavour symmetry manifests itself as the vectorial subgroup of the chiral symmetry group.

Chiral symmetry of the massless theory is *spontaneously broken* by the strong dynamics: the order parameter is a non-zero chiral condensate

$$\langle \bar{q} q \rangle \sim \Lambda_{\mathrm{QCD}}^3. \tag{0.13}$$

The dynamically generated chiral condensate is left invariant by the vectorial flavour subgroup, which thus still classifies the spectrum of the theory. Moreover, the Goldstone theorem predicts the presence in the theory spectrum of $N_{\mathrm{f}}^2 - 1$ massless NG bosons, which are identified with a multiplet of pseudoscalar mesons fitting an adjoint representation of $\mathrm{SU}(N_{\mathrm{f}})$.

In the real world no quark flavours are exactly massless and thus chiral symmetry is broken also *explicitly*. Nevertheless, if the breaking induced by quark masses is small, approximate chiral symmetry is still relevant to the dynamics of the lightest states. In the $N_{\mathrm{f}} = 2$ case, it explains why the isospin triplet of pions is almost an order of magnitude lighter than other mesons and hadrons: they would be massless NG bosons in the limit $m_u, m_d \to 0$, and the smallness of these quark masses compared to Λ_{QCD} constrains the pion masses to be small. In the $N_{\mathrm{f}} = 3$ case, the strange quark mass breaks chiral symmetry in a substantial way, but it is still possible to relate to the chiral dynamics the comparatively small masses, reported in Table 0.1, of the octet of pseudoscalar mesons pictured in Figure 0.2a.

Singlet

In addition to $\mathrm{SU}(N_{\mathrm{f}})_L \times \mathrm{SU}(N_{\mathrm{f}})_R$ chiral symmetry, the massless limit of QCD at the classical level has another global symmetry, described by the phase rotations $\mathrm{U}(1)_L \times \mathrm{U}(1)_R$ acting on the left- and right-handed chiral components in a flavour-independent way. The vector rotation of this flavour-singlet transformation is an exact symmetry of the theory, responsible for the conservation of the baryon number. On the contrary, the $\mathrm{U}(1)_A$ axial rotation has a peculiar phenomenology. Indeed, it is evident from Table 0.1 that the mass of the η' pseudoscalar meson, which is identified as an almost flavour singlet combination, is $\approx 1\,\mathrm{GeV}$ and comparable to the mass of baryons. The effective theory of chiral symmetry breaking is not sufficient to explain the mass of the η'. The missing ingredient is that the $\mathrm{U}(1)_A$ symmetry is *anomalous*: even in the chiral limit it is explicitly broken by quantum effects, not present in the classical theory.

Studying the theory regularized on the lattice is not the only way to explore QCD at the non-perturbative level. *Large-N_{c}* techniques [22] exploit the simplification of QCD when the number of colours N_{c} is sent to infinity. Notably, in this limit the anomaly responsible for the breaking of $\mathrm{U}(1)_A$ symmetry vanishes. Witten [23] and Veneziano [24, 25] used this fact to obtain a formula to compute the anomalous contribution to the flavour-singlet meson mass, up to corrections of $\mathcal{O}(1/N_{\mathrm{c}})$

$$\frac{F_\pi^2}{2N_{\mathrm{f}}} M_{\eta'}^2 = \chi_t^{\mathrm{YM}}. \tag{0.14}$$

The quantity in the r.h.s. of this formula is the *topological susceptibility* of YM theory, i.e. computed in the theory with $N_{\mathrm{f}} = 0$. In Chapter 3 we review the properties of $\mathrm{U}(1)_A$ transformations, both in the continuum and on the lattice, and we give a derivation of the Witten–Veneziano formula in Eq. (0.14).

The original contribution of the thesis

The work of the thesis inserts in a long-time project of verifying the Witten–Veneziano formula using the lattice discretization of QCD. Since the mass of a meson is experimentally accessible, it is possible to check the approximate validity of Eq. (0.14) computing on the lattice the YM topological susceptibility and comparing it with the experimental value of $M_{\eta'}$. However, Eq. (0.14) is an exact field-theoretic relation only in the chiral limit and in the limit $N_c \to \infty$. Clearly, the experimental mass of the η' meson is not given in this limit.

We aim to compute *both sides* of Eq. (0.14) from *first principles*. The validity of the Witten–Veneziano formula relies on quantization, since the anomaly is a pure quantum effect, and falls in the non-perturbative regime of the theory. Therefore, without even assuming QCD, the quantitative verification of Eq. (0.14) would provide a check of quantization of the theory at the non-perturbative level. Clearly, this program is possible only using lattice regularization and simulations. As a bonus, the lattice framework allows the total control of the theory parameters, including the number of colours, the number of quark flavours and quark masses.

The original material of the thesis is dived in two parts: the first part has to do with the topological susceptibility in the r.h.s. of Eq. (0.14), and solves the problem with a computation of $\chi_t^{\rm YM}$ in the large-N_c limit to the percent level. The second part concerns the direct lattice computation of $M_{\eta'}$. With the present techniques, the computational challenge to obtain $M_{\eta'}$ with an acceptable statistical accuracy is not affordable. The thesis addresses this issue conceptually, proposing a new strategy to compute the η' mass.

Part I: the topological susceptibility in Yang–Mills theory

The topological susceptibility is the variance of the distribution of the topological charge, which is quantity of non-perturbative nature. Therefore, this observable is an excellent candidate to be computed with lattice methods. Nevertheless, as explained in Chapter 3, the definition of the topological charge on the lattice is not free of obstruction. The clash between chiral symmetry and lattice regularization [26] implies a complex renormalization of the topological charge and the topological susceptibility that invalidates the Ward identities behind the Witten–Veneziano mechanism. Only in recent years a method to correctly define the topological charge and its cumulants on the lattice has been found [27], based on a Dirac fermion operator satisfying the Ginsparg–Wilson (GW) relation [28]. This lattice discretization of the topological charge and its properties are addressed in Chapter 3. The downside is that this method is very expensive in terms of computational resources.

In Chapter 4 we introduce a different lattice discretization of the topological charge, based on the YM gradient flow [29]. We prove that the cumulants of the topological charge distribution defined at positive flow time have a proper continuum limit which satisfies the anomalous chiral WTIs [1]. Employing this discretization, in Chapter 5 we present the results of a high-statistics computation of the second – i.e. the susceptibility – and fourth cumulants of the topological charge distribution in SU(3) YM theory [1]. In Chapter 6 we extend the computation of the susceptibility at SU(N_c) YM theory, up to $N_c = 6$, in order to extract with confidence the value of the topological susceptibility in the $N_c \to \infty$ limit [3]. This is the proper value to insert in the r.h.s. of Eq. (0.14).

Part II: the flavour-singlet pseudoscalar meson mass

The lattice computation of $M_{\eta'}$ is an open problem. Indeed, it is practically not possible to obtain the mass of the flavour-singlet pseudoscalar with a traditional approach, from the exponential decay rate at asymptotic distances of the proper two-point function. The reason is a very general problem with the signal-to-noise ratio of correlation functions: the statistical error decays asymptotically with a lower rate with respect to the signal. Therefore, with a standard Monte Carlo algorithm, the computational cost to extract the signal at longer distances maintaining the same statistical accuracy increases exponentially. This problem is particularly bad in the case of the flavour-singlet propagator, since it contains disconnected quark diagrams whose error do not decrease at all with distance. In Chapter 7 we state precisely this problem and introduce a method which has been successfully applied to bosonic theories: the multilevel Monte Carlo algorithm [30–32].

Multilevel integration requires the decomposition in space-time domains of the action and the observables, but an exact factorization is not possible if the theory contains fermions. We partially solve this problem introducing an approximate factorization for fermionic observables [2]. The exact result is easily recovered with a simple modification of the multilevel Monte Carlo algorithm. The application of this method to the disconnected contribution of the flavour-singlet propagator is particularly simple. We present the factorization of the disconnected contribution in Chapter 8, including numerical evidence obtained in the quenched theory that shows an exponential gain in the signal-to-noise ratio.

The factorization of the action has been obtained meanwhile in Ref. [4]. Therefore, the combination of the factorization of the propagator presented in Chapter 8 and the action in Ref. [4] paves the way to multilevel simulations with fermions. These in turn are a candidate solution to a broad class of signal-to-noise problems, including the flavour-singlet propagator.

1. Quantum Chromodynamics

QCD is a renormalizable *gauge theory* with SU(3) gauge group. Gauge theories are a special class of QFT in which the Lagrangian is manifestly invariant under a *local* group of fields transformations.

In this chapter, we introduce gauge theories working in Euclidean spacetime, in contrast with usual Minkowski spacetime. For this reason, the expressions introduced in this section differ from the ones in the Introduction. A QFT in Minkowski spacetime is mapped to a Euclidean one by a procedure called *Wick's rotation*, which involves the substitution $x_4 \equiv \mathrm{i}x^0$ and others summarized in Section A.6. Whether or not it is possible to analytically continue a QFT from real to imaginary time is not obvious. The criteria to be satisfied were given by Osterwalder and Schrader [33, 34]. In particular, a QFT in Euclidean spacetime must satisfy the so-called *reflection positivity* to correspond to a well-defined QFT in Minkowiski spacetime. The use of Euclidean spacetime is preparatory to Chapter 2, where the lattice discretization is applied, for technical reasons, to the Euclidean theory.[1]

After imposing gauge symmetry, the structure of the QFT is very constrained. In fact, once a gauge group is chosen and the field content is given, the request of renormalizability completely defines the action of QCD. In Section 1.1 we introduce the action of a non-Abelian SU(N_c) gauge theory with N_f species of Dirac fermions and we quantize it in the path integral formalism. Then, in Section 1.2 we complete the building of this QFT renormalizing it. To conclude the chapter, Section 1.4 is dedicated to chiral symmetry.

1.1. Gauge theories

To introduce the action, we specify the criteria that define the gauge theory. In this thesis, we work with gauge theories with gauge group SU(N_c), where $N_\mathrm{c} \geq 2$ is the unspecified number of colours. This includes the physical case of QCD, which is easily obtained specifying $N_\mathrm{c} = 3$. However, as we will show, it is sometimes useful to consider gauge theories with a different number of colours.

The minimal gauge theory is Yang–Mills theory, which contains only one multicomponent field, $A^a_\mu(x)$. This field possesses a Lorentz index μ and an internal colour index a. In spite of this, A^a_μ is not a proper Lorentz vector [35]. The Lorentz transformation properties are entangled with the internal gauge invariance. A proper Lorentz vector describes a spin-1 massive particle, with three internal degrees of freedom. The massless limit is singular: a massless particle has only two degrees of freedom with *helicity* ± 1. Gauge invariance is necessary for the field $A^a_\mu(x)$ to describe massless particles.

[1] In the following, we use improperly Poincaré symmetry to denote the ISO(4) group of isometries of Euclidean spacetime. Similarly, with Lorentz transformation rules we refer to transformation rules under the homogeneous subgroup SO(4).

From the gauge-group point of view, $A^a_\mu(x)$ transforms under the adjoint N_c^2-1-dimensional representation of the non-Abelian gauge group. This gives rise to a N_c^2-1 multiplicity of massless gauge bosons. In the defining representation, a generic gauge transformation is represented by a unitary $N_c \times N_c$ matrix with unit determinant $\Omega(x)$ for each spacetime point x, which can also be written as the exponential of a Hermitian traceless $N_c \times N_c$ matrix $\lambda(x) \equiv \lambda^a(x) T^a$

$$\Omega(x) = \exp\{\mathrm{i}\lambda^a(x) T^a\}. \tag{1.1}$$

Here, T^a for $a = 1, \dots, N_c^2-1$ are the Hermitian generators of $\mathrm{SU}(N_c)$ introduced in Section A.1, and the $\lambda^a(x)$ parametrize the gauge transformation. In this representation, the gauge field is represented by a Hermitian traceless matrix $A_\mu \equiv A^a_\mu T^a$, which transforms inhomogeneously under the gauge group

$$A_\mu(x) \to A'_\mu(x) = \Omega(x)\big[A_\mu(x) + \mathrm{i}\partial_\mu\big]\Omega^\dagger(x). \tag{1.2}$$

The action is built as a functional of A_μ and derivatives of A_μ on the basis of symmetry considerations. It has to be invariant under local gauge transformations. A proper field derivative is given by the *field strength tensor*

$$F_{\mu\nu} \equiv -\mathrm{i}\big[D_\mu, D_\nu\big] = \partial_\mu A_\nu - \partial_\nu A_\mu - \mathrm{i}\big[A_\mu, A_\nu\big], \tag{1.3}$$

where we introduced the *covariant derivative*

$$D_\mu \equiv \partial_\mu - \mathrm{i}A_\mu. \tag{1.4}$$

As the field itself, $F_{\mu\nu}$ lives in the adjoint representation of $\mathrm{SU}(N_c)$

$$F_{\mu\nu} \equiv F^a_{\mu\nu} T^a, \qquad F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + f^{abc} A^b_\mu A^c_\nu, \tag{1.5}$$

but, differently from A_μ, it transform homogeneously

$$F_{\mu\nu}(x) \to F'_{\mu\nu}(x) = \Omega(x) F_{\mu\nu}(x) \Omega^\dagger(x). \tag{1.6}$$

Using the field strength tensor, terms that are Lorentz-scalars and invariant under the gauge group are easily written. Now, we impose renormalizability as a last criterium. This constrains field monomials in the Lagrangian to be of canonical dimension $d \leq 4$. The resulting YM action is composed of just two $d = 4$ terms[2]

$$S_g[A] = \int \mathrm{d}^4x \left[\frac{1}{2g_0^2}\,\mathrm{tr}\big\{F_{\mu\nu}F_{\mu\nu}\big\} - \frac{\mathrm{i}\theta}{32\pi^2}\epsilon_{\mu\nu\rho\sigma}\,\mathrm{tr}\big\{F_{\mu\nu}F_{\rho\sigma}\big\}\right]. \tag{1.7}$$

[2] We normalize the gauge field in order to have the gauge coupling g_0 appearing at the denominator in front of the YM action. For this reason, there is no coupling in the covariant derivative defined in Eq. (1.4). The perturbative QCD convention is recovered with the substitution

$$A_\mu(x) \to g_0 A_\mu(x).$$

Other terms, such as a mass term $\mathrm{tr}\{A_\mu A_\mu\}$, are not allowed by the requirement of gauge invariance. In contrast with the Abelian case, this is not a free theory: it is evident from Eq. (1.3) that $F_{\mu\nu}$ contains a term quadratic in A_μ. Therefore, the action in Eq. (1.7) implicitly includes interactions between three and four gauge bosons. Remarkably, symmetry alone completely dictates the structure of these interactions. The only free parameters are two dimensionless *bare* couplings, g_0 and θ.

The second term in Eq. (1.7) is the so-called *θ-term*. It has peculiar properties, which will be described in details in Chapter 3. For the moment, it suffices to notice that excluding this last term the Euclidean action is real. The θ-term is purely imaginary instead: this is because it contains an odd number of times derivatives, differently from every other terms, that results in an extra i factor after Wick's rotation. In the physical world, $\theta \lesssim 1 \times 10^{-10}$, as we will motivate in Chapter 3. For this reason, in the following we restrict to the case $\theta = 0$. This implies a real action which can be easily shown to be bounded from below.

1.1.1. The gauge theory with fermions

Physical gauge theories include additional *matter* fields besides A_μ. These can be Dirac or Weyl fermions or scalars, in any representation of the gauge group. In this thesis, we focus to the case that includes QCD: the gauge theory of N_f Dirac fermions transforming under the fundamental representation of $\mathrm{SU}(N_\mathrm{c})$. The N_f species are labelled by *flavour*.

This choice excludes the electroweak sector of the SM, which contains Weyl fermions and scalars. Also, we do not consider representations different from the fundamental for matter fields. These are common in models of physics beyond the SM.

A Dirac fermion transforms under a *reducible*[3] representation of the Lorentz group. It is represented by a multicomponent Dirac spinor field $\psi(x)$. As a Dirac spinor, this field is a four-component vector in Dirac γ-matrices space. The convention used for Dirac γ-matrices is given in Section A.3. On top of this, $\psi(x)$ has internal indices: it is a N_c-component vector transforming under the fundamental $\boldsymbol{N}_\mathrm{c}$ representation of $\mathrm{SU}(N_\mathrm{c})$. Moreover, when $N_\mathrm{f} \geq 2$, we use the notation $\psi(x)$ to denote a vector of N_f fields: $\psi = (\psi_u, \psi_d, \psi_s, \dots)$.

To write the fermion action, it is useful to introduce the Dirac adjoint spinor field

$$\bar{\psi}(x) \equiv \psi^\dagger(x)\gamma_4, \tag{1.8}$$

which transforms under the antifundamental $\bar{\boldsymbol{N}}_\mathrm{c}$ representation of $\mathrm{SU}(N_\mathrm{c})$. Scalar gauge-invariant field monomials are built contracting spinor, Lorentz and gauge indices in bilinear of ψ and $\bar{\psi}$. Ignoring the flavour structure, there are only two of them with $d \leq 4$

$$\bar{\psi}(x)\psi(x), \qquad \bar{\psi}(x)\gamma_\mu\partial_\mu\psi(x). \tag{1.9}$$

The action of free Dirac fermions with a *global* $\mathrm{SU}(N_\mathrm{c})$ symmetry is composed just of these two terms

$$S_\mathrm{f}^\mathrm{free}[\bar{\psi}, \psi] = \int \mathrm{d}^4x\, \bar{\psi}(x)\big[\not{\partial} + M_0\big]\psi(x), \qquad \not{\partial} \equiv \gamma_\mu\partial_\mu, \tag{1.10}$$

where M_0 is a $N_\mathrm{f} \times N_\mathrm{f}$ matrix of bare fermion masses. Other terms, such as four-fermion fields, are excluded either by symmetries or by renormalizability. In view of the path integral approach

[3] The representation is reducible in even spacetime dimensions.

to quantization, it is convenient to consider the action as a functional of independent $\psi(x)$ and $\bar{\psi}(x)$ fields.

Interactions with gauge bosons are introduced requiring ψ and $\bar{\psi}$ to be invariant under *local* gauge transformations

$$\psi(x) \to \psi'(x) = \Omega(x)\psi(x), \tag{1.11a}$$

$$\bar{\psi}(x) \to \bar{\psi}'(x) = \bar{\psi}(x)\Omega^{\dagger}(x). \tag{1.11b}$$

While the first term in Eq. (1.9) is still invariant under transformation rules in Eqs (1.11), the second term is not: the spacetime derivative acting on $\Omega(x)$ generates an extra term. A gauge-covariant derivative field $D_\mu\psi(x)$ is obtained applying the covariant derivative in Eq. (1.4) to $\psi(x)$

$$D_\mu\psi(x) = \partial_\mu\psi(x) - \mathrm{i}A_\mu(x)\psi(x), \qquad D_\mu\psi(x) \to \big[D_\mu\psi\big]'(x) = \Omega(x)D_\mu\psi(x), \tag{1.12}$$

where $\psi(x)$ and $A(x)$ are transformed according to Eqs (1.11a) and (1.2) respectively. Therefore, the gauge-invariant fermion action is

$$S_{\mathrm{f}}[\bar{\psi}, \psi, A] = \int \mathrm{d}^4x\, \bar{\psi}(x)\big[\not{D} + M_0\big]\psi(x), \tag{1.13}$$

where

$$\not{D} \equiv \gamma_\mu D_\mu \tag{1.14}$$

is the interacting *Dirac operator*.

The fermion action in Eq. (1.13) together with the bosonic action in Eq. (1.7) describes a theory of Dirac fermions with non-Abelian interactions

$$S[\bar{\psi}, \psi, A] = S_{\mathrm{g}}[A] + S_{\mathrm{f}}[\bar{\psi}, \psi, A]. \tag{1.15}$$

Like autointeractions between gauge bosons, also the interactions of fermions with one gauge boson are completely determined by symmetry principles alone. Restricting to the $\theta = 0$ case, the only parameters are a dimensionless bare gauge coupling g_0 and the bare fermion mass matrix M_0.

1.1.2. The path integral

We proceed to quantize the theory. The canonical quantization of gauge theories is highly non-trivial. The cleanest way to quantize a QFT with local gauge invariance is called *BRST quantization*, and involves the rewriting of gauge invariance as *global* symmetry, depending on an *anticommuting* parameter and acting on an extended field content. This includes colour-charged scalar fields with Fermi–Dirac statistics c ($\bar{c}$) called *(anti)ghosts*.

In the thesis, we avoid these complications using the *path integral* or functional approach to quantization. We start introducing the formal expression for the path integral of the *partition function*

$$\mathcal{Z} = \int \mathcal{D}[\bar{\psi}, \psi, A]\, \mathrm{e}^{-S[\bar{\psi}, \psi, A]}. \tag{1.16}$$

The symbol $\mathcal{D}[\bullet]$ in Eqs (1.17) and (1.16) denotes the path integral measure, which roughly means 'integrate over all configurations of the field'. It is worth noting that $\psi(x)$ and $\bar{\psi}(x)$ are integrated as independent fields. Moreover, to reproduce Fermi–Dirac statistics, they are *Grassmann number*-valued fields.

The formulae introduced so far are just formal expressions. The correct definition of the path integral and its measure requires a regularization. We postpone this to Chapter 2, in which lattice regularization is introduced. Once a proper regularized path integral is introduced, there is no need to fix the gauge and introduce ghosts to work with it, provided that we do not want to derive Feynman rules for perturbation theory.

In path integral quantization, canonical operators $\hat{O}_i(x)$ are mapped to composite fields $O_i(x) = O_i[\psi(x), \bar{\psi}(x), A(x)]$. Green's functions of time-ordered operators are given by the path-integral expectation values of products of fields

$$\left\langle O_{i_1}(x_1) \cdots O_{i_n}(x_n) \right\rangle = \frac{1}{\mathcal{Z}} \int \mathcal{D}[\bar{\psi}, \psi, A]\, \mathrm{e}^{-S[\bar{\psi},\psi,A]} O_{i_1}(x_1) \cdots O_{i_n}(x_n). \tag{1.17}$$

It is worth noting that in Eqs (1.16) and (1.17) is missing an i factor in the exponential of the action, while it is present in the Minkowski path integral. This factor has been absorbed in a redefinition of the action after Wick's rotation

$$-S \equiv \mathrm{i} S_{\text{Mink.}}\Big|_{\mathrm{i}x^0 \to x_4}. \tag{1.18}$$

Since the action is real and bounded from below, the Euclidean path integral in Eq. (1.16) has, at least formally, the same structure of the partition function of a statistical system. This analogy is of practical use with the introduction of lattice regularization in Chapter 2.

1.2. Renormalization

To complete the definition of a QFT, the path integral must be paired with a set of renormalization conditions. Without these, a path integral defined in terms of bare couplings g_0 and M_0 does not describe a meaningful QFT.

Renormalization is often considered as a two-steps procedure. The first step involves taming mathematically infinities in the process know as *regulatization*. A *regulator* is introduced to this purpose. The predictions of the regulated theory are free of infinities, but the regulator may break some of the symmetries of the theory.

In many cases, the regulator assumes the form of a *cut-off* energy scale a^{-1}. Our choice for the thesis is *lattice regularization*, to be introduced in Chapter 2. This method requires the introduction of a discrete spacetime lattice, with the energy cut-off given by the inverse of the lattice spacing a. This results in the breaking of the continuous Poincaré symmetry, but remarkably it is possible to conserve exact gauge invariance. The lattice-regularized theory is well-suited for non-perturbative calculations, but also Feynman rules for perturbative calculations have been derived.

Dimensional regularization [36] is another well-known regularization procedure. In this case, no dimensionful cut-off is introduced, but correlation functions computed in perturbation

theory are treated as functions of the spacetime dimension D and analytically continued to non-integer $D = 4 - 2\epsilon$.[4] Infinities in $D = 4$ appears in correlation functions as poles for $\epsilon \to 0$. Thanks to the fact that it preserves almost all the symmetries of a gauge theory, dimensional regularization has been used to prove that non-Abelian gauge theories are renormalizable to all order of perturbation theory. A drawback is that dimensional regularization is limited to perturbative calculations.

The process of regularization introduces a dependence on the regulator. The second step of renormalization aims to remove this dependence. In the case of a dimensionful cut-off, this is obtained increasing the separation between the interesting energies Q and the cut-off a^{-1}, sending the latter to infinity: $a^{-1} \to \infty$. To this purpose, a set of *renormalization conditions* is imposed: some observable field-theoretical predictions, e.g. particle masses or scattering amplitudes, are fixed to a definite value. While the regulator is removed, the bare parameters of the theory are tuned in order to reproduce the same renormalization conditions. The bare parameters gain an implicit dependence on the cut-off in this way.

If it is possible to obtain a regulator-independent theory setting a finite number of renormalization conditions, a QFT is said to be *renormalizable*. YM theory and QCD are renormalizable QFTs. This is linked to the absence in the action of coupling with negative mass dimension. For instance, a gauge theory with four-fermion interactions is non-renormalizable.

The procedure of renormalization leads to the introduction of *renormalized* gauge coupling and quark masses, from which any dependence on the regulator has been removed. They are related to the bare parameters by dimensionless renormalization constants Z_i

$$g^2(\mu) = Z_g(g_0, a\mu) g_0^2, \qquad M(\mu) = Z_M(g_0, a\mu) M_0. \tag{1.19}$$

The price to pay to get rid of the regulator is the appearance of the *renormalization scale* μ. Similar Z_i constants are required also for fields

$$A_{\mathrm{R}} = Z_3^{1/2} A, \qquad \psi_{\mathrm{R}} = Z_2^{1/2} \psi, \quad \bar{\psi}_{\mathrm{R}} = Z_2^{1/2} \bar{\psi}. \tag{1.20}$$

Different sets of renormalization conditions define different *renormalization schemes*. While observable quantities must be independent from this choice, generic correlation functions are in general scheme-dependent. Many scheme have been proposed. A very common one is minimal subtractions (MS) [37, 38]. In this scheme, the renormalization conditions are defined indirectly: at any given order of perturbation theory, the divergent part of dimensionally-regularized amplitudes—i.e. poles in ϵ—is subtracted. The related $\overline{\mathrm{MS}}$ removes in addition the universal constant $\ln(4\pi) - \gamma_E$. Other schemes include momentum subtraction (MOM) and finite-volume schemes, which are well-suited for non-perturbative applications.

1.2.1. Renormalization à la Wilson

A different, more modern, point of view on renormalization was introduced by Wilson and others [39]. In their formalism, every theory is regularized assuming the existence of a fundamental cut-off. Then, the *renormalization group* links the change of the physics with the energy

[4] However, scale invariance is broken by a dimensionful parameter μ, which compensates for the dimensionality of the measure $\mathrm{d}^D x$.

scale Q to the change of the parameters of the theory c_i. These can be classified according to their mass dimension $[c_i]$

- $[c_i] > 0$: *irrelevant*, if they are more important at low energies;
- $[c_i] = 0$: *marginal*, if their importance depends only logarithmically on the energy;
- $[c_i] < 0$: *relevant*, if they are more important at high energies.

For instance, in $D = 4$ the gauge coupling is a marginal parameter and a fermion mass is an irrelevant parameter.

With this interpretation, renormalization is just an emergent property. Indeed, following the flow of the renormalization group form the cut-off scale to lower energies, the relevant couplings become less important. In this way, every theory at energies $Q \ll a^{-1}$ looks approximatively like a renormalizable QFT, to become true renormalizable QFT in the limit $a^{-1} \to \infty$.

On the other hand, a non-renormalizable QFT is an effective field theory, i.e. an effective description—valid only up to energies of the order of a^{-1}—of a more complete theory.

1.2.2. The renormalization group

To define the renormalized theory we introduced a renormalization scale μ. Differently from the cut-off, μ can and should be a low-energy scale, of the same order of the processes we are interested in. However, it is still an arbitrary scale, and the physics should not depend on it. Starting from this simple observation, *renormalization group* (RG) methods [40, 41] were developed to relate quantities at different μ.

Consider the gauge-invariant correlation function $G_0^{r,n}(\{x_i\})$ computed in the regularized theory, depending on spacetime coordinates $\{x_i\}$ and involving r gauge fields and n quark-antiquark pairs. It is related to the renormalized correlation function by

$$\left(Z_3^{-1/2}\right)^r \left(Z_2^{-1/2}\right)^{2n} G_{\mathrm{R}}^{r,n}\left(\{x_i\}; \mu, g, M\right) = G_0^{r,n}\left(\{x_i\}; g_0, M_0\right). \tag{1.21}$$

Renormalization group equations [42–44], or Callan–Symanzik equations, follow immediately from the request of μ-independence of $G_0^{r,n}$[5]

$$\mu \frac{\mathrm{d}G_0^{r,n}}{\mathrm{d}\mu} = 0 \quad \Rightarrow \quad \left[\mu\frac{\partial}{\partial\mu} + \beta\frac{\partial}{\partial g} + \tau M \frac{\partial}{\partial M} - r\gamma_3 - 2n\gamma_2\right] G_{\mathrm{R}}^{r,n} = 0, \tag{1.22}$$

where

$$\beta(g) = \mu\frac{\partial g}{\partial\mu}, \qquad M\tau(g) = \mu\frac{\partial M}{\partial\mu}, \tag{1.23a}$$

$$\gamma_3(g) = \frac{1}{2}\mu\frac{\partial}{\partial\mu}\ln Z_3, \qquad \gamma_2(g) = \frac{1}{2}\mu\frac{\partial}{\partial\mu}\ln Z_2. \tag{1.23b}$$

[5] An alternative and equivalent version of RG equations is obtained for correlation function $G_0^{r,n}$ regularized with a physical cut-off a^{-1} requiring the renormalized $G_{\mathrm{R}}^{r,n}$ to be cut-off independent

$$a\frac{\mathrm{d}G_{\mathrm{R}}^{r,n}}{\mathrm{d}a} = 0 \quad \Rightarrow \quad \left[a^{-1}\frac{\partial}{\partial a^{-1}} + \beta(g_0)\frac{\partial}{\partial g_0} + \tau M_0 \frac{\partial}{\partial M_0} + r\gamma_3(g_0) + 2n\gamma_2(g_0)\right] G_0^{r,n} = 0.$$

When renormalization is studied in perturbation theory, the Z_is and their functions have a perturbative expansion in g. For β and τ this expansion is

$$\beta(g) = -g^3 \sum_{k=0}^{\infty} b_k g^{2k}, \qquad \tau(g) = -g^2 \sum_{k=0}^{\infty} d_k g^{2k}. \tag{1.24}$$

The leading coefficients b_0, b_1 and d_0 in these expansion are universal, in the sense that do not depend on the particular regularization and renormalization scheme. In a theory with $\mathrm{SU}(N_\mathrm{c})$ gauge group and N_f quarks transforming under the fundamental representation, they are

$$b_0 = \frac{1}{(4\pi)^2}\left[\frac{11}{3}N_\mathrm{c} - \frac{2}{3}N_\mathrm{f}\right], \tag{1.25a}$$

$$b_1 = \frac{1}{(4\pi)^4}\left[\frac{34}{3}N_\mathrm{c}^2 - \left(\frac{13}{3}N_\mathrm{c} - \frac{1}{N_\mathrm{c}}\right)N_\mathrm{f}\right], \tag{1.25b}$$

$$d_0 = \frac{1}{(4\pi)^2}\frac{3(N_\mathrm{c}^2-1)}{N_\mathrm{c}}. \tag{1.25c}$$

1.2.3. The Λ-parameter

Physical observables $E(\mu, g, M)$ are independent of field renormalization. They obey a simplified RG equation

$$\left[\mu\frac{\partial}{\partial\mu} + \beta\frac{\partial}{\partial g} + \tau M\frac{\partial}{\partial M}\right]E = 0, \tag{1.26}$$

which can be solved in full generality in terms of special solutions.

The first RG-invariant scale which solves Eq. (1.26) is a $d = 1$ quantity $\Lambda(\mu, g)$ independent on quark masses. In the renormalized theory, it must be proportional to the renormalization scale μ times a function of the renormalized coupling $g(\mu)$

$$\Lambda = \mu l\big(g(\mu)\big), \tag{1.27}$$

where $l(g)$ satisfies the simplified RG equation

$$\left[1 + \beta\frac{\partial}{\partial g}\right]l(g) = 0. \tag{1.28}$$

This equation has a formal solution

$$l(g) = \exp\left\{-\int_{g^*}^{g}\frac{\mathrm{d}x}{\beta(x)}\right\}, \tag{1.29}$$

where g^* acts as an integration constant. In principle, any choice of g^* is a valid solution. We would like to set $g^* = 0$, but the fact that $\beta(g)$ vanishes at small g gives rise to a divergent integral. Fortunately, we can use the perturbative expansion of the β-function, valid for small g,

$$\beta(g) = -b_0 g^3\left[1 + \frac{b_1}{b_0}g^2 + \mathcal{O}\big(g^4\big)\right] \quad \Rightarrow \quad \frac{1}{\beta(g)} = -\frac{1}{b_0 g^3}\left[1 - \frac{b_1}{b_0}g^2\right] + \mathcal{O}(g) \tag{1.30}$$

to isolate the μ-independent divergent term and remove it

$$\int_{g^*}^{g} \frac{\mathrm{d}x}{\beta(x)} = \int_{g^*}^{g} \mathrm{d}x \left[\frac{1}{\beta(x)} + \frac{1}{b_0 x^3} - \frac{b_1}{b_0^2 x} \right] + \left[\frac{1}{2b_0 x^2} + \frac{b_1}{2b_0^2} \ln b_0 x^2 \right]_{g^*}^{g}. \tag{1.31}$$

The RG equation (1.26) solution obtained in this way is called the *Λ-parameter*

$$\Lambda = \mu \big(b_0 g^2\big)^{-b_1/2b_0^2} \mathrm{e}^{-1/2b_0 g^2} \exp\left\{ - \int_0^{g(\mu)} \mathrm{d}x \left[\frac{1}{\beta(x)} + \frac{1}{b_0 x^3} - \frac{b_1}{b_0^2 x} \right] \right\}. \tag{1.32}$$

The Λ-parameter is a scheme-dependent quantity: it is sensible to one-loop finite part of Z_g, which is not universal. However, a one-loop computation is sufficient to relate exactly Λ-parameters in different renormalization schemes.

Once the Λ-parameter is fixed, Eq. (1.32) can be solved for the renormalized coupling g at any physical scale $Q = \mu$, to obtain *running coupling*

$$g^2(Q) = \bar{g}^2(t), \qquad t \equiv \ln \frac{Q^2}{\Lambda^2}. \tag{1.33}$$

The dimensionless coupling g at any scale Q is thus completely determined by the dimensionful Λ-parameter. In theories without other dimensionful couplings, such as YM theory or massless QCD, the Λ-parameter alone is sufficient to fix the high energy behaviour of any physical observable $E(Q^2, \mu, g)$

$$E = \Lambda^d \tilde{E}\left(\frac{Q^2}{\mu^2}, g \right) = \Lambda^d \tilde{E}\big(1, \bar{g}(t)\big), \qquad d = [E]. \tag{1.34}$$

In particular, the mass of particle can only be proportional to the Λ-parameter

$$M_X = \Lambda \tilde{M}_X, \tag{1.35}$$

where the dimensionless proportionality coefficient $\tilde{M}_X$ is a *geometric* property of the theory, in the sense that is completely determined by the gauge group and the field content.

At the classical level, the action of YM theory or massless QCD does not have any dimensionful scale. They are thus scale-invariant theories. At the quantum level, *dimensional transmutation* happens: quantization creates the scale Λ. There is a very elegant way to see this: Noether's charge of scale symmetry is the trace of the energy-momentum tensor $T_{\mu\mu}$. In the quantum theory, $T_{\mu\mu}$ is not conserved. Its conservation is broken by an *anomaly* proportional to the β-function. In conclusion, classical scale invariance is broken by a purely quantum effect proportional to the rate of change of the coupling with the energy, encoded in the β-function.

1.2.4. Asymptotic freedom

Using two-loop perturbation theory, the asymptotic behaviour of the running coupling for $Q^2 \to \infty$ is given analytically by

$$\bar{g}^2(t) = \frac{1}{b_0 t} \left[1 - \frac{b_1}{b_0^2 t} \ln t + \mathcal{O}\big(t^{-2}\big) \right], \tag{1.36}$$

where $t = \ln\left(Q^2/\Lambda^2\right)$. This formula shows that the running coupling vanishes logarithmically in the limit of energies $Q^2 \to \infty$. The fact that the coupling itself is small in this region justifies the use of perturbation theory to obtain this result, known as *asymptotic freedom* [14, 15].

In general, the values g^* for which $\beta(g^*) = 0$ are fixed points of the RG flow. The theory at g^* manifests scale invariance. In the neighbourhood of a g^*, the theory flows slowly towards or away from it, with the direction of the flow determined by the sign of β. $g^* = 0$ is a fixed point in the perturbative regime of the theory. It represents the free theory, which is trivially scale invariant. The leading coefficient in the expansion of $\beta(g)$, b_0 in Eq. (1.25a), was computed for a non-Abelian SU(N_c) gauge theory in 1973[6] by Gross and Wilczek [14], and independently by Politzer [15]. They discovered asymptotic freedom showing that, if the number of flavours is not too high

$$N_f < \frac{11}{2} N_c \quad \Rightarrow \quad b_0 > 0. \tag{1.37}$$

This means that $\beta(g)$ is negative for small g and $g^* = 0$ is an UV fixed point, which is attractive at high energies. Non-Abelian gauge theories are the only renormalizable QFTs in four spacetime dimension that show asymptotic freedom.

In Wilson's interpretation of the RG, non-Abelian gauge interactions modify the marginal gauge coupling to be *slightly irrelevant*. The *slightly* means that it changes slowly, with the logarithm of the energy. The free-theory fixed point $g^* = 0$ is not reached at any finite energy but only asymptotically for energies that goes to infinity. For this reason, this high-energy behaviour is called asymptotic freedom.

1.2.5. RGI masses

In QCD with massive quarks there are additional dimensionful coupling. We introduce other special solutions of Eq. (1.26): the RG-invariant masses

$$M^{\mathrm{RGI}} = M(\mu)\theta\big(g(\mu)\big), \tag{1.38}$$

where $\theta(g)$ satisfies the simplified RG equation

$$\left[\beta \frac{\mathrm{d}}{\mathrm{d}g} + \tau(g)\right]\theta(g) = 0. \tag{1.39}$$

With a procedure similar to the Λ-parameter, a solution is[7]

$$M^{\mathrm{RGI}} = M(\mu)\big(2b_0 g^2\big)^{-d_0/2b_0^2} \exp\left\{ -\int_0^{g(\mu)} \mathrm{d}x \left[\frac{\tau(x)}{\beta(x)} - \frac{d_0}{b_0^2 x}\right]\right\}, \tag{1.40}$$

and the *running mass* is given by

$$M(Q) = \bar{M}(t) = \frac{M^{\mathrm{RGI}}}{\theta\big(\bar{g}(t)\big)}. \tag{1.41}$$

[6] b_0 was computed already in 1969 for the SU(2) case, relevant for electroweak interactions, by Khriplovich [45]. He concluded that the interactions become weak at short distances. Gross, Wilczek and Politzer were the first to relate this results to the violation of Bjorken scaling in strong interactions.

[7] There is no universally accepted normalization for RG-invariant masses. The normalization in Eq. (1.40) complies with Gasser and Leutwyler [46–48].

It can be shown that the RG-invariant masses in Eq. (1.40) are scheme- and scale-independent.

The special solutions in Eqs (1.32) and (1.40) allow to solve the RG equation (1.26) in the general case. Consider for instance a dimensionless renormalized observable $E(Q^2, \mu, g, M)$, depending on renormalized quantities and a momentum Q. Since it is dimensionless

$$E = \tilde{E}\left(\frac{Q^2}{\mu^2}, g, \frac{M}{\sqrt{Q^2}}\right). \tag{1.42}$$

Now, solving the RG equation to impose the renormalization scale independence fixes completely the behaviour of E for large Q^2

$$E = \tilde{E}\left(1, \bar{g}(t), \frac{\bar{M}(t)}{\sqrt{Q^2}}\right). \tag{1.43}$$

Dimensionful observable can be proportional to any combination of the two scales Λ and M. Studying the theory in the massless limit, in Section 1.4 we will analyse the scaling of hadron masses with M.

1.2.6. Renormalization of composite fields

Consider the bare correlation function of elementary fields in Eq. (1.21) with the addition of a bare local composite field $O_i(y)$, all localized at different spacetime points

$$G_0^{r,n,i}\big(\{x_i\}, y\big). \tag{1.44}$$

In general, the renormalization of coupling and elementary fields as in Eq. (1.21) is not sufficient to remove all the infinities form these correlation functions. Additional infinities have to be removed renormalizing the composite field. In the simplest situation, a multiplicative renormalization is sufficient

$$O_{\mathrm{R}i} = Z_i O_i \quad \Rightarrow \quad G_{\mathrm{R}}^{r,n,i} = Z_3^{r/2} Z_2^n Z_i G_0^{r,n,i}. \tag{1.45}$$

In this case, the renormalized correlation function $G_{\mathrm{R}}^{r,n,i}$ satisfies the RG equation (for the case of massless quarks)

$$\left[\mu\frac{\partial}{\partial\mu} + \beta\frac{\partial}{\partial g} + \tau M\frac{\partial}{\partial M} - r\gamma_3 - 2n\gamma_2 - \gamma_i\right] G_{\mathrm{R}}^{r,n,i} = 0, \tag{1.46}$$

where

$$\gamma_i(g) = \mu\frac{\partial}{\partial\mu}\ln Z_i. \tag{1.47}$$

More generally, the field mixes with other fields having the same transformation properties under symmetries and equal or lower canonical dimension

$$O_{\mathrm{R}i} = \sum_j Z_{ij} O_j. \tag{1.48}$$

In these cases, the RG equation has to be solved as a matrix equation

$$\left[\left(\mu\frac{\partial}{\partial\mu}+\beta\frac{\partial}{\partial g}+\tau M\frac{\partial}{\partial M}-r\gamma_3-2n\gamma_2\right)\delta_{ij}-\gamma_{ij}\right]G_{\mathrm{R}}^{r,n,j}=0, \tag{1.49}$$

where

$$\gamma_{ij}(g)=\mu\frac{\partial Z_{ik}}{\partial\mu}\left(Z^{-1}\right)_{kj}. \tag{1.50}$$

1.2.7. The operator product expansion

Until now, we have always supposed the fields to stay at a physical distance. However, we will also be interested in correlation functions with fields that tends to small separations, to the point that they coincide. For instance, consider the correlation function involving renormalized local gauge-invariant fields

$$\langle O_a(x)O_b(y)\phi(z_1)\cdots\phi(z_r)\rangle, \tag{1.51}$$

where the z_i are at a physical separation from x and y and from each other. If $x\to y$, the products of two fields at coincident points behaves like a different composite local field. The renormalization of O_a and O_b is not sufficient to guarantee that the correlation function stays finite in the $x\to y$ limit. It is thus possible that *short-distance singularities* arise.

According to Wilson's *operator product expansion* (OPE), the short-distance singularities are described by local operators O_i having the same global symmetries of the product O_aO_b

$$O_b(x)O_b(y)\sim\sum_i C_{ab}^i(x-y)O_i(y),\qquad\text{for } x\to y, \tag{1.52}$$

where $C_{ab}^i(x-y)$ are c-number functions. This expansion has been shown to hold in perturbation theory [49] and it is thought to hold also at the non-perturbative level. Being an *operator* relation, it holds with the same C_{ab}^i when applied to any correlation function.

Dimensional analysis suggests that at short distances

$$C_{ab}^i(x)\sim|x|^{d_i-d_a-d_b},\qquad\text{for } x\to 0, \tag{1.53}$$

where $d_i=[O_i]$ and the coefficient of proportionality is the so-called *Wilson's coefficient*. The most singular behaviour is given by the lowest-dimension O_i field. This is usually a rather simple field, since d_i increases with O_i of increasing complexity.

This simple power-counting argument is modified by renormalization effects. When formulated in terms of fields renormalized at a scale μ, the C_{ab}^i are renormalization-scale dependent and obey the RG equation. If the fields O_a, O_b and O_i are multiplicatively renormalizable, this is

$$\left[\mu\frac{\partial}{\partial\mu}+\beta\frac{\partial}{\partial g}-\gamma_a-\gamma_b+\gamma_i\right]C_{ab}^i(\mu,x)=0, \tag{1.54}$$

This equation has a particularly simple solution in asymptotically free theories like QCD, in which the coupling runs to a trivial fixed point $g(\mu)\to 0$ for $\mu\to\infty$. Remarkably, the leading short-distance behaviour of $C_{ab}^i(x)$ is given by the naïve scaling in Eq. (1.53) up to a power of $\ln|x|$ that is computable in perturbation theory.

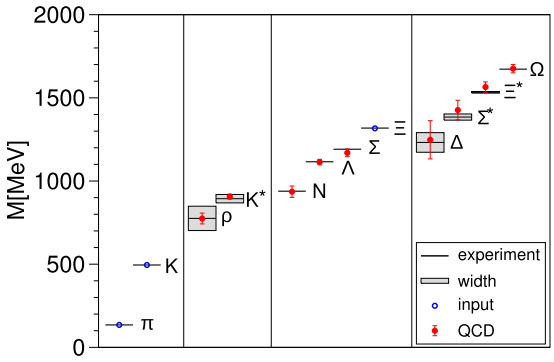


Figure 1.1.: The light hadron spectrum of QCD. Horizontal lines and bands are experimental values with their decay widths. Solid circles are the results of the lattice computation of Ref. [50], with vertical error bars representing combined statistical and systematic errors. π, K and Ξ are used to set the light quark mass, strange quark mass and the overall scale, respectively.

1.3. A couple of non-perturbative results

In this section we present, leaving out many technical details, two results concerning YM theory and QCD obtained using Monte Carlo simulations of the lattice-regularized QFT. They show how lattice techniques allow to study the phenomenology of these theories and obtain quantitative results in a non-perturbative regime, starting from first principles and with minimal experimental input. Lattice regularization in general and some of the tools used to obtain these results are introduced in Chapter 2. A third result, concerning chiral symmetry, is presented in Section 1.4.4.

1.3.1. The spectrum of QCD

In this section we show an impressive result obtained using lattice simulations: an *ab initio* computation of the light hadron spectrum of QCD [50]. It was obtained with Monte Carlo simulations of $N_f = 2 + 1$ QCD. This means that the dynamical quarks present in the action are an isospin doublet of degenerate u and d quarks and a heavier s quark. A number of lattices were simulated, differing for physical volume, lattice spacing, and quark masses. This allowed to study the systematics coming from finite volume, discretization effects and non-physical quark masses, in order to extrapolate to the thermodynamic limit, continuum limit and physical quark masses respectively. The results shown in Figure 1.1 are thus exact predictions, within errors, of QCD.

From the figure it is evident that there are no massless states, thus QCD with physical quark masses has a mass gap. Moreover, most of the states concentrate in the energy region at $\approx 1\,\text{GeV}$ or above. The exceptions are the π and K pseudoscalar mesons, which are lighter. This is understood as a consequence of approximate chiral symmetry and will be the subject of Section 1.4.

At this level of precision, isospin breaking effects start to be relevant. These comes from two competing sources: the different masses and the different electromagnetic interactions of the u

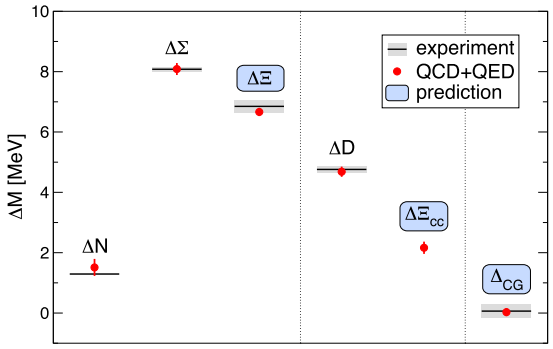


Figure 1.2.: The isospin mass splitting of stable hadrons in QCD+QED. $\Delta_{\mathrm{CG}} = \Delta M_N - \Delta M_\Sigma + \Delta M_\Xi$ is Coleman-Glashow difference [51]. Horizontal lines and bands are experimental values with their error [52]. Solid circles are the results of the lattice computation of Ref. [53]. The blue shaded regions around labels denote predictions of QCD+QED, i.e. quantities that have either not been measured in experiment or are measured with less precision.

and *d* quarks. They are of the same order, thus it makes sense to study them together. In Ref. [53] the mass splitting of the baryon octet masses was computed *ab initio* using $N_{\mathrm{f}} = 1 + 1 + 1 + 1$ QCD+QED lattice simulations. These are simulations with dynamical *u*, *d*, *s* and *c* quarks, all with different masses, and including dynamical photons. The addition of QED is non-trivial since, being a gapless theory, the photon propagator decays only power-like with distance and induces finite volume effects that are suppressed accordingly. Moreover, non-physical values of the QED coupling parameter are often used to exagerate the electromagnetic splitting. The results extrapolated (or interpolated) to the physical value of all parameters are reported in Figure 1.2.

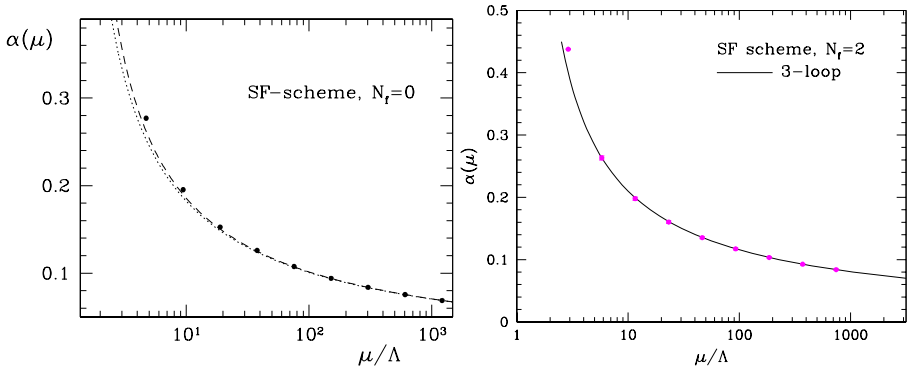
1.3.2. The running coupling in YM theory and $N_{\mathrm{f}} = 2$ QCD.

Lattice simulations are not limited to study the low-energy phenomenology. In this section, we present a first-principles computation of the running of the gauge coupling from the high-energy perturbative regime to low-energy hadronic regime.

Naïvely, it is not possible to accommodate both perturbative and hadronic scales in a single lattice simulation, keeping at the same time finite-volume and cut-off effects under control. A possible solution consists in using a finite-volume renormalization scheme and finite-size scaling techniques as proposed in [56]. A non-perturbative definition of the coupling is needed and one is provided by the *Schrödinger functional* (SF) [57]. In this way, the running of the coupling and other parameters are computed from low to very high energies without relying on perturbation theory.

This has several applications. It allows a precise determination of the Λ-parameter. Moreover, lattice simulations provide the most precise determination of light-quark masses.[8] However, these are scheme-dependent quantities and high-energy perturbative computations are usually

[8] While the *ratios* of light-quark masses are well-described by χPT.



(a) $N_{\rm f} = 0$ data from Ref. [54].

(b) $N_{\rm f} = 2$ data from Ref. [55].

Figure 1.3.: Comparison of the numerically computed values of the running coupling in the SF scheme with perturbation theory. The dotted and dashed curves are obtained from the integration of the 2- and 3-loop expressions of the β-function. The errors on the data are smaller than the symbol size.

interested in the renormalized quark masses in the $\overline{\rm MS}$ scheme at some fixed μ. Using finite-size scaling techniques, the mass computed in the SF-scheme can be matched at low energies with the hadronic non-perturbative scheme and at high energies with the $\overline{\rm MS}$ scheme.

The program was successfully implemented in YM theory in Ref. [54]. The non-perturbative running of the SF coupling

$$\alpha(\mu) \equiv \frac{g^2_{\rm SF}(L)}{4\pi}, \qquad \mu = \frac{1}{L}, \tag{1.55}$$

is obtained from simulations spanning more than two orders of magnitude, as shown in Figure 1.3a,. At the highest scale, perturbation theory is applied and the Λ-parameter is determined. Then, one-loop perturbation theory is sufficient to go from $\Lambda_{\rm SF}$ to $\Lambda_{\overline{\rm MS}}$. At the lowest-scale, a hadronic reference scale such r_0 is computed. This allows to express any hadronic scale in terms of the Λ-parameter

$$r_0 \Lambda^{(0)}_{\overline{\rm MS}} = 0.602(48), \tag{1.56a}$$

$$\sqrt{t_0}\Lambda^{(0)}_{\overline{\rm MS}} = 0.200(16), \tag{1.56b}$$

where we used $\sqrt{8t_0}/r_0 = 0.941(7)$ from Ref. [1], reported also in Chapter 5.

Since YM theory is not physical, any conversion in physical units is somewhat arbitrary. We can supplement the theory with valence quarks and use the value $r_0 = 0.5\,\text{fm}$, without specifying the quenching systematics, to obtain

$$\Lambda^{(0)}_{\overline{\rm MS}} = 238(19)\,\text{MeV}. \tag{1.57}$$

In the same spirit, the running of the quark mass was also computed in Ref. [54].

This finite-size scaling technique applies also to QCD with dynamical quarks. With the choice of a mass-independent scheme, the running is computed in the chiral limit. Thanks to the properties of the Schrödinger functional, it is possible to perform simulations directly at $m = m_{\text{critical}}$. The program was implemented in the $N_{\text{f}} = 2$ case in Ref. [55, 58]. The non-perturbative running of $\alpha(\mu)$ defined in Eq. (1.55) is plotted in Figure 1.3a. The final results are

$$r_0 \Lambda^{(2)}_{\overline{\text{MS}}} = 0.789(52), \tag{1.58a}$$

$$F_K^{-1} \Lambda^{(2)}_{\overline{\text{MS}}} = 2.83(19). \tag{1.58b}$$

The experimentally-measured Kaon decay constant $F_{K^-}^{\text{exp}} \approx 110.1\,\text{MeV}$ [17] (see footnote 13) is the best choice to pass to physical units [9]

$$\Lambda^{(2)}_{\overline{\text{MS}}} = 310(20)\,\text{MeV}. \tag{1.59}$$

1.4. Chiral symmetry

A Dirac fermion field is an object that transforms under a *reducible* representation of SO(4), the group of spacetime rotations. Since at the infinitesimal level $\mathfrak{so}(4) \cong \mathfrak{su}(2) \oplus \mathfrak{su}(2)$, representations are classified by two half-integers. A Dirac spinor belongs to the direct sum $(½, 0) \oplus (0, ½)$, which has dimension four. The two spin-½ representations have opposite *chirality* and are transformed one into each other by a parity transformation. A two-dimensional spinor that transforms in either $(½, 0)$ or $(0, ½)$ is called a *Weyl spinor*.

A Dirac spinor field is decomposed in two Weyl spinor fields with a projection on left- and right-handed chiral components

$$\psi_L = P_- \psi, \quad \bar{\psi}_L = \bar{\psi} P_+, \qquad \psi_R = P_+ \psi, \quad \bar{\psi}_R = \bar{\psi} P_-, \tag{1.60}$$

defined using the projectors

$$P_\pm \equiv \frac{1 \pm \gamma_5}{2}, \qquad P_\pm^2 = P_\pm, \quad P_+ + P_- = \mathbb{1}, \quad P_+ P_- = P_- P_+ = 0. \tag{1.61}$$

Consider the fermion action in Eq. (1.13) with $N_{\text{f}} \geq 2$. In terms of chiral fields, it decomposes

$$S_{\text{f}}[\psi_L, \psi_R, \bar{\psi}_L, \bar{\psi}_R, A] = \int \mathrm{d}^4x \left\{ \bar{\psi}_L(x) \gamma_\mu D_\mu \psi_L(x) + \bar{\psi}_R(x) \gamma_\mu D_\mu \psi_R(x) + \bar{\psi}_L(x) \mathcal{M}^\dagger \psi_R(x) + \bar{\psi}_R(x) \mathcal{M} \psi_L(x) \right\}. \tag{1.62}$$

Here, $\mathcal{M}$ is a matrix in flavour space which is linked to mass matrix M of Dirac fermions by

$$M + \gamma_5 M_5 = P_+ \mathcal{M}^\dagger + P_- \mathcal{M} = \frac{\mathcal{M}^\dagger + \mathcal{M}}{2} + \gamma_5 \frac{\mathcal{M}^\dagger - \mathcal{M}}{2}, \tag{1.63}$$

[9] The small systematics from isospin breaking effects is ignored.

and $M = \mathcal{M}$ if the latter is Hermitian. It is possible to rotate independently the chiral components with two sets of $\mathrm{U}(N_\mathrm{f})$ unitary rotations. We assign them transformation properties under representations of $\mathrm{U}(N_\mathrm{f})_L \times \mathrm{U}(N_\mathrm{f})_R$

$$\psi_L \in (\boldsymbol{N}_\mathrm{f}, \mathbb{1}), \quad \bar{\psi}_L \in (\bar{\boldsymbol{N}}_\mathrm{f}, \mathbb{1}), \qquad \psi_R \in (\mathbb{1}, \boldsymbol{N}_\mathrm{f}), \quad \bar{\psi}_R \in (\mathbb{1}, \bar{\boldsymbol{N}}_\mathrm{f}). \tag{1.64}$$

They imply the *chiral transformation* rules

$$\psi_L \to \psi'_L = U_L \psi_L, \qquad \psi_R \to \psi_R = U_R \psi_R, \tag{1.65a}$$

$$\bar{\psi}_L \to \bar{\psi}'_L = \bar{\psi}_L U_L^\dagger, \qquad \bar{\psi}_R \to \bar{\psi}_R = \bar{\psi}_R U_R^\dagger, \tag{1.65b}$$

with

$$U_L = \mathrm{e}^{\mathrm{i}\epsilon_L^a \tau^a} \in \mathrm{U}(N_\mathrm{f})_L, \qquad U_R = \mathrm{e}^{\mathrm{i}\epsilon_R^a \tau^a} \in \mathrm{U}(N_\mathrm{f})_R, \tag{1.66}$$

where $\epsilon_{R/L}^a$ for $a = 0, \dots, N_\mathrm{f}^2 - 1$ are real coefficients. Here, τ^a for $a = 1, \dots, N_\mathrm{f}^2 - 1$ are the traceless hermitian generator of the *special* unitary Lie group $\mathrm{SU}(N_\mathrm{f})$ supplemented with $\tau^0 = (2N_\mathrm{f})^{1/2}\mathbb{1}$ as the generator of U(1) phase rotations. We use the same conventions given in Section A.1 for T^a.

In the limit in which N_f fermions are massless, $\mathcal{M} = 0$ and the action in Eq. (1.62) is invariant under this transformation. This is unsurprising: the gauge field theory of multiple flavours of massless Dirac fermions is well-known to have large group of global symmetries acting on flavour indices, collectively known as *chiral symmetry*. At the classical level, the group is

$$\mathrm{U}(N_\mathrm{f})_L \times \mathrm{U}(N_\mathrm{f})_R = \mathrm{U}(1)_B \times \mathrm{U}(1)_A \times \mathrm{SU}(N_\mathrm{f})_L \times \mathrm{SU}(N_\mathrm{f})_R. \tag{1.67}$$

The action in Eq. (1.62) is formally invariant for generic $\mathcal{M}$ if we assume $\mathcal{M}$ be a *spurion*, i.e. a constant external field transforming under the $(\bar{\boldsymbol{N}}_\mathrm{f}, \boldsymbol{N}_\mathrm{f})$ representation of $\mathrm{U}(N_\mathrm{f})_L \times \mathrm{U}(N_\mathrm{f})_R$

$$\mathcal{M} \to \mathcal{M}' = U_R \mathcal{M} U_L^\dagger. \tag{1.68}$$

Chiral symmetry is broken as soon as $\mathcal{M}$ is fixed at a non-zero value. However, different choices of $\mathcal{M}$ are still related by symmetry. In particular, it is always possible to choose U_L and U_R in a way to realize a singular value decomposition of $\mathcal{M}$, so that

$$\mathcal{M}' = \mathrm{diag}\big(m_1, \dots, m_{N_\mathrm{f}}\big), \qquad m_i \geq 0. \tag{1.69}$$

Therefore, at the classical level, any choice of $\mathcal{M}$ is equivalent to a diagonal mass matrix with real positive entries.

The extension of these considerations from the classical to the quantum theory requires that these symmetries survive quantization. This holds with the only exception of the $\mathrm{U}(1)_A$ transformation. This classical symmetry develops an anomaly and it is not a symmetry of the theory at the quantum level. As a consequence, the global phase of a non-Hermitian $\mathcal{M}$ cannot be eliminated without consequences. We will study in details the anomalous $\mathrm{U}(1)_A$ symmetry in Chapter 3.

In the rest of this chapter, we focus on the non-singlet sector of chiral symmetry, which is free from anomalies.

1.4.1. Ward–Takahashi identities

Ward–Takahashi identities (WTIs) [59, 60] are relations between correlation functions that derive directly from continuous symmetries of the theory. They are the quantum equivalent of Noether's current conservation laws in classical mechanics. They state that under an infinitesimal variation of the fields

$$\psi \to \psi' = \psi + \delta\psi\,, \quad \bar\psi \to \bar\psi' = \bar\psi + \delta\bar\psi\,, \qquad A \to A' = A + \delta A\,, \tag{1.70}$$

it holds

$$\left\langle O\big\{(-)\mathrm{Tr}\,\delta\mathcal{J} - \delta S\big\}\right\rangle + \left\langle \delta O\right\rangle = 0, \tag{1.71}$$

where δO and δS are the variations of a composite field and of the action respectively, and $\delta\mathcal{J}$ is the variation in the Jacobian $\mathcal{J}$ of the transformation[10] in Eq. (1.70).

In this chapter, we consider transformations with $\mathrm{Det}\,\mathcal{J} = 1$. This excludes *anomalous* transformations such as the flavour-singlet $\mathrm{U}(1)_A$ rotation of fermion fields. Anomalous WTIs will be the topic of Chapter 3. Assuming $\mathrm{Det}\,\mathcal{J} = 1$, Eq. (1.71) simplifies to

$$\langle \delta O\rangle = \langle O\,\delta S\rangle. \tag{1.72}$$

The simplest case is when O is left invariant by a particular choice of the transformation, which implies

$$\langle O\,\delta S\rangle, \qquad \text{if } \delta O = 0. \tag{1.73}$$

Consequences of chiral symmetry in the quantum theory can be studied in terms of chiral WTIs. To obtain them, it is useful to consider the *local* equivalent of chiral transformations in Eqs (1.65), modulated by smooth functions of spacetime which vanish at infinity. Their infinitesimal variations on left- and right-handed components are

$$\delta\psi_L(x) = +\mathrm{i}\tau^a\epsilon^a_L(x)\psi_L(x), \qquad \delta\psi_R(x) = +\mathrm{i}\tau^a\epsilon^a_R(x)\psi_R(x), \tag{1.74a}$$

$$\delta\bar\psi_L(x) = -\mathrm{i}\bar\psi_L(x)\epsilon^a_L(x)\tau^a, \qquad \delta\bar\psi_R(x) = -\mathrm{i}\bar\psi_R(x)\epsilon^a_R(x)\tau^a. \tag{1.74b}$$

These transformations are ofter rewritten in terms of vector and axial-vector transformations, parametrized respectively by

$$\alpha^a(x) = \frac{1}{2}\big[\epsilon^a_R(x) + \epsilon^a_L(x)\big], \qquad \beta^a(x) = \frac{1}{2}\big[\epsilon^a_R(x) - \epsilon^a_L(x)\big], \tag{1.75}$$

which result in infinitesimal variations of Dirac spinor fields

$$\delta\psi(x) = +\mathrm{i}\tau^a\alpha^a(x)\psi(x) + \mathrm{i}\tau^a\beta^a(x)\gamma_5\psi(x), \tag{1.76a}$$

$$\delta\bar\psi(x) = -\mathrm{i}\bar\psi(x)\alpha^a(x)\tau^a + \mathrm{i}\bar\psi(x)\gamma_5\beta^a(x)\tau^a. \tag{1.76b}$$

In the limit of constant $\alpha^a(x)$ and $\beta^a(x)$ these transformations are global symmetries of the massless action: $\delta S\big|_{M=0} = 0$. The variation of the action under local transformations,

$$\delta S\bigg|_{M=0} = \mathrm{i}\int \mathrm{d}^4x\,\big\{\partial_\mu\alpha^a(x)\big\}V^a_\mu(x) + \mathrm{i}\int \mathrm{d}^4x\,\big\{\partial_\mu\beta^a(x)\big\}A^a_\mu(x), \tag{1.77}$$

[10] For a change of variables involving Grassmann-number-valued fields, the *inverse* Jacobian appears and induces a minus sign in Eq. (1.71).

is proportional to the divergence of $\alpha^a(x)$ and $\beta^a(x)$ times, respectively, the vector and axial current

$$V^a_\mu(x) = \bar\psi(x)\tau^a\gamma_\mu\psi(x), \qquad A^a_\mu(x) = \bar\psi(x)\tau^a\gamma_\mu\gamma_5\psi^j(x). \tag{1.78}$$

We include in this notation the $a = 0$ case of flavour-singlet transformations, choosing as normalization $\tau^0 = (2N_\mathrm{f})^{-1/2}\mathbb{1}$.

Assuming that $\alpha^a(x)$ and $\beta^a(x)$ go smoothly to zero outside some bounded domain $\mathcal{B}$, we can integrate by part to obtain

$$\delta S = -\mathrm{i}\int \mathrm{d}^4x\,\alpha^a(x)\big\{\partial_\mu V^a_\mu(x)\big\} - \mathrm{i}\int \mathrm{d}^4x\,\beta^a(x)\big\{\partial_\mu A^a_\mu(x)\big\}. \tag{1.79}$$

Thus, $V^a_\mu(x)$ and $A^a_\mu(x)$ are Noether's currents corresponding to the $\mathrm{U}(N_\mathrm{f})_L\times\mathrm{U}(N_\mathrm{f})_R$ symmetry of the massless action. By Noether's theorem, they are conserved at the classical level in the $M = 0$ theory. The corresponding conserved charges are the zero-momentum projections

$$Q^a_V = \bar V^a_4(x_4) = \int \mathrm{d}^3x\,V^a_4(x), \qquad Q^a_A = \bar A^a_4(x_4) = \int \mathrm{d}^3x\,A^a_4(x). \tag{1.80}$$

More generally, adding to the action a chiral-symmetry-breaking mass term, the action variation is modified into

$$\begin{aligned}\delta S = \mathrm{i}\int \mathrm{d}^4x\,\alpha^a(x)\big\{-\partial_\mu V^a_\mu(x) + \bar\psi(x)\big[M,\tau^a\big]\psi(x)\big\}\\ + \mathrm{i}\int \mathrm{d}^4x\,\beta^a(x)\big\{-\partial_\mu A^a_\mu(x) + \bar\psi(x)\big\{M,\tau^a\gamma_5\big\}\psi(x)\big\}.\end{aligned} \tag{1.81}$$

WTIs follow immediately. For instance, consider the flavour degenerate case $M = m\mathbb{1}$. The first line simply states the conservation of the vector current

$$\big\langle \partial_\mu V^a_\mu(x) O_\mathrm{ext}\big\rangle = 0, \tag{1.82}$$

where O_ext is any composite field localized outside the support $\mathcal{B}$ of $\alpha^a(x)$, so that $\delta O_\mathrm{ext} = 0$. From the second line, we obtain the partially conserved axial current (PCAC) relation

$$\big\langle \partial_\mu A^a_\mu(x) O_\mathrm{ext}\big\rangle = 2m\big\langle P^a(x) O_\mathrm{ext}\big\rangle, \qquad \text{for } a \neq 0, \tag{1.83}$$

where $P^a(x)$ is the pseudoscalar density and we introduce also the scalar density $S^a(x)$

$$S^a(x) = \bar\psi(x)\tau^a\psi(x), \qquad P(x) = \bar\psi(x)\tau^a\gamma_5\psi(x). \tag{1.84}$$

Other interesting WTIs arise if we consider a field $O(y)$ localized inside the support $\mathcal{B}$ of $\alpha^a(x)$ and $\beta^a(x)$, in such a way that

$$\delta O(y) = \mathrm{i}\alpha^a(y)\delta^a_V O(y) + \mathrm{i}\beta^a(y)\delta^a_A O(y). \tag{1.85}$$

From this, for a constant α^a and β^a we obtain the *integrated* WTIs

$$\int \mathrm{d}^4x\,\big\langle \partial_\mu V^a_\mu(x) O(y) O_\mathrm{ext}\big\rangle = -\big\langle \delta^a_V O(y) O_\mathrm{ext}\big\rangle, \tag{1.86a}$$

$$\int \mathrm{d}^4x\,\big\langle \partial_\mu A^a_\mu(x) O(y) O_\mathrm{ext}\big\rangle = 2m\int \mathrm{d}^4x\,\big\langle P^a(x) O(y) O_\mathrm{ext}\big\rangle - \big\langle \delta^a_A O(y) O_\mathrm{ext}\big\rangle, \qquad \text{for } a \neq 0. \tag{1.86b}$$

We can transform the l.h.s. in a boundary integral to obtain, in the chiral limit,

$$\int_{\partial\mathcal{B}} \mathrm{d}\sigma_\mu(x) \left\langle V^a_\mu(x) O(y) O_{\mathrm{ext}} \right\rangle = -\left\langle \delta^a_V O(y) O_{\mathrm{ext}} \right\rangle, \tag{1.87a}$$

$$\int_{\partial\mathcal{B}} \mathrm{d}\sigma_\mu(x) \left\langle A^a_\mu(x) O(y) O_{\mathrm{ext}} \right\rangle = -\left\langle \delta^a_A O(y) O_{\mathrm{ext}} \right\rangle, \qquad \text{for } a \neq 0. \tag{1.87b}$$

This class of WTIs corresponds to the current algebra identities of operators in Minkowski spacetime: the boundary integral gives the commutator of the vector or axial charge with $O(y)$.

1.4.2. Spontaneous symmetry breaking and Goldstone's theorem

A symmetry is *spontaneously broken* if its Noether's currents are conserved but the vacuum is not invariant under the action of the corresponding charges. Symbolically,

$$Q|0\rangle \neq 0. \tag{1.88}$$

The WTIs obtained in the previous section allow us to prove a well-known result of QFT:

Theorem (Goldstone)**.** *Consider a generic continuous symmetry that is spontaneously broken. Then, necessarily, massless scalar particles known as* Nambu–Goldstone (NG) bosons *appear in the spectrum. There is one NG boson for each generator of the symmetry that is broken, i.e. that does not preserve the ground state. The NG mode represent a long-range fluctuation of the corresponding order parameter.*

It is well understood that, in the limit of massless quarks, the $\mathrm{SU}(N_\mathrm{f})_L \times \mathrm{SU}(N_\mathrm{f})_R$ chiral symmetry is spontaneously broken. The order parameter is a quantity of non-perturbative origin: the *chiral condensate*

$$\Sigma \equiv -\frac{2}{\sqrt{2N_\mathrm{f}}} \lim_{m\to 0} \left\langle S^0 \right\rangle = -\frac{1}{N_\mathrm{f}} \lim_{m\to 0} \sum_\mathrm{f} \langle \bar\psi\psi \rangle > 0. \tag{1.89}$$

It transforms under the $(\bar{\boldsymbol{N}}_\mathrm{f}, \boldsymbol{N}_\mathrm{f}) + (\boldsymbol{N}_\mathrm{f}, \bar{\boldsymbol{N}}_\mathrm{f})$ representation of chiral symmetry, just like a flavour-degenerate mass term. Thus, even if the massless action is invariant under the full chiral symmetry group, the vacuum is not. However, the chiral condensate remains invariant under the $\mathrm{SU}(N_\mathrm{f})_F$ vectorial subgroup of flavour symmetry. Therefore, the *chiral symmetry breaking* (χSB) pattern is

$$\underbrace{\mathrm{SU}(N_\mathrm{f})_L \times \mathrm{SU}(N_\mathrm{f})_R}_{\mathcal{G}} \xrightarrow{\Sigma\neq 0} \underbrace{\mathrm{SU}(N_\mathrm{f})_F}_{\mathcal{H}}. \tag{1.90}$$

The $N_\mathrm{f}^2 - 1$ broken generators are in the coset space $\mathcal{G}/\mathcal{H}$. They are odd under parity symmetry and fit into the adjoint representation of the unbroken $\mathrm{SU}(N_\mathrm{f})_F$ flavour group. Thus, according to Goldstone's theorem we expect in the spectrum $N_\mathrm{f}^2 - 1$ NG bosons with $J^P = 0^-$ quantum numbers. Given these properties, they are easily identified as $N_\mathrm{f}^2 - 1$ massless pseudoscalar mesons.

We can prove the existence of these states in the Euclidean spacetime functional formalism. We start from the PCAC relation Eq. (1.83) in the chiral limit and with $O_{\text{ext}} = P^b(y)$

$$\left\langle \partial_\mu A^a_\mu(x) P^b(y) \right\rangle = 0, \qquad \text{for } y \neq x, \tag{1.91}$$

from which, using Lorentz invariance, we have

$$\left\langle A^a_\mu(x) P^b(y) \right\rangle = k^{ab} \frac{x_\mu - y_\mu}{|x-y|^4}. \tag{1.92}$$

On the other hand, let us specialize the integrated WTI in Eqs (1.86) to a non-singlet pseudoscalar current $O(y) = P^b(y)$ and to $O_{\text{ext}} = \mathbb{1}$. Since

$$\delta^a_V P^b(y) = \bar\psi(y)\gamma_5 \left[\tau^b, \tau^a\right]\psi(y) = -\mathrm{i} f^{abc} P^c(y), \tag{1.93a}$$

$$\delta^a_A P^b(y) = \bar\psi(y)\left\{\tau^b, \tau^a\right\}\psi(y) = \frac{2\delta^{ab}}{\sqrt{2N_{\text{f}}}} S^0(y) + d^{abc} S^c(y), \tag{1.93b}$$

from Eq. (1.86b) we obtain[11]

$$\int \mathrm{d}^4x \left\langle \partial_\mu A^a_\mu(x) P^b(y) \right\rangle = 2m \int \mathrm{d}^4x \left\langle P^a(x) P^b(y) \right\rangle - \frac{2\delta^{ab}}{\sqrt{2N_{\text{f}}}} \left\langle S^0(y) \right\rangle. \tag{1.94}$$

We can transform it in a integral over the surface of a ball $\mathcal{B} = \{x : |x| \leq r\}$, centred around $y = 0$. In the chiral limit[12]

$$\int_{|x|=r} \mathrm{d}\sigma_\mu(x) \left\langle A^a_\mu(x) P^b(0) \right\rangle = -\frac{2\delta^{ab}}{\sqrt{2N_{\text{f}}}} \left\langle S^0(0) \right\rangle = \delta^{ab} \Sigma. \tag{1.95}$$

We can now insert Eq. (1.92) to fix k^{ab}

$$k^{ab} \int_{|x|=r} \mathrm{d}\sigma_\mu(x) \frac{x_\mu}{|x|^4} = k^{ab} \Omega^4 = \delta^{ab} \Sigma. \tag{1.96}$$

Remembering that $\Omega^D = 2\pi^{D/2}/\Gamma(D/2)$, we arrive to

$$\left\langle A^a_\mu(x) P^b(0) \right\rangle = \frac{\delta^{ab}}{2\pi^2} \frac{x_\mu}{|x|^4} \Sigma. \tag{1.97}$$

When chiral symmetry is spontaneously broken and $\Sigma \neq 0$, this two-point function shows long-range correlations. The power-like suppression means that the energy spectrum of the theory does not have a mass gap, and massless particles with the quantum numbers of P^a are present. These are the NG bosons of χSB, denoted by $|\pi^a\rangle$, $a = 1, \ldots, N_{\text{f}} - 1$. We arrive at the same conclusion considering the two-point function projected at zero momentum

$$\left\langle \bar A^a_4(x_4) P^b(0) \right\rangle = \int \mathrm{d}^3x \left\langle A^a_4(x) P^b(0) \right\rangle = \frac{\delta^{ab}}{2} \Sigma. \tag{1.98}$$

It does not show the typical exponential suppression at asymptotic time separations given by the mass of a massive particle state.

[11] $\langle S^a \rangle = 0$ for $a \neq 0$, since $\mathrm{SU}(N_{\text{f}})_F$ flavour symmetry is unbroken.

[12] The scalar density S^0 transform under the $(\bar{\boldsymbol{N}}_{\text{f}}, \boldsymbol{N}_{\text{f}}) + (\boldsymbol{N}_{\text{f}}, \bar{\boldsymbol{N}}_{\text{f}})$ representation of chiral symmetry, therefore under renormalization it does not mix with the identity but it can mix with $\mathcal{M}^\dagger + \mathcal{M}$. For this reason the renormalized S^0 is well defined in the chiral limit only.

1.4.3. The GMOR relation

Away from the chiral limit, the $|\pi^a\rangle$ states are no more massless NG bosons. However, for a 'small' breaking, their masses and interactions are still constrained by chiral symmetry. They behave as *pseudo Nambu–Goldstone* (pNG) bosons.

It is possible to derive a relation between the pNG mass M_π and the symmetry breaking parameters. Consider a complete set of properly-normalized one-π states

$$\mathbb{1}_{1\pi} = \int \frac{\mathrm{d}^3 p}{(2\pi)^3 2\mathrm{i}p_4} \big|\pi^a(\vec{p})\rangle\langle\pi^a(\vec{p})\big|, \qquad p_4 = \mathrm{i}E_\pi(\vec{p}) = \mathrm{i}\sqrt{M_\pi^2 + \vec{p}^2}. \tag{1.99}$$

Conventionally, a one-π state with momentum p_μ is created by the pseudoscalar density

$$\big\langle 0\big|P^a(x)\big|\pi^b(\vec{p})\big\rangle \equiv \delta^{ab} G_\pi \mathrm{e}^{\mathrm{i}px}, \tag{1.100a}$$

while Lorentz symmetry dictates its coupling to the axial current[13]

$$\big\langle 0\big|A^a_\mu(x)\big|\pi^b(\vec{p})\big\rangle \equiv \mathrm{i}p_\mu \delta^{ab} F_\pi \mathrm{e}^{\mathrm{i}px}. \tag{1.100b}$$

For any $m \neq 0$, G_π and F_π are related by the PCAC relation in Eq. (1.83)

$$\partial_4 \big\langle \bar{A}^a_4(x_4) P^b(0)\big\rangle = 2m \big\langle \bar{P}^a(x_4) P^b(0)\big\rangle, \qquad \text{for } x_4 \neq 0. \tag{1.101}$$

Close enough to the chiral limit, the long-range behaviour is dominated by the light pNG states. Inserting the identity Eq. (1.99) we get

$$\partial_4 \left[\frac{\delta^{ab}}{2} G_\pi F_\pi \mathrm{e}^{-E_\pi(0)x_4}\right] = 2m \left[\frac{\delta^{ab}}{-2E_\pi(0)} G_\pi^2 \mathrm{e}^{-E_\pi(0)x_4}\right], \tag{1.102}$$

where $E_\pi(0) = M_\pi$ since chiral symmetry is explicitly broken. Thus, we arrive to the exact relation

$$M_\pi^2 F_\pi = 2mG_\pi, \tag{1.103}$$

form which we extract G_π in the chiral limit

$$\lim_{m\to 0} G_\pi = \lim_{m\to 0} \frac{M_\pi^2 F_\pi}{2m}. \tag{1.104}$$

Similarly, inserting the identity Eq. (1.99) in Eq. (1.98), the NG bosons dominate the long-range behaviour

$$\begin{aligned}\big\langle \bar{A}^a_4(x_4) P^b(0)\big\rangle &= \int \frac{\mathrm{d}^3 p}{(2\pi)^3 2\mathrm{i}p_4} \big\langle 0\big|\bar{A}^a_4(x_4)\big|\pi^c(\vec{p})\big\rangle \big\langle \pi^c(\vec{p})\big|P^b(0)\big|0\big\rangle \\ &= \frac{\delta^{ab}}{2} G_\pi F_\pi \mathrm{e}^{-E_\pi(0)x_4} \xrightarrow{|x_4|\to\infty} \frac{\delta^{ab}}{2} G_\pi F_\pi, \end{aligned} \tag{1.105}$$

[13] There are different conventions for F_π (and G_π), differing by $\sqrt{2}$ factors. We denote with f_π the decay constant that correspond to the experimental values $f^{\mathrm{exp}}_{\pi^-} = 130.50(13)\,\mathrm{MeV}$ and $f^{\mathrm{exp}}_{K^-} = 155.7(5)\,\mathrm{MeV}$ [17], while we denote with capital $F_\pi \equiv f_\pi/\sqrt{2}$.

which thus gives

$$G_\pi F_\pi = \Sigma, \qquad \text{for } m = 0. \tag{1.106}$$

Combining Eq. (1.104) with Eq. (1.106), we obtain the *Gell-Mann–Oakes–Renner (GMOR) relation* [61]

$$\lim_{m\to 0} \frac{M_\pi^2 F_\pi^2}{2m} = \Sigma, \tag{1.107}$$

which implies, for a small quark mass,

$$M_\pi^2 = 2m\frac{\Sigma}{F_0^2} + \mathcal{O}\big(m^2\big), \qquad F_0 \equiv \lim_{m\to 0} F_\pi, \tag{1.108}$$

where F_0 and Σ are computed in the chiral limit. Therefore, when chiral symmetry is broken explicitly by a quark mass term, the π states are massive *pseudo Nambu–Goldstone* (pNG) bosons. They have a square mass M_π^2 that is *linear* in the χSB parameter m, times a *low-energy constant* (LEC) proportional to the chiral condensate Σ, plus $\mathcal{O}(m^2)$ corrections.

The $N_\mathrm{f}^2 - 1$ lightest pseudoscalar mesons of QCD with N_f mass-degenerate flavours have thus peculiar properties with respect to all other hadrons. Generically, the mass of hadrons made of light quarks is proportional to the Λ-parameter, with a mild dependence on the quark mass

$$M_\mathrm{had.} \sim \Lambda f\Big(\frac{m}{\Lambda}\Big). \tag{1.109}$$

The $N_\mathrm{f}^2 - 1$ lightest pseudoscalar mesons are pNG bosons and their mass depends strongly on the quark mass according to Eq. (1.108)

$$M_\pi \sim \sqrt{m\Lambda}. \tag{1.110}$$

They become exactly massless NG bosons in the chiral limit.

1.4.4. Chiral symmetry breaking in $N_\mathrm{f} = 2$ QCD

We described how approximate chiral symmetry predicts the lightness of pseudoscalar mesons made only of light quarks, which is observed in nature. However, with a single value for the physical quark masses we do not actually observe the proportionality $M_\pi^2 \sim m$ predicted by the GMOR relation in Eq. (1.108). In this situation, Monte Carlo simulations of the lattice-discretized theory represent an ideal laboratory: changing the parameters of the theory, we can compute the meson masses for arbitrary values of the quark masses.

The GMOR relation has been verified to hold in lattice simulations many times. However, to perform a quantitative test an independent measurement of the chiral condensate Σ is needed. Its direct computation and extrapolation to the chiral limit is problematic due to the divergent renormalization from the mixing with the identity for any $m \neq 0$. This problem is avoided using the Banks–Casher relation [63]

$$\lim_{\lambda\to 0}\lim_{m\to 0}\lim_{V\to\infty} \rho(\lambda, m) = \frac{\Sigma}{\pi}, \tag{1.111}$$

where $\rho(\lambda, m)$ is the *spectral density* of the eigenvalues $\mathrm{i}\lambda$ of the Euclidean massless Dirac operator D. The order of the limits is crucial: the thermodynamic limit has to be taken before

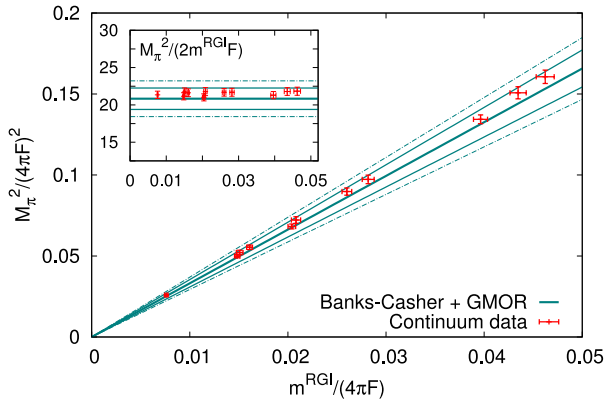


Figure 1.4.: The pion mass squared versus the RGI quark mass from [62], both normalized with $4\pi F$ which is roughly 1 GeV. The red data points are from the direct measurements, extrapolated to the continuum limit. The central line is the GMOR contribution to M_π^2, computed taking the direct measure of the condensate from the spectral density. The upper and lower solid lines show the statistical error and the dotted-dashed ones the total error, the systematic being added in quadrature.

the chiral limit to have spontaneous χSB. Then, the interesting physics is in the spectral density in the neighbourhood of zero modes.

Remarkably, the *mode number*

$$\nu(M,m) \equiv V \int_{-\Lambda}^{+\Lambda} \mathrm{d}\lambda\, \rho(\lambda,m), \qquad \Lambda = \sqrt{M^2 - m^2} \tag{1.112}$$

is a RGI quantity as it stands. Moreover, it is possible to define it also with a regularization which breaks chiral symmetry such as Wilson's discretization of QCD [64, 65]. This allows to compute spectral density on the lattice using the Wilson–Dirac operator. Once all the quantities are properly renormalized, the Banks–Casher relation is expected to hold up to lattice artefacts that vanish in the continuum limit.

In Refs [62, 66] a quantitative study of the Banks–Casher relation resulted in a verification of χSB in $N_\mathrm{f} = 2$ QCD. The authors computed the chiral-limit value of the spectral density at the origin, extracting a value for the chiral condensate, in the $\overline{\mathrm{MS}}$ scheme,

$$\pi\rho^{\overline{\mathrm{MS}}} = \Sigma^{\overline{\mathrm{MS}}} = \big(261(6)(8)\,\mathrm{MeV}\big)^3, \tag{1.113}$$

where the first error is statistical and the second systematic. They compared this number with the prediction of the GMOR relation, extracted form the slope of $M_\pi^2 F_\pi^2/2$ with respect to the quark mass m. To do this, they extracted M_π, F_π and m from the non-singlet pseudoscalar density and axial current. The result is

$$\Sigma^{\overline{\mathrm{MS}}}_{\mathrm{GMOR}} = \big(263(3)(4)\,\mathrm{MeV}\big)^3, \tag{1.114}$$

where again the first error is statistical and the second systematic. The two values are in perfect agreement, as it is also shown in Figure 1.4.

2. Lattice regularization

In this chapter, we introduce the lattice regularization of QCD and techniques related to it. The path integral of a QFT in the continuum is a functional integral that involves integrations over a functional space with an infinite number of degrees of freedom. Mathematically, it is not a well-defined object. Therefore, a first reason to put a field theory on a discrete lattice is to give it a meaning: the lattice path integral is just an ordinary integral with a countable number of degrees of freedom, or even a big but finite number if a finite volume is considered. This leads to mathematically well-defined problems, which at least in principle are solvable and can be approached with computer simulations.

An important feature of the lattice formulation of QFTs is the use of four-dimensional Euclidean spacetime, opposed to the Minkowski spacetime in which physical theories are usually formulated. In the previous Chapter 1, we already defined QCD in Euclidean spacetime. The reason for this choice will become clear in this chapter.

The simplest way to convert a field theory in the continuum to one on a discrete lattice is to take the action and substitute derivatives with finite-difference approximations and spacetime integrals with sums over lattice sites. A pedagogical introduction would require introducing the lattice formalism starting with the theory of a scalar boson field, in which case this simple procedure works without obstructions. For the sake of brevity, in Section 2.1 we introduce directly the lattice theory of massless vector bosons and Dirac fermions, defined in a way to keep exact gauge invariance at finite lattice spacing. Then, in Section 2.2 we deal with the problem of fermion doubling and with the implementation of chiral symmetry on the lattice. In Section 2.3 we use lattice discretization to give a mathematically-sound definition of the path integral. In Section 2.4 we introduce the method of our choice to estimate the path integral: numerical Monte Carlo integration with importance sampling. Finally, Section 2.5 is dedicated to the study of how the continuum theory is recovered for $a \to 0$.

2.1. Gauge theories on the lattice

Formally, a lattice Λ is a subgroup of Euclidean $\mathbb{R}^n$ space which is isomorphic to $\mathbb{Z}^n$. For any basis of $\mathbb{R}^n$, it is given by all linear combinations of the basis vectors with *integer* coefficients. To the purpose of the thesis, we specify to the *hypercubic lattice* of four-dimensional Euclidean spacetime $\mathbb{R}^4$. Choosing the basis vectors as $a\hat{\mu}$, where $\hat{\mu}$ denotes the unit vector in the direction μ and a is the *lattice spacing*, we define

$$\Lambda = \left\{ x = \sum_{\mu=1}^{4} n_\mu a\hat{\mu} \,:\, n_\mu \in \mathbb{Z} \right\}. \tag{2.1}$$

We denote a lattice point with a 4-tuple $n \equiv \sum_\mu n_\mu \hat{\mu} = (n_1, n_2, n_3, n_4)$, which collects the four integers n_μ, or, with an abuse of notation, directly with $x = na \in \Lambda$. The lattice can be viewed

as a regular tiling of $\mathbb{R}^4$ with hypercubic blocks of volume a^4.

This lattice possesses a *discrete* group of isometries, given by the symmetry group of a hypercube together with discrete translational symmetry. In analogy with continuum QFTs, we consider fields that transform under these discrete symmetries. We then build the action as a functional of these fields that is invariant under the lattice isometries.

The discretization of a scalar, Weyl or Dirac field starts with very similar considerations. For the sake of brevity, we present directly the Dirac fermion case. We obtain a lattice field out of a continuum field restricting the four-component spinors $\psi(x)$ and $\bar{\psi}(x)$ to take values only on lattice points $x = na \in \Lambda$,

$$\psi(x) \to \psi(na), \qquad \bar{\psi}(x) \to \bar{\psi}(na). \tag{2.2}$$

The derivative field $\partial_\mu \psi(x)$ is approximated on the lattice by the finite differences

$$\partial_\mu \psi(x) \to \frac{\psi(na + a\hat{\mu}) - \psi(na)}{a} = \sum_y \partial_\mu(x, y)\psi(y), \tag{2.3a}$$

$$\to \frac{\psi(na) - \psi(na - a\hat{\mu})}{a} = \sum_y \partial^*_\mu(x, y)\psi(y), \tag{2.3b}$$

where we introduced the *forward* and *backward lattice derivatives*

$$\partial_\mu(x, y) \equiv \frac{\delta_{x+a\hat{\mu},y} - \delta_{x,y}}{a}, \qquad \partial^*_\mu(x, y) \equiv \frac{\delta_{x,y} - \delta_{x-a\hat{\mu},y}}{a}. \tag{2.4}$$

They connect nearest-neighbour lattice sites. Here and in the following we use the notation $\sum_x$ to denote a sum over lattice points $x \in \Lambda$ including the volume a^4 of the fundamental lattice domain. We regard this as the lattice equivalent of the integration over $x \in \mathbb{R}^4$

$$\int \mathrm{d}^4 x \to \sum_x \equiv a^4 \sum_n . \tag{2.5}$$

Defined in this way, $\sum_x$ in a fixed volume has a finite limit for $a \to 0$.

This simple set of substitutions allows us to write the lattice action for free Dirac fermions

$$S_{\mathrm{f}}^{\mathrm{free}}[\bar{\psi}, \psi] = \sum_{x,y} \bar{\psi}(x)\big\{ D(x, y) + M_0 \delta_{x,y} \big\}\psi(y), \tag{2.6}$$

where $x, y \in \Lambda$ and $D(x, y)$ is the lattice free *Dirac operator*

$$D(x, y) = \frac{1}{2} \sum_\mu \gamma_\mu \big(\partial^*_\mu + \partial_\mu\big)(x, y). \tag{2.7}$$

To have Hermiticity, a symmetrised version of the lattice derivatives in Eqs (2.4) has been used, whose action on a fermion field is the finite difference

$$\frac{1}{2}\big(\partial_\mu + \partial^*_\mu\big)\psi(x) = \frac{\psi(na + a\hat{\mu}) - \psi(na - a\hat{\mu})}{2a}. \tag{2.8}$$

The action in Eq. (2.6) satisfies one of the conditions to be considered a candidate for a regularization of the action of a free fermion field: it has the correct *naïve continuum limit*, i.e. it reduces to Eq. (1.10) in the $a \to 0$ limit.

Given an expression in the continuum, its lattice discretization is not unique: any expression differing by a term that goes to zero for $a \to 0$ has the same naïve continuum limit. At finite lattice spacing, these differences can be relevant. For instance, in Section 2.2.1 we will show that Eq. (2.6) is not satisfactory as a regularization of Dirac fermions and in Section 2.2.2 we will propose a solution given by a term which vanishes for $a \to 0$.

2.1.1. The lattice gauge field

Just as in the continuum, the fields ψ and $\bar{\psi}$ represent a vector of N_c mass-degenerate Dirac fermions, transforming under the (anti)fundamental representation $\boldsymbol{N}_c$ ($\bar{\boldsymbol{N}}_c$) of a SU(N_c) Lie group. Therefore, the action in Eq. (2.6) is invariant under the group of *global* SU(N_c) transformations. We introduce gauge interactions promoting the symmetry to a *local* gauge invariance.

One way to perform this on the lattice is to introduce vector fields $A_\mu^a(x)$ which take values only for $x \in \Lambda$, in complete analogy to what we have done for Dirac fermion fields. However, this necessarily leads to an action that is not gauge invariant. Recognising the importance of gauge invariance to ensure that the vector field correctly describes massless spin-1 particles, with only two helicity degrees of freedom, we prefer not to follow this path.

Fortunately, an elegant way to define gauge fields on the lattice such that gauge invariance is respected for every finite lattice spacing a was proposed by Wilson in 1974 [19]. We start introducing the *parallel transporter* between two points x and y, which in the continuum is defined as the path-ordered exponential of the line integral of the gauge potential along any curve γ such that $\gamma(0) = x$ and $\gamma(1) = y$

$$U_\gamma(x,y) \equiv P\exp\left\{-\mathrm{i}\int_\gamma \mathrm{d}z_\nu\, A_\mu(z)\right\} = P\exp\left\{-\mathrm{i}\int_0^1 \mathrm{d}t\, \gamma_\nu'(t) A_\nu\big(\gamma(t)\big)\right\}. \tag{2.9}$$

$U_\gamma(x,y)$ is an element of the gauge group SU(N_c) and has the property of transforming under a gauge transformation $\Omega(x)$ as

$$U_\gamma(x,y) \to U'_\gamma(x,y) = \Omega(x) U_\gamma(x,y) \Omega^\dagger(y). \tag{2.10}$$

We use parallel transporters to match the gauge transformation laws of the fermion field at different lattice points. For example, with two nearest-neighbour lattice points we perform the following substitutions

$$\bar{\psi}(x)\psi(x+a\hat{\mu}) \to \bar{\psi}(x) U_\gamma(x, x+a\hat{\mu})\psi(x+a\hat{\mu}), \tag{2.11a}$$

$$\bar{\psi}(x+a\hat{\mu})\psi(x) \to \bar{\psi}(x+a\hat{\mu}) U_\gamma(x+a\hat{\mu}, x)\psi(x), \tag{2.11b}$$

where the parallel transporter between two adjacent lattice points can be performed on a straight line ℓ

$$U_\ell(x, x+a\hat{\mu}) = P\exp\left\{-\mathrm{i}a\int_0^1 \mathrm{d}t\, A_\mu(x+ta\hat{\mu})\right\}, \tag{2.12a}$$

$$U_\ell(x+a\hat{\mu}, x) = U_\ell^\dagger(x, x+a\hat{\mu}). \tag{2.12b}$$

Wilson's key idea was to define the lattice gauge field to be made of by gauge group elements $U_\mu(x)$

$$U_{\ell}(x, x + a\hat{\mu}) \to U_\mu(x), \tag{2.13}$$

which transform under local gauge transformations like the parallel transporters in Eq. (2.12) and satisfy

$$U_{-\mu}(x + a\hat{\mu}) = U_\mu^\dagger(x). \tag{2.14}$$

In contrast to the fermion field defined at lattice points, the lattice gauge field is defined on the path linking two nearest-neighbour lattice points. Hence, $U_\mu(x)$ is often referred to as *link variable*, or simply *link*. Forward and backward lattice derivatives in Eq. (2.4) are substituted with their covariant version

$$\partial_\mu(x, y) \to \nabla_\mu[U](x, y) \equiv \frac{U_\mu(x)\delta_{x+a\hat{\mu},y} - \delta_{x,y}}{a}, \tag{2.15a}$$

$$\partial^*_\mu(x, y) \to \nabla^*_\mu[U](x, y) \equiv \frac{\delta_{x,y} - U^\dagger_\mu(x - a\hat{\mu})\delta_{x-a\hat{\mu},y}}{a}. \tag{2.15b}$$

Therefore, a lattice action that is invariant under local gauge transformations is given by Eq. (2.6)

$$S_{\mathrm{f}}[\bar{\psi}, \psi, U] = \sum_{x,y} \bar{\psi}(x)\big\{D[U](x, y) + M_0\delta_{xy}\big\}\psi(y), \tag{2.16}$$

with the lattice Dirac operator in Eq. (2.36) replaced by

$$D[U](x, y) = \frac{1}{2}\sum_\mu \gamma_\mu\big(\nabla_\mu + \nabla^*_\mu\big)[U](x, y). \tag{2.17}$$

Every link can be represented with the exponential of a Lie algebra element

$$U_\mu(x) = \exp\big\{-\mathrm{i}a\mathring{A}_\mu(x)\big\}, \qquad \mathring{A}_\mu(x) = \mathring{A}^a_\mu(x)T^a. \tag{2.18}$$

When the naïve continuum limit $a \to 0$ is taken, the link $U_\mu(x)$ can be expanded around $a = 0$

$$U_\mu(x) = \mathbb{1} - \mathrm{i}a\mathring{A}_\mu(x) + \mathcal{O}\big(a^2\big). \tag{2.19}$$

Comparing this with the small-separation expansion of the parallel transporter in Eq. (2.12) suggests to interpret $\mathring{A}_\mu(x)$, for small a, as the lattice discretization of the gluon field $A_\mu(x)$. With this prescription, the continuum limit of Eq. (2.16) reproduces the continuum gauge-invariant action in Eq. (1.13).

However, it is worth emphasizing that on the lattice the link field $U_\mu(x)$ in the Lie group, rather than the vector field in Lie algebra, is fundamental. Eq. (2.18) is just an exponential representation for $U_\mu(x)$, and at finite a the link is not an infinitesimal deformation of $\mathbb{1}$. Since the group is compact, in a given basis $\mathring{A}^a_\mu(x)$ are angular variables and cover a compact interval: $\mathring{A}^a_\mu(x) \in [-l/a, l/a]$. For this reason, the lattice formulation of QCD is also called *compact QCD*. When $a \to 0$, $\mathring{A}^a_\mu(x)$ is identified with the continuum field $A^a_\mu(x)$ with a suitable renormalization contant and the unbounded range of the vector field is recovered.

2.1.2. Wilson's plaquette action

To complete the construction of the gauge theory on the lattice, we need to find a discrete version of the YM action in Eq. (1.7). This is done considering the ordered product of four links around a *plaquette*, i.e. a unit square in the $(\hat{\mu}, \hat{\nu})$ plane,

$$U_{\mu\nu}(x) \equiv U_\mu(x) U_\nu(x + a\hat{\mu}) U_\mu^\dagger(x + a\hat{\nu}) U_\nu^\dagger(x). \tag{2.20}$$

Being the product of $\mathrm{SU}(N_\mathrm{c})$ elements, $U_{\mu\nu}(x) \in \mathrm{SU}(N_\mathrm{c})$ itself, thus it can be written as an exponential of an algebra element. We make the ansatz

$$U_{\mu\nu}(x) = \exp\{-\mathrm{i}a^2 \mathring{F}_{\mu\nu}(x)\}, \qquad \mathring{F}_{\mu\nu}(x) \in \mathfrak{su}(N_\mathrm{c}). \tag{2.21}$$

Using the Baker–Campbell–Hausdorff formula and the identity

$$\mathring{A}_\mu(x + a\hat{\nu}) = \mathring{A}_\mu(x) + a\partial_\nu \mathring{A}_\mu(x), \tag{2.22}$$

we have

$$\mathring{F}_{\mu\nu}(x) = \partial_\nu \mathring{A}_\mu(x) - \partial_\nu \mathring{A}_\mu(x) - \mathrm{i}\big[\mathring{A}_\mu(x), \mathring{A}_\nu(x)\big] + \mathcal{O}(a). \tag{2.23}$$

Therefore, $\mathring{F}_{\mu\nu}(x)$ is a valid lattice version of the field strength tensor, reducing to the continuum expression in the naïve continuum limit.

Consider the expression

$$\begin{aligned} S_\mathrm{W}[U] &= \frac{1}{g_0^2} \sum_{n,\mu,\nu} \mathrm{tr}\Big\{ \mathbb{1} - \frac{1}{2}\Big[U_{\mu\nu}(x) + U_{\mu\nu}^\dagger(x)\Big] \Big\} \\ &= \mathrm{const} - \frac{\beta}{N_\mathrm{c}} \sum_{n,\mu<\nu} \mathrm{Re}\,\mathrm{tr}\, U_{\mu\nu}(x), \qquad \beta \equiv \frac{2N_\mathrm{c}}{g_0^2}, \end{aligned} \tag{2.24}$$

where g_0 is a dimensionless coupling, the sum extends over the contributions coming from all distinct plaquettes on the lattice[1] and we introduced the *β parameter*. The expression in Eq. (2.24) is the so-called *Wilson's plaquette action*. Expanding $U_{\mu\nu}(x)$ for small a and using the Hermiticity of $\mathring{F}_{\mu\nu}(x)$, it is possible to show that Eq. (2.24) reduces to the continuum action in Eq. (1.7) in the naïve continuum limit.

We introduced the lattice field theory of massless vector bosons starting from a theory of Dirac fermions and imposing local gauge invariance. However, exactly as in the continuum, this is not necessary: a lattice theory of only vector fields with gauge invariance makes perfectly sense, especially in the non-Abelian YM theory case. Wilson's plaquette action in Eq. (2.24) by itself describes precisely $\mathrm{SU}(N_\mathrm{c})$ YM theory discretized on the lattice.

2.2. Lattice fermions and chiral symmetry

Chiral symmetry is crucial to describe the phenomenology of QCD, and it is associated to non-perturbative phenomena such as spontaneous χSB. Therefore it is of extreme importance to

[1] The sum can be restricted to $\mu < \nu$ since $U_{\nu\mu}(x) = U_{\mu\nu}^\dagger(x)$ is not independent of $U_{\mu\nu}(x)$.

be able to define chiral symmetry on the lattice. Beyond QCD, the SM is a *chiral gauge theory*, which contains Weyl fermions with definite chirality.

However, chiral symmetry and Weyl fermions on the lattice are not straightforward to implement, to the point that no lattice regularization of a chiral non-Abelian gauge theory is presently known. In the next sections, we present the profound interplay between lattice discretization and chiral symmetry, starting with the shortcomings of the naïve discretization of the fermion action given in Eq. (2.16).

2.2.1. The fermion doubling problem

The action in Eq. (2.16) reduces to the free action in Eq. (1.13) in the naïve continuum limit, but it has the problem that it describes more than one family of particles. This happens already at the level of free fermions. The easiest way to see this is to rewrite the action in Eq. (2.6) in a momentum basis, performing a Fourier transform

$$\psi(x) = \int_{-\pi/a}^{\pi/a} \frac{\mathrm{d}^4 p}{(2\pi)^4} \mathrm{e}^{\mathrm{i}px} \tilde{\psi}(p), \tag{2.25}$$

where the momentum integral is restricted to the Brillouin zone (BZ) $p_\mu \in [-\pi/a, \pi/a]$. Using the fact that the Dirac operator introduced in Eq. (2.7) is invariant under translations, we have

$$S_{\mathrm{f}}[\tilde{\psi}, \tilde{\bar{\psi}}] = \int_{-\pi/a}^{\pi/a} \frac{\mathrm{d}^4 p}{(2\pi)^4} \tilde{\bar{\psi}}(-p) \big[\tilde{D}(p) + M_0\big] \tilde{\psi}(p). \tag{2.26}$$

where

$$\tilde{D}(p) = \mathrm{i} \sum_\mu \gamma_\mu \mathring{p}_\mu, \qquad \mathring{p}_\mu \equiv \frac{1}{a} \sin a p_\mu. \tag{2.27}$$

From this expression, we can easily read the fermion propagator

$$\begin{aligned} \langle \psi(x) \bar{\psi}(y) \rangle &= \int_{-\pi/a}^{\pi/a} \frac{\mathrm{d}^4 p}{(2\pi)^4} \mathrm{e}^{\mathrm{i}p(x-y)} \big[\tilde{D}(p) + M_0\big]^{-1}(p) \\ &= \int_{-\pi/a}^{\pi/a} \frac{\mathrm{d}^4 p}{(2\pi)^4} \mathrm{e}^{\mathrm{i}p(x-y)} \left[- \frac{\mathrm{i} \sum_\mu \gamma_\mu \mathring{p}_\mu - M_0}{\sum_\mu \mathring{p}_\mu^2 + M_0^2} \right]. \end{aligned} \tag{2.28}$$

Approaching the continuum limit, physical states emerge in regions where the denominator in Eq. (2.28) is small compared to the cut-off. This happens at the minima of $\sum_\mu \mathring{p}_\mu^2 + M_0^2$. However, for each direction μ there are two solutions of $\mathring{p}_\mu = 0$ in the BZ: one at $p_\mu = 0$ the other at $p_\mu = \pm(\pi/a)\hat{\mu}$. Therefore, in $D = 4$ spacetime dimensions the denominator of Eq. (2.28) has $2^D = 16$ degenerate minima, at

$$\bar{p}_\mu^{(\alpha)} = \frac{\pi}{a} n_\mu^{(\alpha)}, \quad n_\mu^{(\alpha)} = \left\lfloor \frac{\alpha}{2^{\mu-1}} \right\rfloor \bmod 2, \qquad \text{for } \alpha = 0, \dots, 2^D - 1. \tag{2.29}$$

This is the origin of the so-called *fermion doubling problem* [67, 68], which is treated extensively in any book on lattice field theory [69–71]. Performing the integration on p_4, with some algebra

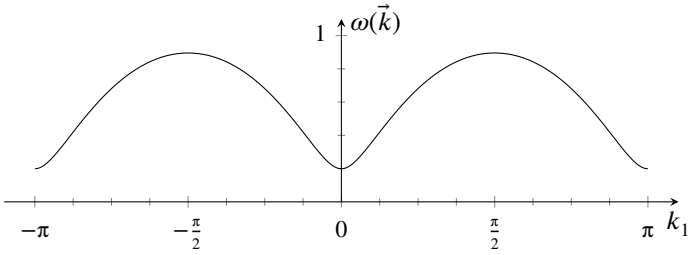


Figure 2.1.: The lattice dispersion relation in Eq. (2.31b) in the BZ. Only a single space dimension is shown, $k_2 = 0$, $k_3 = 0$ and $aM_0 = 0.2$.

it is possible to show that Eq. (2.28) approximate in the continuum limit the propagator of 2^D mass-degenerate free Dirac fermions

$$\langle \psi(x)\bar{\psi}(y)\rangle \xrightarrow{a\to 0} \sum_{\alpha=0}^{2^D-1} \mathrm{e}^{\mathrm{i}\bar{p}^{(\alpha)}x} \int \frac{\mathrm{d}^3k}{(2\pi)^3 2\mathrm{i}k_4} \mathrm{e}^{\mathrm{i}k(x-y)} S_\alpha \left[\mathrm{i}\sum_\mu \gamma_\mu k_\mu - M_0\right] S_\alpha^{-1}, \tag{2.30}$$

where

$$k_i = p_i - \bar{p}_i^{(\alpha)}, \qquad k_4 = \mathrm{i}\lim_{a\to 0}\omega(\vec{k}) = \mathrm{i}\sqrt{M_0^2 + \vec{k}^2}, \tag{2.31a}$$

$$\omega(\vec{k}) = \frac{1}{a}\sinh^{-1}\sqrt{M_0^2 + \textstyle\sum_{j=1}^3 \mathring{k}_j^2}, \tag{2.31b}$$

and S_α is a unitary operator which acts as a change of basis on Dirac γ-matrices, such that

$$S_\alpha \gamma_\mu S_\alpha^\dagger = \gamma_\mu \cos a\bar{p}_\mu^{(\alpha)} \quad \Rightarrow \quad S_\alpha = \prod_\mu \left(\mathrm{i}\gamma_\mu\gamma_5\right)^{n_\mu^{(\alpha)}}. \tag{2.32}$$

There is a deep connection between the doubling problem and chiral symmetry, which will be presented in detail in Section 2.2.3. It is possible to see this already at this level considering a massless Weyl fermion with definite (say left-handed) chirality. It can be represented with a Dirac spinor field projected on half the space

$$\psi_L = \frac{1-\gamma_5}{2}\psi \quad \Rightarrow \quad \bar{\psi}_L = \bar{\psi}\frac{1+\gamma_5}{2}. \tag{2.33}$$

The naïvely-discretized Weyl fermion propagator is

$$\begin{aligned}\langle \psi_L(x)\bar{\psi}_L(y)\rangle &\xrightarrow{a\to 0} \sum_{\alpha=0}^{2^D-1} \mathrm{e}^{\mathrm{i}\bar{p}^{(\alpha)}x} \int \frac{\mathrm{d}^3k}{(2\pi)^3 2\mathrm{i}k_4} \mathrm{e}^{\mathrm{i}k(x-y)} S_\alpha \left[\mathrm{i}\sum_\mu \gamma_\mu k_\mu\right] S_\alpha^{-1}\left(\frac{1-\gamma_5}{2}\right) \\ &= \sum_{\alpha=0}^{2^D-1} \mathrm{e}^{\mathrm{i}\bar{p}^{(\alpha)}x} \int \frac{\mathrm{d}^3k}{(2\pi)^3 2\mathrm{i}k_4} \mathrm{e}^{\mathrm{i}k(x-y)} S_\alpha \left[\mathrm{i}\sum_\mu \gamma_\mu k_\mu\right] \left(\frac{1-\epsilon^\alpha\gamma_5}{2}\right) S_\alpha^{-1},\end{aligned} \tag{2.34}$$

where

$$\epsilon^{\alpha} = (-1)^{\sum_{\mu} n_{\mu}^{(\alpha)}}. \tag{2.35}$$

The result is the propagator of 2^D Weyl fermions. Moreover, it turns out that half of them are *left-handed* Weyl fermions and the other half are *right-handed* Weyl fermions.

There is a deeper message: suppose we want to put on the lattice a chiral gauge theory, with a content of left-handed Weyl fermions. The SM is a prototype of it. Naïve discretization results in a theory with an equal number of left- and right-handed Weyl fermions with the same quantum numbers. The resulting lattice theory is not a chiral gauge theory, and can be recast as a theory of Dirac fermions.

These phenomena are consequences of a theorem from Nielsen and Ninomiya [26], as explained in Section 2.2.3.

2.2.2. Wilson fermions

A possible solution to the fermion doubling problem was given by Wilson [67], who replaced the lattice Dirac operator of Eq. (2.7) with the *Wilson–Dirac operator*, which for free fermions is

$$D_{\mathrm{W}}(x, y) = \frac{1}{2} \sum_{\mu} \gamma_{\mu} \big(\partial_{\mu} + \partial_{\mu}^{*}\big)(x, y) - \frac{a}{2} \sum_{\mu} \partial_{\mu} \partial_{\mu}^{*}(x, y). \tag{2.36}$$

The only difference with respect to the Dirac operator in Eq. (2.7) is an added term that is a discretization of the Laplace operator. It is proportional to a, thus it vanishes in the naïve continuum limit, leading to the usual fermion action in the continuum. To see what happens at finite lattice spacing, we study its Fourier transform

$$\tilde{D}_{\mathrm{W}}(p) = \mathrm{i} \sum_{\mu} \gamma_{\mu} \mathring{p}_{\mu} + \frac{1}{2} a \hat{p}^2, \qquad \hat{p}_{\mu} \equiv \frac{2}{a} \sin \frac{a p_{\mu}}{2}. \tag{2.37}$$

The period of $\hat{p}$ is twice the period of $\mathring{p}$, thus the second term in the Wilson–Dirac operators is non-zero at the corners of the BZ. It acts as an effective mass

$$\omega^{(\alpha)}(\vec{\kappa} = 0) = \frac{1}{a} \ln\left(1 + a M_0 + \sum_{\mu} n_{\mu}^{(\alpha)}\right), \tag{2.38}$$

that is of $\mathcal{O}(a^{-1})$ for every unwanted doubler. Only the solution with $n_{\mu}^{(\alpha)} = 0$, at the centre of the BZ, keeps a sensible continuum limit.

However, the fact that Wilson's term behaves like a mass term has consequences on chiral symmetry, as is discussed in Section 1.4. Indeed, even in absence of an explicit mass term M_0, chiral symmetry is explicitly broken by Wilson fermions at finite lattice spacing.

The interacting case is straightforward. The Dirac operator on a background lattice gauge field U is

$$D_{\mathrm{W}}[U](x, y) = \frac{1}{2} \sum_{\mu} \gamma_{\mu} \big(\nabla_{\mu} + \nabla_{\mu}^{*}\big)[U](x, y) - \frac{a}{2} \sum_{\mu} \nabla_{\mu} \nabla_{\mu}^{*}[U](x, y), \tag{2.39}$$

and action for Wilson fermions is

$$S_{\mathrm{Wf}}[\bar{\psi},\psi,U] = \sum_{x,y} \bar{\psi}(x)\big(D_{\mathrm{W}}[U](x,y) + M_0\delta_{xy}\big)\psi(y). \tag{2.40}$$

The Wilson–Dirac operator is often rewritten as

$$D_{\mathrm{W}}[U](x,y) = C\big(\delta_{x,y} - \kappa H[U](x,y)\big), \qquad \kappa = \frac{1}{2aM_0+8}, \quad C = M_0 + \frac{4}{a} = \frac{1}{2a\kappa}, \tag{2.41}$$

where H collects all the terms in Eq. (2.36) connecting nearest-neighbour sites and is thus called the *hopping matrix*. Explicitly,

$$H[U](x,y) = \sum_{\mu}\Big[(1+\gamma_\mu)U_\mu(x)\delta_{x+a\hat{\mu},y} + (1-\gamma_\mu)U_\mu^\dagger(x-a\hat{\mu})\delta_{x-a\hat{\mu},y}\Big]. \tag{2.42}$$

The constant κ is the *hopping parameter* and it sets the relative strength of the hopping interactions, while C can be easily reabsorbed in the normalization of the fields.

In the following, we suppress the subscript $_{\mathrm{W}}$ from the Wilson–Dirac operator. Unless stated differently, D will always be the Wilson–Dirac operator. Moreover, we introduce the notation D_M to denote the Dirac operator including the mass term

$$D_M[U](x,y) = D[U](x,y) + M_0\delta_{xy}. \tag{2.43}$$

The fact that chiral symmetry is broken by the lattice regulator has a direct impact on the renormalization of the theory, making the renormalization pattern of parameters and fields more complicated than in presence of chiral symmetry, see Section 2.3.4. For instance, the bare mass M_0 gets an additive renormalization on the top of the multiplicative one.

2.2.3. Nielsen–Ninomiya no-go theorem

We already anticipated that the incompatibility between chiral symmetry and lattice regularization is not limited to Wilson fermions, but is more general. This was formalized in 1981 by Nielsen and Ninomiya in their *no-go theorem* [26, 72, 73]. An elegant proof of the theorem is given in Ref. [74].

No-go theorem (Nielsen–Ninomiya)**.** *For a lattice model with an action quadratic in the fields that can be written in momentum space as in Eq.* (2.26)*, the following statements cannot be simultaneously true:*

1. $\tilde{D}(p)$ *is an analytic function of* p_μ *with period* $2\pi/a$*;*
2. *for* $|p_\mu| \ll \pi/a$, $\tilde{D}(p) = \mathrm{i}\sum_\mu \gamma_\mu p_\mu + \mathcal{O}\big(ap^2\big)$*;*
3. $\tilde{D}(p)$ *is invertible everywhere except at* $p_\mu = 0 \pmod{2\pi/a}$*;*
4. $\tilde{D}(p)$ *anticommutes with* γ_5*.*

The condition 1 is necessary for the locality of the coordinate-space Dirac operator $D(x, y)$. The conditions 2 and 3 are required to reproduce a single Dirac fermion in the continuum. The condition 4,

$$D\gamma_5 + \gamma_5 D = 0 \quad \Leftrightarrow \quad DP_\pm = P_\mp D, \quad P_\pm = \frac{1 \pm \gamma_5}{2} \tag{2.44}$$

is equivalent to having a chiral symmetric theory and is obviously true in the continuum.

Therefore, it is not possible to have a chiral-symmetric theory without giving up at least one of conditions 1–3. For example, naïve fermions fail to satisfy condition 3, since $\tilde{D}(p)$ vanishes at the corners of the BZ. Equivalently, any discretization that satisfies conditions 1–3 necessarily violate Eq. (2.44) and thus chiral symmetry. In the case of Wilson fermions,

$$D_\mathrm{W}\gamma_5 + \gamma_5 D_\mathrm{W} = -a\gamma_5 \sum_\mu \nabla_\mu \nabla^*_\mu. \tag{2.45}$$

2.2.4. The Ginsparg–Wilson relation and Lüscher's symmetry

In 1982 Ginsparg and Wilson [28] proposed to replace Eq. (2.44), valid in the continuum, with a lattice version, known as the *Ginsparg–Wilson (GW) relation*

$$D\gamma_5 + \gamma_5 D = \bar{a} D\gamma_5 D, \tag{2.46}$$

where $\bar{a}$ is proportional to the lattice spacing, so that Eq. (2.46) reduces to Eq. (2.44) in the continuum limit. Formally, neglecting the presence of zero modes, the fact that at finite lattice spacing the anticommutator of the quark propagator with γ_5 is just a contact term

$$\left\{D^{-1}, \gamma_5\right\}(x, y) = \bar{a}\gamma_5 \delta_{x,y}, \tag{2.47}$$

suggest a very mild breaking of chiral symmetry.

At the time of its introduction, the GW relation had no immediate consequences since no solutions to Eq. (2.46) in the interacting case were known. In subsequent years, alternative approaches to chiral symmetry were proposed, such as Kaplan's *domain wall fermions* [75, 76] and the *fixed point action* [77]. In the domain wall approach, the physical theory lives on the boundary of a five-dimensional QFT. Chiral symmetry is recovered in the limit of an infinite fifth dimension. In this limit, Neuberger derived an effective four-dimensional description for the domain wall Dirac operator known as *overlap Dirac operator* [78].

In 1997, 15 years after the original paper, it was found that both the fixed point action [79] and the overlap Dirac operator [80] are exact solution of the GW relation, while domain wall fermions satisfy it only approximaitely. Of course, these Dirac operators cannot avoid the no-go theorem conclusion and indeed they break chiral symmetry. However, a way out from the no-go theorem was found in 1998 by Lüscher [81], when he realized that an action employing a Dirac operator that solves Eq. (2.46) conserves a slightly modified, but exact, form of chiral symmetry. This *Lüscher's symmetry* is sufficient to prove that many of the consequences of chiral symmetry in the continuum still holds for GW-discretized fermions at finite lattice spacing.

To construct the action of GW fermions, we introduce the convenient notation

$$\hat{\gamma}_5[U](x, y) = \gamma_5\left\{\delta_{x,y} - \bar{a} D[U](x, y)\right\}, \tag{2.48}$$

where one should keep in mind that $\hat{\gamma}_5$ is not just a constant γ-matrix but an operator depending on the gauge field through the Dirac operator. The fermion field is split in modified chiral components

$$\hat{\psi}_L = \hat{P}_- \psi, \quad \hat{\psi}_R = \hat{P}_+ \psi, \qquad \bar{\psi}_L = \bar{\psi} P_+, \quad \bar{\psi}_R = \bar{\psi} P_-, \tag{2.49}$$

using the projectors in Eq. (1.61) together with

$$\hat{P}_\pm \equiv \frac{1 \pm \hat{\gamma}_5}{2}, \qquad \hat{P}_\pm^2 = \hat{P}_\pm, \quad \hat{P}_+ + \hat{P}_- = \mathbb{1}, \quad \hat{P}_+ \hat{P}_- = \hat{P}_- \hat{P}_+ = 0. \tag{2.50}$$

It is now sufficient to use GW relation to easily show that

$$D\hat{\gamma}_5 + \gamma_5 D = 0 \quad \Leftrightarrow \quad D\hat{P}_\pm = P_\mp D, \tag{2.51}$$

to prove that the massless fermion action built from the modified chiral fields

$$S_{\mathrm{GWf}}\Big|_{M_0=0} = \sum_{x,y} \left[\bar{\psi}_L(x) D(x,y) \hat{\psi}_L(y) + \bar{\psi}_R(x) D(x,y) \hat{\psi}_R(y)\right] \tag{2.52}$$

is invariant, in complete analogy with the continuum, under Lüscher's symmetry transformations

$$\hat{\psi}_L \to \hat{\psi}'_L = U_L \hat{\psi}_L, \qquad \hat{\psi}_R \to \hat{\psi}_R = U_R \hat{\psi}_R, \tag{2.53a}$$

$$\bar{\psi}_L \to \bar{\psi}'_L = \bar{\psi}_L U_L^\dagger, \qquad \bar{\psi}_R \to \bar{\psi}_R = \bar{\psi}_R U_R^\dagger, \tag{2.53b}$$

with the U_L and U_R of Eq. (1.66). The Lüscher's symmetry-breaking mass term is easily introduced

$$\bar{\psi}_L \mathcal{M}_0^\dagger \hat{\psi}_R + \bar{\psi}_R \mathcal{M}_0 \hat{\psi}_L = \bar{\psi} M_0 \left[P_- \hat{P}_- + P_+ \hat{P}_+\right] \psi = \bar{\psi} M_0 \left[\mathbb{1} - \frac{\bar{a}}{2} D\right] \psi, \tag{2.54}$$

thus the massive GW-fermion action is

$$\begin{aligned} S_{\mathrm{GWf}} &= \sum_{x,y} \bar{\psi}(x) \Big\{ D(x,y) + M_0 \Big[\mathbb{1} - \frac{\bar{a}}{2} D\Big](x,y) \Big\} \psi(y) \\ &= \sum_{x,y} \bar{\psi}(x) \big[\omega D(x,y) + M_0 \delta_{x,y}\big] \psi(y), \qquad \omega \equiv \mathbb{1} - \frac{\bar{a} M_0}{2}. \end{aligned} \tag{2.55}$$

2.2.5. Dirac operator spectrum

Consider a Dirac operator that is γ_5-*hermitian*

$$\gamma_5 D \gamma_5 = D^\dagger \quad \Leftrightarrow \quad \gamma_5 D = \big(\gamma_5 D\big)^\dagger. \tag{2.56}$$

It is interesting to study its spectrum

$$D v_\lambda = \lambda v_\lambda. \tag{2.57}$$

Since D is the product of two hermitian matrices, γ_5 and $\gamma_5 D$, for every root λ of the characteristic polynomial $P(\lambda)$, also λ^* is a root. Eigenvalues of D are either real or complex conjugate pairs. Moreover, we have

$$\lambda\big(v_\lambda, \gamma_5 v_\lambda\big) = \big(v_\lambda, \gamma_5 D v_\lambda\big) = \big(v_\lambda, D^\dagger \gamma_5 v_\lambda\big) = \big(D v_\lambda, \gamma_5 v_\lambda\big) = \lambda^*\big(v_\lambda, \gamma_5 v_\lambda\big), \tag{2.58}$$

which means $(\operatorname{Im}\lambda)(v_\lambda, \gamma_5 v_\lambda) = 0$. Thus, only eigenvectors with real eigenvalues can have non-vanishing chirality

$$\big(v_\lambda, \gamma_5 v_\lambda\big) \neq 0 \quad \Rightarrow \quad \lambda \in \mathbb{R}. \tag{2.59}$$

So far, this is true for any γ_5-hermitian Dirac operator, such as the Wilson–Dirac operator. Now, we require it to satisfy the GW relation. Multiplying Eq. (2.46) with γ_5 either on the left or right we arrive at

$$D^\dagger + D = \bar{a} D^\dagger D, \quad D + D^\dagger = \bar{a} D D^\dagger \quad \Rightarrow \quad D^\dagger D = D D^\dagger, \tag{2.60}$$

thus a γ_5-hermitian GW Dirac operator is a *normal operator*, whose eigenvectors form an orthogonal basis. Eq. (2.60) applied to eigenvectors becomes a relations between eigenvalues

$$\lambda^* + \lambda = \bar{a} \lambda^* \lambda, \tag{2.61}$$

which is solved by

$$\lambda = \frac{1}{\bar{a}}\big(1 - \mathrm{e}^{\mathrm{i}\alpha}\big), \qquad \text{for } \alpha \in \mathbb{R}. \tag{2.62}$$

Therefore, eigenvalues are restricted to a circle in the complex plane centred in $1/\bar{a}$ that goes through the origin. D may have zero modes. On this subspace, γ_5 and D commute

$$D v_0 = 0 \quad \Rightarrow \quad \gamma_5 D v_0 = 0 \quad \Rightarrow \quad D \gamma_5 v_0 = 0. \tag{2.63}$$

The zero modes can thus be chosen as eigenvectors of γ_5. These are chiral modes

$$\gamma_5 v_0 = \pm v_0. \tag{2.64}$$

The modes with eigenvalue $+1$ are right-handed, the ones with eigenvalue -1 are left-handed. Exactly in the same way, it is possible to show that also modes with real $\lambda = 2/\bar{a}$ can be chosen to be chiral

$$\gamma_5 v_{2/\bar{a}} = \pm v_{2/\bar{a}}. \tag{2.65}$$

2.2.6. The overlap Dirac operator

To be a bit more specific, we consider a particular Dirac operator that solves the GW relation and implements chiral fermions on the lattice: the *Neuberger–Dirac operator* or *overlap Dirac operator*. While introduced in the early 90s, the modern form of the Neuberger–Dirac operator was presented in 1997 by Neuberger in Ref. [78]. Explicitly it is given by

$$D_{\mathrm{N}} = \frac{1+s}{a}\big[\mathbb{1} + \gamma_5 \operatorname{sgn}\big(H_s\big)\big], \qquad \gamma_5 H_s = aD - \mathbb{1}(1+s), \tag{2.66}$$

where s is a real number with $|s| < 1$ and D is a suitable γ_5-hermitian 'kernel' Dirac operator and

$$\operatorname{sgn}(X) = \frac{X}{\sqrt{X^2}} \tag{2.67}$$

for any Hermitian matrix X. A common choice is to use simply the Wilson–Dirac operator defined in Eq. (2.39). D_{N} is an exact solution of the GW relation [80], with

$$\bar{a} = \frac{a}{1+s}. \tag{2.68}$$

While the Wilson–Dirac operator is a local operator since it connects only nearest-neighbour lattice sites, the Neuberger–Dirac operator includes an inverse square root which leads to a matrix with non-vanishing entries for arbitrary pairs of lattice sites. One may assume that D_{N} is a non-local operator, but this is actually not the case. In the free case, the momentum-space Neuberger–Dirac operator is, for $s = 0$,

$$\tilde{D}_{\mathrm{N}}(p) = \frac{1}{a}\left(1 - \frac{1 - \mathrm{i}a\sum_\mu \gamma_\mu \mathring{p}_\mu - \frac{1}{2}a^2\hat{p}^2}{\sqrt{1 + \frac{1}{2}a^4\sum_{\mu<\nu}\hat{p}_\mu^2\hat{p}_\nu^2}}\right). \tag{2.69}$$

It satisfies conditions 1–3 of the no-go theorem. In particular, the analyticity condition 1 implies that its Fourier transform decays exponentially at large distances, with a rate proportional to a^{-1}. Therefore, the Neuberger–Dirac operator is still local in a more general sense.

This result extends to the interacting case: Hernandez et al. [82] showed analytically that, for sufficient smooth gauge field such that

$$\left\|\mathbb{1} - U_{\mu\nu}(x)\right\| \leq \epsilon, \qquad \text{for any } x, \mu < \nu \text{ and } \epsilon \leq 1/30, \tag{2.70}$$

the Neuberger–Dirac operator decays exponentially at large distances

$$\left\|D_{\mathrm{N}}(x,y)\right\| \leq C\exp\left(-\gamma\frac{|x-y|}{a}\right), \tag{2.71}$$

with C and γ constants independent of the gauge field. Numerical evidence suggests that the same bound holds for all statistically relevant gauge field configurations. In physical units, the interaction distance is a/γ and thus shrinks to zero in the continuum limit, leading to a local Dirac operator.

2.3. The path integral on the lattice

We conclude the introduction of gauge theories on the lattice giving a precice definition of the lattice-regularized path integral. The equivalent of the partition function in Eq. (1.16) on the lattice is

$$\mathcal{Z} = \int \mathcal{D}[\bar{\psi}, \psi, U]\, \mathrm{e}^{-S_{\mathrm{Wf}}[\bar{\psi},\psi,U] - S_{\mathrm{W}}[U]}, \tag{2.72}$$

where S_{Wf} is the action of Wilson fermions in Eq. (2.40) and S_{W} is Wilson's plaquette action in Eq. (2.24). A key point to stress it that, while the path integral is only a formal construction in the continuum, we can give a precise meaning to the expression in Eq. (2.72). To do this, we define what the integration measure symbol $\mathcal{D}[\bullet]$ stands for.

The path integral measure for Dirac fields is

$$\mathcal{D}[\bar{\psi},\psi] = \prod_x \prod_f \prod_i \Big[\mathrm{d}\bar{\psi}^i_{(f)}(x)\, \mathrm{d}\psi^{(f)}_i(x) \Big], \tag{2.73}$$

where the product extends over all lattice points $x \in \Lambda$, flavour indices $f = 1, \dots, N_{\mathrm{c}}$ and colour indices $i = 1, \dots, N_{\mathrm{c}}$. To reproduce the Fermi–Dirac statistics, the ψ and $\bar{\psi}$ fields must satisfy anticommutation relations and cannot be ordinary numbers: they are *Grassmann number*-valued fields. Therefore, each $\mathrm{d}\psi^i_{(f)}(x)$ in the product is the integration measure of a Berezin integral over a Grassmann number. A summary of Grassmann algebra and its rules is given in Section A.4.

Similarly, the path integral measure for the gauge field is

$$\mathcal{D}[U] = \prod_x \prod_\mu \mathrm{d}U_\mu(x)\,. \tag{2.74}$$

Here, for each link variable the integration has to be done on the compact $\mathrm{SU}(N_{\mathrm{c}})$ Lie group. The path integral over U should be invariant under a change of integration variables $U \to U'$

$$\mathcal{Z} = \int \mathcal{D}[U]\, \mathrm{e}^{-S[U]} = \int \mathcal{D}[U']\, \mathrm{e}^{-S[U']} = \int \mathcal{D}[U']\, \mathrm{e}^{-S[U]}, \tag{2.75}$$

where in the last step we have used the gauge-invariance of the action. To have this, the measure $\mathcal{D}[U]$ must be gauge invariant, i.e.

$$\mathrm{d}U_\mu(x) = \mathrm{d}U'_\mu(x) = \mathrm{d}\big\{ \Omega(x) U_\mu(x) \Omega^\dagger(x + a\hat{\mu}) \big\} \tag{2.76}$$

for every gauge link. The proper integration measure which satisfies this property is the so-called *invariant Haar measure*, which is introduced in Section A.5.

Remarkably, there is no need of fixing a gauge to work with the lattice-regularized path integral. This is due to the way the gauge field is introduced. Gauge invariance of the action translates in the existence of *gauge orbits*. These are the equivalence classes of gauge field configurations that leave invariant the action and any gauge-invariant field. When integrating on a gauge orbit, the path integral is not suppressed in regions when the continuum field $A_\mu(x)$ becomes large, since the integrand is a gauge-invariant constant. This gives rise to an infinite factor that has to be removed by gauge fixing. On the contrary, the path integral over a gauge orbit of the lattice field $U_\mu(x)$ is a finite factor: it is just the value of the gauge-invariant integrand times the finite volume of the gauge group. In correlation functions of gauge-invariant fields, this factor automatically cancels between the numerator and the denominator.

2.3.1. The transfer matrix

The connection between the path integral formalism and canonical quantization can be formulated entirely on the lattice: starting from the path integral expression of the partition function,

one can reexpress it in terms of the so-called *transfer matrix* and explicitly build the Hilbert space.

Consider a Euclidean field theory regularized on a lattice with finite and periodic time extent. The transfer matrix represents the operator $\hat{T}$ that evolves from the quantum state at time t to the quantum state at time $t + a$. It allows to define the partition function as

$$\mathcal{Z} = \operatorname{Tr} \hat{T}^{T/a}, \tag{2.77}$$

where T/a is the time extent in lattice units.

In the case of Wilson's discretization of QCD—i.e. Wilson's plaquette action in Eq. (2.24) and the Wilson–Dirac operator in Eq. (2.39)—the explicit form of the transfer matrix is known [67, 83, 84]. A detailed derivation of this transfer matrix, inspired by Ref [85], is given in Appendix C. As an operator on the Hilbert space of the quantum field theory, the transfer matrix is

$$\hat{T} = \hat{P}_{\mathrm{g}} \hat{T}' \hat{P}_{\mathrm{g}} > 0, \tag{2.78}$$

where $\hat{T}'$ is a self-adjoint, bounded operator and $\hat{P}_{\mathrm{g}}$ is a projector on the physical, gauge-invariant sector of the Hilbert space. It is possible to show that the transfer matrix for Wilson fermions, explicitly built in Appendix C, is strictly positive [67, 84]. Moreover, it is unique up to a unitary transformation. It follows that the lattice Hamiltonian

$$\hat{H} \equiv -\frac{1}{a} \ln \hat{T}, \tag{2.79}$$

is self-adjoint and unique.

The expectation value of fields in the path integral formalism can be expressed with corresponding operator insertions at the proper time in the product of transfer matrices. Assuming $x_4 > y_4$,

$$\langle O(x) O(y) \rangle = \frac{1}{\operatorname{Tr} \hat{T}^{T/a}} \operatorname{Tr}\left\{ \hat{T}^{(T-x_4)/a} \hat{O}(\vec{x}) \hat{T}^{(x_4-y_4)/a} \hat{O}(\vec{y}) \hat{T}^{y_4/a} \right\}. \tag{2.80}$$

To finally transform the trace in the vacuum expectation value of operators, we can evaluate the trace over a complete set of eigenstates of the Hamiltonian

$$\hat{H} |k\rangle = E_k |k\rangle \tag{2.81}$$

to obtain

$$\begin{aligned} \langle O(x) O(y) \rangle &= \frac{1}{\sum_k \left\langle k \middle| \hat{T}^{T/a} \middle| k \right\rangle} \sum_k \left\langle k \middle| \hat{T}^{(T-x_4)/a} \hat{O}(\vec{x}) \hat{T}^{(x_4-y_4)/a} \hat{O}(\vec{y}) \hat{T}^{y_4/a} \middle| k \right\rangle \\ &= \frac{1}{\sum_k \mathrm{e}^{-T E_k}} \sum_k \mathrm{e}^{-(T-x_4) E_k} \left\langle k \middle| \hat{O}(\vec{x}) \hat{T}^{(x_4-y_4)/a} \hat{O}(\vec{y}) \middle| k \right\rangle. \end{aligned} \tag{2.82}$$

Let us assume that the theory has a mass gap. That is to say, there is only one state, the vacuum $|0\rangle$, with $E_0 = 0$ and any other state $|k > 0\rangle$ has strictly positive energy $E_k > 0$. Then, taking the $T \to \infty$ limit projects the trace on the vacuum and gives

$$\left\langle O(x) O(y) \right\rangle = \left\langle 0 \middle| \hat{O}(\vec{x}) \hat{T}^{(x_4-y_4)/a} \hat{O}(\vec{y}) \middle| 0 \right\rangle. \tag{2.83}$$

It can be proven that Wilson's formulations of lattice QCD satisfies reflection positivity [86, 87]. This guarantees the positivity of scalar products in Hilbert space.

2.3.2. Spectroscopy

An interesting property of a theory like QCD is its *spectrum*, i.e. the energy spectrum obtained from the eigenvalues of the Hamiltonian of the theory.

The low-energy spectrum can be extracted computing correlation functions of appropriate operators. Consider the two-point function of a composite field O at a physical separation $x - y$

$$C_O(x - y) \equiv \langle O(x)O(y)\rangle. \tag{2.84}$$

Assuming that $O(x)$ and $O(y)$ are local fields at different Euclidean times $x_4 > y_4$, this is the vacuum expectation value of the corresponding operators evolved at different times with the transfer matrix as in Eq. (2.83). If we now insert a complete set of eigenstates of the Hamiltonian between the two operators, the two-point function reads

$$\big\langle O(x)O(y)\big\rangle = \sum_k \big\langle 0\big|\hat{O}(\vec{x})\big|k\big\rangle\big\langle k\big|\hat{O}(\vec{y})\big|0\big\rangle \mathrm{e}^{-E_k(x_4-y_4)}. \tag{2.85}$$

States contributing to the sum are selected among those with quantum numbers compatible with $O(x)$ and $O(y)$. At large separations $x_4 - y_4$, the two-point function is dominated by the exponential of the lowest-energy state. The convergence to this behaviour depends on the separation of the lowest-energy state from the next excited one with the same quantum numbers, assuming they are not degenerate, and on the overlap of $O(x)$ and $O(y)$ with them.

One is usually interested in the mass of one-particle states, i.e. in their rest energy. Therefore, it is common to project the $O(x)$ and $O(y)$ to zero spatial momentum to nullify the overlap with states with lattice momentum $\vec{p} \neq 0$. This is done averaging at least one of the two fields over the time-slice at fixed time

$$\bar{C}_O(x_4 - y_4) = \big\langle \bar{O}(x_4)O(y)\big\rangle = \frac{1}{L^3}\sum_{\vec{x}}\big\langle O(x)O(y)\big\rangle = \sum_k \left(R_k^{\hat{O}}\right)^2 \mathrm{e}^{-M_k x_4}. \tag{2.86}$$

2.3.3. Quark bilinears on the lattice

The low-energy spectrum of non-Abelian gauge theories is composed by glueballs and, if $N_\mathrm{f} \geq 1$, hadrons. Among hadrons, mesons are created and annihilated by quark bilinear interpolating fields. There is an infinite number of fields with the right transformation properties to have a non-zero overlap with a given meson. The simplest are *local* quark bilinears

$$O_\Gamma^{ij}(x) = \bar{\psi}^i(x)\Gamma\psi^j(x). \tag{2.87}$$

Slightly more complex ones are non-local fields built taking ψ and $\bar{\psi}$ at different spacetime points. Gauge invariance is recovered inserting a parallel transporter

$$O_\Gamma^{ij}(x, x') = \bar{\psi}^i(x)\Gamma U(x, x')\psi^j(x'). \tag{2.88}$$

This is easily realized on the lattice using links on any path between x and x'. Or similarly one can build quark bilinear using *smeared* quark fields

$$\psi_S(x) = \sum_y S(x, y)\psi(y), \tag{2.89a}$$

$$\bar{\psi}_S(x) = \sum_y \bar{\psi}(y)S^\dagger(y, x), \tag{2.89b}$$

with some smearing function $S(x, y)$. For instance, smearing on spacetime point on the time-slice can improve the overlap with the lowest-energy state with respect to the higher ones. The extreme form of this smearing are fields averaged on the time-slice, obtained with $S(x, y) = \delta_{x_4 - y_4}$.

2.3.4. Lattice currents and densities

In Section 1.4 we used WTIs induced by chiral symmetry to show some important results regarding the low-energy structure of QCD-like theories. We would like to use WTIs also in the lattice-discretized theory, but this is not unhindered. We have seen in Section 2.2 that the lattice regulator is incompatible with chiral symmetry. As a consequence of no-go theorem, chiral WTIs do not hold on the lattice as they are.

Let us focus on the case of Wilson's discretization of QCD. With a bit of algebra, it is possible to derive chiral WTIs that look like the continuum ones modified by the addition of lattice artefacts [88]. Just as in the continuum, it is possible to define a vector current that is conserved for degenerate quark masses. Consider instead a non-singlet axial transformation. The PCAC relation in Eq. (1.83) is modified in

$$\left\langle \partial^*_\mu A^a_\mu(x) O_{\text{ext}} \right\rangle = 2m \left\langle P^a(x) O_{\text{ext}} \right\rangle + \left\langle X^a(x) O_{\text{ext}} \right\rangle, \tag{2.90}$$

where the discretized axial current and density are

$$A^a_\mu(x) = \frac{1}{2}\left[\bar\psi(x)\tau^a\gamma_\mu\gamma_5 U_\mu(x)\psi(x+a\hat\mu) + \text{h.c.}\right], \tag{2.91a}$$

$$P^a(x) = \bar\psi(x)\tau^a\gamma_5\psi(x), \tag{2.91b}$$

and X^a is a local field coming from the variation of Wilson's term

$$X^a(x) = -\frac{1}{2a}\sum_\mu \Big[\bar\psi(x)\tau^a\gamma_5 U_\mu(x)\psi(x+a\hat\mu) + \bar\psi(x+a\hat\mu)\tau^a\gamma_5 U^\dagger_\mu(x)\psi(x) + (x \to x - a\hat\mu) - 4\bar\psi(x)\tau^a\gamma_5\psi(x)\Big]. \tag{2.92}$$

Therefore, it is not possible to define an axial current that is conserved for vanishing quark masses.

In the continuum limit, chiral symmetry is recovered with a more involved renormalization pattern. In the case of the PCAC in Eq. (2.90), X^a is a $d = 4$ pseudoscalar field, so in general it mixes with $\partial^*_\mu A^a_\mu$ and P^a

$$X^a = -2\bar m P^a - (Z_A - 1)\partial^*_\mu A^a_\mu + X^a_{\text{R}}, \tag{2.93}$$

where the last term is the renormalized field, which vanishes in the continuum limit. $\bar m \sim a^{-1}$ induces the divergent additive renormalization of the quark mass, while Z_A renormalizes the axial current and can be shown to be finite. Both renormalization factors need to be determined non-perturbatively for the WTI in Eq. (1.83) to be useful.

As an alternative, we can work with a discretization of fermions that satisfies the GW relation. Ordinary chiral transformations are substituted with Lüscher's modified ones in Eq. (2.53),

which are exact at finite lattice spacing. Since $\hat{\gamma}_5$ is no more a constant matrix, some care must be taken in defining the local version of these transformations. For instance, Eq. (1.76a) is substituted by

$$\delta\psi(x) = +\mathrm{i}\tau^a \frac{1}{2} \sum_y \left[\alpha^a(x)\delta_{x,y} + \sum_z \hat{\gamma}_5(x,z)\alpha^a(z)\hat{\gamma}_5(z,y) \right] \psi(y) + \mathrm{i}\tau^a \frac{1}{2} \sum_y \left[\hat{\gamma}_5(x,y)\beta^a(y) + \beta^a(x)\hat{\gamma}_5(x,y) \right] \psi(y). \quad (2.94)$$

The massless action with GW fermions in Eq. (2.52) satisfies exact global Lüscher's symmetry. This allows to derive WTIs at finite lattice spacing that look formally identical to the continuum ones in Eqs (1.82) and (1.83). Scalar and pseudoscalar densities are defined by

$$S^a(x) = \bar{\psi}(x)\tau^a \left[\left(\mathbb{1} - \frac{\bar{a}}{2} D \right) \psi \right](x), \quad (2.95a)$$

$$P^a(x) = \bar{\psi}(x)\tau^a \gamma_5 \left[\left(\mathbb{1} - \frac{\bar{a}}{2} D \right) \psi \right](x), \quad (2.95b)$$

while the divergence of the currents is given by the expressions

$$\partial^*_\mu V^a_\mu(x) = \left(1 - \frac{am}{2} \right) \left[\bar{\psi}(x)\tau^a \left(D\hat{\gamma}_5\psi \right)(x) - (\bar{\psi} D)(x)\tau^a \left(\hat{\gamma}_5\psi \right)(x) \right], \quad (2.96a)$$

$$\partial^*_\mu A^a_\mu(x) = \left(1 - \frac{am}{2} \right) \left[\bar{\psi}(x)\tau^a (D\psi)(x) - (\bar{\psi} D)(x)\tau^a \psi(x) \right], \quad (2.96b)$$

which correctly vanish when summed over x. The existence of (partially) conserved currents is guaranteed by the lattice Poincaré lemma [89], but they are ambiguous by terms with vanishing divergence and their explicit construction is non-trivial. Explicit expressions for the massless case are provided in Ref. [90].

Using the form in Eq. (2.95) of the scalar current, the massive GW-fermion action in Eq. (2.55) can be rewritten as

$$S_{\mathrm{GWf}} = S_{\mathrm{GWf}}\Big|_{M_0=0} + (M_0)_{ij} \sum_x S^{ij}(x). \quad (2.97)$$

2.3.5. Approximate methods

Even if on the lattice the path integral is mathematically well defined, its complexity in the interacting theory is too high to hope for an analytical evaluation in closed form. Therefore, one has to resort to approximate methods.

If the coupling parameter is small, it is possible to do a *weak coupling expansion* which is completely analogous to perturbation theory in the continuum. Feynman rules can be derived. This requires the gauge freedom to be fixed, which leads to the introduction of ghosts. In the thesis, we will not present lattice Feynman rules, which are much more involved than the continuum ones. Remarkably, the existence of the continuum limit is rigorously established to all orders [91–93]. This makes perturbation theory a highly valuable tool on the lattice: from a practical point of view, it allows to match results of non-perturbative Monte Carlo simulations on the lattice, in a regime in which lattice perturbation theory is valid, to perturbative calculations

done with other regularizations more commonly used in the continuum. From a theoretical point of view, it is important to give structural insight into the approach to the true continuum limit, as described in Section 2.5.

On the exact opposite of perturbation theory, if dynamical fermions are not present, it is possible to perform an expansion in powers of the inverse coupling parameter, which resembles the high temperature expansion in statistical mechanics. This is the so-called *strong coupling expansion*, which is of historical importance since it was used by Wilson [19] to show analytically that the potential of a quark-antiquark pair grows linearly with their distance, leading to confinement, and that the theory has a mass gap. Unfortunately, the continuum limit cannot be reached in the strong-coupling expansion because it has to be obtained with $g \to 0$. Therefore, Wilson's argument is not sufficient to prove that YM theory confines and has a mass gap in the continuum.

2.4. Monte Carlo simulations

The most powerful method to obtain numerical results from the path integral in Eq. (2.72) is suggested by the similarity between quantum field theory and statistical mechanics. Indeed, the path integral of a quantum field theory defined in four-dimensional Euclidean spacetime is completely analogous to the partition function of a classical statistical mechanics system at equilibrium in four space dimensions. The typical analogy is with a magnetic system: the action of the field theory, as a spacetime integral of Lagrangian density $\mathcal{L}(x)$ corresponds to the space integral of the statistical system Hamiltonian $\mathcal{H}(x)$. The source field $J(x)$ corresponds to an external source like the external magnetic field $H(x)$. The quantum fluctuations of the field $\phi(x)$ around the classical configurations correspond to the statistical fluctuations of the spin field $s(x)$ around its equilibrium value. Part of quantum field theory terminology, such as 'partition function' and 'free energy', is directly borrowed from statistical mechanics. When the quantum field theory is defined on a lattice the analogy is even stronger. Indeed, the spin field $s(x)$ of a magnetic system is naturally defined, in three dimensions, on the lattice of atoms in a crystal, thus a lattice cut-off is natural.

Thanks to this correspondence, a general correlation function in quantum field theory can be interpreted as the expectation value of the corresponding fields over an ensemble of configurations. Denoting with ϕ the elementary fields of the theory,

$$\langle O \rangle = \frac{1}{\mathcal{Z}} \int \mathcal{D}[\phi]\, O[\phi] \mathrm{e}^{-S[\phi]} = \int \mathrm{d}P[\phi]\, O[\phi], \tag{2.98}$$

where the probability density $\mathrm{d}P[\phi]$ is proportional to the *Boltzmann factor* $\mathrm{e}^{-S[\phi]}$

$$\mathrm{d}P[\phi] = \frac{\mathcal{D}[\phi]\, \mathrm{e}^{-S[\phi]}}{\int \mathcal{D}[\phi]\, \mathrm{e}^{-S[\phi]}}. \tag{2.99}$$

It is now clear that, if we want to hope to give a statistical meaning to a quantum field theory, the formulation in Euclidean spacetime instead of Minkowski spacetime is needed to have a Boltzmann factor that can be interpreted as a probability distribution. However, this is not always sufficient: very simple deformations of QCD, such as a baryon chemical potential or

a non-zero θ-angle, result in a non-zero imaginary part of the Euclidean action. These imply an oscillatory part in the Boltzmann factor which does not necessarily spoil the statistical interpretation, but is responsible for a *sign problem* that hinders any practical application.

2.4.1. Importance sampling

Even if an analytical evaluation of Eq. (2.98) is out of question, it can be estimated numerically. Simple numerical quadrature is not affordable. In fact, Eq. (2.98) requires $32(V/a)^4$ integrations for an SU(3) gauge field: 8 components times 4 links times $(V/a)^4$ lattice points. Even for a small 10^4 lattice, this amounts to 320 000 integrations. However, with Monte Carlo methods it is possible to estimate high-dimensional integrals like the one in Eq. (2.98) from a set of configurations randomly generated.

The probability density is negligible on the vast majority of the configurations. Indeed, only a small subset of configuration space has a small action and contributes significantly to the average in Eq. (2.98).

To overcome this, it is convenient to apply *importance sampling*. Typically using Markov chain methods, configurations are generated directly with the probability distribution in Eq. (2.99). In a first phase of *thermalization*, a number of configurations are discarded to get closer to equilibrium. The length of this phase depends on the actual algorithm, on the action of the theory, on the observables measured and also on the bare parameters of the theory. Once equilibrium is achieved, the sequence $\{\phi_i\}$ of n configurations is a representative sample of the path integral ensemble. An estimate for the ensemble average in Eq. (2.98) is given by

$$\bar{O} = \frac{1}{n} \sum_{i=1}^{n} O[\phi_i]. \tag{2.100}$$

However, configurations produced by Markov chain Monte Carlo methods are not statistically independent but correlated, and this has to be taken into account when estimating the statistical error of the estimator in Eq. (2.100).

The actual Markov chain Monte Carlo algorithm used depends on the details of the action. A historical prototypical algorithm, proposed in 1953 and in principle applicable to any system, is the *Metropolis algorithm* [94]. An iteration of this algorithm is composed of two steps: in the first a *candidate configuration* is chosen according to some a priori selection probability, in the second the candidate configuration is accepted as the new configuration according to some acceptance probability. If the candidate is rejected, the new configuration in the Markov chain is set equal to the previous one.

In general, we distinguish between a *local* and a *global* Markov chain Monte Carlo. In local algorithms, the candidate configuration differs locally from the previous one. It make sense to update a configuration locally only if the computation of the action is a local operation too. Usually, many local updates are chained in a sweep over the whole lattice. In global algorithms, a configuration that is globally different from the previous one is proposed as a candidate.

2.4.2. The Monte Carlo for Yang–Mills theory

In YM theory, the action is local. Changing a single link variable, the action is changed by an amount

$$\Delta S = S_{\mathrm{loc}}[U'_\mu(x)] - S_{\mathrm{loc}}[U_\mu(x)], \qquad S_{\mathrm{loc}}[U_\mu(x)] = -\frac{\beta}{N_{\mathrm{c}}} \operatorname{Re}\operatorname{tr}\big\{U_\mu(x) S_\mu(x)\big\}, \tag{2.101}$$

where $S_\mu(x)$ is the sum of six *staples*

$$S_\mu(x) = \sum_{\nu\neq\mu} \Big[U_\nu(x + a\hat{\mu}) U^\dagger_\mu(x + a\hat{\nu}) U^\dagger_\nu(x) + U^\dagger_\nu(x + a\hat{\mu} - a\nu) U^\dagger_\mu(x - a\hat{\nu}) U_\nu(x - a\hat{\nu}) \Big]. \tag{2.102}$$

The computation of the action change is a $\mathcal{O}(1)$ operation with respect to the lattice volume. Therefore, local algorithms work well in YM theory.

In the very simple *heat bath* algorithm [95], the updated link variable is extracted randomly according to the probability distribution

$$\mathrm{d}P[U_\mu(x)] = \mathrm{d}U_\mu(x) \exp\left\{ \frac{\beta}{N_{\mathrm{c}}} \operatorname{Re}\operatorname{tr}\big\{U_\mu(x) S_\mu(x)\big\} \right\}, \tag{2.103}$$

where $S_\mu(x)$ is the sum of staples. The algorithm's name is due to the fact that this probability distribution brings $U_\mu(x)$ in equilibrium with a 'heat bath reservoir' composed by links in the neighbourhood. Differently from the Metropolis algorithm, no acceptance test is needed and the configuration is always changed.

In the SU(2) YM theory case, it is easy to generate a new link according to the probability distribution in Eq. (2.103). In the case of other gauge groups, a prescription to perform a heat bath update was given by Cabibbo and Marinari [96]: update in sequence each SU(2) subgroup. In SU(N_{c}) YM theory, this means updating $N_{\mathrm{c}}(N_{\mathrm{c}} - 1)/2$ subgroups.[2] To generate a new configuration in the Markov chain, the local heat bath update is applied on each link in the lattice in turn, until the entire lattice is swept.

To speed up the motion trough configuration space, the Cabibbo–Marinari algorithm combines a heat bath sweep with one or more sweeps of the *over-relaxation* algorithm [97]. An over-relaxation is a local update which works deterministically, changing the link variable with a new one that has the same action. By itself, this algorithm is not ergodic since it samples only a subspace of constant action of configuration space, namely the microcanonical ensemble in statistical mechanics terms.

2.4.3. The Monte Carlo with fermions

The partition function in Eq. (2.72) contains fermion fields represented by Grassmann numbers, but these anticommuting objects cannot be simulated with Monte Carlo methods. Nevertheless,

[2] Albeit it is sufficient to update $N_{\mathrm{c}} - 1$ subgroups to obtain ergodicity, to minimize the autocorrelation of configurations it is convenient to update all of them.

since the fermion action is bilinear in fermion fields, it is always possible to evaluate the path integral over fermion fields, leaving only the path integral over gauge fields. Using the rules in Section A.4, we have

$$\int \mathcal{D}[\bar{\psi},\psi]\, \mathrm{e}^{-S_{\mathrm{wf}}[\bar{\psi},\psi,U]} = \int \mathcal{D}[\bar{\psi},\psi]\, \mathrm{e}^{-\bar{\psi} D_M[U]\psi} = \mathrm{Det}\, D_M[U], \tag{2.104}$$

where $\mathrm{Det}\, D_M[U] = \mathrm{Det}\{D[U] + M_0\}$ is the *fermion determinant*. At least for the cases in which we are interested, the fermion determinant is a well-behaved positive functional of the background gauge field U. Therefore, it is possible to interpret it as a proper probability distribution for the gauge field and perform numerical simulations. Writing the fermion determinant as an effective (non-local) action, the partition function in Eq. (2.72) is rewritten in a completely-equivalent purely-bosonic form

$$\mathcal{Z} = \int \mathcal{D}[U]\, \mathrm{e}^{-S_{\mathrm{w}}[U] + \ln \mathrm{Det}\, D_M[U]}, \tag{2.105}$$

and the expectation value of a field $O[U]$ depending only on the gauge field is given by

$$\langle O \rangle = \frac{1}{\mathcal{Z}} \int \mathcal{D}[U]\, O[U] \mathrm{e}^{-S_{\mathrm{w}}[U] + \ln \mathrm{Det}\, D_M[U]}. \tag{2.106}$$

The Dirac operator is a sparse matrix of dimension $4N_{\mathrm{c}}(V/a^4) \times 4N_{\mathrm{c}}(V/a^4)$, thus a direct computation of its determinant is out of question. For one or more doublets of mass-degenerate fermions, the determinant can be evaluated introducing in the theory complex scalar fields ϕ with Dirac and colour indices, called *pseudofermions*. Neglecting an irrelevant normalization constant, the partition function is then rewritten as

$$\mathcal{Z} = \int \mathcal{D}[U]\, \mathrm{e}^{-S_{\mathrm{w}}[U] - S_{\mathrm{pf}}[\phi,\phi^\dagger,U]}, \qquad S_{\mathrm{pf}}[\phi,\phi^\dagger,U] = \phi^\dagger \left(D_M^\dagger D_M\right)^{-1}[U]\phi. \tag{2.107}$$

Even if we are able to rewrite the QCD path integral as a path integral over bosonic fields only, there is a price to pay: the pseudofermion action involves the *inverse* of the Dirac operator and it is thus non-local. As a first consequence, local Monte Carlo algorithm are not well-suited to be applied to the path integral in Eq. (2.107). Indeed, when a single link is modified the action has to be recomputed globally. This is an $\mathcal{O}(V)$ operation, which combined with the $\sim V$ number of links results in an unbearable $\mathcal{O}(V^2)$ scaling.

The partition function in Eq. (2.107) can be sampled efficiently with non-local Markov chain Monte Carlo algorithms. These generate a candidate configuration which differs globally with respect to the previous one. The tricky part is to modify the configuration in a way that the acceptance rate is not too small nor too big.

A number of variants of the Metropolis algorithm have been developed over the years to work with non-local actions and be efficient with the non-local fermion determinant. The most advanced ones are variants of the *hybrid Monte Carlo* (HMC) algorithm [98]. In these algorithms, the candidate configuration is obtained integrating a *Molecular Dynamics (MD)* Hamiltonian in a fictitious MD time. Then, a final acceptance step is included to account for errors in the numerical integration.

Nevertheless, these algorithms are 1–2 orders of magnitude more expensive to simulate with respect to the Cabibbo–Marinari algorithm applied to YM theory. The solution of the Dirac equation, needed to evaluate the pseudofermion action, is by itself very expensive, in particular for light quark masses. At the Lattice 2001 conference in Berlin, the cost of $N_{\mathrm{f}} = 2$ QCD simulations, i.e. with a dynamical doublet of mass-degenerate light quarks, was estimated to be described by [99]

$$\text{cost} \approx 2.8 \left[\frac{\#\,\text{conf}}{1000}\right] \left[\frac{M_\pi/M_\rho}{0.6}\right]^{-6} \left[\frac{L}{3\,\text{fm}}\right]^{5} \left[\frac{a^{-1}}{2\,\text{GeV}}\right]^{7} \text{Tflops year}, \tag{2.108}$$

where $M_\pi/M_\rho \sim \sqrt{am}$ encodes the cost of going closer to the chiral limit. Small lattice spacings and quark masses were behind an insurmountable 'Berlin wall' [100]. For this reason, often simulations of *quenched QCD* were performed. This consists in neglecting the fermion determinant completely and generating the configurations distributed according to the YM action. The resulting theory, quenched QCD, is a non-unitary QFT in which virtual quark loops are not present. Clearly, this introduces a significant systematic error, dependent on the specific observable and difficult to estimate with precision.

Starting from the first years of the new millennium, thanks both to the large availability of fast computers and to fundamental advances in algorithmic techniques, Monte Carlo simulations of full QCD have become common. Among the many advancements that made this possible, we cite

- Sexton–Weingarten multiple time-step integration [101], to accelerate the MD integration treating different contributions with different integration step sizes;
- Hasenbusch factorization [102], to split contributions of different frequencies to the fermion determinant in independent pseudofermion actions, and mass preconditioning [103], which combines Hasenbusch factorization with Sexton–Weingarten multiple time-step integration;
- domain decomposition [104, 105] and low-modes deflation [106, 107], to solve efficiently the Dirac equation for small quark masses;
- the rational HMC algorithm [108], to simulate a single quark flavour;
- the introduction of open boundary conditions [109], to avoid the freezing of topological modes at small lattice spacings.

Thanks to these algorithmic advancements, the cost of a $N_{\mathrm{f}} = 2$ QCD simulation has been reduced significantly. See for instance estimates in Refs [110, 111]. Remarkably, with respect to the Berlin 2001 formula, the dependence on quark masses has basically been eliminated. Moreover, the prefactor itself has been made more than two orders of magnitude smaller.

2.4.4. Fermion correlators

Using interpolating fields of the form in Eq. (2.87), the mass of the meson can be extracted from the long range behaviour of the correlator

$$C^a_\Gamma(x_4 - y_4) = \sum_{\vec{x}} \left\langle O^a_\Gamma(x) O^a_\Gamma(y) \right\rangle. \tag{2.109}$$

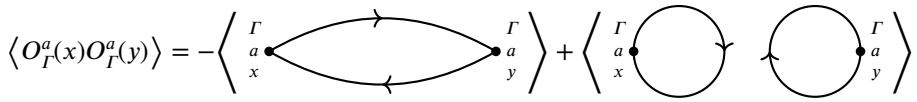


Figure 2.2.: Graphical representation of Wick's theorem application to Eq. (2.113).

However, in Monte Carlo simulations, quark bilinears and their correlators are not computed directly. We have seen in Section 2.4.3 that the fermion path integral is computed analytically. The resulting path integral is purely bosonic, the partition function and correlation functions of gluon fields are given in Eqs (2.105) and (2.106) respectively.

The simplest expectation value involving fermion fields is the *quark propagator* $\langle \psi(y)\bar{\psi}(x)\rangle$. The fermionic part of the path integral can be computed analytically using the rules of Grassmass algebra in Section A.4. For a given background gauge field configuration, this is the basic *Wick's contraction* denoted by a line joining the two fields, as in

$$\langle \psi(y)\bar{\psi}(x)\rangle_{\mathrm{q}} = \overbrace{\psi(y)\bar{\psi}(x)} \equiv S(y,x), \tag{2.110}$$

where $\langle \bullet \rangle_{\mathrm{q}}$ denotes path integral over quark fields only. The quark propagator $S(y,x)$ at fixed U is the inverse of the massive Dirac operator

$$S[U](y,x) = D_M^{-1}[U](y,x) = \left[D[U](x,y) + M_0\delta_{xy}\right]^{-1}. \tag{2.111}$$

It inherits γ_5-hermiticity from the Dirac operator, which relates the forward and backward propagators with

$$\gamma_5 S(y,x)\gamma_5 = S(x,y)^\dagger. \tag{2.112}$$

Usually, correlation functions with more than two fermion fields are needed. To these we can apply the well-known result of QFT, *Wick's theorem*, that in the path-integral formulation is just a theorem of probability theory. The expectation value of an arbitrary string of quark and antiquark fields is reduced to the sum of the product of all possible Wick's contractions between two fields, of the form in Eq. (2.110). Some care must be taken to the minus signs arising from the interchange of anticommuting fields. As a result, the expectation value of any number of quark and antiquark fields reduces to the sum of expectation values of products of quark propagators, to be computed with the purely bosonic partition function in Eq. (2.105). This generalizes the expectation value in Eq. (2.106) to observables depending on quark fields.

As an example, consider the meson correlator in Eq. (2.109). The application of Wick's theorem gives

$$\begin{aligned}
\left\langle O_\Gamma^a(x)O_\Gamma^a(y)\right\rangle &= \left\langle \left[\bar{\psi}(x)\tau^a\Gamma\psi(x)\right]\left[\bar{\psi}(y)\tau^a\Gamma\psi(y)\right]\right\rangle \\
&= \langle \overbrace{\bar{\psi}(x)\tau^a\Gamma\psi(x)\bar{\psi}(y)\tau^a\Gamma\psi(y)}\rangle + \langle \overbrace{\bar{\psi}(x)\tau^a\Gamma\psi(x)}\overbrace{\bar{\psi}(y)\tau^a\Gamma\psi(y)}\rangle \\
&= -\langle \mathrm{tr}\{\tau^a\Gamma S(x,y)\tau^a\Gamma S(y,x)\}\rangle + \langle \mathrm{tr}\{\tau^a\Gamma S(x,x)\}\,\mathrm{tr}\{\tau^a\Gamma S(y,y)\}\rangle.
\end{aligned} \tag{2.113}$$

In general, Wick's contractions result in two classes of contributions. The first in the last line of Eq. (2.113) is the *connected* contribution, because it contains a single trace and can be pictured

as the expectation value of a single quark line which connects y with x and goes back to y. The second contribution is the *disconnected* one, which consists in the correlation between two different traces and can be pictured as the expectation value of two disconnected fermion loops, one around y and the other around x. This graphical representation is shown in Figure 2.2.

The presence of both contributions depends on the flavour structure one is interested in. Consider the $a \neq 0$ case in which O^a_Γ transforms under the adjoint representation of the $\mathrm{SU}(N_\mathrm{f})_F$ flavour group. This couples to the non-singlet mesons in the theory with N_f flavours of mass-degenerate quarks. Then Eq. (2.113) reduces to

$$\left\langle O^a_\Gamma(x) O^a_\Gamma(y) \right\rangle = -\frac{1}{2}\langle \mathrm{tr}\{\Gamma S(x,y)\Gamma S(y,x)\}\rangle = -\frac{1}{2}\left\langle \mathrm{tr}\left\{\gamma_5 \Gamma S(x,y)\Gamma\gamma_5 S(x,y)^\dagger\right\}\right\rangle. \quad (2.114)$$

Thus, in the case of flavour non-singlet mesons disconnected diagrams are not present. On the contrary, in the $a = 0$ flavour-singlet case, so that O^0_Γ couples to flavour-singlet mesons, we have

$$\left\langle O^0_\Gamma(x) O^0_\Gamma(y) \right\rangle = -\frac{1}{2}\langle \mathrm{tr}\{\Gamma S(x,y)\Gamma S(y,x)\}\rangle + \frac{N_\mathrm{f}}{2}\langle \mathrm{tr}\{\Gamma S(x,x)\}\,\mathrm{tr}\{\Gamma S(y,y)\}\rangle. \quad (2.115)$$

This is a linear combination of the same connected contribution of the non-singlet case plus a disconnected contribution.

Beyond mesons, the low-energy spectrum is composed by baryon states. Baryons interpolating fields are given by colour-singlet combinations of N_c quarks (or antiquarks) fields with the proper Lorentz and flavour transformation properties. For instance, in QCD with $N_\mathrm{f} = 2$ mass-degenerate u and d quarks, the *nucleon* is the isospin doublet of the $J^P = 1/2^+$ baryons: the proton and the neutron. A Dirac-spinor field which interpolates for the nucleon is

$$N(x) = P_+ \psi_u(x) \Gamma^A \left[\psi_u^T(x) \Gamma^B \psi_d(x)\right], \qquad \bar{N}(x) = \left[\bar{\psi}_d(x)\Gamma^B \bar{\psi}_u^T(x)\right]\bar{\psi}_u(x)\Gamma^A P_+, \quad (2.116)$$

where $P_+ = (1+\gamma_4)/2$ selects the positive parity states and spin $J = 1/2$ is realized with three possible choices of (Γ^A, Γ^B): $(\mathbb{1}, C\gamma_5)$, (γ_5, C) or $(\mathbb{1}, \mathrm{i}\gamma_4 C\gamma_5)$, with C the charge conjugation matrix. The mass of the nucleon is extracted from the correlator

$$C_N(x_4 - y_4) = \sum_{\vec{x}} \left\langle \bar{N}(x) N(y) \right\rangle, \quad (2.117)$$

where the open Dirac indices of N and $\bar{N}$ are contracted. Applying Wick's theorem, the correlator reduces to a sum of products of quark propagators. In this specific case, only two connected contributions are present.

2.5. The continuum limit

Monte Carlo simulations give results about the lattice-regularized QFT in the form of dimensionless combinations, with no reference to any physical scale. For instance, consider an observable E with positive mass dimension $d = [E]$ that for simplicity does not depend on quark masses. Given a lattice discretization E_L, we can perform a Monte Carlo simulation to obtain, as a function of the bare coupling g_0,

$$\left[a^d E_\mathrm{L}\right](g_0) \quad \Rightarrow \quad E_\mathrm{L}(a, g_0) = a^{-d}\left[a^d E_\mathrm{L}\right](g_0). \quad (2.118)$$

Renormalization group methods constrain E to be proportional to the Λ-parameter times a function of g as in Eq. (1.34). On the lattice, this holds modulo lattice artefacts

$$E_{\mathrm{L}} = \Lambda^d \left\{ \tilde{E}(g) + \mathcal{O}(a^n) \right\}, \tag{2.119}$$

where the order n depends on the discretization of the theory and of the observable. In principle, this is sufficient to obtain E_{L} in physical units. However, we cannot study hadronic observables relying on renormalization in a perturbative scheme.

The most direct way to fix the scale on the lattice is to define a hadronic renormalization scheme. To this purpose, bare quark masses are first tuned at fixed bare coupling using a sufficient number of scales, e.g. M_π/f_π, M_K/f_K, etc., until they match their physical value[3]. This defines a line of constant physics. Then, g_0 is varied and the lattice spacing in function of g_0 is obtained from a *reference scale* f_0

$$a(g_0) = \frac{\left[a f_0\right](g_0)}{f_0\big|_{\mathrm{phys.}}}, \tag{2.120}$$

where f_0 can be any physical scale with the dimension of a mass, for instance a meson decay constant or a hadron mass.

In general, it is convenient to choose a reference scale that can be computed easily to high precision on the lattice, with a well-know continuum limit. For instance, the *Sommer parameter* r_0 has been a typical reference scale for many years. It is defined in Ref. [112] implicitly by

$$r^2 F(r)\Big|_{r=r_0} = 1.65, \tag{2.121}$$

where $F(r)$ is the force between static quarks at distance r. The value 1.65 is chosen such that $r_0 \approx 0.5\,\mathrm{fm}$ in phenomenological models of the QCD potential.

Especially with dynamical quark, r_0 has proven difficult to estimate with high precision. A replacement proposed in recent years is the t_0 scale [29], which will be defined in Eq. (5.6) after introducing the YM gradient flow in Section 4.1.

After choosing a hadronic renormalization scheme, Eq. (2.120) can be inverted in its range of validity to give

$$g_0(a f_0). \tag{2.122}$$

This is the lattice version of dimensional transmutation, which allows one to tune the bare coupling g_0 to realize a specific lattice spacing. Its connection in physical units is given from the direct measurement of f_0, when this is possible.

2.6. Symanzik's effective theory

At first sight, it is not guaranteed that a theory defined on a spacetime lattice, even arbitrarily fine, is a well-behaved approximation of the theory in the continuum. Indeed, relativistic QFTs

[3] or some other value, e.g. predicted by χPT, if one is interested in non-physical quark masses.

are highly symmetric: they have specific transformation properties under continuous spacetime translations and Lorentz transformations. On the other hand, this large group of symmetries is lost on the lattice. The discretized theory remains symmetric only under a discrete subgroup of ISO(4), which accounts for discrete spacetime translation and discrete rotational symmetry of a four-dimensional hypercube. It is thus natural to ask if the $a \to 0$ limit of a lattice QFT correctly leads to the relativistic QFT.

There is numerical evidence from simulations that the answer is positive: the lattice theory approaches criticality, at which the Lorentz symmetry is dynamically restored, when the bare coupling approaches the value $g^* = 0$. However, lattice simulations are limited to small but finite values of lattice spacing a.

Thanks to the work by Symanzik in a series of papers [113, 114], it is possible to derive strong analytical results on how the lattice theory approaches the continuum limit. Symanzik's approach is based on an effective field theory (EFT) in the continuum, known as *Symanzik's effective theory* (SET), which describes the lattice theory at small energy scales with respect to the cut-off

$$Q \ll a^{-1}. \tag{2.123}$$

To this purpose, the action of SET is built including all the local fields with the symmetry of the lattice theory. This includes composite fields of any canonical dimension d that are not necessarily invariant under continuous SO(4) spacetime transformations, but only under the hypercubic subgroup. The result is an EFT that approximates, in the continuum, the low-energy dynamics of the lattice theory. Each field monomial is multiplied by a dimensionless coefficient and the appropriate power of the cut-off to have a dimensionless contribution to the action. Fields are then ordered according to their order in a

$$S_{\text{SET}} = \int \mathrm{d}^4x \left[\mathcal{L}^{(0)} + a\mathcal{L}^{(1)} + a^2\mathcal{L}^{(2)} + \cdots\right]. \tag{2.124}$$

What is remarkable is that, at the lowest order, SET is given just by the renormalizable action of ordinary QCD in the continuum

$$\int \mathrm{d}^4x\, \mathcal{L}^{(0)} = S_{\text{g}} + S_{\text{f}}. \tag{2.125}$$

According to Wilson's interpretation of renormalization, this is the renormalizable theory that emerges when the cut-off is sent to infinity in SET, and thus in the lattice theory at long distances. The continuous ISO(4) Lorentz symmetry is an emergent symmetry of the lattice theory at low energy scales.

The term $\mathcal{L}^{(1)}$ in Eq. (2.124) is a collection of $d = 5$ fields and gives the first lattice correction to the continuum theory. In the case of lattice QCD, a complete list of $d = 5$ scalar fields with

the symmetries of the lattice theory is

$$O_1 = \bar{\psi} i \sigma_{\mu\nu} F_{\mu\nu} \psi, \tag{2.126a}$$

$$O_2 = \bar{\psi} D_\mu D_\mu \psi + \bar{\psi} \overline{D}_\mu \overline{D}_\mu \psi, \tag{2.126b}$$

$$O_3 = M_0 \operatorname{tr}\{F_{\mu\nu} F_{\mu\nu}\}, \tag{2.126c}$$

$$O_4 = M_0 \Big[\bar{\psi} \gamma_\mu D_\mu \psi + \bar{\psi} \overleftarrow{D}_\mu \gamma_\mu \psi\Big], \tag{2.126d}$$

$$O_5 = M_0^2 \bar{\psi} \psi. \tag{2.126e}$$

Given the presence in the theory of a $d = 1$ parameter M_0, some of these fields are obtained from lower-dimension fields. Moreover, they are not all independent: applying the equations of motion it is possible to derive relations between them

$$O_1 - O_2 + 2O_5 = 0, \qquad O_4 + 2O_5 = 0, \tag{2.127}$$

which allows to eliminate two fields, e.g. O_2 and O_4.

Similarly, any renormalized lattice field O_{L} corresponds in the effective theory to an expansion in powers of a

$$O_{\mathrm{SET}} = O^{(0)} + a O^{(1)} + a^2 O^{(2)} + \cdots, \tag{2.128}$$

where the $O^{(i)}$ have the appropriate dimension and the same symmetries of the lattice field they represent. For instance, in the case of the $d = 3$ flavour non-singlet axial current

$$A^a_\mu(x) = \bar{\psi}(x) \gamma_\mu \gamma_5 \tau^a \psi(x), \qquad \text{for } a \neq 0, \tag{2.129}$$

the leading $\mathcal{O}(a)$ corrections are given by the $d = 4$ axial vector fields with flavour structure

$$(O_6)^a_\mu = \bar{\psi} \gamma_5 \tau^a \sigma_{\mu\nu} D_\nu \psi - \bar{\psi} \overline{D}_\mu \sigma_{\mu\nu} \gamma_5 \tau^a \psi, \tag{2.130a}$$

$$(O_7)^a_\mu = \partial_\mu P^a = \partial_\mu \big[\bar{\psi} \gamma_5 \tau^a \psi\big], \tag{2.130b}$$

$$(O_8)^a_\mu = M_0 A^a_\mu = M_0 \bar{\psi} \gamma_\mu \gamma_5 \tau^a \psi. \tag{2.130c}$$

Again, the use of equations of motion allows us to reexpress O_6 as a linear combination of O_7 and O_8.

The effective continuum theory is matched to the lattice theory, at a given order, comparing corresponding correlation functions. For instance, consider a generic renormalized connected n-point correlation function of a local lattice field $O_{\mathrm{L}}(x)$

$$G^n_{\mathrm{R}}(x_1, \ldots, x_n) = Z^n_O G^n_0(x_1, \ldots, x_n) = Z^n_O \langle O_{\mathrm{L}}(x_1) \cdots O_{\mathrm{L}}(x_n) \rangle_{\mathrm{con}}, \tag{2.131}$$

where for simplicity we keep all x_i at physical distance and we have supposed that $O_{\mathrm{L}}(x)$ does not mix with other fields. In SET, the same correlation function is given by

$$\begin{aligned} G_n(x_1, \ldots, x_n) = {} & \big\langle O^{(0)}(x_1) \cdots O^{(0)}(x_n) \big\rangle_{\mathrm{con}} \\ & - a \int \mathrm{d}^4 y \, \big\langle O^{(0)}(x_1) \cdots O^{(0)}(x_n) \mathcal{L}^{(1)}(y) \big\rangle_{\mathrm{con}} \\ & + a \sum_{i=1}^{n} \big\langle O^{(0)}(x_1) \cdots O^{(1)}(x_i) \cdots O^{(0)}(x_n) \big\rangle_{\mathrm{con}} + \mathcal{O}\big(a^2\big), \end{aligned} \tag{2.132}$$

where the expectation values are taken with respect to the action $\mathcal{L}^{(0)}$. In general, the integral over y diverges when $y \to x_i$. A subtraction prescription must be supplied, however, its precise definition is unimportant since the arbitrariness correspond to a local insertion at $y = x_i$, thus to a redefinition of the field $O^{(1)}(x)$.

Not all the lattice effects are accounted for by the explicit factor of a in Eq. (2.132). Indeed, the coefficients of the fields contributing to $O^{(1)}$ and $\mathcal{L}^{(1)}$ can have a slowly-varying dependence on a. If computed in perturbation theory, this dependence is a polynomial in $\ln a$. Moreover, additional lattice effects emerge depending on the specific correlator from short-distance singularities if some of the x_i are allowed to become coincident.

2.6.1. Lattice improvement

While the general structure of continuum SET is determined just by the field content and the lattice symmetries, its coefficients depend on the lattice action used in the matching procedure. It is thus natural to ask if, using the freedom in the definition of a lattice action with the correct naïve continuum limit, one can obtain a SET with smaller or even vanishing coefficients to a given order. For an observable measured in Monte Carlo simulation at finite lattice spacing, this implies higher order lattice discretization effects in Eq. (2.119). In turn, this means a smaller finite lattice systematics, which can be traded for a cheaper simulation at a coarser lattice spacing.

The answer to this question is positive: Symanzik proved that it is in principle possible to design a lattice action that matches a SET with vanishing coefficients up to $\mathcal{O}(a^n)$. This is done adding to the lattice action irrelevant field monomials up to $\mathcal{O}(a^n)$ with properly chosen coefficients. This order-by-order iterative procedure goes under the name of *Symanzik's improvement program*.

From the list in Eq. (2.126a), it is evident that no $d = 5$ fields would be there in YM theory to contribute to $\mathcal{L}^{(1)}$. This is true also for a general bosonic observable. Therefore, lattice YM theory given by Wilson's plaquette action in Eq. (2.24) is naturally $\mathcal{O}(a)$-improved and lattice artefacts emerge as an $\mathcal{O}(a^2)$ effect. Nevertheless, it is possible to improve the lattice pure gauge theory to $\mathcal{O}(a^2)$: a lattice action of this type is the *Lüscher–Weisz action* [115, 116].

Here we concentrate on the possibility of $\mathcal{O}(a)$-improving the action of Wilson fermions. Assuming that the fields O_2 and O_4 have been eliminated using equations of motion, we improve the lattice action adding a term

$$S_{\mathrm{L}}^{(1)} = a \sum_x \big[c_1 O_{\mathrm{L}1}(x) + c_3 O_{\mathrm{L}3}(x) + c_5 O_{\mathrm{L}5}(x) \big], \tag{2.133}$$

where $O_{\mathrm{L}i}$ are lattice discretizations of O_i. The exact choice of the discretization is not important, since the ambiguities are higher-order in a. We choose to represent $\mathrm{tr}\{F_{\mu\nu}F_{\mu\nu}\}$ and $\bar\psi\psi$ with the plaquette and the local scalar density that already appear in the unimproved lattice action in Eqs (2.24) and (2.40). Therefore, the terms proportional to $O_{\mathrm{L}3}$ and $O_{\mathrm{L}5}$ just amount to a rescaling of the bare coupling and mass by factors of the form $1 + \mathcal{O}(aM_0)$. We are left with adding $O_{\mathrm{L}1}$, for instance

$$S_{\mathrm{L}}^{(0)} + S_{\mathrm{L}}^{(1)} = S_{\mathrm{W}} + S_{\mathrm{Wf}} + a c_{\mathrm{SW}} \sum_x \bar\psi(x) \frac{1}{4} \sigma_{\mu\nu} F_{\mathrm{L}\mu\nu}(x) \psi(x), \tag{2.134}$$

where $F_{\mathrm{L}\mu\nu}$ is a lattice discretization of the field strength tensor. A symmetric definition is given by [117, 118]

$$F_{\mathrm{L}\mu\nu} = -\frac{\mathrm{i}}{8a^2}\big[Q_{\mu\nu}(x) - Q_{\nu\mu}(x)\big], \tag{2.135}$$

with

$$\begin{aligned} Q_{\mu\nu}(x) = {} & U_\mu(x)U_\nu(x+a\hat{\mu})U_\mu^\dagger(x+a\hat{\nu})U_\nu^\dagger(x) \\ & + U_\nu(x)U_\mu^\dagger(x-a\hat{\mu}+a\hat{\nu})U_\nu^\dagger(x-a\hat{\mu})U_\mu(x-a\hat{\mu}) \\ & + U_\mu^\dagger(x-a\hat{\mu})U_\nu^\dagger(x-a\hat{\mu}-a\hat{\nu})U_\mu(x-a\hat{\mu}-a\hat{\nu})U_\nu(x-a\hat{\nu}) \\ & + U_\nu^\dagger(x-a\hat{\nu})U_\mu(x-a\hat{\nu})U_\nu(x+a\hat{\mu}-a\hat{\nu})U_\mu^\dagger(x). \end{aligned} \tag{2.136}$$

This improvement term first appeared in a paper by Sheikholeslami and Wohlert [119] and the coefficient c_{SW} is thus called *Sheikholeslami–Wohlert (SW) coefficient*.

It is important to stress that c_{SW} is not a new free parameter of the theory: $\mathcal{O}(a)$-improvement is obtained for a specific value of c_{SW}, which is a function of the bare parameters. As such, its fixing does not require an additional physical observable to be traded for a renormalization condition, but it is based only on consistency relations of the $\mathcal{O}(a)$-improved theory.[4] At tree-level in perturbation theory $c_{\mathrm{SW}} = 1$ [119], while at one-loop [120]

$$c_{\mathrm{SW}}(g_0) = 1 + 0.2659(1)g_0^2 + \mathcal{O}\big(g_0^4\big). \tag{2.137}$$

However, to actually achieve $\mathcal{O}(a)$-improvement at the non-perturbative level, the value of c_{SW} must be tuned appropriately, which is highly non-trivial. Methods have been developed to compute the proper value of c_{SW}. For instance, in the $N_{\mathrm{f}} = 2$ theory, it is well described by [121]

$$c_{\mathrm{SW}}(g_0) \simeq \frac{1 - 0.454g_0^2 - 0.175g_0^4 + 0.012g_0^6 + 0.045g_0^8}{1 - 0.720g_0^2}, \qquad \text{for } g_0 < 1.074 \quad (\beta > 5.2), \tag{2.138}$$

while in the $N_{\mathrm{f}} = 3$ theory [122]

$$c_{\mathrm{SW}}(g_0) \simeq \frac{1 - 0.194785g_0^2 - 0.110781g_0^4 - 0.0230239g_0^6 + 0.137401g_0^8}{1 - 0.460685g_0^2}, \qquad \text{for } g_0 < 1.074 \quad (\beta > 5.2). \tag{2.139}$$

[4] It is convenient to fix c_{SW} through consistency relations in the chiral limit. In this way, the SW coefficient is independent of quark masses and it is a function of the gauge coupling only.

3. The flavour-singlet sector

The main focus of the thesis is on the subgroup of chiral transformations that acts evenly on different flavours, i.e. on the *flavour-singlet sector*. These are two phase rotations of the fermion fields $\mathrm{U}(1)_L \times \mathrm{U}(1)_R$ applied independently on left- and right-handed chiral components. They are usually rearranged in a vectorial phase rotation $\mathrm{U}(1)_B$, which is the symmetry corresponding to the conserved baryon number B, and an axial one $\mathrm{U}(1)_A$. The peculiar property of the $\mathrm{U}(1)_A$ chiral phase rotation is that, in the massless limit, it describes a symmetry of the classical action but not of the partition function of the quantum theory. $\mathrm{U}(1)_A$ describes an *anomalous* symmetry.

In Section 3.1 we start with an unrelated, at first glance, topic: the topology of the gauge field in the continuum. The connection is provided by the fact that the $\mathrm{U}(1)_A$ anomaly is proportional to the *topological charge* of the gauge field. We aim to study this on the lattice. However, the realization on the lattice of topological observables is not straightforward. Indeed, a naïve lattice discretization of the topological charge needs to be combined with an unambiguous renormalization condition. The cumulants of the charge, for instance the susceptibility, require also additional subtractions of short-distance singularities to make them integrable distributions.

In Section 3.2 we prove that $\mathrm{U}(1)_A$ is an anomalous symmetry using a lattice regularization of the QCD path integral that satisfies the GW relation. Then, we show that renormalized cumulants can be defined using a particular definition of the topological charge, based on a Dirac operator that satisfies the GW relation. Moreover, we show that the cumulants can be written as integrated correlation functions of scalar and pseudoscalalar density chains or combination of them [123, 124, 65]. In this form a particular regularization is not required anymore to prove that no renormalization factors or subtractions of short-distance singularities are required. These expressions provide a universal definition of the susceptibility and of the higher cumulants that satisfies the anomalous chiral WTIs [124].

Using these results, in Section 3.3 we derive a precise and unambiguous implementation on the lattice of the Witten–Veneziano formula for the η' mass [27, 125], originally conjectured in 1979 by Witten [23] and Veneziano [24, 25]. Finally, in Section 3.4 we briefly show how the Witten–Veneziano formula is embeded in a more general framework, which combine χPT with the expansion of the theory with $N_\mathrm{c} \to \infty$ in powers of $1/N_\mathrm{c}$.

3.1. The θ-term

The action for the gauge field introduced in Eq. (1.7) depends on an extra coupling θ, multiplying the θ-term in the Lagrangian. We can rewrite it as

$$S_\mathrm{g}[A] = \int \mathrm{d}^4x \left[\frac{1}{2g^2} \operatorname{tr}\{F_{\mu\nu}(x) F_{\mu\nu}(x)\} - \mathrm{i}\theta q(x) \right], \tag{3.1}$$

where we introduced the *topological charge density* field

$$q(x) \equiv \frac{1}{32\pi^2}\epsilon_{\mu\nu\rho\sigma}\,\mathrm{tr}\big\{F_{\mu\nu}(x)F_{\rho\sigma}(x)\big\}. \tag{3.2}$$

Its name suggests a link with the topological properties of the theory. Indeed, $q(x)$ has some interesting properties.

First, we justify the presence in the action of this term: it is a dimension-four gauge-invariant Lorentz-scalar field independent on the kinetic term, thus it has every right to be included in the action. From an EFT point of view, it is actually expected to be there, lacking symmetry reasons to exclude it. However, it is odd under both parity (P) and time reversal (T) transformations, differently from any other term in the action. For this reason, P and T are not symmetries of the theory described by Eq. (3.1) with $\theta \neq 0$.

Applying the CPT theorem, it follows that also CP is broken. CP symmetry breaking is observed in nature in the electroweak sector, but not in strong interactions. A value of $\theta \neq 0$ would imply, for instance, a significant contribution to the neutron electric dipole moment (EDM). However, neutron EDM is strongly constrained by experiment to be [17, 126]

$$\big|d_\mathrm{n}\big| < 3.0 \times 10^{-26}\,\mathrm{e\,cm} \quad (90\,\%\ \mathrm{C.L.}). \tag{3.3}$$

This translates into a very stringent bound on θ [127][1]

$$|\theta| \lesssim 10^{-10}. \tag{3.4}$$

Since no profound symmetry reason why $\theta = 0$ should be preferred are known, this is a fine-tuning problem known as the *strong CP problem*.

In the Euclidean YM action in Eq. (3.1), the θ-term is purely imaginary. This is correct since $q(x)$ is odd under the time reversal transformation, which implies an extra i factor after Wick's rotation. Thus, an imaginary θ-term in Euclidean spacetime correspond to a physical real θ-term in Minkowski spacetime.

This is a problem for Monte Carlo lattice simulations: for $\theta \neq 0$ the corresponding lattice action has an imaginary contribution, thus the statistical interpretation of the partition functions fails. In the thesis, we are interested in the theory with $\theta = 0$. This is justified by the very small value of θ in real-world physics and it saves us from dealing with the problem of a complex action.

3.1.1. The topological charge

The name 'topological charge' hints a connection with the topology of the gauge field. To start with, notice that $q(x)$ is the divergence of a gauge-variant *topological current*

$$q(x) = \partial_\mu K_\mu(x), \qquad K_\mu(x) \equiv \frac{1}{16\pi^2}\epsilon_{\mu\nu\rho\sigma}\,\mathrm{tr}\Big\{A_\nu(x)\Big[F_{\rho\sigma}(x) - \frac{2\mathrm{i}}{3}A_\rho(x)A_\sigma(x)\Big]\Big\}. \tag{3.5}$$

[1] In the electroweak sector, CP symmetry is broken by the CKM matrix phase. This is responsible for a small contribution to the neutron EDM: $|d_\mathrm{n}| \approx 10^{-32}\,\mathrm{e\,cm}$ [127]. Thus, the CKM phase alone would not be able to explain a neutron EDM of the size of the present bound. However, there could be contributions from physics beyond the Standard Model.

Despite this, its spacetime integral, i.e. the *topological charge* Q, does not vanish

$$Q \equiv \int \mathrm{d}^4x\, q(x) \in \mathbb{Z}. \tag{3.6}$$

This is a topological quantity, which is not changed by continuous deformations of $A_\mu(x)$. With the proper normalization, it represents a non-zero integer topological index of the gauge field, also known as *winding number*. Moreover, a result in differential geometry, the *Atiyah–Singer index theorem* [128], relates Q to the index of the Dirac operator, i.e. the difference in the number of left- and right-handed chiral zero modes of the Dirac operator.

Remarkably, at the classical level, there exist gauge field solutions of the Euclidean equations of motion with non-zero Q. These are called *instantons* and are introduced in Appendix E. Beyond the classical level, the partition function can be written as a sum of partition functions $\mathcal{Z}_\nu$ with definite topological charge ν, which in turn can be written as path integrals in the sector of gauge field with $Q = \nu$

$$\mathcal{Z}_\nu = \int \mathcal{D}[\bar\psi, \psi, A]\, \delta_{Q\nu} \mathrm{e}^{-S|_{\theta=0}}. \tag{3.7}$$

Even when interested in the physics at $\theta = 0$, it is useful to leave the θ-term in the action as a source. The partition function and the free energy are thus functions of the parameter θ

$$\mathcal{Z}(\theta) = \mathrm{e}^{-F(\theta)} = \int \mathcal{D}[\bar\psi, \psi, A]\, \mathrm{e}^{-S|_{\theta=0} + \mathrm{i}\theta Q}. \tag{3.8}$$

Using that $Q \in \mathbb{Z}$ allows to write the θ dependence explicitly

$$\mathcal{Z}(\theta) = \sum_{\nu\in\mathbb{Z}} \mathrm{e}^{\mathrm{i}\theta\nu} \mathcal{Z}_\nu. \tag{3.9}$$

From this it is evident that θ is an angular variable and is often called *θ-angle*

$$\mathcal{Z}(\theta + 2\pi) = \mathcal{Z}(\theta) \quad \Rightarrow \quad \theta \in [0, 2\pi]. \tag{3.10}$$

Moreover, a parity transformation maps $\nu \to -\nu$, thus the partition function is an even function of θ

$$\mathcal{Z}(\theta) = \mathcal{Z}(-\theta). \tag{3.11}$$

Remarkably, the path integral expression of $\mathcal{Z}_\nu$ in Eq. (3.7) shows that the $\mathcal{Z}(\theta)$ in Eq. (3.9) is a Fourier series with positive coefficients. Therefore, for $-\pi < \theta < \pi$, $\mathcal{Z}(\theta)$ has an absolute maximum and $F(\theta)$ an absolute minimum at $\theta = 0$. It is common to normalize the partition function such as $F(0) = 0$.[2]

For every different value of θ, the partition function $\mathcal{Z}(\theta)$ describes a different theory with a different vacuum state known as *θ-vacuum*. It is an interesting question to ask if a sensible QFT has to be restricted to a θ-vacuum with a definite θ value, or it could be a superposition like

$$\sum_i c_i \mathcal{Z}(\theta_i). \tag{3.12}$$

[2] The functional dependence of $F(\theta)$ is studied Ref. [129] in the framework of the large-N_c expansion. Some details are given in Appendix D.

For instance,

$$\mathcal{Z}(\theta) + \mathcal{Z}(\theta + \pi) = \sum_{\nu \in \mathbb{Z}} \mathrm{e}^{2\mathrm{i}\theta\nu} \mathcal{Z}_{2\nu}, \quad \text{or} \quad \int_0^{2\pi} \frac{\mathrm{d}\theta}{2\pi} \mathcal{Z}(\theta) = \mathcal{Z}_0. \tag{3.13}$$

It is possible to show that any superposition different from the one in Eq. (3.9) violates the cluster decomposition property of the QFT [35]. This has consequences: consider for instance the theory described by $\mathcal{Z}_\nu$, with fixed ν, in a periodic box with finite volume V and small quark masses M. It turns out that, in the regime in which the product $V\Sigma M$ is small, connected correlation functions in the flavour-singlet sector are long-range [130]. Exceptionally, this long-range correlation is generated by a single spacetime-constant degree of freedom and it is not associated with a massless NG boson resulting from symmetry breaking. Nevertheless, it is sometimes useful to simulate on the lattice the theory at fixed topological charge [131]. The price to pay is to recover the cluster decomposition in the large volume regime with power-law finite volume corrections, instead of exponential ones.

The θ dependence of Eq. (3.8) is useful to study the distribution of the topological charge Q. For instance, the cumulants of this distribution are easily obtained deriving the free energy with respect to θ, before setting it to the physical value $\theta = 0$. We define even cumulants to be

$$\begin{aligned} C_n \equiv \frac{1}{V}\big\langle Q^{2n} \big\rangle_{\text{con}} &= -\frac{(-\mathrm{i})^{2n}}{V} \frac{\partial^{2n}}{\partial\theta^{2n}} F(\theta)\bigg|_{\theta=0} \\ &= \int \mathrm{d}^4x_1 \cdots \mathrm{d}^4x_{2n-1} \, \langle q(x_1) \cdots q(x_{2n-1}) q(0) \rangle_{\text{con}}, \end{aligned} \tag{3.14}$$

while odd cumulants vanish, since the theory is symmetric under $\theta \to -\theta$. We are mostly interested in the second cumulant, which gives the variance of the distribution. This is also the integrated two-point function of the topological charge density field and its called *topological susceptibility*

$$\chi_{\text{t}} \equiv C_1 = \frac{1}{V}\big\langle Q^2 \big\rangle = \frac{1}{V} \frac{\partial^2}{\partial\theta^2} F(\theta)\bigg|_{\theta=0} = \int \mathrm{d}^4x \, \langle q(x) q(0) \rangle, \tag{3.15}$$

where we used that $\langle q(x) \rangle\big|_{\theta=0} = 0$. Since different models predict different shapes for the distribution, higher cumulants also play a rôle in quantifying the deviation from a normal distribution and, as will be explained in the following, may have observable phenomenological consequences.

The topological charge is obviously a flavour-singlet object, given that it is a purely gluonic field. As such, it has a deep connection to the flavour-singlet sector of quarks: as will be justified in Section 3.2, the topological charge density appears as an *anomalous* contribution to the WTIs of flavour-singlet $\mathrm{U}(1)_A$ rotations

$$\big\langle \partial_\mu A^0_\mu(x) O \big\rangle = 2m \big\langle P^0(x) O \big\rangle + \sqrt{2N_{\text{f}}} \big\langle q(x) O \big\rangle. \tag{3.16}$$

3.1.2. The distribution of the topological charge on the lattice

We would like to give sense to Eq. (3.16) in the lattice theory. However, it is far from trivial to extend topological properties of the continuum gauge field to the lattice-discretized gauge

field, since no notion of infinitesimally-small deformations of the field is present on the lattice. Topology is a property of the classical field space of the theory in the continuum only. On the lattice, the field space is connected and the decomposition of the partition function into topological charge sectors is arbitrary to some extent.

Indeed, over the years various definitions of the topological charge were proposed, but it was not always clear if these definitions would have the correct continuum limit. In principle, any $d = 4$ local composite field with the correct symmetry properties would work. The simplest choice is a naïve discretization of the gauge field in $q(x)$ that reduces to Eq. (3.2) in the continuum limit. For future reference and consistency, we define the naïve discretization of the topological charge density to be the continuum-looking field

$$q_{\mathrm{L}}(x) = \frac{1}{32\pi^2}\epsilon_{\mu\nu\rho\sigma}\,\mathrm{tr}\left\{F^{\not\square}_{L\mu\nu}(x)F^{\not\square}_{L\rho\sigma}(x)\right\}. \tag{3.17}$$

where the lattice $F^{\not\square}_{L\mu\nu}$ is given by the traceless symmetric discretization[3]

$$F^{\not\square}_{\mathrm{L}\mu\nu} = F^{a}_{\mathrm{L}\mu\nu}T^{a}, \qquad F^{a}_{\mathrm{L}\mu\nu}(x) = -\frac{\mathrm{i}}{4a^2}\,\mathrm{tr}\Big\{\big[Q_{\mu\nu}(x) - Q_{\nu\mu}(x)\big]T^{a}\Big\}, \tag{3.18}$$

where $Q_{\mu\nu}$ is defined in Eq. (2.136).

Let us study the renormalization properties of $q(x)$. In the continuum with any regularization that preserves chiral symmetry, if $N_{\mathrm{f}} \geq 1$ the anomalous WTI in Eq. (3.16) strictly constrain the mixing pattern. Working in a mass-independent scheme, we have [132, 133]

$$\begin{pmatrix} \sqrt{2N_{\mathrm{f}}}q(x) \\ \partial_\mu A^0_\mu(x) \\ P^0(x) \end{pmatrix}_{\mathrm{R}} = \begin{pmatrix} 1 & Z-1 & 0 \\ 0 & Z & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{2N_{\mathrm{f}}}q(x) \\ \partial_\mu A^0_\mu(x) \\ P^0(x) \end{pmatrix}_{0}, \tag{3.19}$$

where Z is the axial current renormalization factor

$$A^0_{\mathrm{R}\mu}(x) = Z A^0_{0\mu}(x), \tag{3.20}$$

that is computable in perturbation theory and gets the first non-trivial contribution at two loops. In the $\overline{\mathrm{MS}}$,

$$Z = 1 - \frac{g^4}{(2\pi)^4}\frac{3}{8}C_F N_{\mathrm{f}}\frac{1}{\epsilon} + \mathcal{O}\big(g^6\big), \qquad C_F = \frac{N_{\mathrm{c}}^2 - 1}{2N_{\mathrm{c}}}. \tag{3.21}$$

We cannot use the same argument to constrain the renormalization of $q(x)$ in YM theory. We can speculate on the expression of $Z - 1$, which is proportional to $\lambda_{\mathrm{t}}^2 N_{\mathrm{f}}/N_{\mathrm{c}}$ and goes to zero in the $N_{\mathrm{c}} \to \infty$ limit at fixed $\lambda_{\mathrm{t}} = N_{\mathrm{c}}g^2$. Assuming that YM theory is recovered in a smooth way from QCD for $N_{\mathrm{c}} \to \infty$, this suggests that $q(x)$ does not need renormalization in YM theory. In Ref. [1] we provided a rigorous demonstration of this statement, which is reported in the original material of Chapter 4.

[3] Notice that this definition corresponds to the one in Eq. (2.135) with the trace removed

$$F^{\not\square}_{\mathrm{L}\mu\nu}(x) = 2\,\mathrm{tr}\big\{F_{\mathrm{L}\mu\nu}T^{a}\big\}T^{a}.$$

This clean renormalization pattern is preserved on the lattice only if a GW regularization is used. The topological charge density discretized as in Eq. (3.17) does not satisfy any index theorem and it is multiplicatively renormalized even in YM theory. Moreover, when considering the topological susceptibility and higher cumulants, things get even more complicated. The product of fields $q(x)q(0)$ makes sense only as a distribution, defined up to possibly-divergent contact terms. The infinities must be subtracted and the finite part must be fixed before integrating $q(x)q(0)$ over coincident points as in the topological susceptibility and higher cumulants.

3.2. The anomalous $\mathrm{U}(1)_A$

In the derivation of the WTIs presented in Section 1.4.1, we considered only field transformations that leave the path integral measure invariant. This is not always the case. Indeed, there are field transformations that leave the action invariant, and are thus symmetries of the theory at the classical level, but they do not leave the path integral measure invariant. These symmetries do not survive quantization and for this reason are called *anomalous*. The corresponding WTIs are modified by an additional term, the *anomaly*, which accounts for the variation of the path integral measure. In perturbation theory, the anomaly appears at one-loop level in correlation functions of the anomalous currents.

Quantum anomalies in QFT is a vast topic. The first anomaly was discovered in 1969 by Adler [134], Bell and Jackiw [135]. It is known as ABJ anomaly and it has been crucial to explain the short lifetime of the π^0 meson. We refer to standard textbooks for the general theory of anomalies [136]. In the following, we restrict to the anomaly possessed by $\mathrm{U}(1)_A$ symmetry of the classical massless action. As in the ABJ anomaly case, the $\mathrm{U}(1)_A$ anomaly is generated by the non-invariance of the path integral measure under a global axial rotation of chiral fermions. For this reason it is known as *chiral anomaly* or *axial anomaly*.

Providing a useful regulation of the path integral measure [137, 78, 138, 81], it is possible to show that the $\mathrm{U}(1)_A$ anomaly is proportional to the topological charge density

$$\mathcal{A}(x) = 2N_{\mathrm{f}}q(x). \tag{3.22}$$

We would like to derive this result in the lattice-discretized theory, in which the path integral measure is mathematically well-defined. However, Wilson's discretization of QCD is not a convenient regularization in this case: the lattice path integral measure results invariant under $\mathrm{U}(1)_A$ rotations. This is not surprising: chiral symmetry is broken *explicitly* by the Wilson–Dirac operator and the action is not invariant.

In the continuum limit, both classical $\mathrm{U}(1)_A$ symmetry and its anomaly are recovered with a complicated pattern of renormalization. The flavour-singlet equivalent of the axial WTI in Eq. (2.90) has an additional term that, after renormalization, does not vanish in the continuum limit. This term reproduces correctly the anomaly, which is missing from the transformation of the path integral measure.

3.2.1. The anomalous $\mathrm{U}(1)_A$ with Ginsparg–Wilson fermions

Ginsparg–Wilson discetization of QCD breaks naïve chiral symmetry, even if in a mild way controlled by the GW relation. However, it possesses an exact symmetry, given by Eq. (2.53),

that reduces to chiral symmetry in the continuum limit. This makes possible to derive the anomalous axial WTIs on the lattice à la Fujikawa, from the non-invariance of the path integral measure under the flavour-singlet transformation in Eq. (2.53). Consider an infinitesimal chiral transformation, given in Lüscher's version by Eqs (2.94) and (1.76b). Two factors contribute to the linear change of the Jacobian

$$\delta\mathcal{J}_{\psi}(x,y) = \frac{\delta}{\delta\psi(y)}[\delta\psi(x)] = \mathrm{i}\tau^a\frac{1}{2}\left[\alpha^a(x)\delta_{x,y} + \sum_z \hat{\gamma}_5(x,z)\alpha^a(z)\hat{\gamma}_5(z,y)\right] + \mathrm{i}\tau^a\frac{1}{2}\left[\hat{\gamma}_5(x,y)\beta^a(y) + \beta^a(x)\hat{\gamma}_5(x,y)\right], \tag{3.23a}$$

$$\delta\mathcal{J}_{\bar{\psi}}(x,y) = \frac{\delta}{\delta\bar{\psi}(y)}[\delta\bar{\psi}(x)] = -\mathrm{i}\alpha^a(x)\tau^a\delta_{x,y} + \mathrm{i}\gamma_5\beta^a(x)\tau^a\delta_{x,y}. \tag{3.23b}$$

Since ψ and $\bar{\psi}$ are Grassmann-number-valued fields, the change to the path integral measure is given by the *inverse* Jacobian determinant

$$\mathrm{Det}^{-1}\mathcal{J} = \mathrm{Det}^{-1}\{1+\delta\mathcal{J}\} = 1 - \mathrm{Tr}\,\delta\mathcal{J}_{\psi} - \mathrm{Tr}\,\delta\mathcal{J}_{\bar{\psi}}\,. \tag{3.24}$$

It easy to see that the vector transformation contribution cancels between Eq. (3.23a) and Eq. (3.23b). On the contrary, the axial transformation contribution vanishes in Eq. (3.23b) but not in Eq. (3.23a), since $\mathrm{Tr}\,\hat{\gamma}_5 \neq 0$. What is left is

$$\mathrm{Tr}\,\delta\mathcal{J}_{\psi} = \frac{\mathrm{i}}{a^4}\sum_x \mathrm{tr}\{\tau^a\beta^a(x)\hat{\gamma}_5(x,x)\} = -\frac{\mathrm{i}\bar{a}}{a^4}\sum_x \beta^a(x)\,\mathrm{tr}\{\tau^a\gamma_5 D(x,x)\}. \tag{3.25}$$

In the case of flavour non-singlet transformations, $\mathrm{tr}\,\tau^a = 0$ and the anomaly vanishes. The only case in which we obtain a non-zero contribution is the flavour-singlet axial transformation, since $\mathrm{tr}\,\tau^0 = \sqrt{N_{\mathrm{f}}/2}$

$$\mathrm{Tr}\,\delta\mathcal{J}_{\psi} = \mathrm{i}\sqrt{2N_{\mathrm{f}}}\sum_x \beta^0(x) q_{\mathrm{GW}}(x), \tag{3.26}$$

where, by analogy with the continuum, we introduced the GW discretization of the topological charge density [78, 138, 81]

$$q_{\mathrm{GW}}(x) \equiv -\frac{\bar{a}}{2a^4}\,\mathrm{tr}\{\gamma_5 D(x,x)\}. \tag{3.27}$$

At this point, it is still not obvious why this composite lattice field is a valid discretization of the topological charge density in Eq. (3.2). It is possible to show that Eq. (3.27) reduces to Eq. (3.2) in the naïve continuum limit [82, 139, 140]. Indeed, if the gauge field satisfies the smoothness condition in Eq. (2.70), with $\epsilon \leq 1/30$, we have [82]

$$-\frac{\bar{a}}{2a^4}\,\mathrm{tr}\{\gamma_5 D(x,x)\} \xrightarrow{a\to 0} \frac{1}{32\pi^2}\epsilon_{\mu\nu\rho\sigma}\,\mathrm{tr}\{F_{\mu\nu}F_{\rho\sigma} + \mathcal{O}(a^2)\}. \tag{3.28}$$

Moreover, it is worth noting that Eq. (3.27) is a purely bosonic field, which depends implicitly on the gauge field U through the GW–Dirac operator. As a lattice discretization of $q(x)$, this definition is perfectly valid also in the YM theory, where fermions are not present.

For any composite field O_{ext} localized outside the support of $\beta^0(x)$, from the specialization of Eq. (1.71) we derive the anomalous flavour-singlet WTI

$$\big\langle \partial^*_\mu A^0_\mu(x) O_{\text{ext}} \big\rangle = 2m \big\langle P^0(x) O_{\text{ext}} \big\rangle + \sqrt{2N_{\text{f}}} \big\langle q_{\text{GW}}(x) O_{\text{ext}} \big\rangle, \tag{3.29}$$

which looks identical to the correspondent WTI in the continuum theory, in Eq. (3.16). Similarly, we obtain integrated anomalous WTIs

$$\begin{aligned}\sum_x \big\langle \partial^*_\mu A^0_\mu(x) O(y) O_{\text{ext}} \big\rangle = 2m \sum_x \big\langle P^0(x) O(y) O_{\text{ext}} \big\rangle \\ + \sqrt{2N_{\text{f}}} \sum_x \big\langle q_{\text{GW}}(x) O(y) O_{\text{ext}} \big\rangle - \big\langle \delta^0_A O(y) O_{\text{ext}} \big\rangle.\end{aligned} \tag{3.30}$$

3.2.2. Index theorem

In addition to satisfying the WTIs, the discretization of $q(x)$ in Eq. (3.27) has an interesting property: its spacetime integral

$$Q_{\text{GW}} = \sum_x q_{\text{GW}}(x) = -\frac{\bar{a}}{2} \operatorname{Tr}\{\gamma_5 D\} \tag{3.31}$$

satisfies an *index theorem* [138].

To prove this, we express the operator trace as a sum over eigenvectors v_λ and use the properties of the spectrum of a γ_5-hermitian GW–Dirac operator, studied in Section 2.2.5. For any function $f(\lambda)$ bounded on the spectrum of D, we have

$$\operatorname{Tr}\{\gamma_5 f(D)\} = \sum_\lambda \big(v_\lambda, \gamma_5 f(D) v_\lambda\big) = \sum_\lambda f(\lambda) \big(v_\lambda, \gamma_5 v_\lambda\big) = \sum_{\lambda \in \mathbb{R}} f(\lambda) \big(v_\lambda, \gamma_5 v_\lambda\big). \tag{3.32}$$

In the last sum, we used that $(v_\lambda, \gamma_5 v_\lambda) = 0$ if λ is complex. Therefore, either $\lambda = 0$ or $\lambda = 2/\bar{a}$. This proves the identity

$$\operatorname{Tr}\{\gamma_5 f(D)\} = f(0)(n_+ - n_-) + f\Big(\frac{2}{\bar{a}}\Big)(n'_+ - n'_-), \tag{3.33}$$

where $n_\pm$ is the number of right- and left-handed zero modes, and $n'_\pm$ is the number of right- and left-handed $\lambda = 2/\bar{a}$ modes.

Taking $f(\lambda) = \lambda$ proves that the quantity in Eq. (3.31) is an integer

$$Q_{\text{GW}} = n'_- - n'_+ \in \mathbb{Z}. \tag{3.34}$$

More interestingly, we can manipulate Eq. (3.31) into

$$Q_{\text{GW}} = \frac{1}{2} \operatorname{Tr}\{\gamma_5 [2 - \bar{a} D]\} = n_+ - n_- \in \mathbb{Z}, \tag{3.35}$$

that shows that Q_{GW} is the index of the GW-Dirac operator and amounts to the difference between the number of left- and right-handed zero modes. This closely resembles the continuum and guarantees that Q in Eq. (3.31) is an integer at finite lattice spacing. Of course, Q_{GW} is not a topological index at $a \neq 0$ and its topological meaning is recovered only in the continuum limit.

3.2.3. The θ-angle and the fermion mass matrix

According to Eq. (1.71), the non-invariance of the path integral measure under a $\mathrm{U}(1)_A$ rotation can be reinterpreted as an induced variation of the massless action

$$S \to S' = S + \mathrm{i}\sqrt{2N_\mathrm{f}} \sum_x \beta^0(x) q(x). \tag{3.36}$$

This is suspiciously similar to the variation of the action

$$S \to S' = S - \mathrm{i}\,\delta\theta \sum_x q(x), \tag{3.37}$$

induced by a shift in the θ-parameter

$$\theta \to \theta' = \theta + \delta\theta\,. \tag{3.38}$$

Consequently, a possible θ-term can be erased from the action with a global $\mathrm{U}(1)_A$ rotation

$$\beta^0_\theta = \frac{\theta}{\sqrt{2N_\mathrm{f}}}. \tag{3.39}$$

Away from the chiral limit, this is not without consequences: the global $\mathrm{U}(1)_A$ rotation β^0_θ induces a transformation in the quark mass term. At the non-linear level it generates a phase in the mass matrix $\mathcal{M}$

$$\mathcal{M} \to \mathcal{M}' = \mathcal{M}\mathrm{e}^{2\mathrm{i}\beta^0_\theta/\sqrt{2N_\mathrm{f}}} = \mathcal{M}\mathrm{e}^{\mathrm{i}\theta/N_\mathrm{f}}. \tag{3.40}$$

Therefore, the quark mass matrix phase and the θ term are not independent. Only the combination

$$\bar\theta = \theta + \arg\det\mathcal{M} \tag{3.41}$$

is physical. Depending on the situation, it is more useful to work without θ term and a complex mass matrix, or with a θ term and $\det\mathcal{M} = 1$. The physical consequences are in any case the same.

If at least one flavour of quark is massless, $\det\mathcal{M} = 0$, so its phase is not defined. In this situation, any value of the θ parameter can be rotated away without consequences with a non-singlet axial rotation on the massless quark field. This is expected, and its related to the well known fact that the topological susceptibility vanishes if one of the quarks is massless.[4]

3.2.4. Universal definition

The GW relation can be used to prove strong identities that relate topological observables to chains of scalar and pseudoscalar densities [123, 124]. Consider the product of r scalar and pseudoscalar densities as defined in Eq. (2.95), localized at different points in a theory with at least $N_\mathrm{f} = r$ flavours. We can choose them to have non-singlet flavour transformation

[4] Indeed, a massless u quark was a proposed solution to strong CP problem in the 70s. Today, $m_u = 0$ is excluded. [17]

properties in a way that only their full product is flavour-singlet. Using a couple of flavour indices $i, j = 1, \dots, N_{\mathrm{f}}$ to denote them, we have

$$P_{r1}(x_1)S_{12}(x_2)\cdots S_{r-1r}(x_r). \tag{3.42}$$

We can perform the quark path integral and express the spacetime-integrated *density chain* as

$$\sum_{x_1,\dots,x_r}\langle P_{r1}(x_1)S_{12}(x_2)\cdots S_{r-1r}(x_r)\rangle_{\mathrm{q}} \\ = -\operatorname{Tr}\Big\{\gamma_5\Big[\mathbb{1}-\frac{\bar{a}}{2}D\Big]\big(D+m_1\big)^{-1}\cdots\Big[\mathbb{1}-\frac{\bar{a}}{2}D\Big]\big(D+m_r\big)^{-1}\Big\}. \tag{3.43}$$

Working in a discretization which preserves chiral symmetry, we can apply Eq. (3.33) to obtain another expression for Q

$$Q = -m_1\cdots m_r\sum_{x_1,\dots,x_r}\langle P_{r1}(x_1)S_{12}(x_2)\cdots S_{r-1r}(x_r)\rangle_{\mathrm{q}} = n_- - n_+. \tag{3.44}$$

This density chain trick is even more powerful: in a similar way it is possible to show that the topological susceptibility is given on the lattice by

$$\chi_{\mathrm{t}} = m_1\cdots m_s\sum_{x_1,\dots,x_{s-1}}\langle P_{r1}(x_1)S_{12}(x_2)\cdots S_{r-1r}(x_r) \\ \cdot P_{sr+1}(x_{r+1})S_{r+1r+2}(x_{r+2})\cdots S_{s-1s}(0)\rangle_{\mathrm{con}}. \tag{3.45}$$

For this expression to make sense in the continuum limit, it must be free from any non-integrable short distance singularity. We prove that this is the case for $s \geq 5$. If one of the x_j goes close to a x_k, thanks to the non-singlet structure of the densities, the operator-product expansion starts with $d \geq 3$ fields. The leading short-distance singularity are

$$P_{ij}(x)S_{jk}(0) \overset{x\to 0}{\sim} \frac{1}{|x|^3}P_{ik}(0), \qquad S_{ij}(x)S_{jk}(0) \overset{x\to 0}{\sim} \frac{1}{|x|^3}S_{ik}(0), \tag{3.46}$$

and they are integrable. Similarly, when more than two coordinates coincide. If all the coordinates are reduced to a point, the correlator is singular as $|x|^{-3s}$, which is integrable for $s \geq 5$. Moreover, the products $m_s P_{st}$ and $m_s S_{st}$ are related by *non-singlet* WTIs to conserved currents, which do not need renormalization. Therefore,

$$\chi_{\mathrm{t}} = m_1\cdots m_5\sum_{x_1,\dots,x_4}\langle P_{31}(x_1)S_{12}(x_2)S_{23}(x_3)\cdot P_{54}(x_4)S_{45}(0)\rangle_{\mathrm{con}}, \tag{3.47}$$

is a good definition of the topological susceptibility on the lattice, which tends to the proper continuum limit. As described in Ref. [124], it is possible to give a similar definition for the higher cumulants defined in Eq. (3.14) which is valid in the continuum limit using at least five different flavours of quarks.

The requirement on the number of flavours in the theory can be relaxed by introducing valence quarks, whose contribution in the action is neutralized by mass-degenerate pesudofermion fields $\phi(x)$

$$S_{\mathrm{vf}} = \sum_{r=1}^{2N_{\mathrm{vf}}}\sum_{x,y}\bar{\psi}_r(x)\big[D(x,y)+m\delta_{x,y}\big]\psi_r(y) + \sum_{r=1}^{N_{\mathrm{vf}}}\sum_{x}\Big|\sum_{y}\big[D(x,y)+m\delta_{x,y}\big]\phi_r(y)\Big|^2. \tag{3.48}$$

Therefore, using valence quarks for the densities, Eq. (3.47) applies to any number of sea quarks or even to YM theory.

We arrive to the *universal formula* for the cumulants of the topological charge in terms of density chains

$$C_{\mathrm{dc},n} = \frac{1}{V}\big\langle Q_{\mathrm{dc}}^{2n}\big\rangle_{\mathrm{con}}, \qquad Q_{\mathrm{dc}} = -m^3 a^6 \sum_{x_1,x_2,x_3} \langle P_{31}(x_1) S_{12}(x_2) S_{23}(x_3)\rangle_{\mathrm{q}}, \tag{3.49}$$

where up to $6n$ different valence quarks have been used. Here, Q_{dc} is just the triangle fermion loop with one pseudoscalar and two scalar vertices at vanishing external momenta.

If a discretization that conserves chiral symmetry is used, the cumulants of the topological charge defined with index of the Dirac operator as in Eq. (3.31)

$$C_{\mathrm{GW},n} = \frac{1}{V}\big\langle Q_{\mathrm{GW}}^{2n}\big\rangle_{\mathrm{con}}, \qquad Q_{\mathrm{GW}} = \frac{\bar{a}}{2}\,\mathrm{Tr}\big\{\gamma_5 D\big\} \tag{3.50}$$

reduces to Eq. (3.49) as we described. This proves that Eq. (3.50) provides a good lattice discretization of the topological charge and its cumulants.

However, it is important to understand that the universal definition in Eq. (3.49) is not limited to GW fermions. If Wilson's discretization of QCD is employed, Eq. (3.49) is still free of short-distance singularities and correctly normalized, provided that densities and quark masses are renormalized in accordance with the non-singlet chiral WTIs.

3.3. The Witten–Veneziano mechanism

In 1979 Witten [23] and Veneziano [24, 25] proposed a mechanism based on the large-N_{c} expansion through which the $U(1)_A$ problem is solved. Moreover, they derived a formula for the leading contribution in $1/N_{\mathrm{c}}$ to the mass of the η' meson. The exact form of the two original derivations is slightly different. Witten performs an expansion in $1/N_{\mathrm{c}}$ and obtains a result that is exact in the 't Hooft limit $N_{\mathrm{c}} \to \infty$. Veneziano obtains the same result working with an expansion of WTIs in powers of the parameter $u = N_{\mathrm{f}}/N_{\mathrm{c}}$. Here, we follow the argument of Witten, in the form given in [27]. In this section, we work formally with continuum renormalized fields.

Consider the two-point correlation function in Eq. (3.15) at non-zero momentum p

$$\chi_{\mathrm{t}}(p) \equiv \int \mathrm{d}^4x\, \mathrm{e}^{-\mathrm{i}px} \langle q(x) q(0)\rangle. \tag{3.51}$$

As already stated, care must be taken in integrating $\langle q(x)q(0)\rangle$: this operator product has to be thought of as a distribution and in $x = 0$ its definition includes contact terms proportional to the Dirac delta and derivatives of the Dirac delta. Thus, its Fourier transform includes a polynomial in p^2, up to second degree since $\chi_{\mathrm{t}}(p)$ has mass-dimension 4. Using the Källén–Lehmann spectral representation, in Euclidean spacetime version, $\chi_{\mathrm{t}}(p)$ can be written as a three-times-subtracted dispersion relation [125]

$$\chi_{\mathrm{t}}(p) = a_1 + a_2 p^2 + a_3 (p^2)^2 + (p^2)^3 \int_{m^2}^{\infty} \mathrm{d}t\, \frac{\rho(t)}{(p^2+t)t^3}, \tag{3.52}$$

where $\rho(t)\,\mathrm{d}t$ is a positive measure, growing at most as t^2 for $t \to \infty$. For this formula to be useful, it is convenient to extract from Eq. (3.52) the explicit contribution of the η' meson

$$\chi_{\mathrm{t}}(p) = b_1 + b_2 p^2 + b_3 (p^2)^2 - \frac{R_{\eta'}^2}{p^2 + M_{\eta'}^2} + (p^2)^3 \int_{m^2}^{\infty} \mathrm{d}t\, \frac{\rho'(t)}{(p^2+t)t^3}, \tag{3.53}$$

where

$$\big\langle 0 \big| q(x) \big| \eta'(\vec{p}) \big\rangle \equiv R_{\eta'} \mathrm{e}^{\mathrm{i}px} \tag{3.54}$$

gives the residue at the η' pole and $\rho'(t)\,\mathrm{d}t$ is another positive measure that accounts for states different from the η', e.g. pseudoscalar glueballs. The minus sign in front of the η' contribution in Eq. (3.53) is specific to this two-point function that satisfies an unusual form of reflection positivity. The reason is that q is odd under the time reversal transformation. This fact was not immediately realized and for some time Witten's argument was being thought to require the cancellation of contributions of the same sign [141, 125].

At zero momentum, the correlator $\chi_{\mathrm{t}}(0)$ is just the topological susceptibility in Eq. (3.15), which vanishes in the chiral limit. Thus, taking the limit $p \to 0$ in Eq (3.53)

$$\lim_{p\to 0} \chi_{\mathrm{t}}(p) = b_1 - \frac{R_{\eta'}^2}{M_{\eta'}^2} = 0. \tag{3.55}$$

Now we have to connect these quantities of QCD to the topological susceptibility in Yang–Mills theory. To do this, consider the anomalous flavour-singlet WTI of Eq. (3.16), specialized to $O_{\mathrm{ext}} = q(0)$

$$\big\langle \partial_\mu A_\mu^0(x) q(0) \big\rangle = 2m \big\langle P^0(x) q(0) \big\rangle - \sqrt{2N_{\mathrm{f}}} \big\langle q(x) q(0) \big\rangle, \tag{3.56}$$

or better its Fourier transform in the chiral limit $m \to 0$

$$\mathrm{i}p_\mu \int \mathrm{d}^4x\, \mathrm{e}^{-\mathrm{i}px} \big\langle A_\mu^0(x) q(0) \big\rangle + \mathrm{CT}(p) = \sqrt{2N_{\mathrm{f}}}\chi_{\mathrm{t}}(p) + \mathrm{CT}(p). \tag{3.57}$$

To make both sides separately finite, we add a contact term $\mathrm{CT}(p)$, which is a polynomial of 4^{th} degree in p that vanishes at $p = 0$, since the integral of the l.h.s. of Eq. (3.56) is zero as it is. Since at the end we are interested in quantities at $p = 0$, the exact form of $\mathrm{CT}(p)$ does not affect the final result.[5]

The rest of the Witten–Veneziano argument relies on the large-N_{c} expansion of Eq. (3.57) (for further details, see Appendix D), before taking the $p \to 0$ limit. We assume the *smooth quenching hypothesis*, i.e. at $p \neq 0$ all the relevant quantities and observables have a well-defined large-N_{c} expansion and the 't Hooft limit $N_{\mathrm{c}} \to \infty$ is equivalent to neglecting the fermion determinant. The topological susceptibility from Eq. (3.15) has a finite 't Hooft limit, equal to the topological susceptibility of Yang–Mills theory

$$\lim_{N_{\mathrm{c}}\to\infty} \chi_{\mathrm{t}}(p) = \lim_{N_{\mathrm{c}}\to\infty} \chi_{\mathrm{t}}^{\mathrm{YM}}(p). \tag{3.58}$$

[5] It affects, however, derivatives of the susceptibility like $\chi_{\mathrm{t}}'(0) = \mathrm{d}\chi_{\mathrm{t}}(p^2)/\mathrm{d}p^2\big|_{p=0}$, which are relevant [133, 142–144] to the so-called proton spin crisis problem [145].

Taking now the $p \to 0$ limit, $\chi_{\rm t}^{\rm YM}(0)$ is in general different from zero. For this to be possible, the correlation function in the l.h.s. of Eq. (3.57) must have a pole at $p = 0$ in the 't Hooft limit. That is to say, there must exist a meson in the flavour-singlet channel whose mass vanishes in the $N_{\rm c} \to \infty$ limit

$$M_{\eta'} \xrightarrow{N_{\rm c}\to\infty} 0. \tag{3.59}$$

3.3.1. The η' mass formula

To obtain the WV formula for the η' mass, we first express the residue $R_{\eta'}$ in Eq. (3.53). The amplitude for the flavour singlet axial current $A_\mu^0(x)$ to create one-η' states is, as in Eq. (1.100b),

$$\left\langle 0\middle|A_\mu^0(x)\middle|\eta'(\vec{p})\right\rangle \equiv {\rm i}p_\mu F_{\eta'} {\rm e}^{{\rm i}px}, \tag{3.60}$$

where $F_{\eta'}$ is the analogue of F_π for pions. At leading order in $1/N_{\rm c}$ we can take $F_{\eta'} = F_\pi$. We can relate $F_{\eta'}$ to $R_{\eta'}$ using the fact that the η' saturates the anomalous flavour-singlet WTI in Eq. (3.56), in the chiral limit and at long distances,

$$-M_{\eta'}^2 F_{\eta'}\left(-R_{\eta'}\right) = \sqrt{2N_{\rm f}} R_{\eta'}^2. \tag{3.61}$$

Therefore, the residue at $p = 0$ vanishes in the 't Hooft limit

$$R_{\eta'} = \frac{F_{\eta'} M_{\eta'}^2}{\sqrt{2N_{\rm f}}} = \mathcal{O}\left(\frac{1}{\sqrt{N_{\rm c}}}\right). \tag{3.62}$$

The key observation is to notice that the $p \to 0$ and $N_{\rm c} \to \infty$ limits in Eq. (3.53) do not commute. Applying the $N_{\rm c} \to \infty$ limit at some fixed value $p \neq 0$ the contribution of the η' meson vanishes, just as any other mesonic contribution in the flavour-singlet sector. Taking the $p \to 0$ limit after the $N_{\rm c} \to \infty$ limit, we arrive at

$$\chi_{\rm t}^{\rm YM}(0) = \lim_{p\to 0} \lim_{N_{\rm c}\to\infty} \chi_{\rm t}(p) = b_1, \tag{3.63}$$

that combined with Eqs (3.55) and (3.62) finally gives

$$\chi_{\rm t}^{\rm YM}(0) = \frac{F_\pi^2}{2N_{\rm f}} M_{\eta'}^2. \tag{3.64}$$

3.3.2. Higher cumulants

Higher cumulants of the topological charge distribution are obtained from higher θ-derivatives of the free energy as shown in Eq. (3.14). According to large-$N_{\rm c}$ counting rules, it follows that even cumulants

$$C_n = \mathcal{O}\left(N_{\rm c}^{2-2n}\right). \tag{3.65}$$

Only the second cumulant $C_1 = \chi_{\rm t}$ survives in the $N_{\rm c} \to \infty$ limit. This means that the topological charge distribution tends to a normal distribution in the 't Hooft limit.

The fourth cumulant C_2 measures the leading deviation from Gaussianity in the distribution

$$C_2 = \frac{1}{V}\langle Q^4\rangle_{\text{con}} = \frac{1}{V}\Big(\langle Q^4\rangle - 3\langle Q^2\rangle^2\Big) = \mathcal{O}\bigg(\frac{1}{N_\text{c}^2}\bigg). \tag{3.66}$$

It is of phenomenological interest, since it can be related with the Witten-Veneziano mechanism to the anomalous contribution to $\eta'\eta' \to \eta'\eta'$ scattering. Moreover, we can define the ratio between the fourth and the second cumulant

$$R \equiv \frac{C_2}{C_1} = \mathcal{O}\bigg(\frac{1}{N_\text{c}^2}\bigg), \tag{3.67}$$

which is itself suppressed at large-N_c. This is in contrast with the prediction of the semiclassical approximation know as dilute instanton gas (DIG) model, that predicts $R = 1$ (see Appendix E). This model is expected to work only at asymptotically high temperatures, but to fail near and below the temperature of deconfinement. Thus, a measure of the suppression of R with respect to unity quantifies the non-perturbative effects beyond the DIG approximation.

3.4. Large-N_c in chiral perturbation theory

The η' mass formula in Eq. (3.64), rewritten as

$$M_{\eta'}^2 = \frac{2N_\text{f}}{F_\pi^2}\chi_\text{t}^{\text{YM}}(0) + \mathcal{O}\bigg(\frac{1}{N_\text{c}^2}\bigg), \tag{3.68}$$

is reminiscent of the GMOR relation in Eq. (1.108). Indeed, the η' mass squared is proportional to a term, $\chi_\text{t}^{\text{YM}}(0)$, computed in the $N_\text{c} \to \infty$ limit in which the $U(1)_A$ symmetry is restored. Moreover, the η' mass squared is linear in the symmetry breaking parameter $1/N_\text{c}$, implicit in F_π^2, up to higher-order corrections. This suggests an EFT approach inspired to χPT: an expansion of the low-energy theory in powers of momenta p^2, quark masses m and $1/N_\text{c}$

The building of the effective theory closely mirrors the one of χPT presented in Appendix B. For a more complete introduction see Refs [146, 147]. We add to the QCD action an external sources term as in Eq. (B.6)

$$S_{\text{source}}(v_\mu, a_\mu, s, p, \theta) = \int \mathrm{d}^4x\, \big\{-\mathrm{i}\bar\psi\gamma_\mu\big(v_\mu + \gamma_5 a_\mu\big)\psi + \bar\psi\big(s - \mathrm{i}\gamma_5 p\big)\psi + \mathrm{i}\theta q\big\}, \tag{3.69}$$

where we allowed for a spacetime-dependent source $\theta(x)$ to account for the θ dependence. The partition function

$$\mathcal{Z}[v, a, s, p, \theta] = \mathrm{e}^{-F[v,a,s,p,\theta]} = \int \mathcal{D}[\psi, \bar\psi, A]\, \mathrm{e}^{-S\big|_{M=0,\theta=0} - S_{\text{source}}[v,a,s,p,\theta]}, \tag{3.70}$$

is invariant under $\mathrm{U}(N_\text{f})_L \times \mathrm{U}(N_\text{f})_R$ transformations if the sources transform according to Eqs (B.8), but now with $V_L, V_R \in \mathrm{U}(N_\text{f})$. There is one caveat: the anomalous $U(1)_A$ rotation induces a change in the path integral measure. This is accounted for by a change in the θ sources, given by

$$\theta \to \theta' = \theta + \mathrm{i}\ln\det V_R - \mathrm{i}\ln\det V_L. \tag{3.71}$$

The effective theory in the $U(1)_A$ sector is built as the theory of a scalar field ϕ that transforms according to

$$\phi \to \phi' = \phi - \mathrm{i} \ln \det V_R + \mathrm{i} \ln \det V_L. \tag{3.72}$$

This can be combined with the standard χPT introducing a field $U(x) \in \mathrm{U}(3)$ such that the special part is the ordinary χPT Lagrangian field while

$$\det U(x) = \mathrm{e}^{\mathrm{i}\phi(x)}. \tag{3.73}$$

The building blocks are then the same with the addition of the invariant

$$\theta - \mathrm{i} \ln \det U = \theta + \phi. \tag{3.74}$$

This has to be supplemented by a proper power counting for the low-energy expansion. The correct one in given by

$$p = \mathcal{O}\big(\delta^{1/2}\big), \qquad m = \mathcal{O}(\delta), \qquad \frac{1}{N_c} = \mathcal{O}(\delta). \tag{3.75}$$

From this bookkeeping, it follows that

$$U, \phi, \theta = \mathcal{O}(1), \qquad v_\mu, a_\mu = \mathcal{O}\big(\delta^{1/2}\big), \qquad s, p = \mathcal{O}(\delta). \tag{3.76}$$

According to this power counting, the LO Lagrangian at $\mathcal{O}(\delta^0)$ is

$$\mathcal{L}^{(0)}_{\chi\mathrm{PT}} = \frac{F_0^2}{4} \operatorname{tr}\big\{ D_\mu U^\dagger D_\mu U \big\} - \frac{F_0^2}{4} \operatorname{tr}\big\{ U^\dagger \chi + \chi^\dagger U \big\} - \frac{1}{2}\tau(\theta + \phi)^2, \tag{3.77}$$

where $\chi = 2B_0(s + \mathrm{i}p)$, the first and second terms are $\mathcal{O}(N_c p^2)$ and the third is $\mathcal{O}(1)$. At this order, the Lagrangian depends on three LECs: the usual F_0 and B_0 and the new τ. At this order in the expansion, τ is given by the topological susceptibility in YM theory in the large-N_c limit

$$\lim_{N_c \to \infty} \chi_t^{\mathrm{YM}} = -\frac{\partial^2}{\partial \theta^2} F^{(0)}_{\chi\mathrm{PT}} = \tau, \qquad \text{at LO.} \tag{3.78}$$

In real world QCD, also the small quark masses contribute to give mass to the physical η'. The mass of the pNG bosons, including the flavour-singlet one which correspond to the η', is given by EFT Lagrangian at $p^2 = 0$, with the sources set to $s = M$, $v_\mu = a_\mu = p = \theta = 0$

$$-\frac{F_0^2}{4} 2B_0 \operatorname{tr}\big\{ M(U + U^\dagger) \big\} = 2B_0 \left[\text{const} + \operatorname{tr}\big\{ M\pi^2 \big\} - \frac{1}{3F_0^2} \operatorname{tr}\big\{ M\pi^4 \big\} + \mathcal{O}\left(\frac{\pi^6}{F_0^4} \right) \right]. \tag{3.79}$$

In the $N_f = 3$ case, we parametrize the $U(x) = \mathrm{e}^{2\mathrm{i}\pi(x)/F_0} \in \mathrm{U}(3)$ matrix as

$$\pi = \frac{1}{2} \begin{pmatrix} \pi^0 + \frac{1}{\sqrt{3}}\eta_8 + \sqrt{\frac{2}{3}}\eta_1 & \sqrt{2}\pi^+ & \sqrt{2}K^+ \\ \sqrt{2}\pi^- & -\pi^0 + \frac{1}{\sqrt{3}}\eta_8 + \sqrt{\frac{2}{3}}\eta_1 & \sqrt{2}K^0 \\ \sqrt{2}K^- & \sqrt{2}\bar{K}^0 & -\frac{2}{\sqrt{3}}\eta_8 + \sqrt{\frac{2}{3}}\eta_1 \end{pmatrix}, \tag{3.80}$$

where with respect to Eq. (B.21) there is also a ninth component, $\eta_{\mathbf{1}}$, that represents the flavour-singlet field

$$\phi(x) = -\mathrm{i} \ln \det U(x) = \sqrt{6}\eta_{\mathbf{1}}(x). \tag{3.81}$$

Setting $M = \mathrm{diag}(m_u, m_d, m_s)$ gives the flavour-octet pseudoscalar meson masses in Eqs (B.22) together with a new formula for the flavour-singlet pseudoscalar meson mass that supersedes Eq. (3.64)

$$M_{\eta_1}^2 = \frac{2}{3}\big(m_u + m_d + m_s\big) B_0 + 2N_{\mathrm{f}} \frac{\tau}{F_0^2} + \mathcal{O}\bigg(m^2, \frac{m}{N_{\mathrm{c}}}, \frac{1}{N_{\mathrm{c}}^2}\bigg). \tag{3.82}$$

Moreover, the regime of validity of Eqs (B.22) is modified: they are leading order relation in $1/N_{\mathrm{c}}$ with the LEC F_0, B_0 and τ given in the $N_{\mathrm{c}} \to \infty$ limit.

At this order of the expansion, we can solve Eqs (B.22) and (3.82) for the LEC τ

$$\lim_{N_{\mathrm{c}}\to\infty} \chi_{\mathrm{t}}^{\mathrm{YM}} = \tau = \frac{F_0^2}{6}\big(M_{\eta_1}^2 + M_{\eta_8}^2 - 2M_K^2\big) \simeq \frac{F_\pi^2}{6}\big(M_{\eta'}^2 + M_\eta^2 - 2M_K^2\big). \tag{3.83}$$

Using the physical values of the meson masses given in Table 0.1, we arrive to an experimental estimate for the topological susceptibility

$$\chi_{\mathrm{t}}^{\mathrm{YM}} \approx (180\,\mathrm{MeV})^4. \tag{3.84}$$

At the time the Witten–Veneneziano mechanism was formulated, it was not possible to say if this value is actually correct, because 'we can neither measure nor calculate' [23] the topological susceptibility. Nevertheless, in the last years $\chi_{\mathrm{t}}^{\mathrm{YM}}$ has been measured in lattice YM theory Monte Carlo simulations. The results are in good agreement with the value in Eq. (3.84) and thus the conjecture of Witten and Veneziano is confirmed.

One main result of the thesis is the lattice computation of $\chi_{\mathrm{t}}^{\mathrm{YM}}(0)$ to percent level in the SU(3) YM theory (see Chapter 5 and Ref. [1]) and for the first time in the $N_{\mathrm{c}} \to \infty$ limit (see Chapter 6 and Ref. [3]).

4. The topological charge with the Yang–Mills gradient flow

In the last few years, a new definition of the topological charge has been proposed [29], whose cumulants have a finite and unambiguous continuum limit [29, 148]. It is a naïve discretization of the charge evolved with the Yang–Mills gradient flow. It is particularly appealing because its numerical evaluation is significantly cheaper with respect to the GW definition in Eq. (3.27).

In this chapter, we first introduce the YM gradient flow, both in the continuum and on the lattice, and the new topological charge definition. Then, we present the first original result of the thesis, published in Ref. [1]. We prove that, in YM theory, the cumulants of the topological charge at strictly positive flow time coincide, in the continuum limit, with those of the universal definition of Section 3.2.4. Therefore, the topological susceptibility obtained from evolving the naïve discretization with the YM gradient flow is a renormalized quantity and it satisfies the anomalous chiral WTIs of QCD.

4.1. The Yang–Mills gradient flow in the continuum

Starting from the ordinary continuum gauge field

$$B_\mu(t,x)\Big|_{t=0} = A_\mu(x), \tag{4.1}$$

where $A_\mu = A_\mu^a T^a$ (see Appendix A for the generator conventions), the Yang–Mills gradient flow evolves the gauge field as a function of the flow time $t \geq 0$ by solving the differential equation [29]

$$\partial_t B_\mu = D_\nu G_{\nu\mu} + \alpha_0 D_\mu \partial_\nu B_\nu, \tag{4.2a}$$

$$G_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu - \mathrm{i}\big[B_\mu, B_\nu\big], \qquad D_\mu = \partial_\mu - \mathrm{i}\big[B_\mu, \cdot\big], \tag{4.2b}$$

where the term multiplied by $\alpha_0 > 0$ is needed in perturbation theory to damp the gauge modes. Solutions with different α_0 are related by flow time-dependent gauge transformations, which do not affect gauge-invariant observables. Composite fields at positive flow time are built just using the evolved gauge field. Here we focus on the gradient-flow evolution of the topological charge density defined in the continuum in Eq. (3.2)[1]

$$q(t,x) = \frac{1}{32\pi^2}\epsilon_{\mu\nu\rho\sigma}\,\mathrm{tr}\big\{G_{\mu\nu}(t,x)G_{\rho\sigma}(t,x)\big\}, \tag{4.3}$$

[1] Unless stated otherwise, the explicit t on the quantities evolved with the gradient flow is always $t > 0$.

and of the corresponding topological charge

$$Q(t) = \int \mathrm{d}^4x\, q(t,x). \tag{4.4}$$

Under a generic variation δB_μ of a given gauge field configuration, the topological charge density changes as

$$\delta q(t,x) = \partial_\rho \tilde{w}_\rho(t,x), \qquad \tilde{w}_\rho(t,x) = \frac{1}{8\pi^2}\epsilon_{\rho\mu\nu\sigma}\,\mathrm{tr}\big\{G_{\mu\nu}(t,x)\,\delta B_\sigma(t,x)\big\}, \tag{4.5}$$

see for instance Ref. [149]. If we now specify

$$\delta B_\mu(t,x) = \partial_t B_\mu(t,x)\,\delta t\,, \tag{4.6}$$

it is straightforward to show that

$$\partial_t q(t,x) = \partial_\rho w_\rho(t,x), \qquad w_\rho(t,x) = \frac{1}{8\pi^2}\epsilon_{\rho\mu\nu\sigma}\,\mathrm{tr}\big\{G_{\mu\nu}(t,x)D_\alpha G_{\alpha\sigma}(t,x)\big\}, \tag{4.7}$$

where $w_\rho(t,x)$ is a local $d = 5$ gauge-invariant pseudovector field. This in turn implies that for a given gauge field configuration

$$\partial_t Q(t) = 0, \tag{4.8}$$

an equation that reflects the topological nature of $Q(t)$.

When $q(t,x)$ is inserted in a correlation function, Eq. (4.7) implies

$$\begin{aligned}\langle q(t,x)O_{\mathrm{ext}}\rangle &= \langle q(0,x)O_{\mathrm{ext}}\rangle + \int_0^t \mathrm{d}t'\,\partial_{t'}\langle q(t',x)O_{\mathrm{ext}}\rangle \\ &= \langle q(0,x)O_{\mathrm{ext}}\rangle + \partial_\rho \int_0^t \mathrm{d}t'\,\big\langle w_\rho(t',x)O_{\mathrm{ext}}\big\rangle,\end{aligned} \tag{4.9}$$

where O_{ext} is any finite field with support at a physical distance from x. The l.h.s. of Eq. (4.9) is finite thanks to the fact that a gauge-invariant local composite field constructed with the gauge field evolved at positive flow time is finite [29, 148]. Since there are no local composite fields of dimension $d < 5$ with the symmetry properties of $w_\rho(t,x)$, the integrand on the r.h.s. of Eq. (4.9) diverges at most logarithmically when $t' \to 0$. This implies that the quantity

$$\langle q(0,x)O_{\mathrm{ext}}\rangle \equiv \lim_{t\to 0}\langle q(t,x)O_{\mathrm{ext}}\rangle, \tag{4.10}$$

is finite, i.e. the limit on the r.h.s. exists for any finite field O_{ext}. Eq. (4.10) can be taken as the definition of $q(0,x)$, i.e. the renormalized topological charge density operator at $t = 0$. The latter satisfies the proper singlet chiral WTIs when fermions are included, see next section for an explicit derivation. It is worth noting that Eq. (4.9) implies that the small-t expansion of $q(t,x)$ is of the form

$$\langle q(t,x)O_{\mathrm{ext}}\rangle = \langle q(0,x)O_{\mathrm{ext}}\rangle + \mathcal{O}(t). \tag{4.11}$$

In the following we will be interested in the cumulants of the topological charge, as defined in Eq. (3.14) but evolved with the gradient flow

$$C_n(t) = \frac{1}{V}\big\langle Q^{2n}(t)\big\rangle_{\mathrm{con}} = \int \mathrm{d}^4x_1 \cdots \mathrm{d}^4x_{2n-1}\,\langle q(x_1)\cdots q(x_{2n-1})q(0)\rangle_{\mathrm{con}}, \tag{4.12}$$

which, thanks to Eq. (4.8), are expected to be independent of the flow time for $t \geq 0$ with the limit $t \to 0$ which requires some care due to the possible appearance of short-distance singularities. It is the aim of the next section to address this question in the regularized theory on the lattice.

To study the short-distance singularities, we use the technology introduced in Section 3.2.4. We supplement the theory with extra degenerate valence quarks of mass m, and we consider the (integrated) correlator of a topological charge density with a chain made of scalar and pseudoscalar densities [124] defined as

$$\langle q(0,x) \cdot P_{51}(z_1) S_{12}(z_2) S_{23}(z_3) S_{34}(z_4) S_{45}(z_5) \rangle, \tag{4.13}$$

where S_{ij} and P_{ij} are the scalar and the pseudoscalar renormalized densities with flavour indices i and j. Power counting and the operator product expansion predict that there are no non-integrable short-distance singularities when the coordinates of two or more densities in Eq. (4.13) tend to coincide among themselves or with x. When only one of the densities is close to $q(0,x)$, the operator product expansion predicts the leading singularity to be

$$q(0,x) S_{ij}(0) \xrightarrow{x \to 0} c(x) P_{ij}(0) + \cdots , \tag{4.14}$$

where

$$c(x) \overset{|x| \to 0}{\sim} \frac{1}{|x|^4} \tag{4.15}$$

and the dots indicate sub-leading contributions. An analogous expression is valid for the pseudoscalar density. Being the leading short-distance singularity in the product of fields $q(0,x) S_{ij}(0)$, its Wilson coefficient $c(x)$ can be computed in perturbation theory.

By using Eq. (4.9), to all orders in perturbation theory[2] we can write

$$\left\langle q(t,x) S_{ij}(0) O_{\text{ext}} \right\rangle = \left\langle q(0,x) S_{ij}(0) O_{\text{ext}} \right\rangle + \partial_\rho \int_0^t \mathrm{d}t' \left\langle w_\rho(t',x) S_{ij}(0) O_{\text{ext}} \right\rangle, \tag{4.16}$$

where again O_{ext} is any finite field localized at a physical distance from 0 and x. When $t > 0$, the l.h.s. of Eq. (4.16) has no singularities for $|x| \to 0$. The first term in the r.h.s. is integrable over spacetime: to all orders in perturbation theory we can substitute $q(x) = \partial_\mu K_\mu(x)$ and perform the integral on a large spherical surface. Similarly, the second term can be rewritten by Gauss law as an integral on a large spherical surface. Since $w_\rho(0,x)$ is a gauge-invariant current, the integral goes to zero at infinity. Therefore, any singularity in the first term in the r.h.s. must be of the form

$$c(x) = \partial_\rho u_\rho(x), \tag{4.17}$$

and cannot contribute to the integral (over all coordinates) of the correlation function Eq. (4.13).

[2] Since the function $|x|^{-4} \ln\left(x^2\right)^{-p}$ is integrable for $p > 1$, the singularity needs to be determined only up to some finite order.

4.2. Cumulants of the topological charge on the lattice

On the lattice the YM gradient-flow equation can be written as a first-order differential equation [29]

$$\partial_t V_\mu(t,x) = -g_0^2\{\partial_{x,\mu} S_{\rm W}[V]\}V_\mu(t,x), \qquad V_\mu(t,x)\Big|_{t=0} = U_\mu(x), \tag{4.18}$$

where $S_{\rm W}$ is Wilson's plaquette action of Eq. (2.24) and the link differential operators $\partial_{x,\mu}$ are defined in Section A.2. The gauge field evolved at positive flow time $V_\mu(x)$ is smooth on the scale of the cut-off. When inserted at a physical distance, the gauge-invariant local composite fields constructed with the evolved gauge field are finite as they stand. Remarkably their universality class is determined only by their asymptotic behaviour in the classical continuum limit [29, 148]. At $t > 0$ any decent definition of the topological charge density is therefore finite. The same line of argumentation applies to the cumulants of the topological charge. At $t > 0$ short-distance singularities cannot arise because of the exponential damping of the high-frequency components of the fields enforced by the flow evolution.

It remains to be shown, however, that the cumulants of the topological charge distribution defined at $t > 0$ satisfy the proper singlet chiral WTIs when fermions are included in the theory. To show this it is sufficient to work with a particular discretization of the topological charge, and then appeal to the above mentioned universality argument for the other definitions. The GW discretization of $q_{\rm GW}(x)$ in Eq. (3.27) has a privileged rôle: in Section 3.2 we showed that at $t = 0$ the lattice bare cumulants are finite, and they satisfy the singlet chiral WTIs when fermions are included in the theory.

4.2.1. Ginsparg–Wilson definition of the charge density

We introduce the GW definition of the topological charge density at positive flow time

$$q_{\rm GW}(t,x) = \frac{\bar{a}}{2a^4}\,\mathrm{tr}\big\{\gamma_5 D[V](x,x)\big\}, \tag{4.19}$$

which is given by Eq. (3.27) where $D[V](x,y)$ is a GW-satisfying Dirac operator, such as the Neuberger–Dirac operator, in which each link variable $U_\mu(x)$ is replaced by the corresponding evolved one $V_\mu(t,x)$ when $t > 0$. Since there are no other operators of dimension $d \le 4$ which are pseudoscalar and gauge-invariant,

$$\lim_{a\to 0} Z_q \langle q_{\rm GW}(t,0) q_{\rm GW}(0,x)\rangle = \text{finite}, \tag{4.20}$$

where Z_q is a renormalization constant which is at most logarithmically divergent, while $q_{\rm GW}(t,0)$ is finite as it stands. This in turn implies that

$$\lim_{a\to 0} Z_q \sum_x \langle q_{\rm GW}(t,0) q_{\rm GW}(0,x)\rangle = \text{finite}, \tag{4.21}$$

since there are no short-distance singularities that contribute to the integrated correlation function because $q_{\rm GW}(t,0)$ is evolved at positive flow time. Now, we substitute the $t = 0$

integrated $q_{\rm GW}$ in Eq. (4.21) with a density-chain expression in Eq. (3.44) of the topological charge [124], using at least five flavours of valence quarks of mass m

$$\sum_x \langle q_{\rm GW}(t,0) q_{\rm GW}(0,x)\rangle = -m^5 \sum_{z_1,\dots,z_5} \langle q_{\rm GW}(t,0) P_{51}(z_1) S_{12}(z_2) S_{23}(z_3) S_{34}(z_4) S_{45}(z_5)\rangle. \quad (4.22)$$

As argued in Section 3.2.4, power counting and the operator product expansion predict that there are no non-integrable short-distance singularities when the coordinates of two or more densities coincide. The r.h.s. of Eq. (4.22) is finite as it stands, and it converges to the continuum limit with a rate proportional to a^2. This in turn implies that the limits on the l.h.s. of Eqs (4.20) and (4.21) are reached with the same rate if Z_q is set to any fixed (g_0-independent) value. Since in the naïve continuum limit the GW definition in Eq. (4.19) has the same asymptotic behaviour of the definition in Eq. (4.3) [139, 140], we may set $Z_q = 1$. In this case

$$\lim_{a\to 0} \big\langle q_{\rm GW}(0,x) O_{\rm L,ext}\big\rangle = \big\langle q(0,x) O_{\rm ext}\big\rangle, \quad (4.23)$$

where $O_{\rm L,ext}$ is a discretization of the generic finite continuum field $O_{\rm ext}$. Once inserted in correlation functions at a physical distance from other (renormalized) fields, $q_{\rm GW}(0,x)$ does not require any renormalization in the Yang–Mills theory. It is finite as it stands, and it satisfies the singlet WTIs when fermions are included in the theory. It is interesting to note that Eqs (4.11) and (4.23) imply

$$\big\langle q(t,x) O_{\rm ext}\big\rangle = \big\langle q_{\rm GW}(0,x) O_{\rm L,ext}\big\rangle + \mathcal{O}\big(a^2\big) + \mathcal{O}(t), \quad (4.24)$$

where in general discretization effects depend on t.

With some extra assumptions, we could have arrived to Eq. (4.23) by following a procedure analogous to the one in the continuum, see Eqs (4.7)–(4.11). With no loss of generality, we can specify the Dirac operator in $q_{\rm GW}$ to be Neuberger's operator in Eq. (2.66). Then, to all orders in perturbation theory, or in general when the gauge field satisfies the smoothness condition in Eq. (2.70) [82], the change of the topological charge density with respect to the flow time can be written, analogously to Eq. (4.7), as [89, 150] (see also Ref. [151])

$$\partial_t q_{\rm GW}(t,x) = \partial^*_\rho w_{{\rm GW},\rho}(t,x), \quad (4.25)$$

where $w_{{\rm GW},\rho}(t,x)$ is a discretization of the $d = 5$ gauge-invariant pseudovector operator defined in Eq. (4.7).

4.2.2. Ginsparg–Wilson definition of cumulants

The GW definition of the topological charge cumulants at positive flow time is

$$C_{{\rm GW},n}(t) = \sum_{x_1,\dots,x_{2n-1}} \langle q_{\rm GW}(t,x_1)\cdots q_{\rm GW}(t,x_{2n-1}) q_{\rm GW}(t,0)\rangle_{\rm con}. \quad (4.26)$$

As discussed in Section 3.2, for $t = 0$ the cumulants have an unambiguous universal continuum limit as they stand and, when fermions are included, they satisfy the proper singlet chiral WTIs [27, 123, 124]. They are the proper quantities to be inserted in the Witten–Veneziano

relations for the mass and scattering amplitudes of the η' meson in QCD [23, 24, 27, 125]. It is far from being obvious that $C_{\mathrm{GW},n}(t=0)$ coincide with those defined at positive flow time, since the two definitions may differ by additional finite contributions from short-distance singularities.

For the clarity of the presentation we start by focusing on the lowest cumulant, the topological susceptibility $C_{\mathrm{GW},1}(t)$. At $t=0$, by replacing one of the two $q_{\mathrm{GW}}(t=0)$ with its density-chain expression [124], we obtain

$$\sum_x \langle q_{\mathrm{GW}}(0,0) q_{\mathrm{GW}}(0,x)\rangle = -m^5 \sum_{z_1,\dots,z_5} \langle q_{\mathrm{GW}}(0,0) P_{51}(z_1) S_{12}(z_2) S_{23}(z_3) S_{34}(z_4) S_{45}(z_5)\rangle. \quad (4.27)$$

When the susceptibility is written in this form, the discussion toward the end of Section 4.1 and in particular Eq. (4.17) guarantee that there are no contributions from short-distance singularities. This result, together with the fact that $Z_q = 1$, implies that

$$\lim_{t\to 0}\lim_{a\to 0} \sum_x \langle q_{\mathrm{GW}}(t,x) q_{\mathrm{GW}}(0,0)\rangle = \lim_{a\to 0} \sum_x \langle q_{\mathrm{GW}}(0,x) q_{\mathrm{GW}}(0,0)\rangle. \quad (4.28)$$

By replacing on the l.h.s. $q_{\mathrm{GW}}(0,0)$ with the evolved one, no further short-distance singularities are introduced and we arrive to the final result

$$\lim_{t\to 0}\lim_{a\to 0} \sum_x \langle q_{\mathrm{GW}}(t,x) q_{\mathrm{GW}}(t,0)\rangle = \lim_{a\to 0} \sum_x \langle q_{\mathrm{GW}}(0,x) q_{\mathrm{GW}}(0,0)\rangle. \quad (4.29)$$

By replacing $2n-1$ of the charges in the n^{th} cumulant with their density-chain definitions, the very same line of argumentation can be applied. The Eq. (4.29), together with the independence up to harmless discretization effects of $C_{\mathrm{GW},n}(t)$ from the flow time for $t>0$ [29], implies that the continuum limit of $C_{\mathrm{GW},n}(t)$ coincides with the one of $C_{\mathrm{GW},n}(t=0)$. The cumulants of the topological charge distribution defined at $t>0$ thus satisfy the proper singlet chiral WTIs when fermions are included [27, 123, 124]. They are the proper quantities to be inserted in the Witten–Veneziano relations for the mass and scattering amplitudes of the η' meson in QCD [23, 24, 27, 125].

4.2.3. Universality at positive flow time

For $t>0$, different lattice definitions of the topological charge density belong to the same universality class if they share the same asymptotic behaviour in the classical continuum limit [29, 148]. Starting from now and for the rest of the thesis, we are interested in the naïve discretization of the topological charge density in Eq. (3.17), evolved with the YM gradient flow

$$q_{\mathrm{L}}(t,x) = \frac{1}{32\pi^2}\epsilon_{\mu\nu\rho\sigma}\,\mathrm{tr}\big\{G_{\mu\nu}(x) G_{\rho\sigma}(x)\big\}, \quad (4.30)$$

and the corresponding cumulants

$$C_{\mathrm{L},n}(t) = \sum_{x_1,\dots,x_{2n-1}} \langle q_{\mathrm{L}}(t,x_1)\cdots q_{\mathrm{L}}(t,x_{2n-1}) q_{\mathrm{L}}(t,0)\rangle_{\mathrm{con}}. \quad (4.31)$$

As already mentioned, at $t=0$ the naïve lattice discretization $q_{\mathrm{L}}(0,x)$ in YM theory requires a multiplicative renormalization constant[3] when inserted in correlation functions at a physical

[3] This renormalization constant can be fixed by enforcing the analogous of Eqs (4.20) and (4.23).

distance from other operators [152]. Moreover, the cumulants of the corresponding topological charge $Q_{\mathrm{L}}(t=0)$ have additional short-distance power-divergent singularities, and they do not have a well defined continuum limit.

The situation is different at $t > 0$. The density $q_{\mathrm{L}}(t,x)$ in Eq. (4.30) shares with $q_{\mathrm{GW}}(t,x)$ the same asymptotic behaviour in the classical continuum limit [139, 140]. Since for $t > 0$ short-distance singularities cannot arise, $C_{\mathrm{GW},n}(t)$ in Eq. (4.26) and $C_{\mathrm{L},n}(t)$ in Eq. (4.31) tend to the same continuum limit. The results in the previous section then imply that the continuum limit of the naïve definition of $C_{\mathrm{L},n}(t)$, *at positive flow time*, coincides with the universal definition which satisfies the chiral WTIs when fermions are added [27, 123, 124]. It is interesting to note, however, that at fixed lattice spacing there can be quite some differences. For instance, the topological susceptibility defined at $t > 0$ with the naïve definition is not guaranteed to go to zero in the chiral limit at finite lattice spacing in presence of fermions [153].

5. Lattice computation of SU(3) topological susceptibility and non-Gaussianity

By implementing the gradient-flow definition of the topological charge density in Eq. (4.30), in this chapter we describe our original computation of the topological susceptibility in SU(3) YM theory. Preliminary results were presented in Ref. [5], published in final form in Ref. [1] and presented in Ref. [6]. We obtained a value of χ_t[1] in the continuum limit with a precision five times better than the reference computation with Neuberger's definition [154]. Moreover, we determined the ratio R defined in Eq. (3.67) of the forth cumulant over the second one in the continuum limit by keeping for the first time all systematics, especially finite volume effects, negligible with respect to the statistical errors. As a by-product we also performed an interesting universality test at the permille level by comparing the values of the topological susceptibility at different flow times.

5.1. Numerical setup

For the numerical computation we discretize the SU(3) Yang–Mills theory with standard Wilson's plaquette action on a finite four-dimensional lattice with spacing a, with the same L/a size in all four space-time directions, and with periodic boundary conditions imposed on the gauge fields, see Appendix A for details. The basic Monte Carlo update of each link variable implements the Cabibbo–Marinari scheme [96], by sweeping the full lattice with one heat bath update followed by $L/(2a)$ sweeps of over-relaxation updates [155].

5.1.1. Ensembles generated

We have simulated three series of lattices in order to estimate and remove the systematic effects due to the finiteness of the lattice spacing and volume, see Table 5.1 for details. In the first series $\{A_1, B_1, \dots, F_1\}$ the inverse coupling $\beta = 6/g_0^2$ is kept fixed so that the lattice spacing is approximatively 0.1 fm, while the physical volume increases from $(1.0\,\mathrm{fm})^4$ to $(1.6\,\mathrm{fm})^4$. The number N_{conf} of independent gauge configurations generated scales with L^8 to ensure that the relative statistical error on R, see Eq. (5.2), is always at the 10 % level [156]. In the second series $\{B_1, \dots, B_4\}$ the physical volume is kept approximatively fixed, while the spacing is decreased down to 0.068 fm. The volume is always $(1.2\,\mathrm{fm})^4$ to guarantee that finite-size effects on R are within the statistical errors, while the computational cost remains affordable. In

[1] To simplify the notation, in this chapter χ_t always denotes the topological susceptibility of YM theory.

Table 5.1.: Overview of the ensembles and statistics used in this study. For each lattice we give the label, $\beta = 6/g_0^2$, the reference scale t_0/a^2, the spatial extent of the lattice, the lattice spacing, the number N_{conf} of independent configurations generated, the number of sweeps `nit` required to space them, and the tolerances `eerr`, `q2err` and `q4err` on the primary observables considered (see main text).

Lattice	β	t_0/a^2	L/a	L [fm]	a [fm]	N_{conf}	`nit`	`eerr`	`q2err`	`q4err`
A_1	5.96	2.79	10	1.0	0.102	36 000	30	0.19	0.0005	0.0024
B_1			12	1.2		144 000		0.45		0.005
C_1			13	1.3		280 000		0.42		0.0068
D_1			14	1.4		505 000		0.74		0.01
E_1			15	1.5		880 000		0.89		0.012
F_1			16	1.6		1 440 000		1.04		0.015
B_2	6.05	3.78	14	1.2	0.087	144 000	60	0.31	0.0005	0.005
D_2			17	1.5		144 000		0.045		0.01
B_3	6.13	4.87	16	1.2	0.077	144 000	90	0.25	0.0005	0.005
D_3			19	1.5		144 000		0.058		0.01
B_4	6.21	6.20	18	1.2	0.068	144 000	250	0.20	0.0005	0.005
D_4			21	1.4		144 000		0.042		0.01

the third series $\{D_1, \dots, D_4\}$ is again the physical volume which is kept approximatively fixed, always at least $(1.4\,\text{fm})^4$, to guarantee that finite-size effects in the reference scale t_0 and in the topological susceptibility χ_t, see Eq. (5.2), are within their (smaller) statistical errors. In both cases the measurements at the four lattice spacings are used to estimate discretization effects in the observables, and to extrapolate them away in the continuum limit.

5.1.2. Computation of the observables

The primary observables that we have computed on each configuration at $t \geq 0$ are: the energy density, averaged over the four-volume,

$$E_{\text{L}}(t) = \frac{1}{4V} \sum_x e_{\text{L}}(t,x), \qquad e_{\text{L}}(t,x) = F^a_{\text{L}\mu\nu}(t,x) F^a_{\text{L}\mu\nu}(t,x), \tag{5.1}$$

where $F^a_{\text{L}\mu\nu}(t,x)$ is defined in Eq. (3.18); and the topological charge $Q_{\text{L}}(t)$, defined from the density in Eq. (4.30). The quantum averages we are interested in are

$$\langle E_{\text{L}}(t) \rangle, \qquad \chi_t(t) = \frac{\langle Q^2_{\text{L}}(t) \rangle}{V}, \qquad R(t) = \frac{\langle Q^4_{\text{L}}(t) \rangle_{\text{con}}}{\langle Q^2_{\text{L}}(t) \rangle}. \tag{5.2}$$

To numerically integrate the Yang–Mills gradient flow we have implemented a fourth order *Runge–Kutta–Munthe-Kaas (RKMK) method* [157–159]. It is a *structure-preserving* Runge–Kutta (RK) integrator, designed to exactly preserve the Lie group structure of the gradient flow

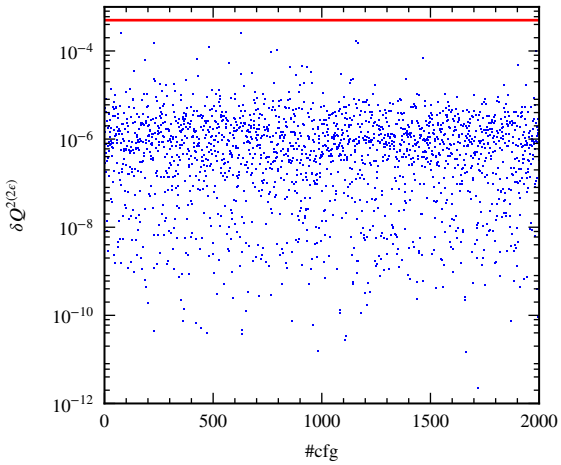


Figure 5.1.: History plot of the systematic error $\delta Q_{\mathrm{L}}^{2,(2\epsilon)}$ at $t \simeq t_0$ for the first 2000 configurations of D_4. The red line indicates the bound `q2err` $= 0.0005$ of the systematic error enforced on all configurations.

equation, see Appendix F for details. On each lattice the field has been evolved approximatively up to $t = 1.2t_0$, where t_0 is the reference flow time value defined below. The observables in Eq. (5.2) have been computed with a flow-time resolution of $0.08a^2$ or smaller. The numerical integration of the flow equation introduces a systematic error in the gauge field values at positive flow time t, and thus in each observable. In our case, at small values of the RK step size, it is proportional to ϵ^4. There are, however, large fluctuations in the pre-factor among the various gauge configurations, see Figure 5.1. A reliable estimate of this systematics is achieved by monitoring the error configuration by configuration, and occasionally adapt the step ϵ. To do so we integrate the flow equation two times with steps ϵ and 2ϵ, where in our case $\epsilon = 0.08a^2$. Denoting with $E_{\mathrm{L},j}^{(\epsilon)}$ and $Q_{\mathrm{L},j}^{(\epsilon)}$ the basic observables E_{L} and Q_{L} respectively computed on the j^{th} field configuration evolved with step size ϵ, at small enough ϵ the error is given by

$$\delta E_{\mathrm{L},j}^{(2\epsilon)} = \left| E_{\mathrm{L},j}^{(\epsilon)} - E_{\mathrm{L},j}^{(2\epsilon)} \right|, \qquad \delta Q_{\mathrm{L},j}^{(2\epsilon)} = \left| Q_{\mathrm{L},j}^{(\epsilon)} - Q_{\mathrm{L},j}^{(2\epsilon)} \right|, \tag{5.3}$$

with both observables obviously measured at the same flow time. By applying linear propagation, the error on the average over all configurations is bounded by

$$\delta \bar{E}_{\mathrm{L}}^{(2\epsilon)} \leq \max_j \left(\delta E_{\mathrm{L},j}^{(2\epsilon)} \right), \tag{5.4a}$$

$$\delta \bar{Q}_{\mathrm{L}}^{2,(2\epsilon)} \leq \max_j \left(\left| 2Q_{\mathrm{L},j} \right| \delta Q_{\mathrm{L},j}^{(2\epsilon)} \right), \qquad \delta \bar{Q}_{\mathrm{L}}^{4,(2\epsilon)} \leq \max_j \left(\left| 4Q_{\mathrm{L},j}^3 \right| \delta Q_{\mathrm{L},j}^{(2\epsilon)} \right), \tag{5.4b}$$

$$\delta \bar{R}^{(2\epsilon)} \leq \frac{1}{\bar{Q}_{\mathrm{L}}^2} \max \left(\max_j \left(\left| 4Q_{\mathrm{L},j}^3 \right| \delta Q_{\mathrm{L},j}^{(2\epsilon)} \right), \frac{\bar{Q}_{\mathrm{L}}^4 + 3(\bar{Q}_{\mathrm{L}}^2)^2}{\bar{Q}_{\mathrm{L}}^2} \max_j \left(\left| 2Q_{\mathrm{L},j} \right| \delta Q_{\mathrm{L},j}^{(2\epsilon)} \right) \right). \tag{5.4c}$$

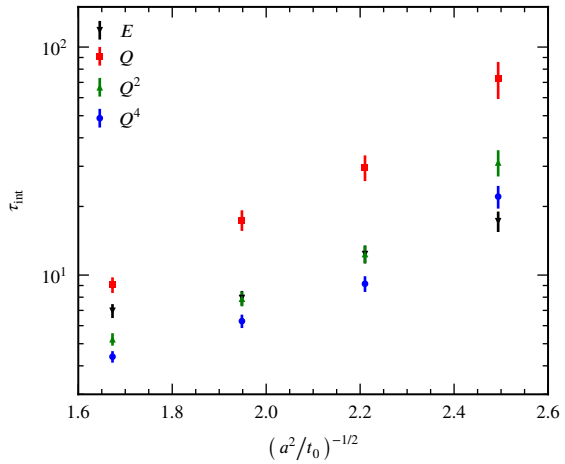


Figure 5.2.: The integrated autocorrelation times τ_{int} of the primary observables as a function of $(a^2/t_0)^{-1/2}$.

At run-time, for each configuration and each flow time the systematic errors of the observables E_{L}, Q_{L}^2 and Q_{L}^4 are compared with the given tolerances `eerr`, `q2err` and `q4err` respectively. If one of the tests fails, the flow evolution is re-computed for that configuration with a new step size $\epsilon' = (1/2)\epsilon$ and new observables data, along with old $\epsilon = 2\epsilon'$ data, are used to estimate the systematic errors and compare them with the tolerances. If the test fails again, the field is evolved with $\epsilon'' = (1/2)\epsilon'$, and so on. This ensures that

$$\delta\bar{E}_{\mathrm{L}}^{(2\epsilon)} \leq \texttt{eerr}, \qquad \delta\bar{Q}_{\mathrm{L}}^{2,(2\epsilon)} \leq \texttt{q2err}, \qquad \delta\bar{Q}_{\mathrm{L}}^{4,(2\epsilon)} \leq \texttt{q4err}. \tag{5.5}$$

The parameters `eerr`, `q2err` and `q4err` are chosen as a function of the target statistical error on the corresponding observables. If we set the upper limit for the systematic error to be roughly 10 times smaller than the statistical one, this condition is readily translated into a limit for `eerr`, `q2err` and `q4err`, see Table 5.1 for the values chosen for each lattice. The Eqs (5.5) put bounds on the systematic errors for the coarser evolution, but the data evolved with the finer step size ϵ are actually those used in the final analysis. This choice is rather conservative in our case, being the actual error more than one order of magnitude smaller. For the quantity $E(t)$ the actual error turns out to be more than two orders of magnitude smaller with respect to the bound in Eq. (5.5), see Figure F.1 in Appendix F. We have therefore chosen larger values for `eerr` with respect to one given by the bound in Eq. (5.5).

5.1.3. Autocorrelation times

To measure the autocorrelation time of the various observables, we perform a dedicated run for each lattice $\{B_1, \ldots, B_4\}$ where the gauge field configurations are separated by a single iteration of the update algorithm. Each series is replicated 36 times to increase statistical accuracy. The integrated autocorrelation times τ_{int} of the observables E_{L}, Q_{L}, Q_{L}^2, and Q_{L}^4, estimated as in

Table 5.2.: Integrated autocorrelation times of the various observables in units of a single sweep of the update algorithm. They have been measured on dedicated runs made of 36 series of 1000 sweeps each.

Lattice	N_{conf}	t/a^2	$\tau_{\text{int}}^{E_{\text{L}}}$	$\tau_{\text{int}}^{Q_{\text{L}}}$	$\tau_{\text{int}}^{Q_{\text{L}}^2}$	$\tau_{\text{int}}^{Q_{\text{L}}^4}$
B_{1a}	36×1000	3.36	7.0(5)	9.1(7)	5.2(3)	4.39(25)
B_{2a}	36×1000	4.64	7.9(6)	17.4(18)	7.9(6)	6.3(4)
B_{3a}	36×1000	6.08	12.4(11)	30(4)	12.4(11)	9.2(7)
B_{4a}	36×1000	7.68	17.2(18)	73(13)	31(4)	22.1(25)

Ref. [160], are reported in Table 5.2. In the range of β values considered, Q has the largest autocorrelation time which increases rapidly toward the continuum limit [161]. To ensure that the measurements in the main runs are statistically independent, we have spaced them by `nit` sweeps of the lattice, see Table 5.1.

5.2. Physics results

A first analysis of the data reveals the effectiveness of the gradient flow in splitting the field space of the lattice theory into different topological sectors. In Figure 5.3 we plot the histograms of the topological charge Q measured at different flow times on the lattice D_4. In the plot on the top-left corner, the topological charge distribution at $t = 0$ is a smooth function over non-integer values. By increasing the flow time, the configurations with charge close to integers become more and more probable. The spikes in the bottom-right plot turn out to be slightly shifted towards zero with respect to the integer values due to discretization effects. In Figure 5.4 we plot the same histograms measured at the same value of flow time $t = 1/2t_0$, but at different lattice spacings. These plots show that the configuration space with non-integer topological charge gets suppressed and topological sectors emerge dynamically when $a \to 0$.

5.2.1. Scale setting

The reference flow time t_0 is defined through the implicit equation [29]

$$t^2\langle E(t)\rangle\Big|_{t=t_0} = 0.3. \tag{5.6}$$

In the region of interest $t^2\langle E_{\text{L}}(t)\rangle$ grows approximatively as a linear function of t. Since we have computed $\langle E_{\text{L}}(t)\rangle$ at flow times spaced by finite steps, we have solved equation (5.6) by interpolating linearly the two data points closest to t_0. The results are reported in Table 5.3, with the systematic error due to the interpolation being negligible. By comparing the values of t_0/a^2 obtained on the lattices $\{A_1, \dots, F_1\}$, finite-size effects are not visible at the level of 0.1 permille in the statistical precision for $L \geq 1.4\,\text{fm}$. We thus fix the lattice spacing at all values of β from t_0/a^2 determined on the lattices $\{D_1, \dots, D_4\}$.

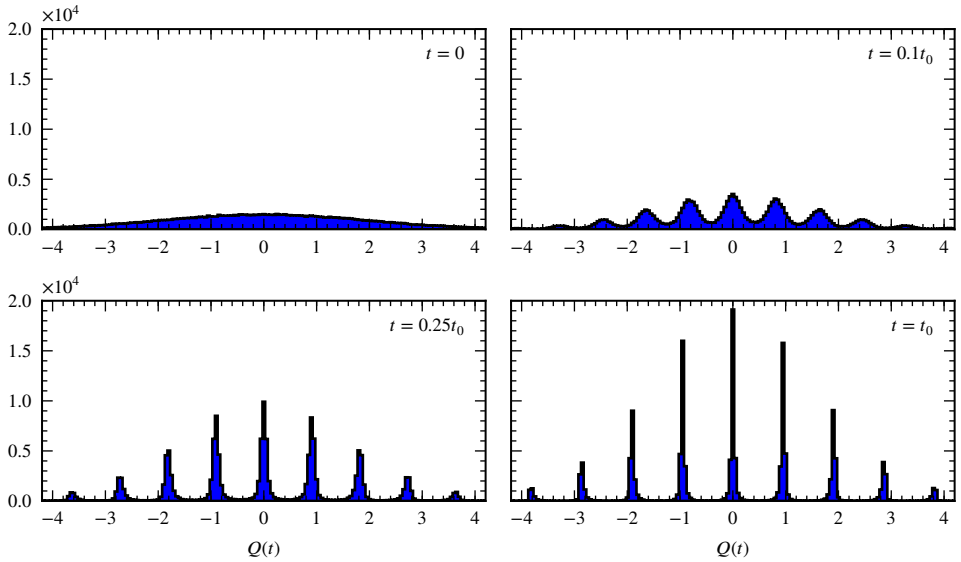


Figure 5.3.: Histograms of the topological charge distribution measured on the lattice D_4 at different flow times.

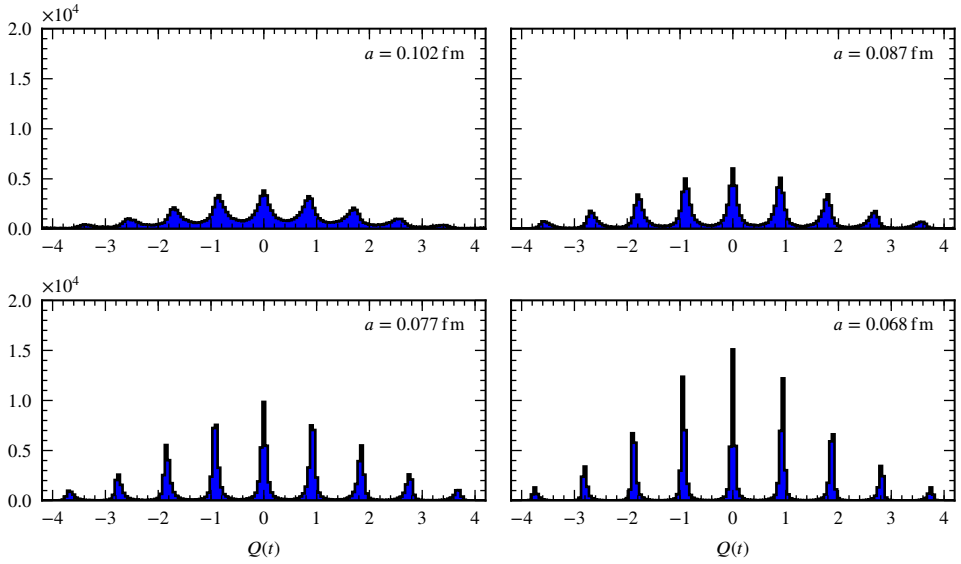


Figure 5.4.: Histograms of the topological charge distribution measured on lattices D_i at flow time $t = 1/2t_0$. Only the first 144 000 configurations of D_1 are included. The figure is adapted from Ref. [162].

Table 5.3.: Results for the reference flow time t_0/a^2 and the ratio t_0/r_0^2. The error on the latter is dominated by the 0.3–0.6 % relative error on r_0/a quoted in [163].

Lattice	t_0/a^2	t_0/r_0^2
A_1	2.995(4)	0.1195(9)
B_1	2.7984(9)	0.1117(9)
C_1	2.7908(5)	0.1114(9)
D_1	2.7889(3)	0.1113(9)
E_1	2.788 92(23)	0.1113(9)
F_1	2.788 67(16)	0.1113(9)

Lattice	t_0/a^2	t_0/r_0^2
B_2	3.7960(12)	0.1114(9)
B_3	4.8855(15)	0.1113(10)
B_4	6.2191(20)	0.1115(11)
D_2	3.7825(8)	0.1110(9)
D_3	4.8722(11)	0.1110(10)
D_4	6.1957(14)	0.1111(11)

In Table 5.3 the values of t_0/r_0^2, where r_0 is the Sommer scale defined in Eq. (2.121) and computed in Ref. [163], are also reported. As shown in Figure 5.5, discretization effects in this ratio are indeed negligible with respect to the statistical errors dominated by the 0.3–0.6 % error on r_0/a. By extrapolating the results linearly in a^2/t_0, we obtain in the continuum limit

$$\frac{\sqrt{8t_0}}{r_0} = 0.941(7), \tag{5.7}$$

which corresponds to $t_0/r_0^2 = 0.1108(17)$. As discussed in Section 1.3.2, using the value of the Λ-parameter computed in Ref. [54] we also have

$$\sqrt{t_0}\Lambda_{\overline{\text{MS}}} = 0.200(16). \tag{5.8}$$

Since YM theory is not realized in Nature, any conversion of this result to physical units is a matter of convention. For the sake of clarity in the presentation, however, it is useful to assign a physical value to t_0, which we choose to be

$$\sqrt{t_0} = 0.166\,\text{fm}. \tag{5.9}$$

This is motivated by Eq. (5.7) together with the conventional value of the Sommer scale $r_0 = 0.5\,\text{fm}$ [112]. We use this value of t_0 to express the lattice sizes and lattice spacings in physical units.[2]

5.2.2. Topological susceptibility

The full set of results for the topological charge moments and cumulants are given in Table 5.4. They are computed[3] at the reference flow time t_0 by linearly interpolating the numerical data as described in the previous section.

[2] As an alternative, we can supplement the theory with quenched quarks. The value of $F_K r_0 = 0.293(7)$ from Ref. [164] together with $F_K = 109.6\,\text{MeV}$ leads to

$$\sqrt{t_0} = 0.176(4)\,\text{fm},$$

the error being dominated by the one on $F_K r_0$.

[3] Unless explicitly indicated, the gradient flow time at which the topological quantities are computed throughout this and the next sections is $t = t_0$.

Table 5.4.: Results for the various topological observables measured at flow time t_0 on all lattices simulated.

Lattice	$\langle Q_{\mathrm{L}}^2 \rangle$	$\langle Q_{\mathrm{L}}^4 \rangle$	$\langle Q_{\mathrm{L}}^4 \rangle_{\mathrm{con}}$	R
A_1	0.701(6)	1.75(4)	0.273(20)	0.39(3)
B_1	1.617(6)	8.15(7)	0.30(4)	0.187(24)
C_1	2.244(6)	15.50(10)	0.40(5)	0.177(23)
D_1	3.028(6)	28.14(14)	0.63(7)	0.209(23)
E_1	3.982(6)	48.38(18)	0.81(9)	0.202(23)
F_1	5.167(6)	80.90(22)	0.81(11)	0.157(22)
B_2	1.699(7)	9.07(9)	0.41(5)	0.24(3)
D_2	3.686(14)	41.6(4)	0.83(19)	0.22(5)
B_3	1.750(7)	9.58(9)	0.39(5)	0.22(3)
D_3	3.523(13)	37.8(3)	0.56(17)	0.16(5)
B_4	1.741(7)	9.44(9)	0.35(5)	0.20(3)
D_4	3.266(12)	32.7(3)	0.68(15)	0.21(5)

In Figure 5.6 we show the values of the topological susceptibility $\chi_{\mathrm{t}} = \langle Q_{\mathrm{L}}^2 \rangle / V$ from the lattices $\{A_1, \ldots, F_1\}$ as a function of the linear extension of the lattice. For $L \geq 1.4\,\mathrm{fm}$, finite-size effects turn out to be below our target statistical error of approximatively 0.5 %. The continuum value of $t_0^2 \chi_{\mathrm{t}}$ can thus be obtained by extrapolating the results from the lattices $\{D_1, \ldots, D_4\}$, see left plot of Figure 5.7. Symanzik's effective theory analysis predicts discretization errors to start at $\mathcal{O}(a^2)$, and indeed the four data points are compatible with a linear behaviour in a^2. A linear fit of all of them gives as intercept $t_0^2 \chi_{\mathrm{t}} = 6.75(4) \times 10^{-4}$ with a significance of $\chi^2/\mathrm{dof} = 1.26$. A quadratic fit gives $t_0^2 \chi_{\mathrm{t}} = 6.49(18) \times 10^{-4}$ with $\chi^2/\mathrm{dof} = 0.38$, and with a coefficient of the quadratic term compatible with zero within the statistical errors. By restricting the linear fit to the three points at the finer lattice spacings, we obtain

$$t_0^2 \chi_{\mathrm{t}} = 6.67(7) \times 10^{-4}, \tag{5.10}$$

with $\chi^2/\mathrm{dof} = 0.88$, which is our best result for this quantity. It is five times more precise than the determination which uses Neuberger's definition of the topological charge [154].

The cumulants of the topological charge are expected to be t-independent in the continuum limit. In the right plot of Figure 5.7 we show the topological susceptibility computed at various flow times normalized to its value at t_0. The data points have statistical errors which range from 0.1 to 1 permille due to the correlation between the numerator and the denominator. At finite lattice spacing discretization effects are clearly visible, and they depend on t. When each set of data is extrapolated to the continuum limit with a quadratic function in a^2/t_0, the intercepts are all compatible with 1 within the statistical errors which, depending on t, range from 0.5 to 5 permille. We can also compare our result in Eq. (5.10) with the one obtained almost 10 years ago with Neuberger's definition of the topological charge [154]. If we use Eqs (5.7) and (5.10), we obtain

$$r_0^4 \chi_{\mathrm{t}} = 0.0544(18), \tag{5.11}$$

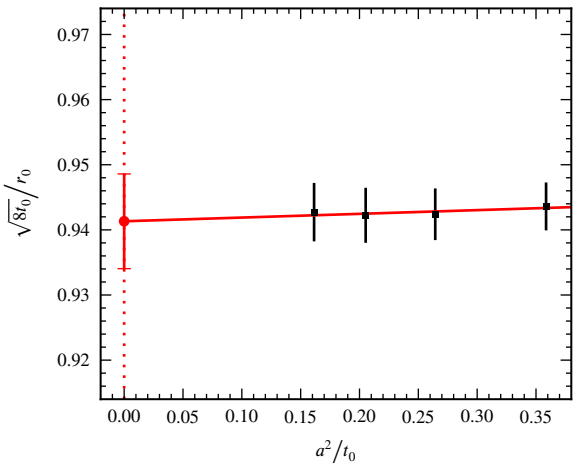


Figure 5.5.: Continuum limit extrapolation of $\sqrt{8t_0}/r_0$ computed on the lattices $\{D_1, \ldots, D_4\}$. The errors are dominated by the 0.3–0.6 % relative error on r_0/a quoted in Ref. [163].

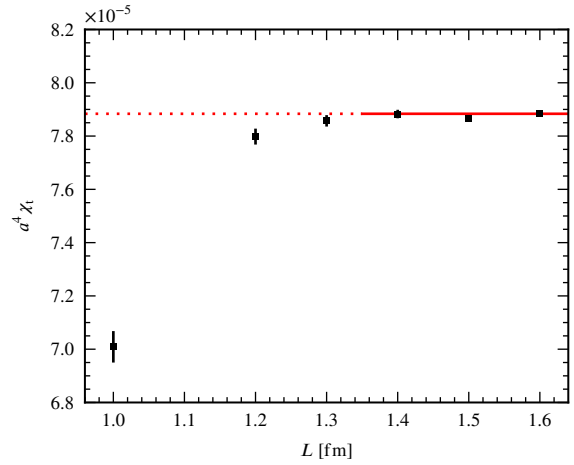


Figure 5.6.: Values of $a^4\chi_t$ as a function of L for the series $\{A_1, \ldots, F_1\}$.

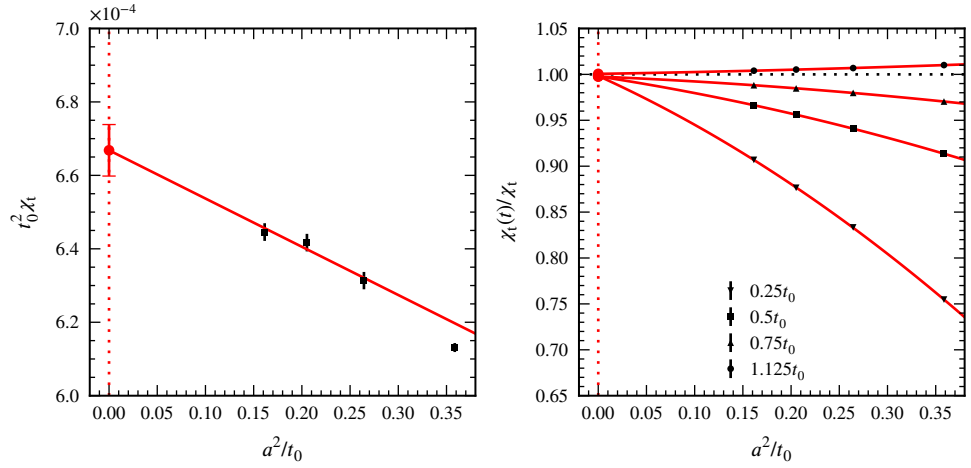


Figure 5.7.: Left: the dimensionless quantity $t_0^2\chi_t$ as a function of a^2/t_0, and its extrapolation to the continuum limit. Right: the ratio $\chi_t(t)/\chi_t$ (errors are smaller than symbols) as a function of a^2/t_0 for several values of the flow time t, and its extrapolation to the continuum limit.

which differs by less than 1.5 standard deviations[4] from the result in Eq. (11) of Ref. [154]. It is interesting to note that after ten years from the first computation of χ_t in the continuum limit [154], we moved from an unsolved problem to a universality test at the permille level.[5] The same result is given in units of the Λ-parameter using Eq. (5.8), albeit with a big error

$$\frac{\chi_t^{1/4}}{\Lambda_{\overline{\text{MS}}}} = 0.80(6). \tag{5.12}$$

By using Eq. (5.9), the value of χ_t in physical units is given by[6]

$$\chi_t = (191.0(5)\,\text{MeV})^4. \tag{5.13}$$

If we use the physical value of t_0 estimated using quenched quarks, we obtain a value of χ_t in physical units which differs downwards by $\sim 10\,\%$ per linear dimension, while if we use the physical value of t_0 determined in QCD with $N_f = 2$ and $N_f = 2 + 1$ flavours [167, 168], χ_t differs upwards by approximatively the same amount. This is the size of the ambiguity which is expected when results of the Yang–Mills theory are expressed in physical units.

[4] This value takes into account the fact that the same determination of r_0 is used in the two computations.

[5] A first test of universality for χ_t was already presented in Ref. [165] with statistical errors more than one order of magnitude larger than those obtained here. Results with similar large statistical errors were recently obtained in Ref. [166].

[6] Note that in Ref. [154], $F_K r_0 = 0.293(7)$ [164] and $F_K = 113.1$ MeV were used to set the scale in physical units.

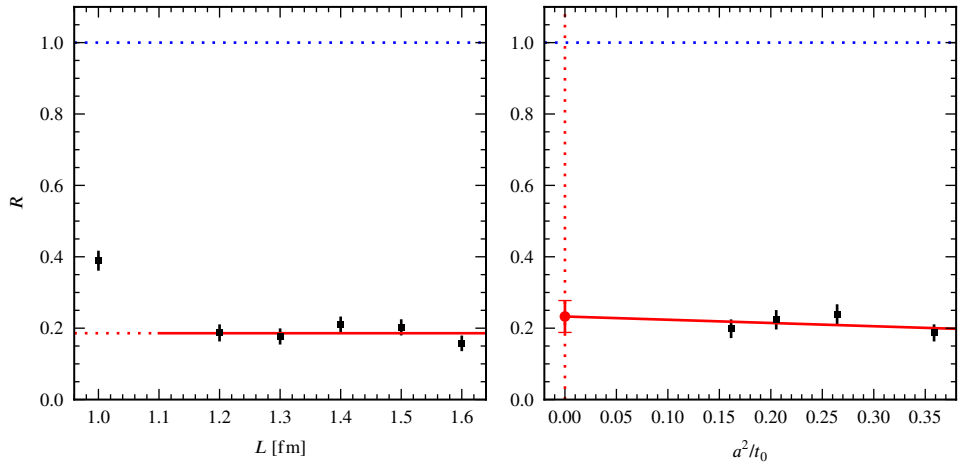


Figure 5.8.: Left: values of R at flow time t_0 versus L for the series $\{A_1, \ldots, F_1\}$. Right: the quantity R as a function of a^2/t_0 and its extrapolation to the continuum limit; the dotted blue line is the dilute instanton gas model prediction $R = 1$.

5.2.3. The ratio R

The values of $R = \langle Q_{\mathrm{L}}^4 \rangle_{\mathrm{con}} / \langle Q_{\mathrm{L}}^2 \rangle$ from the lattices $\{A_1, \ldots, F_1\}$ are shown in the left plot of Figure 5.8 as a function of L. Since our target statistical error is approximatively 10 %, a linear extension of $L \geq 1.2\,\mathrm{fm}$ is enough for finite-size effects to be within errors. Given the increase with L^8 of the computational cost of R, we have chosen to determine its continuum limit by extrapolating the data from the lattices $\{B_1, \ldots, B_4\}$, see left plot of Figure 5.8. Also in this case Symanzik's effective theory analysis predicts discretization errors to start at $\mathcal{O}(a^2)$, and indeed the four data points are compatible with a linear behaviour in a^2. A fit to a constant of all of them gives $R = 0.210(13)$ with a significance of $\chi^2/\mathrm{dof} = 0.83$. A linear fit in a^2/t_0 gives

$$R = 0.233(45), \tag{5.14}$$

which is our best result for this quantity. The significance of the fit is $\chi^2/\mathrm{dof} = 1.1$, and the slope is compatible with zero.

The value in Eq. (5.14) is compatible with the one obtained with Neuberger's definition in Ref. [156], albeit with an error 2.5 times smaller. It is also relevant to note that a full-fledged systematic study of finite-size effects was not carried out in Ref. [156], and finite-size effects were estimated and added to the final error.

5.3. Conclusions

The θ-dependence of the vacuum energy, or equivalently the functional form of the topological charge distribution, is a distinctive feature of the ensemble of gauge configurations that dominate

the path integral of a Yang-Mills theory. The value of $R = 0.233(45)$ in Eq. (5.14) rules out the θ-behaviour predicted by the dilute instanton gas model. Its large distance from 1 implies that, in the ensemble of gauge configurations that dominate the path integral, the fluctuations of the topological charge are of quantum *non-perturbative* nature. The large N_c expansion does not provide a sharp prediction for R. Its small value, however, is compatible with being a quantity suppressed as $1/N_c^2$ in the limit of large number of colours N_c. The value of R found here is related via the Witten–Veneziano mechanism to the leading anomalous contribution to the η'–η' elastic scattering amplitude in QCD. It is one of the low-energy constants which enter the effective theory of QCD when its Green's functions are expanded simultaneously in powers of momenta, quark masses and $1/N_c$.

The Yang–Mills gradient flow provides a reference scale and a sensible definition of the topological charge which are cheap to be computed numerically. With a modest numerical effort by today standards, it allowed us to compute the dimensionless ratio $t_0^2\chi = 6.67(7) \times 10^{-4}$ with a relative error of roughly 1 % in the continuum limit, i.e. five times smaller than the one of the previous reference computation with Neuberger's definition.

6. Lattice computation of $\mathrm{SU}(N_\mathrm{c})$ topological susceptibility and $N_\mathrm{c} \to \infty$ limit

The limit of large number of colours N_c has proved to be a fruitful tool in the study of $\mathrm{SU}(N_\mathrm{c})$ Yang–Mills theories [22]. As we discussed in Section 3.3, the Witten-Veneziano formula—explaining the large value of the mass of the η' meson in the chiral limit—relies on the large-N_c expansion. The WV formula is exact in the $N_\mathrm{c} \to \infty$ limit. However, the value of $\chi_\mathrm{t}^{\mathrm{YM}}$ found in SU(3) YM theory [154] is large enough to solve the $\mathrm{U}(1)_A$ problem in QCD. The fact that both $\chi_\mathrm{t}^{\mathrm{YM}}$ and $M_{\eta'}$ receive large-N_c corrections makes extremely interesting to study their values in the 't Hooft limit. In this chapter, we focus on the $N_\mathrm{c} \to \infty$ limit of the YM topological suggestibility.

Exploratory computations with cooling techniques at large N_c have a long tradition on the lattice [169, 161, 170], with quoted errors for the topological susceptibility at the 10 % level. These results, however, reflect the short-comings of the techniques available at the time. In particular, a theoretically sound definition of the topological susceptibility with a well-defined and universal continuum limit had not been used. Only Ref. [171] opted for the theoretically clean but expensive definition via the index of a chiral Dirac operator, and was therefore limited to a very coarse lattice spacing and small statistics.

The second problem affecting all simulations concerned with topological quantities is the quickly freezing topological charge as the continuum limit is approached. At large values of N_c this makes it exceedingly hard to perform reliable simulations at small lattice spacings, since the number of updates needed rises dramatically with the inverse lattice spacing [161, 172]. This comes on top of the increase of the cost of the updates growing with N_c^3, such that it cannot be overcome by a brute force approach.

Taking advantage of the conceptual, algorithmic and technical developments of the last decade, we are in the position to improve significantly over these results. The exceptional slowing down of the topological modes can be avoided by using open boundary conditions in time [109]. With the introduction of the gradient flow, a theoretically clean and numerically cheap definition of the topological charge has become available [29, 165]. In the continuum limit the corresponding topological susceptibility satisfies the singlet chiral Ward identities when fermions are included, and is the proper quantity to be inserted in the Witten–Veneziano formula, as proven in Chapter 4.

In this chapter, we present our original computation, published in Ref. [3], of the topological susceptibility in the large-N_c and continuum limits with percent accuracy. The same result was presented in Ref. [7]. We measured χ_t[1] for the groups SU(4), SU(5) and SU(6), and combined

[1] As in the previous chapter, in the following χ_t denotes the topological susceptibility of YM theory.

the results with previous ones for SU(3) reported in Chapter 5. Since leading corrections are expected to be $\mathcal{O}(N_c^{-2})$, this gives us a factor of four in their size. For each group the three lattice spacings simulated ranged from 0.096 fm to 0.065 fm with leading $\mathcal{O}(a^2)$ discretization effects decreasing by more than a factor 2 in size. This coverage of parameter space allowed for a robust extrapolation of the results to the large-N_c and continuum limits.

This chapter starts generalizing the definition of the t_0 scale to $N_c \neq 3$ in Section 6.1, followed by the details of the lattice setup in Section 6.2. The extrapolations to the continuum and large-N_c limit, giving the final results, are presented in Section 6.3 before some concluding remarks.

6.1. Definition of the reference scale t_0

In order to relate results in theories with different N_c, we need to define a reference scale in terms of which the observables are expressed. While different choices are logically possible, it is desirable to choose a quantity which is a (non-zero) constant at leading order in $1/N_c$, and that can be computed with high numerical precision. We opt for generalizing t_0 defined in Eq. (5.6) to arbitrary values of N_c, by requiring

$$t^2\langle E(t)\rangle\Big|_{t=t_0} = 0.1125\frac{N_c^2-1}{N_c}, \tag{6.1}$$

such that the right hand side attains the canonical value of 0.3 for SU(3). At small t, perturbation theory gives

$$t^2\langle E(t)\rangle = \frac{3}{128\pi^2}\frac{N_c^2-1}{N_c}\lambda_t(q)\big[1+c_1\lambda_t(q)+\mathcal{O}\big(\lambda_t^2\big)\big], \tag{6.2}$$

where $\lambda_t(q) = N_c g^2(q)$ at the scale $q = (8t)^{-1/2}$ is the renormalized 't Hooft coupling, and $c_1 = (1/16\pi^2)[(11/3)\gamma_E + 52/9 - 3\ln 3]$. The sub-leading term on the r.h.s. of Eq. (6.1) has been included following the indication of the perturbative expression.

Since SU(N_c) YM theory is not realized in Nature, any conversion to physical units is purely conventional. Following Chapter 5, we use the value of $\sqrt{t_0} = 0.166\,\text{fm}$ in Eq. (5.9) for all values of N_c to express the lattice sizes and lattice spacings in physical units. In Chapter 5 we used $\sqrt{t_0}\Lambda_{\overline{\text{MS}}} = 0.200(16)$ for $N_c = 3$. To our knowledge, a computation of the Λ-parameter for $N_c > 3$ is missing. It would be desirable in the future to compute $\Lambda_{\overline{\text{MS}}}$ at higher N_c, and eventually take the $N_c \to \infty$ limit.

6.2. Lattice details

The standard discretization of SU(N_c) YM theory on four-dimensional lattices of size $T \times L^3$ and lattice spacing a is used throughout this study. Differently from the SU(3) study in Chapter 5, we opted for open boundary conditions in time as implemented in Ref. [109], while spatial directions are periodic. This is due to the long autocorrelation time of the topological charge with periodic boundary conditions, as explained in Section 6.2.4. We use Wilson's plaquette

Table 6.1.: Parameters of the simulation. For each of the three gauge groups SU(N_c) we give the inverse coupling β, the inverse of the 't Hooft coupling $\lambda_0 = g_0^2 N_c$ to four significant digits, the dimensions of the lattice, the approximate lattice spacing using $\sqrt{t_0} = 0.166\,\mathrm{fm}$ followed by the number of measurements and their separation in Cabibbo–Marinari updates of the lattice.

Lattice	N_c	β	$1/\lambda_0$	T/a	L/a	$a\,[\mathrm{fm}]$	#meas.	#it.
$A(4)_1$	4	10.92	0.3413	64	16	0.096	22 000	40
$A(4)_2$	4	11.14	0.3481	80	20	0.078	41 000	80
$A(4)_3$	4	11.35	0.3547	96	24	0.065	21 000	160
$A(5)_1$	5	17.32	0.3464	64	16	0.095	15 000	120
$A(5)_2$	5	17.67	0.3534	80	20	0.077	27 000	240
$A(5)_3$	5	18.01	0.3602	96	24	0.064	14 000	480
$A(6)_1$	6	25.15	0.3493	64	16	0.095	30 000	250
$A(6)_2$	6	25.68	0.3567	80	20	0.076	17 000	500
$A(6)_3$	6	26.15	0.3632	96	24	0.063	16 000	450

action as defined in Eq. (2.24)

$$S_W[U] = \beta \sum_{n,\mu<\nu} w_{\mu\nu}(x)\left(1 - \frac{1}{N_c}\operatorname{Re}\operatorname{tr} U_{\mu\nu}(x)\right), \qquad \beta \equiv \frac{2N_c}{g_0^2} = \frac{2N_c^2}{\lambda_0}, \tag{6.3}$$

where λ_0 is the bare 't Hooft coupling and

$$w_{\mu\nu}(x) = \begin{cases} 1/2 & x = \{0, T-a\}, \quad \mu \neq 4, \quad \nu \neq 4, \\ 1 & \text{otherwise.} \end{cases} \tag{6.4}$$

The parameters of the simulation are collected in Table 6.1: for each of the three gauge groups SU(4), SU(5) and SU(6), three values of β are chosen such as to give approximately the same t_0/a^2. Using $\sqrt{t_0} = 0.166\,\mathrm{fm}$, they correspond to lattice spacings of approximately 0.096, 0.078 and 0.065 fm. The size of the boxes have been scaled such that $L \approx 1.5\,\mathrm{fm}$, while the temporal extent is chosen to be $T = 4L$, so that a sufficiently large bulk region with negligible boundary effects is available for the measurements.

6.2.1. Gradient flow observables

We employ the standard discretization of the YM gradient flow. It is integrated with the third-order Runge–Kutta integrator defined in Ref. [29] with an integration step size such that the integration error is well below the statistical accuracy of the observables. The two primary observables $e_L(t, x)$ and $q_L(t, x)$ are measured with a $0.04a^2$ resolution in the flow time t and interpolated quadratically from the neighbouring points to get the observables at arbitrary values of t.

Following again Ref. [29], the discretized $e_L(t, x)$ and $q_L(t, x)$ are defined through the standard

'clover' field strength tensor in Eq. (3.18) and immediately summed over the spatial directions

$$\bar{e}_{\mathrm{L}}(t,x_4)=\sum_{\vec{x}} e_{\mathrm{L}}(t,\vec{x},x_4), \qquad \bar{q}_{\mathrm{L}}(t,x_4)=\sum_{\vec{x}} q_{\mathrm{L}}(t,\vec{x},x_4). \tag{6.5}$$

Because of the open boundary conditions, time translation invariance is broken and some care must be taken when averaging over the x_4 coordinate. A plateau range needs to be determined, where boundary effects can be neglected. To this end, for each observable we first perform a fit to the symmetrized data using the contribution of one excited state $f(x_4)=A+B\mathrm{e}^{-x_4 m}$ in a region where this ansatz describes the data well.

With this result, we determine the minimal distance of the plateau fit from the boundary requiring that $|f(d)-A|<\sigma/4$, with σ being the average error of the measurement for $x_4>d$. Using this criterion, the choice of $d=9.5\sqrt{t_0}$ guarantees that boundary effects in $\bar{e}_{\mathrm{L}}(t,x_4)$ at $t=t_0$ are negligible with our statistics, and therefore we define

$$\langle E_{\mathrm{L}}(t)\rangle=\frac{1}{(T-2d)L^3}\sum_{x_4=d}^{T-a-d}\langle\bar{e}_{\mathrm{L}}(t,x_4)\rangle. \tag{6.6}$$

6.2.2. Topological susceptibility

For the topological susceptibility, we use the approach of Ref. [153]. The topological charge correlator is to be averaged over the bulk region given by a minimal distance d from the boundaries

$$\bar{C}(t,\Delta)=\frac{1}{(T-2d-\Delta)L^3}\sum_{x_4=d}^{T-a-d-\Delta}\langle\bar{q}_{\mathrm{L}}(t,x_4)\bar{q}_{\mathrm{L}}(t,x_4+\Delta)\rangle. \tag{6.7}$$

Again we determine d such that for all values of Δ boundary effects are negligible. Using the same strategy as for the energy density above, $d=7.5\sqrt{t_0}$ turns out to be a conservative choice for all ensembles.

An estimator of the topological susceptibility is then obtained by truncating the sum over Δ with a cut-off r

$$\chi_{\mathrm{t}}^{\mathrm{corr}}(t,r)=a\bar{C}(t,0)+2\sum_{\Delta=a}^{r}\bar{C}(t,\Delta), \tag{6.8}$$

where r has to be chosen such that the contribution of the neglected tail is insignificant compared to the statistical accuracy of the result. Such an r can always be found, because the correlator converges exponentially to zero for large separations Δ. Unfortunately, the combination of the smoothing by the gradient flow and the numerical errors obscure this behaviour in the actual data.

Due to the smoothing, the correlation function $\bar{C}(t,\Delta)$ is positive for small values of Δ and would be expected to turn negative before exponentially converging to zero for $\Delta\gg\sqrt{8t_0}$. We cannot resolve this latter feature due to the numerical uncertainties of our data, the correlator being zero within errors typically from $\Delta=5\sqrt{t_0}$ on.

In order to get a better handle on the contribution of the tail, we use high precision data in SU(3) from Chapter 5. Assuming that the relative contribution of the tail does not change drastically with N_c, and given the accuracy of our data, cutting the summation over Δ at $r=7\sqrt{t_0}$ is a conservative choice, leading to a negligible systematic error.

6.2.3. Finite volume

By their nature, lattice simulations are done in a finite volume, which can distort the results. For a large enough box size, these systematic effects are exponentially suppressed, but we need to verify that they are negligible given the target accuracy.

The lattices employed in this study are slightly larger than the ones used for SU(3) in Chapter 5. Despite significantly smaller statistical errors, no significant finite size effects could be detected in the SU(3) study. In order to avoid relying only on the independence of these finite volume effects on N_c, we also generated lattices with $L = 1.1\,\mathrm{fm}$ and $2.3\,\mathrm{fm}$ for SU(4) and SU(5) at the smallest values of β. These lattices bracket the $L = 1.5\,\mathrm{fm}$ used in our analysis. With a numerical accuracy matching our target, no significant differences between the three sizes are found, such that we conclude that also this systematics is under control.

6.2.4. Autocorrelations

Simulations like the one presented here are known to be challenging due to a rapid rise of the autocorrelation times τ_{int}, in particular of topological observables. Numerical evidence suggests they increase with a very high power or even exponentially with $1/a$ and N_c when periodic boundary conditions are implemented [161].

In our study, the gauge field is updated with the Cabibbo–Marinari scheme [96]: one update consists of a heat bath sweep of the full lattice followed by $n_{\mathrm{ov}} \sim a^{-1}$ overrelaxation sweeps. Both the heat bath and the overrelaxation sweeps update all the $N_c(N_c-1)/2$ SU(2) subgroups of a given SU(N_c) link. The number of these updates between measurements is given in Table 6.1 and chosen such that for all our observables autocorrelations are hardly detectable. We take them into account in our analysis using the standard methods of Ref. [160].

To study the effect of the open boundaries, we have computed τ_{int} for the coarser lattices $A(4)_1$, $A(5)_1$ and $A(6)_1$ with dedicated runs in the presence of periodic and open boundaries, putting fewer updates between measurements for increased sensitivity.

In units of updates with $n_{\mathrm{ov}} = L/(2a)$, τ_{int} of χ_t for the periodic lattices is 16(2), 54(6) and 187(19) for $N_c = 4$, 5 and 6, respectively. With open boundaries the corresponding values are 12(1), 46(6) and 111(10). For all values of N_c we observe a reduction in τ_{int} for open compared to periodic boundary conditions. It is most significant at $N_c = 6$ and hardly statistically significant for the other values of N_c.

Going to finer lattices, this advantage is expected to be more pronounced. The exponential scaling observed in Ref. [161] with periodic boundary conditions would suggest a value of τ_{int} one or two orders of magnitude larger than the one we observe with open boundaries. Therefore the finer lattice spacings would not have been feasible with PBCs and our computer resources.

6.3. Results

The results for the observables for the three gauge groups are listed in Table 6.2. At finite lattice spacing, we reach accuracies on the percent level for χ_t and below the permille level for t_0. Here and in the following χ_t is evaluated at $t = t_0$ if not otherwise stated. The values for the dimensionless product $t_0^2\chi_t$ are displayed in the left plot of Figure 6.1, where we also add the

Table 6.2.: Results for t_0, $t_0^2\chi_t$ and its fourth root.

Lattice	t_0/a^2	$10^4 t_0^2\chi_t$	$t_0^{1/2}\chi_t^{1/4}$
$A(4)_1$	2.9900(7)	6.61(6)	0.1603(4)
$A(4)_2$	4.5207(8)	6.54(5)	0.1599(3)
$A(4)_3$	6.4849(16)	6.68(7)	0.1607(4)
$A(5)_1$	3.0636(7)	6.47(7)	0.1595(4)
$A(5)_2$	4.6751(8)	6.73(7)	0.1611(4)
$A(5)_3$	6.8151(17)	6.62(8)	0.1604(5)
$A(6)_1$	3.0824(4)	6.57(6)	0.1601(4)
$A(6)_2$	4.8239(9)	6.81(8)	0.1615(5)
$A(6)_3$	6.9463(13)	6.80(7)	0.1615(4)

SU(3) results from Ref. [1]. It is clear that, both, the effects of finite N_c and finite cut-off a are roughly at the level of our statistical errors.

In order to extrapolate the raw data to the continuum and the $N_c \to \infty$ limit, we use the functional form

$$t_0^2\chi_t\left(\frac{1}{N_c}, a\right) = t_0^2\chi_t(0,0) + c_1\frac{1}{N_c^2} + c_2\frac{a^2}{t_0}, \tag{6.9}$$

which takes into account the leading corrections dictated by Symanzik's and the large-N_c expansion. This is motivated by the observation that both corrections are small, given the statistical accuracy of our data. The fact that the N_c-dependence of the $\mathcal{O}(a^2)$ term can be neglected within our precision is further supported by the observation that discretization effects in the ratio $\chi_t(t)/\chi_t(t_0)$, which can be captured to exceedingly high accuracy, turn out to be independent of N_c.

Our main result is obtained by fitting Eq. (6.9) to the two finer points of SU(4), SU(5) and SU(6) data together with the two finer data points for SU(3), where the latter is only used to constrain the coefficient c_2. Discarding the coarser lattice points and the smallest N_c reduces the assumptions made on the scaling region of our results. This fit renders

$$t_0^2\chi_t(0,0) = 7.03(13)\times 10^{-4}, \tag{6.10}$$

i.e. a 2 % accuracy is reached. The fit quality is excellent with a $\chi^2/\mathrm{dof} = 0.94$. In the continuum limit the fit gives $t_0^2\chi_t(1/N_c, 0) = 6.68(12)\times 10^{-4}$, $6.81(11)\times 10^{-4}$ and $6.87(11)\times 10^{-4}$ for $N_c = 4$, 5 and 6 respectively, see right plot of Figure 6.1.

To get a better handle on possible systematic effects of this result, many other fits to the data have been tried, all of them leading to similar results. Among them the most obvious modification is to include also the third finest point of the SU(3) data determining the discretization effects. This changes the result to $t_0^2\chi_t(0,0) = 7.13(10)\times 10^{-4}$ with $\chi^2/\mathrm{dof} = 1.1$, compatible with the above number. If the three finest SU(3) points are globally fitted with the two finer points of the other groups, the results is $t_0^2\chi_t(0,0) = 7.09(7)\times 10^{-4}$ with an excellent value of $\chi^2/\mathrm{dof} = 1.0$. A global fit of Eq. (6.9) to all data, including the three finer SU(3) ones, adding an a^2/N_c^2 term to Eq. (6.9), gives $t_0^2\chi_t(0,0) = 7.02(13)\times 10^{-4}$ with a $\chi^2/\mathrm{dof} = 1.7$. Performing the continuum

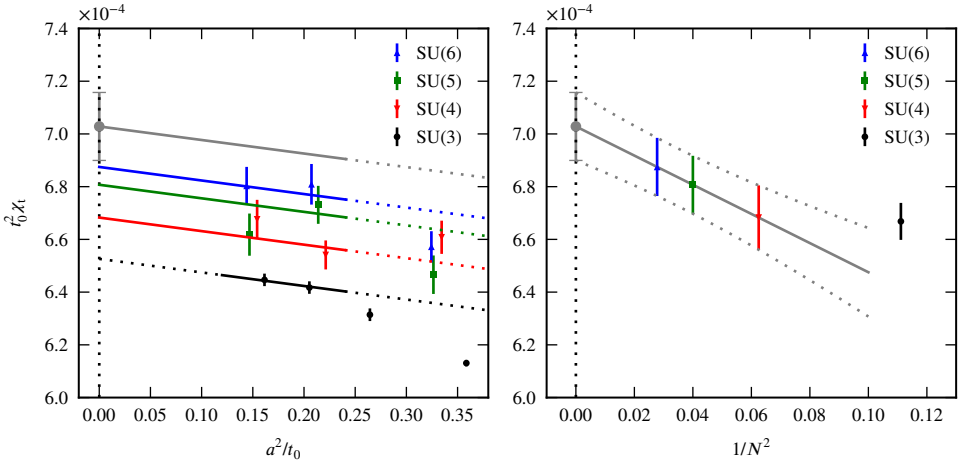


Figure 6.1.: Left: Combined continuum and large-N_c extrapolation of the topological susceptibility by fitting the data in the region indicated by the solid lines to Eq. (6.9). Right: Same data as in left plot as a function of $1/N_c^2$. The data points indicate the continuum result of the global fit for $N_c > 3$. For $N_c = 3$ the data is taken from Chapter 5.

limit group-by-group and applying the large-N_c extrapolation only in a second step also gives a compatible result. An extended discussion of these fits can be found in Ref. [7].

From these analyses we conclude that the systematic effects coming from the continuum and large-N_c extrapolations are under control within the errors quoted.

6.4. Conclusions

This is the first investigation of the large-N_c behaviour of the topological susceptibility in pure Yang–Mills theory using a theoretically sound definition of χ_t, and small lattice spacings which allow for control over the continuum limit. As a final result we quote for $N_c \to \infty$

$$t_0^2\chi_t(0,0) = 7.03(13) \times 10^{-4}. \tag{6.11}$$

This result proves that the leading anomalous contribution to the η' mass is large enough to solve the $U(1)_A$ problem in QCD. The bulk of the mass of the pseudoscalar singlet meson is generated by the anomaly through the Witten–Veneziano mechanism. The $1/N_c^2$ corrections that we have found in $t_0^2\chi_t(0,0)$ are at most of the expected size (even a bit smaller), with no large prefactor in the expansion. This explains why the $N_c = 3$ result, $t_0^2\chi_t = 6.67(7) \times 10^{-4}$, in Eq. (5.10) is already large enough to explain the large value of the η' mass in Nature. The difference with the $N_c \to \infty$ value is barely visible within errors, despite their high accuracy.

In the Yang–Mills theory, it will be challenging to improve significantly on these results by brute force. Discretization effects and large-N_c effects are roughly of the same level. The much higher accuracy needed to resolve higher order effects in the large-N_c expansions will therefore

require significantly smaller lattice spacings. These are still computationally very expensive, even with the open boundary conditions, which make those used in the present study possible.

The accuracy presented here is certainly sufficient for the completion of the proof of the Witten-Veneziano relation in Eq. (3.64). It will need to be matched by the one on the hadronic quantities entering the relation to be computed in the large-N_c limit of QCD.

7. Noise reduction with multilevel integration

In this chapter we present some preparatory material, mostly taken from the literature, that is needed to understand the original work presented in Chapter 8.

As discussed in Section 2.4.3, in the last decades the availability of faster computers and important algorithmic advancements made it possible simulations of lattices larger in physical units and with quark masses closer to or at the physical point. This allowed two-point functions to be computed at larger distances, in a regime in which the systematics from excited states is highly reduced. One may expect that this resulted in an increased accuracy in the extraction of the hadronic spectrum. However, with the current state-of-the-art techniques, the numerical computation of hadronic correlation functions in Monte Carlo simulations is characterized by a signal-to-noise ratio (S/N) decreasing exponentially with the time separation of the sources, as it was first recognized in Refs [173, 174]. It means that an exponential increase of the number of samples is needed to maintain the same statistical accuracy at a larger separation. This represent an unsolved *exponential noise problem* that afflicts propagators, with a notable exception being the propagators of *nonsinglet* pseudoscalar mesons.

Simulations try to balance between systematic effects at short distances and poor statistics at long distances. It is possible to reduce the systematics with complex or multiple interpolating fields, which minimize the overlap with unwanted states. Still, a satisfactory compromise is not always found. For instance, we would like to compute the mass of the η' meson. Or, more generally, the mass of the lightest pseudoscalar state in the flavour-singlet sector of a $N_{\mathrm{f}} \geq 2$ theory. The exponential noise problem is particularly severe for this specific correlator. Indeed, its variance includes a vacuum contribution, which is not suppressed at all with the separation of the two points of the correlator.

Analogous problems afflict the computation of correlators in a large variety of quantum systems, from the harmonic oscillator to YM theory. In some cases, multilevel algorithms have been proposed. They lead to an impressive acceleration of the simulations [175, 31, 32, 176–178]. They take advantage of the fact that, when the action and the observables depend locally on the integration variables, the S/N problem can be solved by independent measurements of the local building blocks of the observables. So far, these ideas have been restricted to bosonic theories.

In this chapter, in Section 7.1 we introduce the problem with the S/N of correlators at asymptotic separations. Then, in Section 7.2 we describe how the domain decomposition and multilevel integration solve the problem in bosonic theories. Finally, in Section 7.3 we review an application of the technique to the computation of Wilson loops in Yang–Mills theory [31].

7.1. Signal-to-noise ratio of Monte Carlo observables

As discussed in Section 2.3.2, the properties of the spectrum of a QFT are contained in the two-point functions

$$C_O(x_4 - y_4) = \langle O(x_4)O(y_4)\rangle = \sum_k \left|\left\langle 0\middle|\hat{O}\middle|k\right\rangle\right|^2 \mathrm{e}^{-E_k|x_4-y_4|}. \tag{7.1}$$

In theories with a mass gap, the exponential decay at large Euclidean time separations $|x_4 - y_4|$ singles out the lightest state compatible with the symmetry transformation properties of the field O. If the energy of this state is M_O,

$$C_O(x_4 - y_4) \overset{|x_4-y_4|\to\infty}{\sim} \exp\{-M_O|x_4 - y_4|\}. \tag{7.2}$$

The correlator C_O is usually computed numerically with a Monte Carlo algorithm. As described in Section 2.4, the field-theoretic expectation value is estimated by sampling the path integral on a finite set of n gauge field configurations U_i, $i = 1, \dots, n$. An estimator of the correlator is thus

$$\bar{C}_O(x_4 - y_4) = \frac{1}{n}\sum_{i=1}^{n} O[U_i](x_4)O[U_i](y_4). \tag{7.3}$$

For a large number n of configurations, which we assume to be statistically uncorrelated, $\bar{C}_O$ approximates the field-theoretic expectation value with an error of order $\sqrt{\sigma^2(C_O)/n}$ [173, 174]. The variance $\sigma^2(C_O)$ of C_O has a field-theoretic expression

$$\sigma^2(C_O)(x_4 - y_4) = \left\langle O^2(x_4)O^2(y_4)\right\rangle - \left\langle O(x_4)O(y_4)\right\rangle^2. \tag{7.4}$$

The first term in the r.h.s. of this last equation can be treated in the same way as the correlator inserting a full set of eigenstates of the Hamiltonian. Therefore, it is expected to decay exponentially at large $|x_4 - y_4|$ with the energy of the lightest state compatible with the quantum numbers of O^2

$$\sigma^2(C_O)(x_4 - y_4) \overset{|x_4-y_4|\to\infty}{\sim} \exp\{-M_{O^2}|x_4 - y_4|\}, \qquad M_{O^2} \le 2M_O, \tag{7.5}$$

where M_{O^2} is in general smaller that twice M_O. As a consequence, the signal-to-noise ratio

$$S/N = \frac{\langle C_O\rangle\sqrt{n}}{\sqrt{\sigma^2(C_O)}} \overset{|x_4-y_4|\to\infty}{\sim} \exp\left\{-\left(M_O - \frac{1}{2}M_{O^2}\right)|x_4 - y_4|\right\} \tag{7.6}$$

decays exponentially with $|x_4 - y_4|$.

The exponential decay of S/N with separation $|x_4 - y_4|$ is currently the main limitation to the accuracy with which the spectrum and the matrix elements can be studied. Indeed, most of the systematic effects, such as finite volume and finite lattice spacing, can be solved with a polynomial increase of the simulation cost. On the contrary, to deal with excited states systematics one needs to go to longer separation, but this require an *exponential* increase of the number of configurations n, and thus of simulation cost, to have the same statistical accuracy.

7.1.1. Variance of fermionic correlators

In the case of fermionic interpolating field, since the quark path integral is computed exactly, the Monte Carlo variance is due only to fluctuations of the gauge field.

The pion, being the pseudo-NG boson of χSB, is the lightest state and play a special rôle. Its mass can be extracted by the large Euclidean time behaviour of the zero-momentum correlator of two pseudoscalar densities P^a, with $a \neq 0$. After integrating out fermionic fields, the correlator reduces to the bosonic expectation value of quark propagators, according to Eq. (2.114) of Section 2.4.4,

$$C_\pi(x_4 - y_4) = \sum_{\vec{x}} \langle P^a(x) P^a(y) \rangle = -\frac{1}{2} \sum_{\vec{x}} \Big\langle \mathrm{tr}\big\{ S(x,y) S(x,y)^\dagger \big\} \Big\rangle, \tag{7.7}$$

which is expected to decay as

$$C_\pi(x_4 - y_4) \overset{|x_4 - y_4| \to \infty}{\sim} \exp\big\{ -M_\pi \big| x_4 - y_4 \big| \big\}. \tag{7.8}$$

Its gauge field variance is, a priori,

$$\sigma^2(C_\pi)(x_4 - y_4) = \frac{1}{4} \sum_{\vec{x},\vec{x}'} \Big\langle \mathrm{tr}\big\{ S(x,y) S(x,y)^\dagger \big\}\, \mathrm{tr}\big\{ S(x',y) S(x',y)^\dagger \big\} \Big\rangle - \big[C_\pi(x_4 - y_4) \big]^2. \tag{7.9}$$

The second term in the r.h.s. is just the square of the pion propagator, so it decays with twice the pion mass. The field-theoretic interpretation of the first term in the r.h.s. is not manifest. It is however Wick's contraction of

$$\Big\langle P^a(x) P^b(x') P^a(y) P^b(y) \Big\rangle, \qquad \text{with } a \neq b, \tag{7.10}$$

which propagates a two-π state between x (x') and y. In absence of bound states or a continuum whose energy is smaller than $2M_\pi$, the first term in Eq. (7.9) is thus expected to decay asymptotically with $\exp\{-2M_\pi |x_4 - y_4|\}$. Therefore,

$$\sigma^2(C_\pi)(x_4 - y_4) \overset{|x_4 - y_4| \to \infty}{\sim} \exp\big\{ -2M_\pi \big| x_4 - y_4 \big| \big\}. \tag{7.11}$$

Both the signal and the error of the pion propagator decrease with distance with the same rate: there is no exponential S/N degradation in the pion propagator. This is a direct consequence of the fact that the quark path integral is computed exactly. Moreover, since both the mean and the width of the distribution of $\mathrm{tr}\{S(x,y)S(x,y)^\dagger\}$ on gauge field configurations fall off with the same exponential, we expect that the quark propagator, on every representative background gauge field, decay exponentially at asymptotic separations as

$$\mathrm{tr}\big\{ S(x,y) S(x,y)^\dagger \big\}^{\frac{1}{2}} \sim \exp\Big\{ -\frac{1}{2} M_\pi |x - y| \Big\}. \tag{7.12}$$

Things are less good for the propagator of the flavour-singlet pseudoscalar meson, e.g. the η' in the $N_\mathrm{f} = 3$ theory. Because of the anomalous contribution, the flavour-singlet pseudoscalar meson is heavier than the non-singlet ones, thus the propagator is expected to decrease with a

faster rate. According to Eq. (2.115), after integrating out fermions, the propagator is given by two different contributions

$$C_0 = \sum_{\vec{x}} \left\langle P^0(x) P^0(y) \right\rangle = -\frac{1}{2} \sum_{\vec{x}} \left\langle \mathrm{tr}\left\{ S(x,y) S(x,y)^\dagger \right\} \right\rangle + \frac{N_\mathrm{f}}{2} \sum_{\vec{x}} \left\langle \mathrm{tr}\left\{\gamma_5 S(x,x)\right\} \mathrm{tr}\left\{\gamma_5 S(y,y)\right\} \right\rangle. \quad (7.13)$$

The connected one decays exponentially with M_π, thus the disconnected one must decay with the same exponential in order to cancel the leading exponential in the linear combination in Eq. (7.13)

$$C_{\mathrm{dis.}}(x_4 - y_4) = \sum_{\vec{x}} \left\langle \mathrm{tr}\left\{\gamma_5 S(x,x)\right\} \mathrm{tr}\left\{\gamma_5 S(y,y)\right\} \right\rangle \overset{|x_4 - y_4| \to \infty}{\sim} \exp\left\{ -M_\pi \left| x_4 - y_4 \right| \right\}. \quad (7.14)$$

However, the disconnected contribution is a lot more noisier that the connected one. Its variance is

$$\sigma^2(C_{\mathrm{dis.}}) = \sum_{\vec{x},\vec{x}'} \left\langle \mathrm{tr}\left\{\gamma_5 S(x,x)\right\} \mathrm{tr}\left\{\gamma_5 S(y,y)\right\} \mathrm{tr}\left\{\gamma_5 S(x',x')\right\} \mathrm{tr}\left\{\gamma_5 S(y,y)\right\} \right\rangle - C_{\mathrm{dis.}}^2. \quad (7.15)$$

In this formula, the first term contains four disconnected loops of quark propagators. Intuitively, there is not a quark propagator line which connects x (x') with y, thus the size of this contribution is not expected to decay with the separation between x_4 and y_4. Indeed, Eq. (7.15) is Wick's contraction of

$$\left\langle P^{00}(x) P^{11}(x') P^{22}(y) P^{33}(y) \right\rangle \quad (7.16)$$

in the theory with $N_\mathrm{f} \geq 4$, eventually adding valence quarks as in Section 3.2.4. In general, $\langle P^{ii} P^{jj} \rangle \neq 0$ since it has an overlap with the vacuum. Therefore, the expectation value in Eq. (7.15) is not a connected correlation function in the field-theoretical sense and the first term in the r.h.s. goes to a finite value in the limit $|x_4 - y_4| \to \infty$.

Therefore, the disconnected contribution has a huge problem in the S/N: a signal decaying with $\exp\{-M_\pi |x_4 - y_4|\}$ must be extracted from a noise constant in $|x_4 - y_4|$. With ordinary Monte Carlo techniques, this means that the number of configurations must be exponentially increasing

$$n \sim \exp\left\{ 2 M_\pi \left| x_4 - y_4 \right| \right\} \quad (7.17)$$

to obtain a signal at a distance $|x_4 - y_4|$. An exponential scaling is simply too much to be dealt with a pure brute force approach, even with the help of the continuously increasing computer power, and is not satisfactory from a purely conceptual point of view.

In general, any propagator with the exclusion of the non-singlet pseudoscalar one develops an exponential S/N degradation. A notable example is the nucleon propagator. Indeed, the signal is obtained contracting in the proper way three quark propagators from x to y

$$C_N(x_4 - y_4) = \sum_{\vec{x}} \left\langle \bar{N}(x) N(y) \right\rangle \overset{|x_4 - y_4| \to \infty}{\sim} \exp\left\{ -M_N \left| x_4 - y_4 \right| \right\}. \quad (7.18)$$

The variance is thus given by six quark propagators. These are arranged to contribute to the propagator of a three-π states

$$\sigma^2(C_N)(x_4 - y_4) \overset{|x_4-y_4|\to\infty}{\sim} \exp\{-3M_\pi|x_4 - y_4|\}. \tag{7.19}$$

Therefore, the S/N of the nucleon propagator is exponentially decaying

$$S/N = \frac{\langle C_N\rangle\sqrt{n}}{\sqrt{\sigma^2(C_N)}} \overset{|x_4-y_4|\to\infty}{\sim} \exp\Big\{-\Big(M_N - \frac{3}{2}M_\pi\Big)|x_4 - y_4|\Big\}, \tag{7.20}$$

with a rate as big as $M_N - 3M_\pi/2 \approx 3.7\,\mathrm{fm}^{-1}$ at the physical point.

7.2. Domain decomposition and multilevel integration

The domain decomposition and multilevel integration technique has proven useful to attack the exponential S/N problem in bosonic theories. In this section, we introduce this technique for a general local quantum field theory [32] and we describe the implementation of a multilevel Monte Carlo algorithm [31, 32] for a two-point function in the simple case of pure YM theory. It results in an exponential error reduction at asymptotic separations.

We first define what we mean for a quantum field theory to be local in an abstract way. Following Ref. [32], let us divide a field configuration ϕ into mutually disjoint subset ϕ_{Γ_0}, ϕ_{Γ_1} and $\phi_{\partial\Gamma}$, with Γ_0, Γ_1 and $\partial\Gamma$ being their (disjoint) support in the (continuous) spacetime manifold $\mathcal{M}$. Then, suppose that any continuous path $\gamma : [0,1] \to \mathcal{M}$ going from $\gamma(0) \in \Gamma_0$ to $\gamma(1) \in \Gamma_1$ necessary pass through $\partial\Gamma$. The theory is local if the probability density in the path integral can be written as

$$\mathrm{d}P[\phi_{\Gamma_0\cup\Gamma_1}] = \int_{\partial\Gamma} \mathrm{d}P[\phi_{\partial\Gamma}]\mathrm{d}P_0[\phi_{\Gamma_0}]\mathrm{d}P_1[\phi_{\Gamma_1}], \tag{7.21}$$

i.e. there exist P_0 and P_1 such that 'ϕ_{Γ_0} and ϕ_{Γ_1} influence each other only through $\phi_{\partial\Gamma}$' [32]. This definition of locality automatically realizes a domain decomposition that factorizes the theory into sectors that are not directly influenced.

The locality condition is satisfied by any QFT defined through a Lagrangian density with finitely many derivatives. With a suitable notion of 'continuity', the definition is extended lattice actions. Eq. (7.21) is satisfied if the action can be written as

$$S[\phi] = S_0[\phi_{\bar\Gamma_0}] + S_\partial[\phi_{\partial\Gamma}] + S_1[\phi_{\bar\Gamma_1}], \tag{7.22}$$

where $\bar\Gamma_i = \Gamma_i \cup \partial\Gamma$, setting

$$\mathrm{d}P[\phi_{\partial\Gamma}] = \frac{\mathcal{Z}_0\mathcal{Z}_1}{\mathcal{Z}}\,\mathcal{D}[\phi_{\partial\Gamma}]\,\mathrm{e}^{-S_\partial[\phi_{\partial\Gamma}]}, \tag{7.23a}$$

$$\mathrm{d}P_0[\phi_{\Gamma_0}] = \frac{1}{\mathcal{Z}_0}\,\mathcal{D}[\phi_{\Gamma_0}]\,\mathrm{e}^{-S_0[\phi_{\bar\Gamma_0}]}, \qquad \mathrm{d}P_1[\phi_{\Gamma_1}] = \frac{1}{\mathcal{Z}_1}\,\mathcal{D}[\phi_{\Gamma_1}]\,\mathrm{e}^{-S_1[\phi_{\bar\Gamma_1}]}. \tag{7.23b}$$

For instance, YM theory discretized on the lattice with Wilson's plaquette action is local in this sense. A domain decomposition in thick time-slices is explicitly constructed in Section 7.3.

7.2.1. Factorization of the path integral

Consider two fields O and O' localized respectively in $\bar{\Gamma}_0$ and $\bar{\Gamma}_1$. As a direct consequence of locality, their correlator has a factorized path integral expression

$$\langle C_{OO'} \rangle = \langle O[\phi] O'[\phi] \rangle = \int \mathrm{d}P[\phi] O[\phi] O'[\phi] \\ = \int_{\partial\Gamma} \mathrm{d}P[\phi_{\partial\Gamma}] \langle\!\langle O[\phi_{\bar{\Gamma}_0}] \rangle\!\rangle_{\Gamma_0} \langle\!\langle O'[\phi_{\bar{\Gamma}_1}] \rangle\!\rangle_{\Gamma_1}, \quad (7.24)$$

with

$$\langle\!\langle O \rangle\!\rangle_{\Gamma_0}[\phi_{\partial\Gamma}] = \int_{\Gamma_0} \mathrm{d}P_0[\phi_{\Gamma_0}] O[\phi_{\bar{\Gamma}_0}], \quad (7.25a)$$

$$\langle\!\langle O' \rangle\!\rangle_{\Gamma_1}[\phi_{\partial\Gamma}] = \int_{\Gamma_1} \mathrm{d}P_1[\phi_{\Gamma_1}] O'[\phi_{\bar{\Gamma}_1}]. \quad (7.25b)$$

Thus, the path integral average is factorized into independent averages in disjoint domains Γ_0 and Γ_1 with boundary conditions fixed in $\partial\Gamma$, times an average over the 'boundary' $\partial\Gamma$. In Eqs (7.25), we introduced the notation $\langle\!\langle \bullet \rangle\!\rangle_\Gamma$ to denote the field-theoretical *sub-lattice* expectation value in the domain Γ at fixed boundary conditions. Its worth noting that this expectation value is still a functional of the field on the boundary.

We can perform a further step on Eq. (7.24): re-expressing $\mathcal{Z}_0$ and $\mathcal{Z}_1$ in path-integral form and using Eq. (7.21), we obtain a standard path integral on the whole lattice

$$\langle C_{OO'} \rangle = \int \mathrm{d}P[\phi] \langle\!\langle O \rangle\!\rangle_{\Gamma_0}[\phi_{\partial\Gamma}] \langle\!\langle O' \rangle\!\rangle_{\Gamma_1}[\phi_{\partial\Gamma}] = \Big\langle \langle\!\langle O \rangle\!\rangle_{\Gamma_0} \langle\!\langle O' \rangle\!\rangle_{\Gamma_1} \Big\rangle, \quad (7.26)$$

where the sub-lattice expectation values are treated as observables depending on the boundary field only.

There are many different ways in which this factorization can be iterated. One possibility is to carry out a domain decomposition on Γ_0 and Γ_1, to obtain a factorization in more than two domains that allows to treat the general n-point function. A second possibility is to iterate the factorization on $\langle\!\langle O \rangle\!\rangle_{\Gamma_0}$ and $\langle\!\langle O' \rangle\!\rangle_{\Gamma_1}$ to obtain a nested expression with three or more levels.

7.2.2. The multilevel Monte Carlo algorithm

The two-point function in Eq. (7.24) has a simple realization in terms of a multilevel Monte Carlo algorithm [31]. In the simple case of a two-domain two-level realization:

level-0: a number n_0 of field configurations $\{\phi_i\}$, $i = 1, \dots, n_0$ defined on the whole lattice is generated, using a standard Monte Carlo;

level-1: starting from every level-0 field configuration, a number n_1 of field configurations $\{\phi_{i,j}\}$, $j = 1, \dots, n_1$ is generated updating independently fields in the two regions and keeping fixed fields in the boundary region $\partial\Gamma$.

This requires computer time comparable to the generation of $n_0 \cdot n_1$ configurations with a standard single-level algorithm.

The two-point function is estimated as two-level averaging process

$$\bar{C}_{OO'} = \frac{1}{n_0} \sum_{i=1}^{n_0} \left\{ \left[\frac{1}{n_1} \sum_{j=1}^{n_1} O[\phi_{i,j}] \right] \left[\frac{1}{n_1} \sum_{j=1}^{n_1} O'[\phi_{i,j}] \right] \right\}. \tag{7.27}$$

It is important to notice that every level-1 field configuration in the domain Γ_0 combined with every level-1 field configuration in the domain Γ_1 is an independent configuration for both O and O' fields. This means that at level-1 we have effectively n_1^2 configurations.

Eq. (7.27) is an improved estimator of $\langle C_{OO'} \rangle$. Suppose that O and O' show no dependence at all on the boundary field. Then, each term in the square parenthesis in Eq. (7.27) estimates the field-theoretical expectation value with an error of order $\sqrt{\sigma^2(O)/n_1}$. This, combined with the level-0 average, results in an error scaling with $(n_0 n_1^2)^{-1/2}$, instead of the $(n_0 n_1)^{-1/2}$ of the standard case. In general, O and O' retain, at least indirectly, a dependence on the quantum fluctuations of the boundary field. This induces in the estimator in Eq. (7.27) a contribution to the error that scales with ${n_0}^{-1/2}$. This scaling will dominate for $n_1 \to \infty$, but how large n_1 can be before it becomes dominant depends on the details of the observable and the domain decomposition. With a sensible setup, the dependency is small and one can take quite large n_1 before the boundary contribution becomes relevant.

To recap, the validity of Eq. (7.24) relies on some necessary conditions:

- the theory must be local, i.e. it is possible to factorize the action of the theory in domains as described in Eq. (7.21);
- the fields in the expectation value must be localized to a domain, or can be written as a product of localized fields.

Only if these conditions are satisfied the multilevel Monte Carlo integration of Eq. (7.27) make sense. Moreover, we anticipate that for the algorithm to be effective in obtaining an exponential gain in the S/N of Eq. (7.27), it is needed that:

- the fluctuations on the boundary have a much smaller influence than those occurring in the region updated at level-1, such that one can scale n_1 to the optimal value;
- no phase transition occurs [31] due to the small volume and the boundary conditions of the level-1 domains.

7.3. An application of multilevel integration to Yang–Mills theory

The multilevel Monte Carlo algorithm described in the previous section is directly applicable to SU(N_c) YM theory [31, 32]. Here, we describe how it is possible to exploit it to exponentially reduce the variance in expectation values of large Wilson loops [31].

Consider an oriented closed curve $\mathcal{C}$ on the lattice and denote with $U_{\mathcal{C}}$ the ordered product of plaquettes along the curve. The associated Wilson loop is

$$W(\mathcal{C}) = \operatorname{tr} U_{\mathcal{C}}. \tag{7.28}$$

Its expectation value is of interest because it satisfies, for large Wilson loops, an *area law*

$$\langle W(C)\rangle \sim \mathrm{e}^{-\sigma A(C)}, \tag{7.29}$$

where $A(C)$ is the area enclosed by the path C and the $d = 2$ constant σ is the *string tension*. However, when computed in Monte Carlo simulations, this expectation value has a variance of order unity, independently on the area A. Therefore, it is exponentially more expensive to compute the expectation value in Eq. (7.29) for larger loops.

7.3.1. Domain decomposition of the Yang–Mills partition function

To be specific, we consider a domain decomposition of the lattice Λ in *thick time slices* denoted by Γ_α. The transfer matrix expression for the path integral introduced in Section 2.3.1 suggests a natural way to factorize the YM action in thick time slices. Using Eqs (C.3), we can write the partition function restricted to the thick time-slice $\Gamma = \{x : x_4 \in [0, d]\}$, where d is the thick-time-slice thickness, as the path integral

$$\mathcal{Z}_\Gamma[U_k(0), U_k(d)] = \int \mathcal{D}[U_\Gamma]\, \mathrm{e}^{-S_{\mathrm{W}}^\Gamma[U_\Gamma]}, \tag{7.30}$$

where, using the notation from Appendix C,

$$\mathcal{D}[U_\Gamma] = \prod_{x_4=a}^{d-a} \mathcal{D}\big[U_k(x_4)\big] \cdot \prod_{x_4=0}^{d-a} \mathcal{D}\big[U_4(x_4)\big]\,, \tag{7.31a}$$

$$S_{\mathrm{W}}^\Gamma[U_\Gamma] = \sum_{x_4=0}^{d-a} \Big\{ K_{\mathrm{g}}(x_4 + a, x_4) + \frac{1}{2} W_{\mathrm{g}}(x_4) + \frac{1}{2} W_{\mathrm{g}}(x_4 + a) \Big\}. \tag{7.31b}$$

The result of Eq. (7.30) is a functional of space-oriented links on the two time-slice at $x_4 = 0$ and d which bound the domain, which are not integrated over. For any composite field O localized in Γ, we define its sub-lattice expectation value as the functional of the boundary fields

$$\langle\!\langle O \rangle\!\rangle_\Gamma[U_k(0), U_k(d)] = \frac{1}{\mathcal{Z}_\Gamma} \int \mathcal{D}[U_\Gamma]\, O[U_\Gamma] \mathrm{e}^{-S_{\mathrm{W}}^\Gamma[U_\Gamma]}. \tag{7.32}$$

In the minimal case, a single time slice is sufficient as boundary between domains to have a factorized gauge field dependence. For an arbitrary number of adjacent thick time slices Γ_α of thickness d

$$\Gamma_\alpha = \big\{x : x_4 \in [\alpha d, \alpha d + d]\big\}, \tag{7.33}$$

the action is just the sum of the thick-time-slices actions

$$S_{\mathrm{W}}[U] = \sum_\alpha S_{\mathrm{W}}^{\Gamma_\alpha}[U_{\Gamma_\alpha}], \tag{7.34}$$

with boundary action $S_{\partial\Gamma} = 0$. The full partition function is then the path integral over boundaries of $\mathcal{Z}_{\Gamma_\alpha}$ in Eq. (7.30)

$$\mathcal{Z} = \int \mathcal{D}[U]\, \mathrm{e}^{-S_{\mathrm{W}}[U]} = \prod_\alpha \int \mathcal{D}\big[U_k(\alpha d)\big]\, \mathcal{Z}_{\Gamma_\alpha}[U_k(\alpha d), U_k(\alpha d + d)]. \tag{7.35}$$

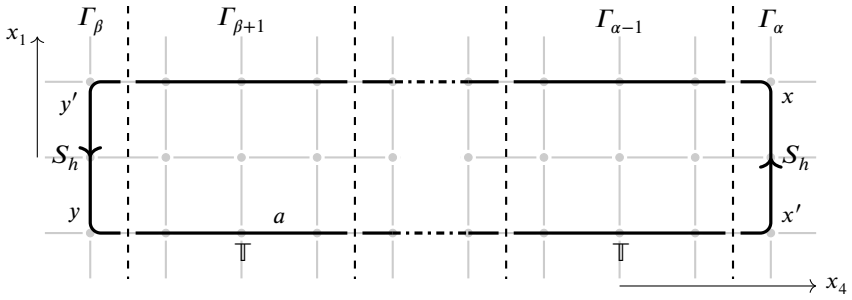


Figure 7.1.: Sketch of the domain decomposition of the lattice in thick time slices Γ_α and the factorized Wilson loop $W(x, y)$, with $x, x' \in \Gamma_\alpha$ and $y, y' \in \Gamma_\beta$.

7.3.2. Factorization of the Wilson loop

Following Ref. [31], we consider a rectangular Wilson loop $W(x, y)$ in the $(\hat{1}, \hat{4})$ plane as pictured in Figure 7.1. Its size is $r = |\vec{x} - \vec{y}|$ in the spatial direction and $t = |x_4 - y_4|$ in the time direction and the four corners are denoted by $x = (\vec{x}, x_4)$, $x' = (\vec{x} + r\hat{1}, x_4)$, $y = (\vec{x} + r\hat{1}, y_4)$ and $y' = (\vec{x}, y_4)$. As it is, this observable is not a local field depending on gauge links belonging to a single domain. It is easy however to write it as a product of local fields

$$W(x, y) = S_h(x, x')\big[\mathbb{T}(x - a\hat{4}, x' - a\hat{4}) \cdots \mathbb{T}(y + a\hat{4}, y' + a\hat{4})\mathbb{T}(y, y')\big] S_h(y, y'). \tag{7.36}$$

In this expression, $S_h(x, x + r\hat{1})$ is a tensor in colour space in the $\bar{\boldsymbol{N}}_c \otimes \boldsymbol{N}_c$ representation

$$S_h(x, x + r\hat{1})_{\alpha\beta} = \big[U_1(x) \cdots U_1(x + (r - a)\hat{1})\big]_{\alpha\beta}, \tag{7.37a}$$

while the two-links field $\mathbb{T}(x, x')$ acts as a matrix on the colour tensor S_h

$$\mathbb{T}(x, x')_{\alpha\beta,\gamma\delta} = U(x)^*_{\alpha\gamma} U(x')_{\beta\delta}. \tag{7.37b}$$

$W(x, y)$ has a nice field-theoretic interpretation in YM theory supplemented with *static* fermions, i.e. in QCD in the limit of infinite-mass quarks. To see this, it is better to perform a gauge fixing: we work in the *temporal gauge*, which on the lattice is given by $U_4(x) = \mathbb{1}$. After gauge fixing, $\mathbb{T}(x_4) = \mathbb{1}$, thus we obtain the two-point function

$$W(x, y) \xrightarrow{U_4(x)=\mathbb{1}} S_h(x, x + r\hat{1}) S_h(y, y - r\hat{1}). \tag{7.38}$$

The non-gauge-invariant field $S_h(x, x')$ has the same transformation properties under gauge transformation of a quark-antiquark bilinear $\psi(x)\bar{\psi}(x')$. Moreover, it is possible to show that in the heavy quark limit, using the hopping parameter expansion and with a proper renormalization of quark fields,

$$S(x, x') = \overbrace{\psi(x)\bar{\psi}(x')} \overset{M\to\infty}{\sim} S_h(x, x'). \tag{7.39}$$

This justifies the statement that the corresponding operator $\hat{S}_h(x, x')$ has non-vanishing overlap with states describing static colour charges, corresponding to a quark-antiquark pair, at distance

r. Therefore, the expectation value of the rectangular Wilson loop in Eq. (7.36) decays at large Euclidean time separations according to

$$\langle W(x,y)\rangle \overset{|x_4-y_4|\to\infty}{\sim} \mathrm{e}^{-V(|\vec{x}-\vec{y}|)|x_4-y_4|}, \tag{7.40}$$

where $V(r)$ is the *static potential*

$$V(r) = A + \frac{B}{r} + \sigma r. \tag{7.41}$$

To extract the string tension, the interesting region is for large $r = |\vec{x}-\vec{y}|$. Using a standard Monte Carlo algorithm with n measurements, the S/N of the estimate is

$$\frac{\langle W\rangle}{\sqrt{\sigma_W^2/n}} \sim \sqrt{n}\,\mathrm{e}^{-\sigma r|x_4-y_4|}, \tag{7.42}$$

thus the number of measurement has to be scaled exponentially, as $n \sim \mathrm{e}^{2\sigma r|x_4-y_4|}$, to have a good signal for large Wilson loops.

7.3.3. Multilevel integration

Combining the decomposition of the partition function in Eq. (7.35) and the factorization of the Wilson loop in Eq. (7.36), we can write its expectation value in a compact form as

$$\langle W(x,y)\rangle = \Big\langle \Big\langle\!\Big\langle S_h(x,x+r\hat{1})\cdots\mathbb{T}\Big\rangle\!\Big\rangle_{\Gamma_\alpha} \Big\langle\!\Big\langle \mathbb{T}\cdots\mathbb{T}\Big\rangle\!\Big\rangle_{\Gamma_{\alpha-1}} \cdots \Big\langle\!\Big\langle \mathbb{T}\cdots\mathbb{T}\Big\rangle\!\Big\rangle_{\Gamma_{\beta+1}} \Big\langle\!\Big\langle \mathbb{T}\cdots S_h(y,y-r\hat{1})\Big\rangle\!\Big\rangle_{\Gamma_\beta}\Big\rangle, \tag{7.43}$$

where we have supposed that $x \in \Gamma_\alpha$ and $y \in \Gamma_\beta$.

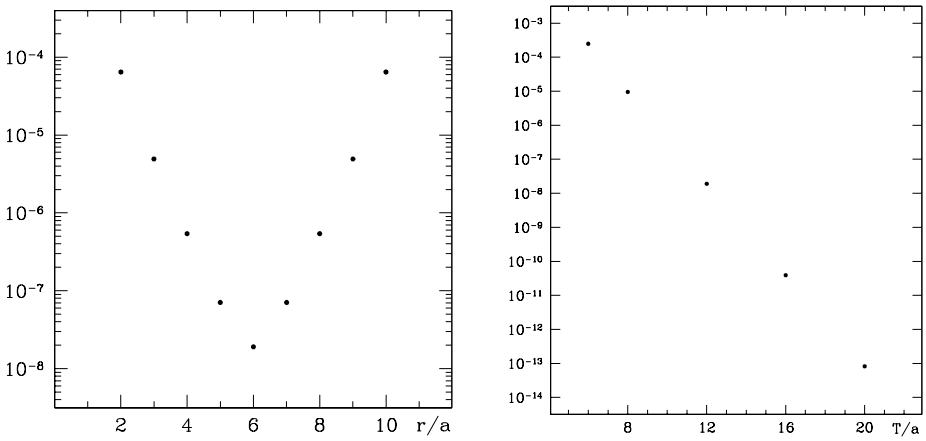
In this form, the expectation value can be evaluated with a two-level Monte Carlo algorithm:

level-0: a number n_0 of field configurations $\{U_i\}$, $i = 1,\dots,n_0$ defined on the whole lattice is generated, using a standard Monte Carlo;

level-1: starting from every level-0 field configuration, a number n_1 of field configurations $\{U_{i,j}\}$, $j = 1,\dots,n_1$ is generated updating independently in the Γ_α domains and keeping the spatial links of the boundary time slices fixed.

The estimator of $\langle W(x,y)\rangle$ is

$$\bar{W}(x,y) = \frac{1}{n_0}\sum_{i=1}^{n_0}\Bigg\{\left[\frac{1}{n_1}\sum_{j=1}^{n_1}\big(S_h(x,x+r\hat{1})\cdots\mathbb{T}\big)[U_{i,j}]\right]\left[\frac{1}{n_1}\sum_{j=1}^{n_1}\big(\mathbb{T}\cdots\mathbb{T}\big)[U_{i,j}]\right]\cdots \cdots\left[\frac{1}{n_1}\sum_{j=1}^{n_1}\big(\mathbb{T}\cdots\mathbb{T}\big)[U_{i,j}]\right]\left[\frac{1}{n_1}\sum_{j=1}^{n_1}\big(\mathbb{T}\cdots S_h(y,y-r\hat{1})\big)[U_{i,j}]\right]\Bigg\}. \tag{7.44}$$



(a) $\langle P^* P \rangle$ vs. distance r, at $T = 12a$.

(b) $\langle P^* P \rangle$ vs. lattice time extent T, at $r = 6a$.

Figure 7.2.: Polyakov loop correlation function $\langle P(x)^* P(x + r\hat{1}) \rangle$ from Ref. [31] on a $L = 12a$ lattice at $\beta = 5.7$ ($a \approx 0.17\,\mathrm{fm}$), with variable time extent L up to $20a$. Statistical errors, not visible on this scale, decrease exponentially with the signal for up to almost ten order of magnitude.

Let us now study the variance of this multilevel Monte Carlo estimate. The sub-lattice expectation value $\langle\!\langle \mathbb{T} \cdots \mathbb{T} \rangle\!\rangle_\Gamma$ is a matrix in colour tensor space. Taking the matrix norm, it looks similar to the correlator of two Polyakov loops at space separation r at finite-temperature, only with fixed boundary conditions instead of thermal ones. For sufficiently large thickness d to be in the confinement phase, the sub-lattice expectation value decays exponentially with Polyakov loops separation r

$$\langle\!\langle \mathbb{T} \cdots \mathbb{T} \rangle\!\rangle_\Gamma \sim \mathrm{e}^{-m_0 r}, \qquad m_0 \lesssim \sigma d. \tag{7.45}$$

Then, suppose to chose a number of level-1 measurements $n_1 = \mathrm{e}^{2m_0 r}$. The noise on the sub-lattice estimator

$$\frac{1}{n_1} \sum_{j=1}^{n_1} (\mathbb{T} \cdots \mathbb{T})[U_{i,j}] \tag{7.46}$$

is reduced to the same order of the signal. As a consequence, each term contributing to the level-0 average is of order $\mathrm{e}^{-m_0 r |x_4 - y_4|/d}$ and the variance is reduced to this level. This means that the S/N after n_0 level-0 measurements is

$$\frac{\langle W \rangle}{\sqrt{\sigma_W^2 / n_0}} \sim \sqrt{n_0} \mathrm{e}^{-(\sigma - m_0/d) r |x_4 - y_4|} = \sqrt{n_0 \cdot n_1^{|x_4 - y_4|/d}} \mathrm{e}^{-\sigma r |x_4 - y_4|}, \tag{7.47}$$

at a computational cost comparable to $n_0 \cdot n_1$ measurements of a standard algorithm. Here, $|x_4 - y_4|/d$ is the number of thick-time-slice domains between x and y. If d is kept fixed in physical units and it is sufficiently large such that $m_0 \to \sigma d$, the S/N dependence on the Wilson loop size is, in principle, gone.

In Figure 7.2 we report the result of the study performed in Ref. [31]. The authors studied the correlator of two Polyakov loops. It can be considered as limit of the rectangular Wilson loop in Eq. (7.40) for $|x_4 - y_4| = T$ on a lattice with periodic time boundary conditions. It is also factorized easily

$$\left\langle P(x)^* P(x + r\hat{1}) \right\rangle = \left\langle \mathrm{tr} \left\{ \langle\!\langle \mathbb{T} \cdots \mathbb{T} \rangle\!\rangle_{\Gamma_{T/d-1}} \cdots \langle\!\langle \mathbb{T} \cdots \mathbb{T} \rangle\!\rangle_{\Gamma_1} \langle\!\langle \mathbb{T} \cdots \mathbb{T} \rangle\!\rangle_{\Gamma_0} \right\} \right\rangle, \tag{7.48}$$

where the trace contracts the $\mathbb{T}$ at $x_4 = T - a$ across the periodic time boundary with the the one at $x_4 = 0$. Using a quite coarse $a \approx 0.17\,\mathrm{fm}$ lattice spacing for demonstration purpose, the domain thickness was chosen as $d = 2a = 0.34\,\mathrm{fm}$.

The results shown in Figure 7.2 are quite impressive. At $r = 6a$ and $T = 12a$, the efficiency of the multilevel algorithm is estimated to be better by a factor $\approx 3 \times 10^5$ with respect to a standard Monte Carlo. For a fixed r and number of measurements, the S/N in the Polyakov loop correlator increases *linearly* with T in the considered range. Therefore, the exponential noise problem is completely solved.

8. Domain decomposition with fermions

It is not straightforward to implement the domain decomposition and multilevel integration program of Section 7.2 to systems with fermions. Obstacles arise because, once fermions have been analytically integrated out in the path integral, as needed for Monte Carlo simulations, the manifest locality of the action and of the observables is lost. Its easy to identify two problems: one with the action, another with observables.

1. The probability density, derived from the action in Eq. (2.105)

$$S[U] = S_{\mathrm{W}}[U] - \ln \mathrm{Det}\big\{ D_{\mathrm{W}}[U] + M_0 \big\}, \tag{8.1}$$

   is a non-local functional of the gauge field, with the non-locality coming from the fermion determinant. Therefore, it is not possible for this theory to satisfy the locality condition of Eq. (7.21) and decompose the partition function in spacetime domains.

2. The expectation values of fermion fields are mapped by Wick's theorem to expectation values in the bosonic theory of quark propagators $S(y, x)$. While the Dirac operator is a sparse matrix with non-zero elements only between nearest neighbour lattice sites, its inverse $S(y, x)$ is not and it depends on links arbitrarily far away from the source x and sink y. It is not possible to write $S(y, x)$ as a product of terms with a localized gauge field dependence.

In order to make lattice computations with fermions amenable for multilevel algorithms, these two problems need a solution.

In this chapter, we report our solution to the second problem, published in Ref. [2] and presented in Ref. [8]. Factorizations of fermionic correlation functions cannot be achieved with the exact inverse of the Dirac operator, because each of its elements depends on the gauge field over the all lattice. However, a series of approximations can be found which exhibit the various degrees of non-locality in the propagator. In Ref. [2] we pursued two strategies, adapted to different types of correlation functions, both based on domain decomposition techniques [179, 104]. In the first, which leads to an efficient algorithm for disconnected correlation functions, we define a succession of domains $\Gamma_0 \subset \Gamma_1 \subset \Gamma_2 \subset \cdots$ which are larger and larger and which contain the two end points x and y of the propagator. The latter can then be expanded in series as

$$S(y, x) = S^{(0)}(y, x) + \big[S^{(1)} - S^{(0)}\big](y, x) + \big[S^{(2)} - S^{(1)}\big](y, x) + \cdots , \tag{8.2}$$

where $S^{(i)}(y, x)$ is the inverse of the Dirac operator restricted to the i^{th} domain, and depends on the values of the gauge field in Γ_i only. The larger the domain Γ_i, the smaller the corresponding term $[S^{(i+1)} - S^{(i)}]$, the less local is its dependence on the gauge field. By inserting $S^{(0)}$ in Wick's contraction of the disconnected contribution of two pseudoscalar densities located in different domains, it is clear that the gauge-field dependence in the product of the two traces is

factorized. The average can then be computed by a two-level Monte Carlo integration scheme. This first strategy is of direct interest for the thesis: it is a step in the direction of implementing a multilevel integration scheme to measure the pseudoscalar flavour-singlet meson mass directly on the lattice.

The second approach in Ref. [2] leads one step further, and demonstrates that also connected hadron correlation functions can be factorized so that multilevel algorithms can be used. This case is beyond the scope of the thesis.

We stress again that an analogous factorization of the fermion determinant is also needed to apply multilevel methods to the QCD case. Once this thesis was almost complete, we proposed a way to factorize the fermion determinant along the same lines of the factorization presented here. For further details see Ref. [4].

In Section 8.1 we introduce in very general terms the locality properties of the domain-decomposed quark propagator. In Section 8.2 we specialize the multilevel integration technique introduced in Chapter 7 to the case of the disconnected quark propagator. Finally, in Section 8.3 we present a numerical test of these ideas.

8.1. Quark propagator and locality

Let us take a lattice Γ with open boundary conditions in the time direction [109], and define the domains

$$\Gamma_0 \subset \Gamma_1 \subset \Gamma_2 \subset \cdots \subset \Gamma. \tag{8.3}$$

We adopt here the same block terminology as in Ref. [104]. We choose the Γ_i to be a hyper-cubic domain of lattice points. Its exterior boundary $\partial\Gamma_i$ is defined to be the set of all points that have distance 1 from Γ_i. Each exterior boundary point has a closest 'partner' point in the block. The interior boundary $\partial\Gamma_i^*$ of Γ_i consists of all these points. The set of points that are not in the block is denoted by

$$\Gamma_i^* = \Gamma \setminus \Gamma_i. \tag{8.4}$$

For a given domain decomposition of the lattice, the Wilson–Dirac operator defined in Eq. (2.39), being a sparse matrix in position space, assumes the block-diagonal form[1]

$$D = \begin{pmatrix} D_{\Gamma_i} & D_{\partial\Gamma_i} \\ D_{\partial\Gamma_i^*} & D_{\Gamma_i^*} \end{pmatrix}. \tag{8.6}$$

The operator D_{Γ_i} acts on quark fields on Γ_i in the same way as D, except that all terms involving the exterior boundary points $\partial\Gamma_i$ are set to zero (which is equivalent to impose Dirichlet boundary conditions on $\partial\Gamma_i$).

We can apply the LU decomposition to the block matrix in Eq. (8.6)

$$D = \begin{pmatrix} \mathbb{1} & 0 \\ D_{\partial\Gamma_i^*} D_{\Gamma_i}^{-1} & \mathbb{1} \end{pmatrix} \times \begin{pmatrix} D_{\Gamma_i} & D_{\partial\Gamma_i} \\ 0 & S_{\Gamma_i^*} \end{pmatrix}, \tag{8.7}$$

[1] In this chapter we denote simply with D the massive Wilson–Dirac operator D_M

$$D(x, y) = D_{\mathrm{W}}(x, y) + M_0 \delta_{x,y}. \tag{8.5}$$

where the *Schur complement* is defined as

$$S_{\Gamma_i^*} = D_{\Gamma_i^*} - D_{\partial\Gamma_i^*} D_{\Gamma_i}^{-1} D_{\partial\Gamma_i}. \tag{8.8}$$

The block inverse of D is then straightforward to write

$$D^{-1} = \begin{pmatrix} D_{\Gamma_i}^{-1} + D_{\Gamma_i}^{-1} D_{\partial\Gamma_i} S_{\Gamma_i^*}^{-1} D_{\partial\Gamma_i^*} D_{\Gamma_i}^{-1} & -D_{\Gamma_i}^{-1} D_{\partial\Gamma_i} S_{\Gamma_i^*}^{-1} \\ -S_{\Gamma_i^*}^{-1} D_{\partial\Gamma_i^*} D_{\Gamma_i}^{-1} & S_{\Gamma_i^*}^{-1} \end{pmatrix}. \tag{8.9}$$

It is worth noting that $S_{\Gamma_i^*}^{-1}$ is the exact block in the block inverse of D. By putting the Schur complement in the upper-left block, the analogous formula can be written as

$$D^{-1} = \begin{pmatrix} S_{\Gamma_i}^{-1} & -S_{\Gamma_i}^{-1} D_{\partial\Gamma_i} D_{\Gamma_i^*}^{-1} \\ -D_{\Gamma_i^*}^{-1} D_{\partial\Gamma_i^*} S_{\Gamma_i}^{-1} & D_{\Gamma_i^*}^{-1} + D_{\Gamma_i^*}^{-1} D_{\partial\Gamma_i^*} S_{\Gamma_i}^{-1} D_{\partial\Gamma_i} D_{\Gamma_i^*}^{-1} \end{pmatrix}, \tag{8.10}$$

with S_{Γ_i} defined as in Eq. (8.8) but with $\Gamma \leftrightarrow \Gamma^*$.

By using the decomposition in Eq. (8.9), the exact quark propagator between the points $x, y \in \Gamma_i$ is given by

$$S(y,x) = S^{(i)}(y,x) + \sum_{w_1,w_2 \in \partial\Gamma_i^*} S^{(i)}(y,w_1) \Big[D_{\partial\Gamma_i} D^{-1} D_{\partial\Gamma_i^*} \Big](w_1,w_2) S^{(i)}(w_2,x), \tag{8.11}$$

where

$$S^{(i)}(y,x) = D_{\Gamma_i}^{-1}(y,x) \tag{8.12}$$

depends on the values of the gauge field in the block Γ_i only. It is rather clear at this point that we can generate a succession of approximations $S^{(i)}$ which, by construction, converges to the exact propagator when Γ_i gets larger and larger. For a typical gauge configuration, when the sink and the source of the two $S^{(i)}$ in the sum on the r.h.s. of Eq. (8.11) are at asymptotically large distances, it holds

$$\mathrm{tr}\big\{ S^{(i)}(y,x) S^{(i)}(y,x)^\dagger \big\}^{1/2} \sim \mathrm{tr}\big\{ S(y,x) S(y,x)^\dagger \big\}^{1/2} \sim \mathrm{e}^{-\frac{1}{2} M_\pi |y-x|}, \tag{8.13}$$

with M_π the mass of the corresponding pseudoscalar meson made of degenerate quarks. A rough estimate of the distance between the exact and the approximated propagator is

$$\mathrm{tr}\big\{ \big[S(y,x) - S^{(i)}(y,x) \big] \big[S(y,x)^\dagger - S^{(i)}(y,x)^\dagger \big] \big\}^{1/2} \sim \mathrm{e}^{-M_\pi d_i}, \tag{8.14}$$

with d_i the average of the distances of x and y from the interior boundaries of Γ_i.

8.2. Multilevel integration of the disconnected pseudoscalar propagator

The decomposition in Eq. (8.11) calls for a multilevel integration of disconnected contributions to correlation functions. We test the idea in SU(3) YM theory with open boundary conditions [109] supplemented with a doublet of quenched quarks, u and d, degenerate in mass. Both

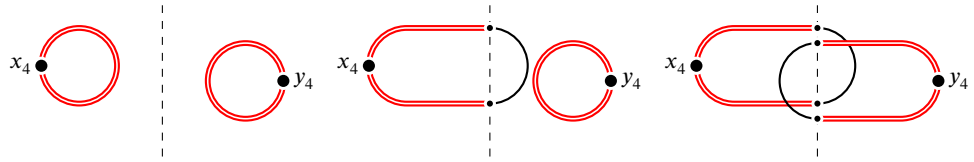


Figure 8.1.: The three type of contributions to the disconnected pseudoscalar propagator in Eqs (8.17)–(8.19b). Black (single) lines are full propagators, red (double) ones are those within a domain.

fermions are discretized with the Wilson–Dirac operator, so that isospin symmetry is exactly preserved. We compute the correlator of two different flavour-diagonal pseudoscalar densities (the generalization to other cases being straightforward)

$$C_{\text{dis.}}(y_4, x_4) = \frac{1}{L^3} \sum_{\vec{x},\vec{y}} \left\langle \bar{d}(y)\gamma_5 d(y)\bar{u}(x)\gamma_5 u(x) \right\rangle = \frac{1}{L^3} \sum_{\vec{x}\vec{y}} \langle W_{\text{dis.}}(y, x)\rangle, \tag{8.15}$$

where $W_{\text{dis.}}(y, x)$ indicates Wick's contraction of the fermion fields, and L is the length of the lattice in the spatial directions[2]. In a standard Monte Carlo simulation, the statistical error of $C_{\text{dis.}}(y_4, x_4)$ is constant as a function of $|y_4 - x_4|$ while its expectation value decreases proportionally to $\exp\{-M_\pi|y_4 - x_4|\}$ at large time separations. The number of configurations n_0 required to reach a given relative statistical error thus grows exponentially with the time distance of the densities, i.e. $n_0 \sim \exp\{2M_\pi|y_4 - x_4|\}$

The depletion of the signal-to-noise ratio is particularly severe at large time-distances. To take advantage of the locality of the theory, it is therefore natural to divide the lattice Γ in two non-overlapping thick time-slices Γ_0 and Γ_0^*. The first time coordinate of Γ_0^*, x_4^{cut}, is chosen approximatively in the middle between the two densities (see Figure 8.1). By using the first and the second diagonal elements in Eqs (8.9) and Eqs (8.10) respectively, Wick's contraction can be decomposed as

$$W_{\text{dis.}}(y, x) = W_{\text{dis.}}^{(\text{f})}(y, x) + W_{\text{dis.}}^{(\text{r})}(y, x), \tag{8.16}$$

where

$$W_{\text{dis.}}^{(\text{f})}(y, x) = \text{tr}\left\{ \gamma_5 D_{\Gamma_0}^{-1}(x, x) \right\} \times \text{tr}\left\{ \gamma_5 D_{\Gamma_0^*}^{-1}(y, y) \right\}, \tag{8.17}$$

$x \in \Gamma_0$, and $y \in \Gamma_0^*$. The rest of the contraction is given by

$$W_{\text{dis.}}^{(\text{r})}(y, x) = \left[W_{\text{dis.}}^{(\text{r}_1)}(y, x) + (\Gamma_0, x) \leftrightarrow (\Gamma_0^*, y) \right] + W_{\text{dis.}}^{(\text{r}_2)}(y, x), \tag{8.18}$$

[2] In this chapter, dimensionful quantities are always expressed in units of the lattice spacing a unless explicitly specified.

with

$$W^{(\mathrm{r}_1)}_{\mathrm{dis.}}(y,x) = \mathrm{tr}\Big\{\gamma_5 D^{-1}_{\Gamma_0}(x,\cdot)[D_{\partial\Gamma_0} D^{-1} D_{\partial\Gamma_0^*}](\cdot,\cdot) D^{-1}_{\Gamma_0}(\cdot,x)\Big\} \times \mathrm{tr}\Big\{\gamma_5 D^{-1}_{\Gamma_0^*}(y,y)\Big\}, \tag{8.19a}$$

$$W^{(\mathrm{r}_2)}_{\mathrm{dis.}}(y,x) = \mathrm{tr}\Big\{\gamma_5 D^{-1}_{\Gamma_0}(x,\cdot)[D_{\partial\Gamma_0} D^{-1} D_{\partial\Gamma_0^*}](\cdot,\cdot) D^{-1}_{\Gamma_0}(\cdot,x)\Big\} \times \mathrm{tr}\Big\{\gamma_5 D^{-1}_{\Gamma_0^*}(y,\cdot)[D_{\partial\Gamma_0^*} D^{-1} D_{\partial\Gamma_0}](\cdot,\cdot) D^{-1}_{\Gamma_0^*}(\cdot,y)\Big\}. \tag{8.19b}$$

When the spatial gauge links at x_4^{cut} are kept frozen, the dependence of the action and of $W^{(\mathrm{f})}_{\mathrm{dis.}}(y,x)$ on the remaining link variables is factorized.

8.2.1. Two-level integration

When an observable depends only on the link variables in a given sub-lattice and the action of the theory is local, it is useful to define its expectation value restricted to that domain. This is a function of the link variables at the boundary of the sub-lattice only, and does not depend on the gauge field values elsewhere. For the trace of the Wilson–Dirac operator that we are interested in, it reads

$$\Big\langle\!\Big\langle \mathrm{tr}\Big\{\gamma_5 D^{-1}_{\Gamma_0}(x,x)\Big\}\Big\rangle\!\Big\rangle = \frac{1}{\mathcal{Z}_{\Gamma_0}} \int \mathcal{D}[U_{\Gamma_0}]\, \mathrm{e}^{-S[U_{\Gamma_0}]}\, \mathrm{tr}\Big\{\gamma_5 D^{-1}_{\Gamma_0}(x,x)\Big\}, \tag{8.20}$$

where $\mathcal{D}[U_{\Gamma_0}]$ and $S[U_{\Gamma_0}]$ are the invariant Haar measure and the action restricted to the domain Γ_0, and the sub-lattice partition function is fixed by requiring that $\langle\!\langle \mathbb{1} \rangle\!\rangle = 1$. By following the standard line of argumentation in multilevel integration technique presented in Chapter 7, it follows that (see also Refs [175, 31, 32, 176–178])

$$\Big\langle \mathrm{tr}\Big\{\gamma_5 D^{-1}_{\Gamma_0}(x,x)\Big\}\, \mathrm{tr}\Big\{\gamma_5 D^{-1}_{\Gamma_0^*}(y,y)\Big\}\Big\rangle = \Big\langle \Big\langle\!\Big\langle \mathrm{tr}\Big\{\gamma_5 D^{-1}_{\Gamma_0}(x,x)\Big\}\Big\rangle\!\Big\rangle \Big\langle\!\Big\langle \mathrm{tr}\Big\{\gamma_5 D^{-1}_{\Gamma_0^*}(y,y)\Big\}\Big\rangle\!\Big\rangle\Big\rangle. \tag{8.21}$$

This suggests that the mean value of $W^{(\mathrm{f})}_{\mathrm{dis.}}(y,x)$ can be computed with a two-level algorithm. First, for each level-0 gauge field (spatial components) on the boundary x_4^{cut}, the averages of the two traces are carried out independently on n_1 level-1 configurations generated independently in the two thick time-slices. Then the average over the level-0 configurations of the product of the two means is performed by updating the gauge links over the entire lattice. The crucial question, to be answered numerically in Section 8.3, is whether one can choose n_1 large enough to profit from the level-1 averaging, or if instead the variance of the factorized contribution is dominated by the fluctuations of the spatial components of the gauge field at the boundary. If n_1 can be taken large enough such that the product of the (level-1) mean values is proportional to $\exp\{-M_\pi|y_4 - x_4|\}$, then a good statistical precision is reached with a number of updates $n_0 \cdot n_1 \sim \exp\{M_\pi|y_4 - x_4|\}$. Notice that the factor in the exponent is halved with respect to the standard Monte Carlo.

The contribution from $W^{(\mathrm{r}_1)}_{\mathrm{dis.}}(y,x)$ is expected to be suppressed, for a typical configuration, by a factor $\exp\{-M_\pi|x_4 - x_4^{\mathrm{cut}}|\}$ at large time separations. Measuring it over the $n_0 \cdot n_1$ configurations generated in the two-level update, by blocking the results and averaging over the

n_0 of them, may be enough to reduce the error at the same level of the one of $W^{(\mathrm{f})}_{\mathrm{dis.}}(y,x)$ (up to a pre-factor that has to be quantified numerically). The last contribution, $W^{(\mathrm{r}_2)}_{\mathrm{dis.}}(y,x)$, is expected to be already proportional to $\exp\{-M_\pi|y_4-x_4|\}$. This is of the same order of the expected signal, and therefore the standard level-0 average is adequate.

8.3. Numerical tests for the disconnected pseudoscalar propagator

We test the ideas discussed in Section 8.2 in the quenched approximation of QCD. We discretize gluons and fermions with the Wilson action, and we impose open and periodic boundary conditions in the time and spatial directions respectively [109, 180]. The inverse coupling constant is fixed to $\beta = 6/g_0^2 = 6.0$, the length of each spatial direction to $L = 24$, and the time extent to $T = 64$. The lattice spacing is $a = 0.093\,\mathrm{fm}$ as fixed by assuming a physical value of 0.5 fm for the Sommer scale $r_0/a = 5.368$ [163]. The up and down quarks are degenerate. Their masses are fixed by the hopping parameter value $\kappa = 0.1560$, corresponding to a pion of approximatively 455 MeV [181].

Numerical simulations have been carried out with a modified version of the openQCD code, version 1.4 [182, 180]. We have generated $n_0 = 200$ level-0 independent gauge field configurations spaced by 400 molecular-dynamics units (MDUs) with the Hybrid Monte Carlo (HMC). Following Section 8.2, the lattice has been split at $x_4^{\mathrm{cut}} = 32$ in two domains of equal size Γ_0 and Γ_0^*. For all level-0 background gauge fields, $n_1 = 100$ level-1 configurations were generated by updating the two regions independently while keeping fixed the spatial links at $x_4^{\mathrm{cut}} = 32$. Also for these updates we used the HMC by skipping 400 MDUs between measurements, a very conservative choice for which the generation of the level-1 configurations is still cheaper than the computation of Wick's contractions. Within this setup, the correlator in Eq. (8.15) is naturally decomposed as

$$C_{\mathrm{dis.}}(y_4,x_4) = C^{(\mathrm{f})}_{\mathrm{dis.}}(y_4,x_4) + C^{(\mathrm{r}_1)}_{\mathrm{dis.}}(y_4,x_4) + C^{(\mathrm{r}_2)}_{\mathrm{dis.}}(y_4,x_4). \tag{8.22}$$

The fully factorized contribution can be written as

$$C^{(\mathrm{f})}_{\mathrm{dis.}}(y_4,x_4) = \frac{1}{L^3}\left\langle \left\langle\!\!\left\langle \sum_{\vec{x}} \mathrm{tr}\left\{\gamma_5 D^{-1}_{\Gamma_0}(x,x)\right\}\right\rangle\!\!\right\rangle \left\langle\!\!\left\langle \sum_{\vec{y}} \mathrm{tr}\left\{\gamma_5 D^{-1}_{\Gamma_0^*}(y,y)\right\}\right\rangle\!\!\right\rangle \right\rangle. \tag{8.23}$$

The other two terms are given by

$$C^{(\mathrm{r}_1)}_{\mathrm{dis.}}(y_4,x_4) = \frac{1}{L^3}\sum_{\vec{x},\vec{y}}\left\langle W^{(\mathrm{r}_1)}_{\mathrm{dis.}}(y,x) + (\Gamma_0,x)\leftrightarrow(\Gamma_0^*,y)\right\rangle \tag{8.24a}$$

$$C^{(\mathrm{r}_2)}_{\mathrm{dis.}}(y_4,x_4) = \frac{1}{L^3}\sum_{\vec{x},\vec{y}}\left\langle W^{(\mathrm{r}_2)}_{\mathrm{dis.}}(y,x)\right\rangle, \tag{8.24b}$$

where $W^{(\mathrm{r}_1)}_{\mathrm{dis.}}(y,x)$ and $W^{(\mathrm{r}_2)}_{\mathrm{dis.}}(y,x)$ are defined in Eqs (8.19a) and (8.19b). If the Wilson–Dirac operator is written as

$$2\kappa D = \mathbb{1} - \kappa D_{\mathrm{hop}}, \tag{8.25}$$

the trace can be re-expressed as[3] [183–185].

$$\mathrm{tr}\left\{\gamma_5 D^{-1}\right\} = \kappa^p \,\mathrm{tr}\left\{\gamma_5 D^p_{\mathrm{hop}} D^{-1}\right\}, \qquad \text{if } p \leq 8. \tag{8.26}$$

By choosing $p = 8$, all the traces appearing in the contributions in Eqs (8.23)–(8.24b) have been estimated stochastically, e.g.

$$\sum_{\vec{x}} \mathrm{tr}\left\{\gamma_5 D^{-1}(x,x)\right\} \longrightarrow \frac{1}{n_{\mathrm{src}}} \sum_{i=1}^{n_{\mathrm{src}}} \sum_{\vec{x}} \eta_i^\dagger(x) \left[\kappa^8 D^8_{\mathrm{hop}} D^{-1} \gamma_5 \eta_i\right](x), \tag{8.27}$$

by inverting the various Dirac operators on the very same $n_{\mathrm{src}} = 100$ Gaussian random sources η_i [186, 187], defined on the whole spacetime volume[4], and by contracting the solution with a time-slice of η_i. The hopping parameter expansion, used in Eq. (8.26), reduces the variance of the stochastic estimator significantly. Other techniques [188–191] may further reduce the cost of the computation, but we prefer to keep it simple and focus on factorization.

The $C^{(\mathrm{f})}_{\mathrm{dis.}}$ contribution is estimated by first averaging, for each of the level-0 configurations, the two traces independently over the n_1 level-1 background fields. The expectation value of the product of the two means is then computed by averaging over the n_0 configurations. The other two contributions are computed as if the n_1 (subset of) configurations, generated for each of the level-0 boundary fields, were correlated level-0 ones. The measures of $C^{(\mathrm{r}_1)}_{\mathrm{dis.}}$ and $C^{(\mathrm{r}_2)}_{\mathrm{dis.}}$ are thus grouped in bins of n_1, and the expectation values and their errors are determined as usual by treating the bins as n_0 independent measurements.

8.3.1. Numerical results

The numerical results for $C^{(\mathrm{f})}_{\mathrm{dis.}}$, $C^{(\mathrm{r}_1)}_{\mathrm{dis.}}$, and $C^{(\mathrm{r}_2)}_{\mathrm{dis.}}$ are plotted in Figure 8.2 as a function of the time separation of the pseudoscalar densities. The central values and their errors are shown in the plots on the left and right columns respectively. The best estimate of $C_{\mathrm{dis.}}$ (the sum of the three) is also shown in each plot on the left for comparison. In all cases x_4 and y_4 belong to different domains, $y_4 > x_4$, and they are chosen to be as much as possible equidistant from x_4^{cut}.

The statistical error on $C^{(\mathrm{f})}_{\mathrm{dis.}}$, top-right plot of Figure 8.2, is a flat function of $|y_4 - x_4|$ with sizeable deviations near the boundaries of the domains. Error bars are smaller than the symbols. Up to the largest value that we have, $n_1 = 100$, the error decreases as n_1^{-1}, i.e. the two-level Monte Carlo works at full potentiality. The mean value of $C^{(\mathrm{f})}_{\mathrm{dis.}}$, top-left plot, is compatible with zero. The correlation between $C_{\mathrm{dis.}}$ and $C^{(\mathrm{f})}_{\mathrm{dis.}}$ goes from 0.9 to 1.0 when $|y_4 - x_4|$ varies from 15 to 50, a value which collapses toward zero when the multilevel is switched on.

The statistical error on $C^{(\mathrm{r}_1)}_{\mathrm{dis.}}$, middle-right plot of Figure 8.2, shows a strong dependence on $|y_4 - x_4|$. It is compatible with an exponential behavior of the form $\exp\{-M|y_4 - x_4|/2\}$ with an effective mass $M = 0.14$, i.e. lighter than expected and roughly 2/3 of the pion mass[5]. It decreases as $n_1^{-1/2}$ up to $n_1 = 100$ and, at fixed time distance, it becomes the dominant

[3] For the $\mathcal{O}(a)$-improved Wilson-Dirac operator, Eq. (8.26) holds if $p \leq 2$.

[4] For the factorized contribution the η_i acts effectively as two independent random sources, one for each domain. The estimate of the two traces is thus obtained with a single global inversion per random source.

[5] We did not attempt to study the dependence of this parameter on the finite size or other sources of systematics.

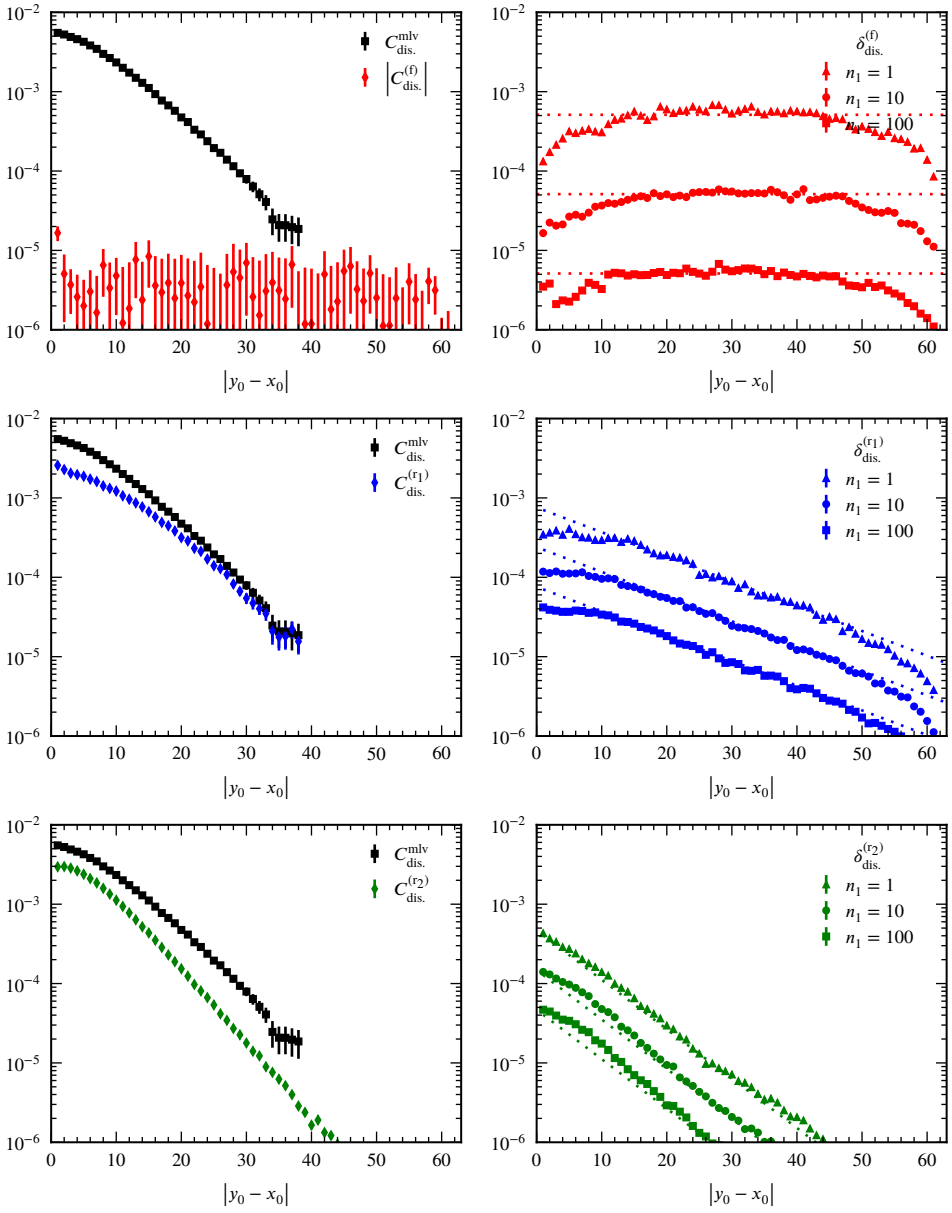


Figure 8.2.: Left-column plots: the three contributions on the r.h.s. of Eq. (8.22) are shown, together with the best estimate of the full correlator (the sum of the three), as a function of the time separation $|y_4 - x_4|$. Right-column plots: the errors of the various contributions are shown as a function of the time distance for various values of n_1.

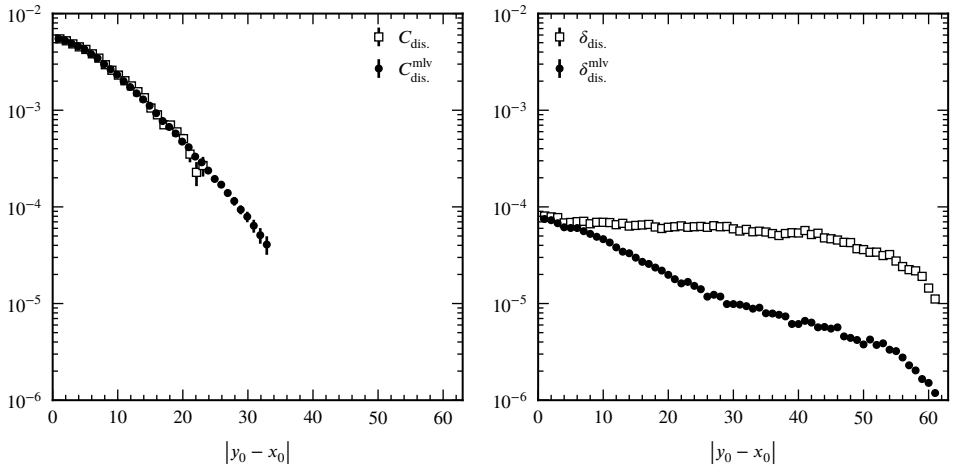


Figure 8.3.: The best estimate of $C_{\text{dis.}}(y_4, x_4)$ (left) and of its error (right) are shown as a function of the time distance, with and without two-level integration of the factorized contribution. In the latter case the n_1 (subset of) configurations, generated for each of the level-0 boundary fields, are treated as if they were correlated level-0 ones. The n_1 measures are thus binned together, and the mean and its error are computed as usual by treating the bins as independent.

contribution to the error of $C_{\text{dis.}}$ once a large enough number n_1 of level-1 updates have been carried out. The mean value of $C^{(\text{r}_1)}_{\text{dis.}}$ is roughly 2/3 of the full correlator at $|y_4 - x_4| = 15$, and it becomes the dominant contribution up to $|y_4 - x_4| = 33$, after which the signal is lost. The statistical errors of $C^{(\text{r}_2)}_{\text{dis.}}$ decreases exponentially as $\exp\{-M|y_4 - x_4|\}$, and it scales as $n_1^{-1/2}$.

A clear picture emerges from the above analysis. At large time distances, the statistical error on the standard estimate of the disconnected pseudoscalar propagator is dominated by the one on $C^{(\text{f})}_{\text{dis.}}$. The second largest contribution is the statistical error on $C^{(\text{r}_1)}_{\text{dis.}}$ which, however, is exponentially suppressed as $\exp\{-M|y_4 - x_4|/2\}$. Once the two-level integration is switched on, the error on $C^{(\text{f})}_{\text{dis.}}$ decreases as n_1^{-1}, while the one on $C^{(\text{r}_1)}_{\text{dis.}}$ continues to scale as[6] $n_1^{-1/2}$. The parameter n_1 can thus be tuned, up to a prefactor of $\mathcal{O}(1)$, so that $n_1 \sim \exp\{Md\}$ with d being the maximum temporal distance in which one is interested in[7]. This way the error on the factorized contribution is reduced to the level of (or below) the uncertainty on $C^{(\text{r}_1)}_{\text{dis.}}$ at the same cost of generating $n_0 \cdot n_1$ global configurations. The net computational gain is therefore $\sim n_1$, and a good statistical precision is reached with a number of updates $n_0 \cdot n_1 \sim \exp\{M_\pi|y_4 - x_4|\}$. Notice that the factor at the exponent is halved with respect to the standard Monte Carlo.

Our best estimate of the disconnected pseudoscalar propagator is shown in Figure 8.3, where

[6] A two-level algorithm can be used to further reduce the statistical error on $C^{(\text{r}_1)}_{\text{dis.}}$ by a domain decomposition of the exact inverse D^{-1} in Eq. (8.19a) with the cut at, for instance, $x_4 = 40$. This is an improvement which goes beyond the exploratory numerical study of this work.

[7] The contributions $C^{(\text{f})}_{\text{dis.}}$ and $C^{(\text{r}_1)}_{\text{dis.}}$ can be computed with different number of sources, different value of p for the HPE, etc. The prefactor of this estimate can thus change depending on the details of the computation.

also the result without the multilevel is reported for comparison. Using the two-level algorithm, the signal-to-noise ratio remains larger than 1 for ten additional time slices. This is better seen in the right plot, where the statistical error is shown in the two cases. With the standard Monte Carlo the error is approximatively flat, while for the two-level algorithm decreases exponentially. The reduction is evident from distance 15, and becomes approximatively $n_1^{1/2} = 10$ at the point $x_4 = 30$ that was taken to fix n_1. For $n_1 = 100$, the overall gain in the computational cost is approximatively 50 since we have to invert two time the Wilson–Dirac operator on each random source.

8.4. Conclusions

Formulating multilevel integration schemes in systems with fermions is not as straight forward as for bosons. The gauge-field dependence of the fermion determinant and of the propagator need to be judiciously factorized before integrating Wick's contractions. Here we have shown that this can be achieved in (quenched) QCD by properly combining the ideas of multilevel integration and domain decomposition of the quark propagator.

The numerical tests that we carried out for the disconnected correlator of two flavour-diagonal pseudoscalar densities (and for a nucleon two-point function, see [2]) show indeed that the signal-to-noise ratio increases exponentially with the time distance of the sources when a two-level integration is at work. In the simple setup that we implemented, the number of configurations needed to reach a given statistical precision is proportional to the square root of those required in the standard case.

9. Conclusions and outlooks

In the thesis, we outlined the program of verifying the Witten–Veneziano as a solution of the $U(1)_A$ problem of QCD, working in the framework of lattice-regularized quantum field theory. This program can be divided in two parts: the computation of the topological susceptibility in Yang–Mills theory and the computation of the η' meson mass.

It is well known that the YM gradient flow is a powerful tool to investigate the topological properties of YM theory. It is, at the same time, theoretically sound and computationally cheap. In Chapter 4 (see Ref. [1]) we studied the cumulants of the topological charge defined on configurations evolved with YM gradient flow. We proved explicitly that the cumulants at positive flow time reduce in the continuum limit to the universal ones, that satisfy the anomalous WTIs.

Exploiting the YM gradient flow definition, in Chapter 5 (see Ref. [1]) we computed the topological susceptibility of SU(3) YM theory. We obtained the result

$$t_0^2\chi_{\mathrm{t}} = 6.67(7) \times 10^{-4}, \qquad \text{for } N_{\mathrm{c}} = 3, \tag{9.1}$$

with a 1 % statistical accuracy on a dimension-4 quantity. We estimated systematic effects from finite volume, continuum limit extrapolation and flow numerical integration to be negligible with respect to the quoted statistical error.

In Chapter 6 (see Ref. [3]) we extended the study of topological properties to $SU(N_{\mathrm{c}})$ YM theory with number of colours $N_{\mathrm{c}} > 3$, for the first time using the theoretically-sound YM gradient flow definition and small lattice spacings. We computed the topological susceptibility with $N_{\mathrm{c}} = 4$, 5 and 6. This allowed us to extrapolate with confidence to the $N_{\mathrm{c}} \to \infty$ limit to obtain the result

$$t_0^2\chi_{\mathrm{t}} = 7.03(13) \times 10^{-4}, \qquad \text{for } N_{\mathrm{c}} \to \infty, \tag{9.2}$$

with a 2 % statistical accuracy and all the systematics under control. The result differs from the $N_{\mathrm{c}} = 3$ value by only (5 ± 2) %, that explain why the SU(3) value is sufficient to saturate the anomalous contribution to the experimental η' meson mass and solve the $U(1)_A$ problem. This small difference is nevertheless in agreement with the expected $\mathcal{O}(1/N_{\mathrm{c}}^2)$ correction. The $N_{\mathrm{c}} \to \infty$ result is the appropriate value to verify the Witten–Veneziano formula, which is exact in the $N_{\mathrm{c}} \to \infty$ limit. In addition, it is one of the LECs of the effective theory obtained expanding simultaneously QCD correlation functions in powers of momenta, quark masses and $1/N_{\mathrm{c}}$.

The accuracy of our result is sufficient to the purpose of testing on the lattice the Witten–Veneziano mechanism. It has to be matched by an equivalent precision in others hadronic quantities entering the Witten–Veneziano formula.

The second part of the thesis concerns the mass of the η' meson. Today, with state-of-the-art techniques, it is not possible to obtain a value of $M_{\eta'}$ with a satisfying accuracy. Available techniques are affected by the exponential degradation of the signal-to-noise ratio of the propagator

with Euclidean time separation between source and sink. For this reason, in the second part we focus on new computational strategies and technical developments.

In Chapter 8 (see Ref. [2]) we developed a technique to factorize the gauge-field dependence of the quark propagator. We applied the factorization to the case of the disconnect contribution to the meson propagator. This is necessary for the application of multilevel Monte Carlo algorithms, which are known to solve the degradation of the signal-to-noise ratio in bosonic theories. We tested the domain decomposition in quenched QCD simulations and we observed an exponential gain applying a multilevel algorithm.

To apply multilevel integration to unquenched QCD, an analogous factorization of the quark determinant is needed. When the thesis was already written, in Ref. [4] we proposed a domain decomposition of the determinant that reuses the same ideas applied to the quark propagator. We tested this factorization with an exploratory study of the disconnected contribution to the meson propagator.

This solves conceptually the problem of computing the disconnected contribution to the flavour-singlet propagator. We remind however that, to extract the flavour-singlet meson mass, the disconnected contribution must be combined with the connected contribution. The two almost cancel to give a steeper exponential decay, and this represents an additional exponential signal-to-noise problem, yet to be solved.

Eventually, the mass of the flavour-singlet meson, either in the chiral limit or its anomalous contribution away from the chiral limit, can be computed with a set of multilevel lattice simulations that allows to extrapolate to the continuum limit. The program of verifying the Witten–Veneziano solution of the $\mathrm{U}(1)_A$ problem would be concluded with the extrapolation of $M_{\eta'}$ to the $N_\mathrm{c} \to \infty$ limit. To our knowledge, this is free of additional obstructions, albeit it is undoubtedly computationally more expensive by a few orders of magnitude with respect to the large-N_c YM topological susceptibility computation.

9.1. Additional results

Besides the results in Refs [1, 3] that meet the subject of the thesis, we obtained additional, but not secondary, results.

We used the YM gradient flow to set the scale. With a modest numerical effort, we computed the t_0 reference scale in YM theory with very high accuracy for $N_\mathrm{c} = 3$ and, generalising appropriately the definition, for $N_\mathrm{c} > 3$. This was instrumental in reporting the continuum limit results with a percent-level statistical accuracy. In the SU(3) case, the value of t_0 in units of the Sommer scale r_0 scale and the Λ-parameter is know. This allowed us to report results in other units commonly used in the literature, albeit with a lowered statistical accuracy.

As a by-product of studying the topological charge at different flow times, we provided a permille-level numerical test of universality of the YM gradient flow definition, reported in Chapter 5.

In Ref. [1] we aimed to probe the non-Gaussianity of the topological charge distribution in SU(3) YM theory. To this purpose, we computed the ratio R between the fourth and second cumulants of the distribution. We performed for the first time a full-fledged study of systematic

effects and obtained the result, reported in Chapter 5,

$$R = 0.233(45). \tag{9.3}$$

Since is not compatible with 1, this value rules out the θ-dependence of the vacuum energy predicted by the semiclassical approximation known as dilute instanton gas model. It implies that fluctuations of the topological charge are of quantum non-perturbative nature. Moreover, R is another LEC of the large-N_c chiral effective theory.

The freezing of topological charge is a technical obstacle to the correct sampling of topological sectors in Monte Carlo simulations. Topology freezes both in the continuum limit and in the $N_c \to \infty$ limit. Open boundary conditions in the time dimension are a known solution for the continuum limit case. In Chapter 6 we showed that OBCs improve also the critical slowing down with respect to N_c.

Ref. [2] contains an additional application of the propagator domain decomposition to the baryon propagator. This is technically more involved: when the source and sink of the quark propagator are in different domains, a projection of internal quark lines on a smaller subspace is needed in practice. We solved this technicality and tested a multilevel setup. For additional details, we refer to the original paper [2].

9.2. Future developments

The techniques developed in Refs [1, 3] suggest a range of possible future developments. It would be interesting to extend the proof of correctness of the YM gradient flow topological charge definition to the theory with quarks. This is directly applicable to a very interesting physical case: if they exist, axions may contribute significantly to the universe energy density through the vacuum realignment mechanism. They would be a candidate for cold dark matter. The axion contribution is determined at a time in cosmological history when the universe had a temperature of a few GeV. While the topological susceptibility of QCD at zero or small temperatures can be computed in chiral perturbation theory, this effective theory fails at temperatures higher that the deconfining transition. Therefore, numerical lattice QCD simulations are the preferred framework to study the topological charge distribution at temperatures of $\mathcal{O}(1\,\mathrm{GeV})$.

The obtained factorization of the quark propagator and determinant allows the application of multilevel Monte Carlo algorithm to a large class of computations. This would lead to an impressive acceleration of computations afflicted by signal-to-noise ratio degradation, opening new perspectives in lattice QCD. Notable examples are generic meson and baryon correlators with and without Wick's contraction, static-light correlators, etc.

Among these applications, we point out the detection of string breaking in QCD. In presence of quarks in the fundamental representation, a flattening of the quark-antiquark potential is expected when the distance exceeds a certain threshold. This happens because the energy stored in the glue string makes favourable the creation of a valence quark-antiquark pair out of the vacuum. The direct observation of this effect in lattice QCD simulations is hindered, among other problems, by a strong exponential worsening with distance of the signal-to-noise ratio. This problem can be solved with a multilevel integration algorithm. In this specific case, only the factorization of the determinant is needed. It can be combined with the known factorization of Wilson or Polyakov loops.

A. Conventions

In this appendix we report the principal conventions that we used in the thesis.

A.1. The special unitary group $\mathrm{SU}(N)$

The *special unitary group* $\mathrm{SU}(N)$ is the Lie group of unitary $N \times N$ matrices with unit determinant. It is a real Lie group of dimension $N^2 - 1$. Topologically, it is compact and simply connected. It is also a simple Lie group, i.e. it does not have nontrivial connected normal subgroups.

A.1.1. The Lie algebra $\mathfrak{su}(N)$

The Lie algebra of $\mathrm{SU}(N)$, denoted by $\mathfrak{su}(N)$, can be identified with linear space traceless Hermitian $N \times N$ matrices, with the regular commutator as Lie bracket. At the abstract level, an element $g \in \mathrm{SU}(N)$ can be represented with an exponential mapping

$$g = \mathrm{e}^{\mathrm{i}tX}, \qquad \text{for } t \in \mathbb{R}, \tag{A.1}$$

to an element $X \in \mathfrak{su}(N)$. A representation R of the Lie algebra and the Lie group is given by a set of traceless Hermitian generators[1]

$$\operatorname{tr} T^a = 0, \qquad T^{a\dagger} = T^a, \quad a = 1, \dots, N^2 - 1. \tag{A.2}$$

They are normalized according to[2]

$$T^a T^a = C_R \mathbb{1}_R, \qquad \operatorname{tr}\{T^a T^b\} = T_R \delta^{ab}, \tag{A.3}$$

where C_R is the quadratic Casimir of the representation and T_R is

$$T_R = \frac{\dim_R}{N^2 - 1} C_R. \tag{A.4}$$

The generators satisfy

$$\left[T^a, T^b\right] = \mathrm{i} f^{abc} T^c, \tag{A.5a}$$

where the *structure constants* f^{abc} are real and antisymmetric in all indices, and

$$\left\{T^a, T^b\right\} = \frac{2T_R}{\dim_R} \delta^{ab} \mathbb{1}_R + d^{abc} T^c, \tag{A.5b}$$

[1] While in the mathematical context generators are usually anti-Hermitian, here we follow the physical convention of using Hermitian generators.

[2] Summation is understood over all repeated indices.

where d^{abc} are real and symmetric in all indices.

It is sometimes useful to define

$$T^0 = \sqrt{\frac{T_R}{\dim_R}}\mathbb{1}_R. \tag{A.6}$$

In this way, Eqs (A.5) can be rewritten

$$\left[T^a, T^b\right] = \mathrm{i}\tilde{f}^{abc}T^c, \qquad \left\{T^a, T^b\right\} = \tilde{d}^{abc}T^c, \tag{A.7}$$

where $a = 0, \ldots, N^2 - 1$ and

$$\tilde{f}^{0bc} = 0, \qquad \tilde{d}^{0bc} = 2\sqrt{\frac{T_R}{\dim_R}}\delta^{bc}. \tag{A.8}$$

A.1.2. The fundamental representation

In the defining *fundamental representation F*, often denoted by $\boldsymbol{N}$ to make explicit the dimension, the generators are $N \times N$ traceless Hermitian matrices. It is conventional to normalize them such as

$$C_F = \frac{N^2 - 1}{2N}, \qquad T_R = \frac{1}{2}. \tag{A.9}$$

The standard basis for the $N = 2$ case is

$$T^a = \frac{\sigma^a}{2}, \tag{A.10}$$

where σ^a are the Pauli matrices

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma^2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \qquad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{A.11}$$

In the $N = 3$ case, is standard to choose

$$T^a = \frac{\lambda^a}{2}, \tag{A.12}$$

where λ^a are the Gell-Mann matrices

$$\begin{aligned}
\lambda^1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \lambda^2 = \begin{pmatrix} 0 & -\mathrm{i} & 0 \\ \mathrm{i} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \lambda^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\
\lambda^4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \qquad \lambda^5 = \begin{pmatrix} 0 & 0 & -\mathrm{i} \\ 0 & 0 & 0 \\ \mathrm{i} & 0 & 0 \end{pmatrix}, \\
\lambda^6 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \lambda^7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -\mathrm{i} \\ 0 & \mathrm{i} & 0 \end{pmatrix}, \qquad \lambda^8 = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.
\end{aligned} \tag{A.13}$$

Strictly related to the fundamental representation is the anti-fundamental representation, usually denoted by $\bar{\boldsymbol{N}}$, which has as generators the complex conjugate of the fundamental generators. The $N = 2$ case is special, since the $\mathbf{2}$ and $\bar{\mathbf{2}}$ are actually the same representation, with generators related by a change of basis.

A.1.3. The adjoint representation

The adjoint representation A is obtained directly from the structure constants, interpreting them as $N^2 - 1$ traceless Hermitian $(N^2 - 1) \times (N^2 - 1)$ matrices

$$T^a{}_{bc} = -\mathrm{i} f_{abc}. \tag{A.14}$$

We have

$$C_A = N, \qquad T_A = N. \tag{A.15}$$

The adjoint is always a real representation, thus there not exist an 'anti-adjoint' representation.

A.2. The link differential operator

The link differential operators acting on functions $f(U)$ of the gauge field are

$$\partial^a_{x,\mu} f(U) = \frac{\mathrm{d}}{\mathrm{d}s} f\left(\mathrm{e}^{sX} U\right)\Big|_{s=0}, \qquad X_\nu(y) = \begin{cases} T^a, & \text{if } (y,\nu) = (x,\mu), \\ 0, & \text{otherwise.} \end{cases} \tag{A.16}$$

While these depend on the choice of the generators T^a, the combination

$$\partial_{x,\mu} f(U) = T^a \partial^a_{x,\mu} f(U) \tag{A.17}$$

can be shown to be basis-independent.

A.3. Dirac γ-matrices

Dirac γ-matrices are four 4×4 matrices which, in Euclidean spacetime, satisfy

$$\left\{\gamma_\mu, \gamma_\nu\right\} = 2\delta_{\mu\nu} \mathbb{1}. \tag{A.18}$$

We choose the γ-matrices to be Hermitian and unitary

$$\gamma_\mu^\dagger = \gamma_\mu = \gamma_\mu^{-1}, \qquad \gamma_\mu^2 = \mathbb{1}. \tag{A.19}$$

It is convenient to define

$$\gamma_5 \equiv -\gamma_1 \gamma_2 \gamma_3 \gamma_4, \tag{A.20}$$

which is also Hermitian and unitary, and anti-commutes with all other γ-matrices

$$\left\{\gamma_5, \gamma_\mu\right\} = 0. \tag{A.21}$$

The six matrices

$$\sigma_{\mu\nu} = -\frac{\mathrm{i}}{4}\big[\gamma_\mu, \gamma_\nu\big], \tag{A.22}$$

with $\mu \neq \nu$, are the generators in the four-dimensional spinor representation, up to a sign, of SO(4), the group of rotations in Euclidean spacetime. For an infinitesimal rotation $R \in \mathrm{SO}(4)$, $R = \mathbb{1} + \omega + \mathcal{O}(\omega^2)$, their actions is

$$\psi(x) \to \psi'(x) = \Lambda\psi(R^{-1}x), \qquad \Lambda = \mathrm{e}^{\mathrm{i}\frac{1}{2}\omega_{\mu\nu}\sigma_{\mu\nu}}. \tag{A.23}$$

This is a reducible representation, since $[\Lambda, \gamma_5] = 0$.

Moreover, the set of sixteen independent matrices

$$\Gamma = \big\{\mathbb{1}, \gamma_\mu, 2\sigma_{\mu\nu}, \mathrm{i}\gamma_\mu\gamma_5, \gamma_5\big\} \tag{A.24}$$

are a basis of the space of Hermitian 4×4 matrices. They all satisfies $\Gamma^2 = \mathbb{1}$, $\mathrm{tr}\,\Gamma = 0$ except for $\Gamma = \mathbb{1}$, $\mathrm{tr}\,\Gamma\Gamma' = 0$ for $\Gamma \neq \Gamma'$.

A.3.1. Chiral representation

In the so-called *chiral representation*, where γ_5 is diagonal, the γ-matrices in Euclidean spacetime are

$$\gamma_{1,2,3} = -\sigma_2 \otimes \sigma_k = \begin{pmatrix} 0 & \mathrm{i}\sigma_{1,2,3} \\ -\mathrm{i}\sigma_{1,2,3} & 0 \end{pmatrix}, \qquad \gamma_4 = \sigma_1 \otimes \mathbb{1}_2 = \begin{pmatrix} 0 & \mathbb{1}_2 \\ \mathbb{1}_2 & 0 \end{pmatrix}, \tag{A.25}$$

$$\gamma_5 = \sigma_3 \otimes \mathbb{1}_2 = \begin{pmatrix} \mathbb{1}_2 & 0 \\ 0 & -\mathbb{1}_2 \end{pmatrix}, \tag{A.26}$$

where σ_i are the Pauli matrices.

A.3.2. Charge conjugation

The charge-conjugation matrix C is a matrix in Dirac γ-matrix space which satisfies

$$C\gamma_\mu C^{-1} = -\gamma_\mu^T. \tag{A.27}$$

In the chiral representation, the explicit form of C is

$$C = -\mathrm{i}\gamma_2\gamma_4 = \sigma_3 \otimes \sigma_2 = \begin{pmatrix} \sigma_2 & 0 \\ 0 & -\sigma_2 \end{pmatrix}. \tag{A.28}$$

A.4. Grassmann numbers

A Grassmann number is a mathematical constructions which allows the definition of a path-integral representation for fermionic variables. Given a collection of n Grassmann variables θ_i, they satisfy the *Grassmann algebra*

$$\theta_i\theta_j = -\theta_j\theta_i. \tag{A.29}$$

That is to say, two Grassmann variables anticommutes. In particular, the square of a Grassmann variable vanishes

$$\theta_i^2 = 0. \tag{A.30}$$

Any function of a single Grassmann variable is then either constant or linear

$$f(\theta) = f_0 + f_1\theta, \qquad \text{with } f_0, f_1 \in \mathbb{R}. \tag{A.31}$$

Differentiation with respect to a Grassmann variable extracts the linear term in θ

$$\frac{\mathrm{d}}{\mathrm{d}\theta} f(\theta) = f_1. \tag{A.32}$$

To apply the path integral formalism, we need to integrate over Grassmann variables. Requiring it to satisfy linearity and partial integration properties

$$\int \mathrm{d}\theta\, [a f(\theta) + b g(\theta)] = a \int \mathrm{d}\theta\, f(\theta) + b \int \mathrm{d}\theta\, g(\theta), \tag{A.33a}$$

$$\int \mathrm{d}\theta \left[\frac{\mathrm{d}}{\mathrm{d}\theta} f(\theta)\right] = 0, \tag{A.33b}$$

defines unambiguously the *Berezin integral*

$$\int \mathrm{d}\theta\, f(\theta) = f_1. \tag{A.34}$$

To define the integral over multiple variables θ_i, we fix the convention of performing the innermost integral first

$$\int \mathrm{d}\theta_2\, \mathrm{d}\theta_1\, \theta_1\theta_2 = \int \mathrm{d}\theta_2\, \theta_2 = +1, \tag{A.35}$$

and we use the symbol $\mathrm{d}^n\theta$ to denote the order of the integrations

$$\int \mathrm{d}^n\theta\, f(\theta) = \int \mathrm{d}\theta_n \left(\cdots \int \mathrm{d}\theta_2 \left(\int \mathrm{d}\theta_1\, f(\theta) \right) \cdots \right). \tag{A.36}$$

Suppose we want to perform the change of variables $\theta_i = \theta_i(\xi_1, \dots, \xi_n)$, with the Jacobian

$$\boldsymbol{J}_{ij} = \frac{\partial \theta_i}{\partial \xi_j}. \tag{A.37}$$

To be consistent with the rules of integration, we have

$$\int \mathrm{d}^n\theta\, f(\theta) = \int \mathrm{d}^n\xi \det \boldsymbol{J}^{-1} f(\theta(\xi)), \tag{A.38}$$

which is the opposite with respect to the transformation of the measure of $\mathbb{R}^n$. This allows to compute Gaussian integrals over Grassmann variables

$$\int \mathrm{d}^n\theta\, \mathrm{e}^{-\frac{1}{2}\theta^T A\theta} = \begin{cases} \sqrt{\det A} & \text{even } n, \\ 0 & \text{odd } n, \end{cases} \tag{A.39a}$$

$$\int \mathrm{d}^n[\bar{\theta}\theta] \mathrm{e}^{-\bar{\theta}^T M\theta} = \det \boldsymbol{M}, \tag{A.39b}$$

where $A = -A^T$ and M are $n \times n$ matrices.

An useful formula is also the generating functional

$$W[\bar{\eta}, \eta] = \int \mathrm{d}^n[\bar{\theta}\theta] \mathrm{e}^{-\bar{\theta}^T M \theta + \bar{\eta}^T \theta + \bar{\theta}^T \eta} = \det M \exp\left\{\bar{\eta}^T M^{-1} \eta\right\}, \tag{A.40}$$

where M is an invertible $n \times n$ matrix. It allows to derive

$$\left\langle \theta_i \bar{\theta}_j \right\rangle = \int \mathrm{d}^n[\bar{\theta}\theta] \theta_i \bar{\theta}_j \mathrm{e}^{-\bar{\theta}^T M \theta} = -\frac{1}{\det M} \frac{\partial}{\partial \bar{\eta}_i} \frac{\partial}{\partial \eta_j} W[\bar{\eta}, \eta] \bigg|_{\eta = \bar{\eta} = 0} = \left(M^{-1}\right)_{ij}. \tag{A.41}$$

A.4.1. Wick's theorem

General expectation values of products of Grassmann numbers is given by Wick's theorem, in its probability theory form, properly generalized to account for anticommutation rules. We have

$$\begin{aligned} \left\langle \theta_{i_1} \bar{\theta}_{j_1} \cdots \theta_{i_m} \bar{\theta}_{j_m} \right\rangle &= \int \mathrm{d}^n[\bar{\theta}\theta] \theta_{i_1} \bar{\theta}_{j_1} \cdots \theta_{i_m} \bar{\theta}_{j_m} \mathrm{e}^{-\bar{\theta}^T M \theta} \\ &= \sum_{P_{1,2,\ldots,m}} \mathrm{sgn}(P) \left(M^{-1}\right)_{i_1 j_{P_1}} \left(M^{-1}\right)_{i_2 j_{P_2}} \cdots \left(M^{-1}\right)_{i_m j_{P_m}}, \end{aligned} \tag{A.42}$$

where the sum runs over all the permutations $P_{1,2,\ldots,m}$ and $\mathrm{sgn}(P)$ is the sign of the permutation P. This formula can be proven using the generating functional in Eq. (A.40) and noting that

$$\left\langle \theta_{i_1} \bar{\theta}_{j_1} \cdots \theta_{i_m} \bar{\theta}_{j_m} \right\rangle = (-)^m \frac{1}{\det M} \frac{\partial}{\partial \bar{\eta}_{i_1}} \frac{\partial}{\partial \eta_{j_1}} \cdots \frac{\partial}{\partial \bar{\eta}_{i_m}} \frac{\partial}{\partial \eta_{j_m}} W[\bar{\eta}, \eta] \bigg|_{\eta = \bar{\eta} = 0}. \tag{A.43}$$

A.5. The Haar measure

Given a function $f(g)$ defined on elements of a compact Lie group $g \in \mathcal{G}$, the *Haar measure* is the proper measure to integrate $f(g)$ over the Lie group. We require it to be invariant under both left and right multiplication with another group element $h \in \mathcal{G}$

$$\mathrm{d}g = \mathrm{d}(gh) = \mathrm{d}(hg) \tag{A.44}$$

and to be normalized so that integration on the whole $\mathcal{G}$

$$\int \mathrm{d}g = 1. \tag{A.45}$$

It can be proven that these conditions define the Haar measure unambiguously. The explicit form for $g = \mathrm{SU}(N)$ can be found in any textbook of lattice gauge theories [69, 70].

A.6. Wick's rotation

Minkowski spacetime is the natural space in which Einstein's theory of special relativity is formulated. It is a four-dimensional real vector space equipped with a metric $\eta_{\mu\nu} = \mathrm{diag}(-1, 1, 1, 1)$, of signature $(-, +, +, +)$, so that the scalar product of two vector x^μ and y^μ is

$$x^\mu y_\mu = x^\mu \eta_{\mu\nu} y^\nu = -x^0 y^0 + x^1 y^1 + x^2 y^2 + x^3 y^3. \tag{A.46}$$

Consequently, the norm of a vector is not positive-definite. This divides vectors in three classes: spacelike vectors if $x^\mu x_\mu > 0$, timelike vectors if $x^\mu x_\mu < 0$ and null or lightlike vectors if $x^\mu x_\mu = 0$. The group of isometries of Minkowski spacetime is the 10-dimensional Poincaré group, which is the semi-direct product of four translations and the 6-dimensional Lorentz group of Lorentz transformations.

Quantum field theory, being the relativistic generalization of quantum mechanics, is naturally formulated in Minkowsi spacetime. However, some mathematical problems are easier to solve in Euclidean spacetime, i.e. a four-dimensional real vector space with a metric given by a Kronecker delta $\delta_{\mu\nu} = \mathrm{diag}(1, 1, 1, 1)$. The scalar product is thus

$$x_\mu y_\mu = x_\mu \delta_{\mu\nu} y_\nu = x_1 y_1 + x_2 y_2 + x_3 y_3 + x_4 y_4, \tag{A.47}$$

where all spacetime indices are subscripted since it is not necessary to distinguish between tangent and cotangent vectors. Moreover, Euclidean spacetime is fundamental to make clear the connection between quantum field theory and statistical mechanics.

If spacetime were a complex vector space rather than a real one, the inner product would be indefinite for both Minkowski and Euclidean metrics. This suggests a way to associate to a theory defined in Minkowski spacetime an equivalent theory in Euclidean spacetime. The time component x^0 of a Minkowski vector $x^\mu_{\mathrm{M}} = (x^0, x^1, x^2, x^3)$ is first allowed to take complex values, by analytic continuation of functions of x^0, and it is then restricted to assume only pure imaginary values $x^0 = \mathrm{i}x_4$. In this way, x_4 is the fourth 'imaginary time' component of a vector $x_\mu = (x_1, x_2, x_3, x_4)$ in a four-dimensional Euclidean spacetime. This procedure is called *Wick's rotation*, since when representing complex numbers on a plane multiplication by i is equivalent to a rotation by $\pi/2$.

In practice, Wick's rotation is implement with the replacements

$$\begin{aligned} &x^i \to x_i, && \partial^i \to \partial_i, && \mathrm{d}x^i \to \mathrm{d}x_i, \\ &x^0 \to \mathrm{i}x_4, && \partial^0 \to -\mathrm{i}\partial_4, && \mathrm{d}x^0 \to \mathrm{i}\mathrm{d}x_4, \end{aligned} \tag{A.48}$$

thus in particular

$$\int \mathrm{d}^4 x \to \mathrm{i} \int \mathrm{d}^4 x\,. \tag{A.49}$$

B. Chiral perturbation theory

In many situations, it has proven useful to describe the low-energy phenomena of an underlying physical theory using an EFT, which reproduces the phenomenology using the appropriate degrees of freedom and a non-renormalizable 'phenomenological' Lagrangian. This is the case also in QCD: Weinberg pioneered the use of an EFT to describe the low-energy dynamics of mesons [192, 193]. The theory is built around the assumption that mesons interaction are constrained by chiral symmetry and realizes in a modern form the predictions of current algebra. The degrees of freedom are identified with the lightest pseudoscalar mesons. In the massless limit, these are the NG bosons of spontaneous χSB and as a consequence they have only derivative interactions. For this reason, the momenta p^2 are treated effectively as coupling constants. Moreover, effects breaking explicitly chiral symmetry, such as small quark masses, are taken into account. Weinberg showed that this EFT, known as *chiral perturbation theory* (χPT), has a simultaneous expansion in powers of momenta and quark masses.

The connection with the underlying theory and its symmetries was systematized by Gasser & Leutwyler [47, 48]. Correlation functions of local fields of QCD are associated to correlations functions of different local fields in χPT. This low-energy theory can be used to derive results such as the GMOR relation in Eq. (1.108) and similar relations in a systematic way.

The construction of χPT is based only on symmetry and its breaking pattern. The elementary fields in the χPT action, denoted by π, represent the NG bosons of the full theory and thus they live in the coset space $\mathcal{G}/\mathcal{H}$, where $\mathcal{G} = \mathrm{SU}(N_\mathrm{f})_L \times \mathrm{SU}(N_\mathrm{f})_R$ and $\mathcal{H} = \mathrm{SU}(N_\mathrm{f})_F$. Choosing as coset representative $\bar{\xi}(\pi) = (\xi_L(\pi), \xi_R(\pi)) \in \mathcal{G}$, under a chiral transformation $g = (V_L, V_R) \in \mathcal{G}$ we have

$$\xi_L(\pi) \to V_L \xi_L(\pi) h^\dagger(\pi, g), \qquad \xi_R(\pi) \to V_R \xi_R(\pi) h^\dagger(\pi, g), \tag{B.1}$$

with the same $h(\pi, g) \in \mathcal{H}$, since the two chiral sector are related by a parity transformation which leaves $\mathcal{H}$ invariant. This allows to combine the two ξ_L, ξ_R in

$$U(\pi) \equiv \xi_R(\pi) \xi_L^\dagger(\pi), \qquad U(\pi) \to V_R U(\psi) V_L^\dagger. \tag{B.2}$$

Explicitly, $U(\pi)$ is parametrized through an exponential mapping

$$U(x) = \exp\{2\mathrm{i}\,\pi(x)/F_0\} = \exp\{2\mathrm{i}\,\pi^a(x)\tau^a/F_0\}. \tag{B.3}$$

The field $U(x)$ is the basic building block of the chiral Lagrangian. Building Lorentz invariant from it, it is easy to write a very simple Lagrangian

$$\begin{aligned}\mathcal{L}_{\chi\mathrm{PT}} &= \frac{F_0^2}{4}\,\mathrm{tr}\{\partial_\mu U^\dagger \partial_\mu U\} \\ &= \mathrm{tr}\{(\partial_\mu \pi)^2\} + \frac{2}{3F_0^2}\,\mathrm{tr}\{(\pi\partial_\mu\pi)^2 - \pi^2(\partial_\mu\pi)^2\} + \mathcal{O}(\pi^4(\partial\pi)^2),\end{aligned} \tag{B.4}$$

which describes NG bosons and their low-energy interactions in the chiral-symmetry theory.

B.1. QCD with sources

This minimal chiral Lagrangian is not very interesting by itself, since in real-world QCD a number of terms which explicitly break chiral symmetry are present. These are quark masses and electroweak interactions, which are usually described by quark-antiquark densities and currents in the full theory. We would like to map these quark-level fields into corresponding fields in the EFT. This is done in an elegant way using external fields. Consider the partition function of massless QCD with the addition of classical fields acting as external sources

$$\mathcal{Z}[v,a,s,p] = \mathrm{e}^{-F[v,a,s,p]} = \int \mathcal{D}[\psi,\bar{\psi},A]\,\mathrm{e}^{-S|_{M=0}-S_{\text{source}}[v,a,s,p]}, \tag{B.5}$$

with

$$\begin{aligned} &S_{\text{source}}[v,a,s,p] \\ &\quad= \int \mathrm{d}^4x\,\left\{-\mathrm{i}\bar{\psi}_L\gamma_\mu l_\mu\psi_L - \mathrm{i}\bar{\psi}_R\gamma_\mu r_\mu\psi_R + \bar{\psi}_L(s-\mathrm{i}p)\psi_R + \bar{\psi}_R(s+\mathrm{i}p)\psi_L\right\} \\ &\quad= \int \mathrm{d}^4x\,\left\{-\mathrm{i}\bar{\psi}\gamma_\mu\big(v_\mu+\gamma_5 a_\mu\big)\psi + \bar{\psi}\big(s-\mathrm{i}\gamma_5 p\big)\psi\right\}, \end{aligned} \tag{B.6}$$

where the fields v_μ, a_μ, s, p and the linear combination

$$r_\mu = v_\mu + a_\mu, \qquad l_\mu = v_\mu - a_\mu \tag{B.7}$$

are Hermitian matrices in flavour space. This theory described by Eq. (B.5) is invariant under *local* $\mathrm{SU}(N_\mathrm{f})_L \times \mathrm{SU}(N_\mathrm{f})_R$ chiral transformations, provided the external fields transform according to

$$r_\mu(x) \to V_R(x)\big[r_\mu(x) + \mathrm{i}\partial_\mu\big]V_R^\dagger(x), \tag{B.8a}$$

$$l_\mu(x) \to V_L(x)\big[l_\mu(x) + \mathrm{i}\partial_\mu\big]V_L^\dagger(x), \tag{B.8b}$$

$$s(x) + \mathrm{i}p(x) \to V_R(x)[s(x) + \mathrm{i}p(x)]V_L^\dagger(x). \tag{B.8c}$$

In particular, $s+\mathrm{i}p$ transform just like the spurion mass term $\mathcal{M}$ in Eq. (1.68). Thus, the massive theory is obtained breaking the symmetry and setting $v_\mu = a_\mu = 0$ and $s + \mathrm{i}p = \mathcal{M}$.

B.2. The chiral Lagrangian

The χPT is build as an EFT which approximate the partition function of QCD at low energies, where the physics of the NG bosons described by $U(x)$ dominates

$$\mathcal{Z}[v,a,s,p] = \mathcal{Z}_{\chi\text{PT}}[v,a,s,p] + \text{truncation errors}, \tag{B.9}$$

where

$$\mathcal{Z}_{\chi\text{PT}}[v,a,s,p] = \mathrm{e}^{-F_{\chi\text{PT}}[v,a,s,p]} = \int \mathcal{D}[U]\,\mathrm{e}^{-\int \mathrm{d}^4x\mathcal{L}_{\chi\text{PT}}[U;v,a,s,p]}. \tag{B.10}$$

The *chiral Lagrangian* $\mathcal{L}_{\chi\text{PT}}$ is constructed as the most general Lagrangian for the field U and external sources consistent with Lorentz invariance and *local* chiral symmetry. To this purpose we introduce the covariant derivative for U

$$D_\mu U = \partial_\mu U - \mathrm{i} r_\mu U + \mathrm{i} U l_\mu, \tag{B.11}$$

which transform as U under local chiral transformations.

Of course, an infinite number of terms can be built from these fields. We write the chiral Lagrangian as an expansion in powers of momenta. To do this, we need power-counting rules for the source fields. Eq. (B.11) suggests $l_\mu, r_\mu = \mathcal{O}(p)$, moreover we postulate $s + \mathrm{i}p = \mathcal{O}(p^2)$. With this power counting, to leading order in momenta the chiral Lagrangian is composed by two terms

$$\mathcal{L}^{(0)}_{\chi\text{PT}} = \frac{F_0^2}{4} \operatorname{tr}\left\{ D_\mu U^\dagger D_\mu U \right\} - \frac{F_0^2}{4} \operatorname{tr}\left\{ U^\dagger \chi + \chi^\dagger U \right\}, \tag{B.12}$$

where

$$\chi = 2B_0(s + \mathrm{i}p) \tag{B.13}$$

contains the coupling constant of the second term. The external sources select a special direction in flavour space, thus they break explicitly chiral symmetry. However, it's important that symmetry is broken in the same way as in the QCD Lagrangian, which includes short-distance effects.

The lowest-order Lagrangian depends on two low-energy constants (LEC): F_0 and B_0. Their physical meaning can be understood from correlators obtained deriving with respect to sources the generating functional F in Eq. (B.10) and comparing with the same correlators in QCD. At at leading order in momenta we are considering the theory at the tree-level, therefore the LO $F_{\chi\text{PT}}$ reduces to the classical action

$$F^{(0)}_{\chi\text{PT}} = S^{(0)}_{\chi\text{PT}}. \tag{B.14}$$

The left and right chiral currents are

$$J_{L\mu} = \frac{\delta S^{(0)}_{\chi\text{PT}}}{\delta l_\mu} = \frac{1}{2}\mathrm{i} F_0^2 D_\mu U^\dagger U = F_0 D_\mu \pi - \mathrm{i}(\pi \overleftrightarrow{D}_\mu \pi) + \mathcal{O}\left(\frac{\pi^3}{F_0}\right), \tag{B.15a}$$

$$J_{R\mu} = \frac{\delta S^{(0)}_{\chi\text{PT}}}{\delta r_\mu} = \frac{1}{2}\mathrm{i} F_0^2 D_\mu U U^\dagger = -F_0 D_\mu \pi - \mathrm{i}(\pi \overleftrightarrow{D}_\mu \pi) + \mathcal{O}\left(\frac{\pi^3}{F_0}\right), \tag{B.15b}$$

thus the axial current is

$$A_\mu = \frac{J_{L\mu} - J_{R\mu}}{2} = F_0 D_\mu \pi + \mathcal{O}\left(\frac{\pi^3}{F_0}\right) \tag{B.16}$$

and F_0 at $\mathcal{O}(p^2)$ is the decay constant of pseudoscalar mesons

$$\left\langle 0 \middle| A^a_\mu \middle| \pi^b \right\rangle = \mathrm{i} p_\mu \delta^{ab} F_0. \tag{B.17}$$

Similarly, taking the derivative with respect to the scalar source field s,

$$S^{ji} = \bar{\psi}^j \psi^i = \frac{\delta S^{(0)}_{\chi\mathrm{PT}}}{\delta s^{ji}} = -\frac{F_0^2}{4} 2B_0 \big(U^{\dagger ij} + U^{ji}\big), \tag{B.18}$$

we relates the constant B_0 to the quark condensate Σ

$$\Sigma \delta^{ij} = \left\langle \bar{\psi}^j \psi^i \right\rangle = -F_0^2 B_0 \delta^{ij}. \tag{B.19}$$

It is now easy to compute how the symmetry breaking from quark masses generates a mass for the NG bosons. An Hermitian quark mass matrix M correspond to an external source $s = M$ to the massless QCD Lagrangian, along with $p = v_\mu = a_\mu = 0$. With this choice of sources, expanding U in powers of π, the terms not involving derivatives are

$$-\frac{F_0^2}{4} 2B_0 \operatorname{tr}\big\{M(U + U^\dagger)\big\} = 2B_0 \left[\text{const} + \operatorname{tr}\big\{M\pi^2\big\} - \frac{1}{3F_0^2} \operatorname{tr}\big\{M\pi^4\big\} + \mathcal{O}\left(\frac{\pi^6}{F_0^4}\right)\right]. \tag{B.20}$$

From the $\mathcal{O}(\pi^2)$ term we can read the pNG masses. In the $N_\mathrm{f} = 3$ case, an explicit parametrization in terms of the physical pseudoscalar mesons of Table 0.1 is

$$\pi = \frac{1}{2} \begin{pmatrix} \pi^0 + \frac{1}{\sqrt{3}}\eta_8 & \sqrt{2}\pi^+ & \sqrt{2}K^+ \\ \sqrt{2}\pi^- & -\pi^0 + \frac{1}{\sqrt{3}}\eta_8 & \sqrt{2}K^0 \\ \sqrt{2}K^- & \sqrt{2}\bar{K}^0 & -\frac{2}{\sqrt{3}}\eta_8 \end{pmatrix}. \tag{B.21}$$

Choosing $M = \operatorname{diag}(m_u, m_d, m_s)$ gives

$$M^2_{\pi^\pm} = 2m_{ud}B_0, \qquad M^2_{\pi^0} = 2m_{ud}B_0 - \epsilon + \mathcal{O}\big(\epsilon^2\big), \tag{B.22a}$$

$$M^2_{K^\pm} = \big(m_u + m_s\big)B_0, \qquad M^2_{K^0} = \big(m_u + m_s\big)B_0, \tag{B.22b}$$

$$M^2_{\eta_8} = \frac{2}{3}\big(m_{ud} + 2m_s\big)B_0 + \epsilon + \mathcal{O}\big(\epsilon^2\big), \tag{B.22c}$$

with

$$m_{ud} = \frac{1}{2}\big(m_u + m_d\big), \qquad \epsilon = \frac{B_0}{4}\frac{\big(m_u - m_d\big)^2}{m_s - m_{ud}}. \tag{B.23}$$

These formulae generalize the GMOR relation in Eq. (1.108), with which they agree at leading order.

B.2.1. Beyond leading order

Moreover, the extension to NLO and beyond of this EFT framework is straightforward. More independent terms are included in the Lagrangian, up to a given order. Since every term has its own LEC, the number of LECs grows rapidly. At NLO, $\mathcal{O}(p^4)$, the chiral Lagrangian includes

ten additional LECs L_i[1]

$$\begin{aligned}\mathcal{L}^{(1)}_{\chi\mathrm{PT}} =& - L_1 \operatorname{tr}\{D_\mu U^\dagger D_\mu U\}^2 - L_2 \operatorname{tr}\{D_\mu U^\dagger D_\nu U\} \operatorname{tr}\{D_\mu U^\dagger D_\nu U\} \\ &+ L_3 \operatorname{tr}\{D_\mu U^\dagger D_\mu U D_\nu U^\dagger D_\nu U\} + L_4 \operatorname{tr}\{D_\mu U^\dagger D_\mu U\} \operatorname{tr}\{U^\dagger \chi + \chi^\dagger U\} \\ &+ L_5 \operatorname{tr}\{D_\mu U^\dagger D_\mu U (U^\dagger \chi + \chi^\dagger U)\} - L_6 \operatorname{tr}\{U^\dagger \chi + \chi^\dagger U\}^2 \\ &- L_7 \operatorname{tr}\{U^\dagger \chi - \chi^\dagger U\}^2 - L_8 \operatorname{tr}\{\chi^\dagger U \chi^\dagger U + U^\dagger \chi U^\dagger \chi\} \\ &+ \mathrm{i} L_9 \operatorname{tr}\{r^{\mu\nu} D_\mu U D_\nu U^\dagger + l^{\mu\nu} D_\mu U^\dagger D_\nu U\} + L_{10} \operatorname{tr}\{U^\dagger r^{\mu\nu} U l_{\mu\nu}\} \\ &+ H_1 \operatorname{tr}\{r_{\mu\nu} r^{\mu\nu} + l_{\mu\nu} l^{\mu\nu}\} - H_2 \operatorname{tr}\{\chi^\dagger \chi\} + \mathcal{L}_{\mathrm{WZW}},\end{aligned} \tag{B.24}$$

where we used the field strength tensors

$$X^{\mu\nu} = \partial_\mu X_\nu - \partial_\nu X_\mu - \mathrm{i}[X_\mu, X_\nu], \qquad X_\mu = l_\mu, r_\mu. \tag{B.25}$$

Starting from this order, quantum effects are taken into account. Tree-level diagrams from $\mathcal{L}^{(1)}_{\chi\mathrm{PT}}$ and one-loop diagrams of vertices in $\mathcal{L}^{(0)}_{\chi\mathrm{PT}}$ contribute to the same order in the NLO generating functional. The loop effects renormalizes the LECs, introducing a renormalization-scale dependence: $L_i = L_i(\mu)$.

The LECs are universal: they do not depend on the process considered. χPT is fully predictive, to a given order, once a sufficient number of LECs is fixed matching the low-energy theory with the underlying fundamental theory, i.e. QCD. Since, in the regime of validity of χPT, QCD cannot be studied in perturbation theory, lattice simulations are the preferred framework to compute the LECs.

Most of the LECs up to $\mathcal{O}(p^4)$ have been determined. A subset of them can be determined experimentally to a better accuracy than on the lattice. See for instance Ref. [194]. However, the convergence of the χPT expansion depends on the observable considered, and it not always satisfying at the NLO level, especially in the $N_\mathrm{f} = 3$ case at the physical strange quark mass.

χPT is widely used in lattice simulation to extrapolate results at different quark masses. To this purpose, various modifications have been introduced, for instance to include discretization effects, or *quenched* and *partially quenched* χPT to account for sea quarks with different masses from the valence quarks.

[1] In the $N_\mathrm{f} = 2$ case not all these terms are independent.

C. The transfer matrix

In this appendix we present a detailed derivation of the transfer matrix of Wilson's discretization of QCD [67, 83, 84], i.e. Wilson's plaquette action in Eq. (2.24) and the Wilson–Dirac operator in Eq. (2.39). We remember that, for a Euclidean field theory regularized on a lattice with time extent T and lattice spacing a, the partition function is

$$\mathcal{Z} = \operatorname{Tr} \hat{T}^{T/a}, \tag{C.1}$$

where $\hat{T}$ is the transfer matrix that evolves a quantum state by a time a. The derivation follows the lines of Ref [85, 195].

We start with a clever decomposition of the plaquette action in time-slice components

$$S_{\mathrm{W}}[U] = \sum_{x_4=0}^{T-a} \Big\{ K_{\mathrm{g}}(x_4 + a, x_4) + \frac{1}{2} W_{\mathrm{g}}(x_4) + \frac{1}{2} W_{\mathrm{g}}(x_4 + a) \Big\}, \tag{C.2}$$

where

$$K_{\mathrm{g}}(x_4 + a, x_4) = \beta \sum_{\vec{x}} \sum_{k} \left[1 - \frac{1}{N_{\mathrm{c}}} \operatorname{Re} \operatorname{tr} U_{0k}(x) \right], \tag{C.3a}$$

$$W_{\mathrm{g}}(x_4) = \beta \sum_{\vec{x}} \sum_{j<k} \left[1 - \frac{1}{N_{\mathrm{c}}} \operatorname{Re} \operatorname{tr} U_{jk}(x) \right]. \tag{C.3b}$$

Similarly, the Wilson–Dirac operator is split in

$$D[U] = \sum_{x_4=0}^{T-a} \big\{ B(x_4) + C(x_4) - K^{-}(x_4, x_4 + a) - K^{+}(x_4 + a, x_4) \big\}, \tag{C.4}$$

where

$$B(x_4) = 1 + m_0 - \frac{1}{2} \nabla_k^* \nabla_k, \qquad C(x_4) = \frac{1}{2} \gamma_k \big(\nabla_k^* + \nabla_k \big), \tag{C.5a}$$

$$K^{-}(x_4, x_4 + a) = P_{-} \big(1 + \nabla_4 \big), \qquad K^{+}(x_4 + a, x_4) = P_{+} \big(1 + \nabla_4^* \big). \tag{C.5b}$$

Here, we introduced the projectors

$$P_{\pm} = \frac{1 \pm \gamma_4}{2}, \tag{C.6}$$

which can be used to decompose the Dirac four-spinor fields into bi-spinor fields

$$\chi = P_{+} B^{1/2} \psi, \qquad \xi^{\dagger T} = P_{-} B^{1/2} \psi, \tag{C.7a}$$

$$\chi^{\dagger} = \bar{\psi} B^{1/2} P_{+}, \qquad \xi^{T} = -\bar{\psi} B^{1/2} P_{-}, \tag{C.7b}$$

which in turns allow the fermion action to be written as

$$S_{\mathrm{Wf}} = \sum_{x_4=0}^{T-a} \left\{ K_{\mathrm{g}}(x_4+a, x_4) + W_{\mathrm{g}}(x_4) + \bar{W}_{\mathrm{g}}(x_4+a) + \chi^\dagger(x_4)\chi(x_4) + \xi^\dagger(x_4)\xi(x_4) \right\}, \quad \text{(C.8)}$$

where

$$\begin{aligned} K_{\mathrm{g}}(x_4+a, x_4) &= \xi^T(x_4)\mathrm{e}^{-M(x_4)}K^-(x_4, x_4+a)\mathrm{e}^{-M(x_4+a)}\xi^{\dagger T}(x_4+a) \\ &\quad - \chi^\dagger(x_4+a)\mathrm{e}^{-M(x_4+a)}K^+(x_4+a, x_4)\mathrm{e}^{-M(x_4)}\xi(x_4), \end{aligned} \quad \text{(C.9a)}$$

$$W_{\mathrm{g}}(x_4) = -\xi^T(x_4)\mathrm{e}^{-M(x_4)}C(x_4)\mathrm{e}^{-M(x_4)}\chi(x_4), \quad \text{(C.9b)}$$

$$\bar{W}_{\mathrm{g}}(x_4) = \chi^\dagger(x_4)\mathrm{e}^{-M(x_4)}C(x_4)\mathrm{e}^{-M(x_4)}\xi^{\dagger T}(x_4), \quad \text{(C.9c)}$$

with $M(x_4) = (1/2)\ln B(x_4)$.

Using these decompositions of the action, the partition function path integral can then be rewritten as

$$\mathcal{Z} = \int \prod_{x_4=0}^{T-a} \mathcal{D}\left[U_k, \chi, \chi^\dagger, \xi, \xi^\dagger\right](x_4)\mathrm{e}^{-\chi^\dagger(x_4)\chi(x_4)-\xi^\dagger(x_4)\xi(x_4)}T(x_4+a, x_4), \quad \text{(C.10)}$$

where $T(x_4+a, x_4)$ is the field

$$\begin{aligned} &T(x_4+a, x_4) = T[U(x_4+a), \chi^\dagger(x_4+a), \xi^\dagger(x_4+a); U(x_4), \chi(x_4), \xi(x_4)] \\ &= \exp\left\{ \operatorname{tr} M(x_4+a) + \operatorname{tr} M(x_4) - \frac{1}{2}W_{\mathrm{g}}(x_4) - \frac{1}{2}W_{\mathrm{g}}(x_4+a) - \bar{W}_{\mathrm{g}}(x_4+a) - W_{\mathrm{g}}(x_4) \right\} \\ &\qquad \times \int \mathcal{D}\left[U_4(x_4)\right] \exp\left\{ -K_{\mathrm{g}}(x_4+a, x_4) - K_{\mathrm{g}}(x_4+a, x_4) \right\}, \end{aligned} \quad \text{(C.11)}$$

and the path integral measure has been decomposed

$$\mathcal{D}[U, \psi, \bar{\psi}] = \left\{ \prod_{x_4=0}^{T-a} \mathcal{D}\left[U_k, \chi, \chi^\dagger, \xi, \xi^\dagger\right](x_4) \right\} \times \left\{ \prod_{x_4=0}^{T-a} \mathcal{D}\left[U_4(x_4)\right] \right\}, \quad \text{(C.12a)}$$

$$\mathcal{D}\left[U_k(x_4)\right] = \prod_{\vec{x}} \prod_{k=1}^{3} \mathrm{d}U_k(\vec{x}, x_4), \qquad \mathcal{D}\left[U_4(x_4)\right] = \prod_{\vec{x}} \mathrm{d}U_4(\vec{x}, x_4). \quad \text{(C.12b)}$$

Notice that the path integral over temporal direction links $U_4(x)$ is treated differently. We can exploit gauge invariance, define $U_4(\vec{x}, x_4) = \Omega^\dagger(\vec{x}, x_4)\Omega'(\vec{x}, x_4+a)$ and transform it as

$$\begin{aligned} &\int \mathcal{D}\left[U_4(x_4)\right] \exp\left\{ -K_{\mathrm{g}}(x_4+a, x_4) - K_{\mathrm{g}}(x_4+a, x_4) \right\} \\ &\qquad = \int \mathcal{D}\left[\Omega(x_4)\right] \mathcal{D}\left[\Omega'(x_4+a)\right] \exp\left\{ -K_{\mathrm{g}}(x_4+a, x_4) - K_{\mathrm{g}}(x_4+a, x_4) \right\}. \end{aligned} \quad \text{(C.13)}$$

C.1. Quantum field theory Hilbert space

To pass to the transfer matrix operator we need to introduce the Hilbert space of the lattice QFT, which is the tensor product of pure gauge-field Hilbert space and a pure fermion space of states

$$\mathcal{H} = \mathcal{H}_{\rm g} \otimes \mathcal{H}_{\rm f}. \tag{C.14}$$

The pure gauge Hilbert space is just the space of square integrable complex functions acting on the $3(L/a)^3$ spatial links at fixed time

$$\mathcal{H}_{\rm g} = \left[L^2(\mathcal{G})\right]^{3(L/a)^3}, \qquad \mathcal{G} = {\rm SU}(N_{\rm c}). \tag{C.15}$$

Introducing a complete set of states $|U(x_4)\rangle$ in Schrödinger picture, on which the field operator $\hat{U}_k(\vec{x})$ is diagonal

$$\big|U(x_4)\rangle = \prod_{\vec{x},m} \big|U_k(\vec{x},x_4)\rangle, \quad m=1,2,3, \qquad \hat{U}_k(\vec{x})\big|U(x_4)\rangle = U_k(\vec{x},x_4)\big|U(x_4)\rangle, \tag{C.16}$$

the scalar product of two states at the same time $|\Psi(x_4)\rangle, |\Phi(x_4)\rangle \in \mathcal{H}_{\rm g}$ is given by

$$\langle\Psi(x_4)\big|\Phi(x_4)\rangle = \int \mathcal{D}\big[U_k(x_4)\big]\, \Psi^*(U,x_4)\Phi(U,x_4), \tag{C.17}$$

where the wave function $\Psi(U,x_4)$ is

$$\Psi(U,x_4) = \langle U(x_4)\big|\Psi(x_4)\rangle. \tag{C.18}$$

Moreover, the decomposition of the identity operator holds

$$\mathbb{1} = \int \mathcal{D}\big[U_k(x_4)\big]\, \big|U(x_4)\rangle\langle U(x_4)\big|. \tag{C.19}$$

Only gauge-invariant states which satisfies $|\Psi\rangle = |\Psi^{\Omega}\rangle$, where Ω denotes a gauge transformation, are physical. We introduce the projector $\hat{P}_{\rm g}$ over the physical, gauge-invariant sector of the Hilbert space. Its action on a generic state is

$$\big\langle U(x_4)\big|\hat{P}_{\rm g}\big|\Psi(x_4)\big\rangle \equiv \int \mathcal{D}[\Omega]\,\Psi\big(U^{\Omega^\dagger},x_4\big), \qquad U_k^{\Omega}(\vec{x},x_4) = \Omega(\vec{x})U_k(\vec{x},x_4)\Omega^\dagger(\vec{x}+a\hat{k}). \tag{C.20}$$

As for the fermionic space of states $\mathcal{H}_{\rm f}$, it is the Fock space generated by the action of the quark and antiquark field operators $\hat{\chi}(\vec{x})$ and $\hat{\bar{\xi}}(\vec{x})$ on the vacuum $|0\rangle$. They satisfies an anticommuting algebra and are the canonical-formalism equivalent of the Grassmann-number-valued bi-spinor fields in Eqs (C.7). Any operator in $\mathcal{H}_{\rm f}$ can be represented as an integral operator in an associated Grassmann algebra. In the following, we avoid to express explicitly the transfer matrix action on the fermionic wave functions. We refer to Ref. [85] and the standard literature for the explicit expression.

C.2. Transfer matrix operator

Inserting T/a times Eq. (C.19) in Eq. (C.1) and identifying $|U(T)\rangle = |U(0)\rangle$, the partition function is rewritten as a product of transfer matrix elements between $|U\rangle$ states

$$\mathcal{Z} = \int \prod_{x_4=0}^{T-a} \mathcal{D}\big[U_k(x_4)\big] \, \big\langle U(x_4+a)\big|\hat{T}\big|U(x_4)\big\rangle. \tag{C.21}$$

Comparing the latter with Eq. (C.10), the transfer matrix element is

$$\begin{aligned}\big\langle U(x_4+a)\big|\hat{T}\big|U(x_4)\big\rangle &= \big\langle U(x_4+a)\big|\hat{P}_{\mathrm{g}}\hat{T}'\hat{P}_{\mathrm{g}}\big|U(x_4)\big\rangle = \mathrm{e}^{-\frac{1}{2}\left(W_{\mathrm{g}}(x_4)+W_{\mathrm{g}}(x_4+a)\right)} \\ \times \int \mathcal{D}\big[\Omega(x_4)\big] \, \mathcal{D}\big[\Omega'(x_4+a)\big] &\exp\big\{-K_{\mathrm{g}}(x_4+a,x_4)\big\}\hat{T}_{\mathrm{f}}^{\dagger}[U^{\Omega'}(x_4+a)]\hat{T}_{\mathrm{f}}[U^{\Omega}(x_4)],\end{aligned} \tag{C.22}$$

where $\hat{T}_{\mathrm{f}}$ is an operator in fermion space. Using the bi-spinor field operators $\hat{\chi}$ and $\hat{\bar{\xi}}$, it is possible to show that $\hat{T}_{\mathrm{f}}$ is given by

$$\hat{T}_{\mathrm{f}} = \mathrm{e}^{\frac{1}{2}\operatorname{tr} M} \exp\big\{-\hat{\chi}^{\dagger} M \hat{\chi} - \hat{\bar{\xi}}^{T} M \hat{\bar{\xi}}^{\dagger T}\big\} \exp\big\{\hat{\bar{\xi}}^{T}\mathrm{e}^{-M} C \mathrm{e}^{-M}\hat{\chi}\big\}. \tag{C.23}$$

The resulting transfer matrix is the self-adjoint, bounded operator

$$\hat{T} = \hat{P}_{\mathrm{g}}\hat{T}'\hat{P}_{\mathrm{g}} > 0. \tag{C.24}$$

It automatically projects on the physical sector of gauge-invariant states and it is possible to show that it is strictly positive [67, 84]. Moreover, it is unique up to an unitary transformation. It follows that the lattice Hamiltonian

$$\hat{H} \equiv -\frac{1}{a}\ln\hat{T}, \tag{C.25}$$

is self-adjoint and unique.

D. The $\frac{1}{N_c}$ expansion

In this appendix, we study how gauge theories change when the rank of the gauge group—the N in either SO(N) or SU(N)—is modified. The result is that 'More variables usually means greater complexity, but not always' [196, 197]. Indeed, there are gauge theories that become simpler as N becomes larger. These theories have an expansion in powers of $1/N$, known as *large-N expansion*.

As 't Hooft showed in Ref. [22], SU(N_c) gauge theory, where N_c is the number of colours, with or without Dirac fermions, possess an expansion in powers of $1/N_c$. In the limit $N_c \to \infty$, the theory is much simpler: for example, in this limit the anomaly vanishes, the $\mathrm{U}(1)_V \times \mathrm{U}(1)_A$ symmetry in the chiral limit becomes spontaneously broken and a flavour singlet NG boson appears. Moreover, the large-N_c expansion if relevant to the physical case: there are indications the part of the QCD phenomenology can be understood considering $N_c = 3$ close to the $N_c \to \infty$ limit!

The correct $N_c \to \infty$ limit, known as *'t Hooft limit*, is obtained treating the coupling g_0 as a quantity of $\mathcal{O}(1/\sqrt{N_c})$. To this purpose, we define the *'t Hooft coupling* which is $\mathcal{O}(1)$ in $1/N_c$

$$\lambda_0 \equiv N_c g_0^2. \tag{D.1}$$

In term of this new coupling, the action in Eqs (1.7) and (1.13) is rewritten as

$$S[\bar\psi, \psi, A] = N_c \int \mathrm{d}^4x \left(\frac{1}{2\lambda_0} \operatorname{tr}\{F_{\mu\nu}F_{\mu\nu}\} - \frac{\mathrm{i}\theta}{N_c} q + \bar\psi \left[\not{D} + M_0 \right] \psi \right). \tag{D.2}$$

The analysis needed for $1/N_c$ expansion is purely combinatoric, thus the extension to the renormalized theory is trivial. Furthermore, we consider only one fermion flavour.

With a N_c factor in front of the action in Eq. (D.2), Feynman rules have a basic N_c-dependence: every interaction vertex is proportional to N_c and every propagator is proportional to $1/N_c$. Moreover, to count N_c powers of a general Feynman diagram we must follow colour indices. The quark propagator is

$$\overbrace{\psi^i(x)\bar\psi_j}(y) = \delta^i_j S(x-y), \tag{D.3}$$

where we have written explicitly colour indices $i, j = 1, \dots, N_c$ and $S(x-y)$ is the propagator of a single Dirac field. Thus a quark propagator simply identifies a single colour index between two vertices. The gluon propagator is

$$\overbrace{[A_\mu]^i_j(x)[A_\nu]^k_l}(y) = \left(\delta^i_l \delta^k_j - \frac{1}{N_c} \delta^i_j \delta^k_l \right) D_{\mu\nu}(x-y), \tag{D.4}$$

where $D_{\mu\nu}(x-y)$ is the propagator of a single massless vector field. The term proportional to $1/N_c$ is there because the gluon field A_μ is a traceless matrix. If the gauge group were U(N_c),

Figure D.1.: Left: A planar vacuum Feynman diagram with one quark loop. Right: The corresponding double-line diagram, showing the colour structure.

this term would not be present. However, being it proportional to $1/N_c$, it is subleading in the $N_c \to \infty$ limit and it can be dropped if we are interested only in the leading order. Thus, the gluon propagator identifies a couple of colour indices between two vertices: at leading order, it has the same colour structure of a quark-antiquark pair propagator. Therefore, to every Feynman diagram is associated a double-line graph where each line represent a colour index instead of a virtual particle.

Consider only connected vacuum diagrams, i.e. connected diagrams without external lines. Then, every index line must form a closed index loop. This can be thought as the perimeter of a polygon. The double-line graph is thus a two-dimensional surface made of polygons, where two edges are identified if they lie on the same gluon propagator. As the index lines have an arrow on them, the perimeter of a polygon is oriented and, by right-hand rule, also the polygon itself gets an orientation. The index lines on a gluon propagator are always oppositely directed, thus the orientation is consistent over the whole surface and double-line graphs are identified with *oriented* surfaces. Each of these surfaces is characterized by its number of vertices V, edges E and faces F. Each vertex is an interaction vertex in the corresponding Feynman diagram, and carries a factor of N_c. Each edge is a propagator, and carries a factor of $1/N_c$. Each face is an index loop, and carries a factor of N_c from the sum over all index values. Therefore, a double-line graph is associated to the power of N_c

$$N_c^{V-E+F} = N_c^{\chi}, \tag{D.5}$$

where $\chi = V - E + F$ is a topological invariant, the *Euler characteristic* of the surface. In general, the Euler characteristic of an oriented surface is given by the formula

$$\chi = 2 - 2G - B, \tag{D.6}$$

where the B is the number of *boundary* of the surface—the number of holes the surface has—and G is the *genus* of the surface—in simpler terms, the number of 'handles' attached to the surface. For example, a sphere has no handles and no holes: its Euler characteristic is 2. A torus is topologically a sphere with an handle on it, so has $\chi = 0$. A disk is a sphere with a hole and flattened on a plane, so it has $\chi = 1$. A cylindrical surface is a sphere with two holes, so it has $\chi = 0$.

A hole in the surface can be made removing a polygonal face. In terms of Feynman diagrams, this means to substitute a gluon loop with a quark loop. Moreover, a graph on a surface with no

handles is *planar*, which means that it can be projected on a plane so that edges intersect only at a vertex. Therefore, a first result is:

1. leading order connected vacuum diagrams are $\mathcal{O}(N_c^2)$.
   They are planar diagrams with only gluons.

Next, we want to consider diagrams including quarks. A quark loop must be the boundary of an hole in the surface. When we project the graph on a plane, we can always choose the quark loop to be the outer loop, with all the gluon propagators inside the quark loop. Thus, a second result is:

3. leading order connected vacuum diagrams with quarks are $\mathcal{O}(N_c)$.
   They are planar diagrams with a single quark loop that forms the boundary of the graph.

At higher order in the expansion, things get more complicated. For example, diagrams $\mathcal{O}(1)$ in N_c are given by planar graphs with two quark loops, or by non-planar graphs made only of gluons. Moreover, we can no longer neglect the subleading term in the gluon propagator. Nevertheless, a large part of hadrons phenomenology requires only these two leading order results.

D.1. Hadrons in the $1/N_c$ expansion

According to the naïve quark model, colour singlet hadrons can be made from quark-antiquark pairs (mesons) or singlet combinations of N_c quarks (baryons). Other models include exotic mesons, like particles made of two quark-antiquarks pairs, exotic baryons and glueballs, made only of gluons. Baryons are difficult to treat in the $1/N_c$ expansion because it takes an infinite number of quarks to form a colour singlet state in the $N_c \to \infty$ limit. A theory for baryons in the $1/N_c$ expansion was formulated by Witten [198]. Here we focus only on mesons and glueballs, which are much simpler to treat.

We need to generalize the analysis for vacuum diagrams of the preceding section to diagrams with external lines. Consider a local quark bilinear $B(x)$, i.e. a local gauge-invariant operator involving one $\bar{\psi}(x)$ and one $\psi(x)$, and possibly some gluon fields, at the same spacetime point. Let

$$\langle B_1(x_1) B_2(x_2) \cdots B_n(x_n) \rangle_{\text{con}} \tag{D.7}$$

be the connected correlation function of a string of quark bilinears. This vacuum expectation value can be obtained from the path integral appending to the action source terms in the form

$$S \to S + N_c \sum_i b_i B_i, \tag{D.8}$$

and deriving with respect to the b_i the free energy $F = -\ln \mathcal{Z}$

$$\langle B_1(x_1) B_2(x_2) \cdots B_n(x_n) \rangle_{\text{con}} = N_c^{-n} \left. \frac{\partial^n F}{\partial b_1 \cdots \partial b_n} \right|_{b_i = 0}. \tag{D.9}$$

With the modified action, we have now vacuum diagrams with a new type of vertex inserted along quark lines. Each vertex carries a factor of N_c and splits a quark propagator in two, with a

factor of $1/N_c$, hence the power of N_c of the whole diagram is not changed. As before, leading order vacuum diagrams with one quark loop are proportional to N_c. The connected correlation functions are obtained from Eq. (D.9) and thus

$$\langle B_1(x_1)B_2(x_2)\cdots B_n(x_n)\rangle_{\text{con}} \sim N_c^{1-n}. \tag{D.10}$$

In a similar way, a relation for local gauge-invariant operators $G(y)$ made only from gluon fields can be obtained. For a correlation function of mixed B and G we have

$$\langle B_1(x_1)\cdots B_n(x_n)G_1(y_1)\cdots G_m(y_m)\rangle_{\text{con}} \sim N_c^{1-n-m}, \tag{D.11}$$

whereas, if the correlation function contains only G operators, we have

$$\langle G_1(y_1)\cdots G_m(y_m)\rangle_{\text{con}} \sim N_c^{2-m}, \tag{D.12}$$

since in this case no quark lines are needed and leading order vacuum diagrams are thus proportional to N_c^2.

These B and G operators are important because they create respectively meson and glueball states when applied to the vacuum. This is true only if we suppose that the theory confines quarks and gluons for arbitrarily large-N_c, in such a way that only states composed by colour singlet particles are created. Regarding QCD, this is supported by experiment, but for larger N_c this is only an assumption. Quark bilinears B are not properly normalized to create mesons with N_c-independent amplitude. Therefore, we renormalize them defining

$$B_i' \equiv \sqrt{N_c} B_i \tag{D.13}$$

and thus Eq. (D.10) is replaced by

$$\left\langle B_1'(x_1)B_2'(x_2)\cdots B_n'(x_n)\right\rangle_{\text{con}} \sim N_c^{\frac{2-n}{2}}. \tag{D.14}$$

Meson scattering amplitudes can be extracted from these correlation functions using LSZ reduction formula, thus a scattering amplitude with n legs is proportional to $N_c^{(2-n)/2}$. The three-point interaction in proportional to $N_c^{-1/2}$, and this can be viewed as an effective coupling for meson scattering. One consequence is that for large-N_c mesons interact weakly. Also, it can be demonstrated that to the leading order quark bilinears B' create only stable one-meson states.

For glueballs, the G operators are correctly normalized, thus from Eq. (D.12) a glueball scattering amplitude with m legs is proportional to N_c^{2-m} and N_c^{-1} is a effective coupling for glueball scattering. Therefore, glueballs interact even more weakly than mesons in the $N_c \to \infty$ limit. Moreover, instead of Eq. (D.11) we have

$$\left\langle B_1'(x_1)\cdots B_n'(x_n)G_1(y_1)\cdots G_m(y_m)\right\rangle_{\text{con}} \sim N_c^{1-\frac{n}{2}-m}, \tag{D.15}$$

hence meson-glueball mixing vanishes in $N_c \to \infty$ limit as $N_c^{-1/2}$.

D.2. The θ-angle dependence in the $1/N_c$ expansion

The large-N_c expansion allows to study the θ-angle dependence in a regime of strong coupling, in which instanton calculus is not valid. We start assuming that there is θ dependence in the 't Hooft limit. In the presence of light quarks, it can be shown that, if χSB survives in the $N_c \to \infty$, so does the θ dependence. Moreover, the WV argument requires the θ dependence to be present in the free energy at leading order in $1/N_c$.

Then, the form of Eq. (D.2) suggests, in analogy with the 't Hooft coupling, to define a new 'ϑ-angle'

$$\vartheta \equiv \frac{\theta}{N_c} \tag{D.16}$$

that is kept constant in the $N_c \to \infty$. The free energy, which is $\mathcal{O}(N_c^2)$, depends as any other function on ϑ

$$F(\theta) = N_c^2 h\left(\vartheta = \frac{\theta}{N_c}\right). \tag{D.17}$$

However, this is apparently in contrast with the fact that θ is an angular variable

$$F(\theta) = F(\theta + 2\pi). \tag{D.18}$$

A way out relies on the existence of multiple candidates vacuum states that all become stable, but not degenerate in energy, in the $N_c \to \infty$ limit [129]. The free energy of the k^{th} vacuum would be

$$F_k(\theta) = N_c^2 h\left(\frac{\theta + 2\pi k}{N_c}\right). \tag{D.19}$$

For any value of θ, the true stable vacuum is the one that minimizes the $F(\theta)$ with respect to k

$$F(\theta) = N_c^2 \min_k h\left(\frac{\theta + 2\pi k}{N_c}\right). \tag{D.20}$$

Therefore, the free energy is periodic in θ, as expected, but, if h is not constant, it is not smooth: at some values of θ there is a change in the vacuum that minimize the free energy.

Using the fact that $F(\theta)$ is an even function with a minimum in $\theta = 0$, we can write for small ϑ

$$h(\vartheta) = C\vartheta^2 + \mathcal{O}\big(\vartheta^4\big), \qquad C \geq 0. \tag{D.21}$$

Assuming that $C \neq 0$, we insert this in Eq. (D.20) to obtain

$$F(\theta) = C \min_k (\theta + 2\pi k)^2 + \mathcal{O}\left(\frac{1}{N_c}\right). \tag{D.22}$$

In the limit of large-N_c, the constant C is of $\mathcal{O}(1)$ and gives a non-vanishing topological susceptibility

$$\chi_t(\theta) = \frac{1}{V}\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} F(\theta) = \frac{C}{V}, \tag{D.23}$$

while higher powers of θ in the free energy are suppressed and thus the topological charge distribution tends to a normal one. In particular, higher cumulants are of order

$$C_n(\theta) = -\frac{(-\mathrm{i})^{2n}}{V}\frac{\mathrm{d}^{2n}}{\mathrm{d}\theta^{2n}} F(\theta) = \mathcal{O}\big(N_c^{2-2n}\big). \tag{D.24}$$

E. Instantons

An *instanton* is a non-trivial solution to the equations of motion of a classical field theory in Euclidean spacetime. In Minkowski spacetime, instantons are interpreted as quantum tunnelling between topologically inequivalent vacua. In this appendix we briefly introduce instantons in the context of gauge theories and give some useful formula. We follow the pedagogical introduction of Refs [199, 197]. See also Ref. [200].

Yang–Mills theory admits instanton solutions: these are topologically non-trivial field configurations that absolutely minimize the action within their topological sector. Let us start considering the YM action in Eq. (1.7) with SU(2) gauge group and $\theta = 0$

$$S_{\mathrm{g}}[A] = \int \mathrm{d}^4x \, \frac{1}{2g_0^2} \, \mathrm{tr}\big\{ F_{\mu\nu}(x) F_{\mu\nu}(x) \big\}. \tag{E.1}$$

We are interested in gauge field configurations of finite action (not necessarily solutions of the equations of motion). This is because infinite action configurations give zero if used as saddle point for a functional Gaussian integration in a semiclassical approximation of the path integral. For the integral in Eq. (E.1) to converge, $F_{\mu\nu}$ must fall off faster than r^{-2}, for large radial coordinates $r = |x|$, thus we require $F_{\mu\nu} = \mathcal{O}(r^{-3})$. It follows that A_μ is $\mathcal{O}(r^{-2})$ except for a gauge transformation of the trivial configuration. We can write A_μ in the form

$$A_\mu = -\mathrm{i}g \partial_\mu g^{-1} + \mathcal{O}\big(r^{-2}\big), \tag{E.2}$$

where g is a function from spacetime to the gauge group of $\mathcal{O}(1)$ in r, i.e. a function of angular coordinates only: $g : S^3 \to \mathrm{SU}(2)$. A gauge transformation $\Omega(x)$ acts on the field A_μ in Eq. (E.2) according to Eq. (1.2) and leads to a transformation law for g

$$g \to g' = \Omega g + \mathcal{O}\big(r^{-2}\big). \tag{E.3}$$

If it were possible to choose Ω to be equal to g^{-1} as $r \to \infty$, g would be set equal to one and eliminated from Eq. (E.2) by a gauge transformation. In general, this is not possible, as Ω needs to be a continuous function all over spacetime, not only on the S^3 hypersurface at infinity. It needs to be continuous on each S^3 hypersurface as the radial coordinate go from zero to infinity. At $r = 0$, $\Omega(x)$ is a constant as here S^3 shrinks to a point, therefore Ω restricted on S^3 at infinity is such a function that can be obtained by continuous deformation from a constant function. Moreover, a constant gauge transformation is obtained with continuity from identity, since SU(2) is connected. Thus, Ω restricted on S^3 at infinity is such a function that can be obtained by continuous deformation from $h = \mathbb{1}$.

In topology, two continuous functions from one topological space to another are called *homotopic* if one can be continuously deformed into the other. This is an equivalence relation, thus an *homotopy class* is the equivalence class of all homotopic functions. g as functions from

S^3 to SU(3) are all in the homotopy class of the identity. A gauge transformation applied to g can transform it to any function homotopic to g, but not to a function in another homotopy class. In conclusion, there is a gauge inequivalent A_μ field configuration for each homotopy class of functions $g: S^3 \to \mathrm{SU}(2)$.

$g \in \mathrm{SU}(2)$ is a unitary 2×2 matrix with unit determinant, and it is well known it can be written

$$g = \mathrm{i}\vec{a} \cdot \vec{\sigma} + a_4 \mathbb{1}_2, \qquad \text{with } |a| = 1, \tag{E.4}$$

where $\vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ are the Pauli matrices, $\vec{a} = (a_1, a_2, a_3)$ and the condition $|a| = 1$ is imposed. This identifies SU(2) with a S^3 sphere embedded in $\mathbb{R}^4$. Thus, as a topological space, SU(2) is S^3 and we have to study homotopy classes of functions from S^3 to S^3. Some standard functions of this type are the trivial mapping

$$g^{(0)}(x) = \mathbb{1}_2 \tag{E.5}$$

and the identity mapping

$$g^{(1)}(x) = \frac{\mathrm{i}\vec{x} \cdot \vec{\sigma} + x_4 \mathbb{1}_2}{r}. \tag{E.6}$$

Both are elements of the family of mappings

$$g^{(\nu)}(x) = \left(g^{(1)}(x)\right)^\nu, \tag{E.7}$$

where ν is called the *winding number* or *Pontryagin index*. It can be shown that every function from S^3 to S^3 is homotopic to one of these. A winding number can be associated with every mapping by the formula

$$\nu = -\frac{1}{24\pi^2} \int \mathrm{d}\theta_1 \, \mathrm{d}\theta_2 \, \mathrm{d}\theta_3 \, \epsilon^{ijk} \, \mathrm{tr}\left\{(g\partial_i g^{-1})(g\partial_j g^{-1})(g\partial_k g^{-1})\right\} \tag{E.8}$$

where θ_1, θ_2 and θ_3 are three angles that parametrize S^3. This quantity is a homotopic invariant. To show this, consider an infinitesimal transformation of g

$$g \to g' = g[\mathbb{1} + \mathrm{i}\lambda(x)], \qquad \lambda(x) = \lambda^a(x) T^a. \tag{E.9}$$

The variation is

$$\delta g = \mathrm{i} g \lambda \quad \Rightarrow \quad \delta(g\partial_k g^{-1}) = -\mathrm{i} g(\partial_k \lambda) g^{-1} \tag{E.10}$$

and, because the three derivatives in Eq. (E.8) make equal contributions,

$$\delta\nu \sim \int \mathrm{d}\theta_1 \, \mathrm{d}\theta_2 \, \mathrm{d}\theta_3 \, \epsilon^{ijk} \, \mathrm{tr}\left\{(g\partial_i g^{-1})(g\partial_j g^{-1}) g(\partial_k \lambda) g^{-1}\right\}. \tag{E.11}$$

Using the identity

$$0 = \partial_j(g g^{-1}) = g\partial_j g^{-1} + (\partial_j g) g^{-1}, \tag{E.12}$$

we get

$$\delta\nu \sim \int \mathrm{d}\theta_1 \, \mathrm{d}\theta_2 \, \mathrm{d}\theta_3 \, \epsilon^{ijk} \, \mathrm{tr}\left\{(\partial_i g^{-1})(\partial_j g)(\partial_k \lambda)\right\}, \tag{E.13}$$

which vanishes if integrated by parts. Invariance under infinitesimal deformations implies that ν is left invariant by finite continuous deformations.

Furthermore, Eq. (E.8) coincides with the exponent in mappings in Eq. (E.7). We show this for $g^{(1)}$, for which the integrand in equation (E.8) is constant, thus we are free to evaluate it at the 'North pole' of S^3: $x_4 = 1$, $x_i = 0$, where the angles can be chosen as $\theta_i = x_i$. This gives $g\partial_i g^{-1} = -\mathrm{i}\sigma_i$ and, using $-\mathrm{i}\sigma_1\sigma_2\sigma_3 = \mathbb{1}_2$,

$$\epsilon^{ijk}\,\mathrm{tr}\big\{(g\partial_i g^{-1})(g\partial_j g^{-1})(g\partial_k g^{-1})\big\}\big|_{g=g^{(1)}} = -12, \tag{E.14}$$

that, since the area of S^3 is $2\pi^2$, is the desired result $\nu = 1$. For other mappings in Eq. (E.7), it suffices note that

$$g = g_1 g_2 \quad \Rightarrow \quad \nu = \nu_1 + \nu_2. \tag{E.15}$$

This is because it is possible to continuously deform g_1 to be constant in half of S^3 and g_2 to be constant in the other half of S^3, without changing ν_1 and ν_2. Then the integral in Eq. (E.8) is the sum of a part, due to g_1, vanishing in one half and a part, due to g_2, vanishing in the other half.

Actually, the topological charge Q introduced in Section 3.1.1 coincides with the winding number ν. Indeed, the integral of the topological charge density $q(x)$ in Eq. (3.2) can be written by Gauss law

$$Q = \int \mathrm{d}^4x\, q(x) = \int \mathrm{d}\sigma_\mu\, K_\mu, \tag{E.16}$$

where the integral in the r.h.s. is performed on a large S^3 surface and K_μ is the topological current defined in Eq. (3.5)

$$K_\mu(x) = \frac{1}{16\pi^2}\epsilon_{\mu\nu\rho\sigma}\,\mathrm{tr}\Big\{A_\nu(x)\Big[F_{\rho\sigma}(x) - \frac{2\mathrm{i}}{3}A_\rho(x)A_\sigma(x)\Big]\Big\}. \tag{E.17}$$

The first term in the right-hand side is $\mathcal{O}(r^{-4})$ and makes no contribution to the integral, the second term gives

$$Q = \int \mathrm{d}^4x\, q(x) = -\frac{\mathrm{i}}{24\pi^2}\int \mathrm{d}\sigma_\mu\,\epsilon_{\mu\nu\rho\sigma}\,\mathrm{tr}\big\{A_\nu A_\rho A_\sigma\big\}, \tag{E.18}$$

which, with A_μ for large r given by the expression in Eq. (E.2), is the integral formula in Eq. (E.8) for the winding number. Therefore,

$$Q = \int \mathrm{d}^4x\, q(x) = \nu. \tag{E.19}$$

E.1. The BPST instanton

We can derive an explicit instanton solution. Introducing the dual tensor

$$\tilde{F}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}F_{\rho\sigma}, \tag{E.20}$$

we notice that $\tilde{F}_{\mu\nu}\tilde{F}_{\mu\nu} = F_{\mu\nu}F_{\mu\nu}$, from which follows

$$0 \le \frac{1}{2}\,\mathrm{tr}\Big\{\big(\tilde{F}_{\mu\nu} \pm F_{\mu\nu}\big)^2\Big\} = \mathrm{tr}\big\{F_{\mu\nu}F_{\mu\nu}\big\} \pm \mathrm{tr}\big\{F_{\mu\nu}\tilde{F}_{\mu\nu}\big\}. \tag{E.21}$$

The l.h.s. is non-negative, thus

$$S_{\mathrm{g}}[A] = \frac{1}{2g_0^2}\int \mathrm{d}^4x\,\mathrm{tr}\{F_{\mu\nu}F_{\mu\nu}\} \geq \frac{1}{g_0^2}\left|\int \mathrm{d}^4x\,\mathrm{tr}\{F_{\mu\nu}\tilde{F}_{\mu\nu}\}\right| = |\nu|S_0, \tag{E.22}$$

where we defined

$$S_0 \equiv \frac{8\pi^2}{g_0^2}. \tag{E.23}$$

For any winding number, this is an absolute lower bound on the action, derived for the first time in 1975 by Belavin, Polyakov, Schwartz and Tyupkin [201]. The bound is saturated for (anti-)self-dual solutions

$$F_{\mu\nu} = \pm\tilde{F}_{\mu\nu}, \tag{E.24}$$

that represent instanton solutions. Moreover, being also absolute minima for a given ν, they dominate the path integral in the quantum theory and other solution in the same topological sector are negligible. The solution with $\nu = 1$ is the so-called *BPST instanton*

$$A_\mu(x) = \frac{\mathrm{i}r^2}{r^2+\rho^2}\left[g^{(1)}(x)\right]^{-1}\partial_\mu g^{(1)}(x) + \mathcal{O}\Big(\frac{1}{r^2}\Big), \tag{E.25}$$

where $g^{(1)}$ is the mapping in Eq. (E.6) and ρ is an arbitrary constant, the *instanton size*. The existence of instantons of any size is a consequence of scale invariance of the classical theory. Other $\nu = 1$ solutions can be obtained from this applying the symmetries of the theory: scale transformations, rotations, translations, special conformal transformations and gauge transformations. Scale transformations change the instanton size. Rotations can always be undone by constant gauge transformations, thus they do not generate new solutions. Spacetime translations generate a four-parameter family of solutions, labelled by the *position of the centre of the instanton* x_0. Special conformal transformations can be undone by gauge transformations and translations. Finally, gauge transformations change the solutions in Eq. (E.25). Indeed, because $g^{(1)}$ is a function of the angular coordinates only, the radial component of A_μ, A_r, vanishes. Under a general gauge transformation $\Omega(x)$, one has

$$A_r \to A_r' = -\mathrm{i}\Omega(A_r + \partial_r)\Omega^{-1} = -\mathrm{i}\Omega\partial_r\Omega^{-1}. \tag{E.26}$$

Hence, for A_r to be invariant, Ω must be independent of r, i.e. a constant gauge transformation. But the only constant gauge transformation that leaves the solutions in Eq. (E.25) unchanged is the identity, while the others generate new solutions. In conclusion, constant gauge transformations generate a three-parameter family of solutions, resulting in an eight-parameter family: one from scale transformations, four form spacetime translations, three from constant gauge transformations. Explicitly, the family of solutions is [202]

$$A^a_\mu(x) = 2\frac{\eta^a_{\mu\nu}(x-x_0)_\nu}{(x-x_0)^2+\rho^2}, \tag{E.27}$$

where $\eta^a_{\mu\nu}$ is the *'t Hooft symbol*

$$\eta^a_{\mu\nu} = \begin{cases} \epsilon_{a\mu\nu} & \text{for } \mu,\nu = 1,2,3, \\ -\delta_{a\nu} & \text{for } \mu = 4, \\ \delta_{a\mu} & \text{for } \nu = 4, \\ 0 & \text{for } \mu = \nu = 4. \end{cases} \tag{E.28}$$

Since the winding number is odd under parity transformation, it must exist also a family of $\nu = -1$ solutions: its elements are *BPST anti-instantons*, given again by Eq. (E.27) with $\eta^a_{\mu\nu}$ replaced by the anti-'t Hooft symbol $\bar{\eta}^a_{\mu\nu}$

$$\bar{\eta}^a_{\mu\nu} = \begin{cases} \epsilon_{a\mu\nu} & \text{for } \mu,\nu = 1,2,3, \\ \delta_{a\nu} & \text{for } \mu = 4, \\ -\delta_{a\mu} & \text{for } \nu = 4, \\ 0 & \text{for } \mu = \nu = 4. \end{cases} \tag{E.29}$$

Moreover, the BPST solutions are the only ones with unit winding number [203].

E.2. Other gauge groups

We have worked only with a SU(2) YM theory, but all these results can be generalized to an arbitrary gauge group. Every continuous function from S^3 into a simple Lie group $\mathcal{G}$ can be continuously deformed to a function from S^3 into a SU(2) subgroup of $\mathcal{G}$, which is characterized by its winding number. Furthermore, functions from S^3 into U(1) are continuously deformable to the trivial mapping, and there is no analogue to the winding number for U(1). As a general compact Lie group is the direct product of an U(1) factor and a series of simple Lie group factors, there is a different winding number for each simple group factor.

In particular, the SU(2) BPST instanton is explicitly embedded [204] into a gauge field $A_\mu = A^a_\mu T^a \in \mathrm{SU}(N_\mathrm{c})$ choosing three out of $N^2 - 1$ fields A^a_μ to be given by Eq. (E.27), while the others are set to zero. This results in an instanton solution of $\mathrm{SU}(N_\mathrm{c})$ YM theory with $\nu = 1$. It is worth noting that with respect to the $N_\mathrm{c} = 2$ case the arbitrariness in the choice of the three non-zero components results in additional $4(N_\mathrm{c} - 2)$ parameters for the instanton solution. Therefore, a $\mathrm{SU}(N_\mathrm{c})$ YM theory has a $4N_\mathrm{c}$-parameter family of $\nu = 1$ instanton solutions. For the physically relevant case $N_\mathrm{c} = 3$, we have thus a twelve-parameter family of solutions.

E.3. The dilute instanton gas model

In Ref. [202] 't Hooft introduced the so-called *dilute instanton gas model*, consisting in a semiclassical approximation of the quantum gauge theory around instanton solutions of the gauge fields. This model is powerful enough to completely determine, up to a single parameter, the θ dependence of the free energy. To show this, motivated by the fact that it is a semiclassical approximation, we first write the gauge field as the sum of two terms

$$A_\mu(x) = A^{\mathrm{cl}}_\mu(x) + g_0 A^{\mathrm{qu}}_\mu(x). \tag{E.30}$$

We choose $A_\mu^{\mathrm{cl}}(x)$ to be a fixed background field that solves the equations of motion, i.e. an instanton solution, while the remainder. With this decomposition, the action can be expanded around the background field

$$S = |\nu| S_0 + \int \mathrm{d}^4x \left(A_\mu^{\mathrm{qu}}(x) \Delta_{\mu\nu}^{\mathrm{gl}} A_\nu^{\mathrm{qu}}(x) + \bar{\psi}(x)(\not{D} + M_0)\psi(x) + \text{ghosts} \right) + \mathcal{O}\big(g_0^2\big), \quad \text{(E.31)}$$

where the first term on the left-hand side is the action of the instanton field with S_0 given in Eq. (E.23). The other terms are given by second functional derivatives of the action with respect to the quantum fields, and thus are bilinear in the fields, while higher derivatives are negligible in this small coupling approximation. It is worth noting that, if we do not set $\hbar = 1$, the exponential in the path integral actually is $\mathrm{e}^{-S/\hbar}$, thus two powers of g_0 are equivalent to one power of $\hbar$, and the small coupling approximation is indeed a semiclassical small-$\hbar$ approximation. However, this is different from ordinary perturbation theory: instantons effects, like quantum tunnelling, are exponentially small in g_0^2 or $\hbar$, therefore, since they vanish more rapidly than any power of g_0^2 or $\hbar$, they do not appear at any fixed order in perturbation theory.

Within this approximation, the path integral can be evaluated with Gaussian functional integration rules. For example, for a BPST instanton background field, the path integral evaluates to

$$\mathcal{Z}_1 \simeq KV\mathrm{e}^{-S_0}, \quad \text{(E.32)}$$

where the spacetime volume V comes from the integral on instanton position, S_0 is the BPST instanton action in Eq. (E.23) and K contains functional determinants of the action part bilinear in the fields and integrals on the remaining parameters on which the instanton solution depends. Similarly, for the BPST anti-instanton

$$\mathcal{Z}_{-1} \simeq KV\mathrm{e}^{-S_0}. \quad \text{(E.33)}$$

Approximate instanton solutions with general winding number ν can be build from m BPTS instantons and $\bar{m}$ anti-instantons such that $\nu = m - \bar{m}$, putting instantons centres at arbitrary large distances. We apply the so-called *dilute instanton gas* approximation, which is reminiscent of the dilute gas approximation of statistical mechanics, and assume that the partition function with given ν is given by

$$\mathcal{Z}_\nu \simeq \sum_{m,\bar{m}} K^{m+\bar{m}} \frac{V^{m+\bar{m}}}{m!\bar{m}!} \mathrm{e}^{-(m+\bar{m})S_0} \delta_{\nu,m-\bar{m}}, \quad \text{(E.34)}$$

where the factor of $m!\bar{m}!$ comes from the possible permutations of instantons and anti-instantons. Then, according to Eq. (3.9) the partition function of θ-vacua is given by the Fourier series

$$\begin{aligned}\mathcal{Z}(\theta) &= \sum_\nu \mathrm{e}^{i\theta\nu}\mathcal{Z}_\nu \simeq \sum_{m,\bar{m}} \frac{\big(KV\mathrm{e}^{-S_0}\big)^{m+\bar{m}}}{m!\bar{m}!} \mathrm{e}^{i(m-\bar{m})\theta} \\ &= \sum_m \frac{\big(KV\mathrm{e}^{-S_0}\mathrm{e}^{i\theta}\big)^m}{m!} \sum_{\bar{m}} \frac{\big(KV\mathrm{e}^{-S_0}\mathrm{e}^{-i\theta}\big)^{\bar{m}}}{\bar{m}!} = \exp\big\{KV\mathrm{e}^{-S_0}\mathrm{e}^{i\theta}\big\} \exp\big\{KV\mathrm{e}^{-S_0}\mathrm{e}^{-i\theta}\big\} \\ &= \exp\big\{2KV\mathrm{e}^{-S_0}\cos\theta\big\}. \quad \text{(E.35)}\end{aligned}$$

This leads to the conclusion that the free energy density depends on θ according to

$$\frac{F(\theta)}{V} = -2K\mathrm{e}^{-\frac{8\pi^2}{g_0^2}}(\cos\theta - 1), \tag{E.36}$$

where we have fixed an irrelevant additive constant in order to have $F(0) = 0$.

No exact instanton solutions of higher winding number are known. However, suppose a solution with $\nu = 2$, called *binstanton*, exists. Then, we have to sum over a dilute gas of instantons, anti-instantons, binstantons and anti-binstantons. As a result

$$\frac{F(\theta)}{V} = -2K\mathrm{e}^{-S_0}(\cos\theta - 1) - 2K'\mathrm{e}^{-S_0'}(\cos 2\theta - 1), \tag{E.37}$$

with primed quantities referred to binstanton. But, since $S_0' = 2S_0$, the binstanton contribution is exponentially small. Therefore, higher winding number solutions can be ignored.

To complete the argument, we should need an expression for K. 't Hooft evaluated it in Ref. [202]. Without going into details of his elaborate derivation, we summarize some properties of K. First, K is proportional to the coupling g_0 to the inverse power of the number of parameters on which the BPST instanton depends

$$K \sim g_0^{-4N_\mathrm{c}}. \tag{E.38}$$

These parameters are treated as global variables and integrated over. The integral over the position of instanton centre has already been done and gives the V factor. The integral over the gauge group is a finite constant since the group is compact. Eventually, there is the integral over the instanton size ρ. Classically, this is divergent, because the theory has scale invariance and there are instantons of any size. When we consider the quantum theory, renormalization has to be taken into account and scale invariance is lost. The free energy density written in terms of the renormalized coupling is

$$\frac{F(\theta)}{V} = -\mathrm{e}^{-\frac{8\pi^2}{g^2}}(\cos\theta - 1)g^{-4N_\mathrm{c}}\int_0^\infty \frac{d\rho}{\rho^5} f(\rho\mu), \tag{E.39}$$

where f is a unknown function, μ an arbitrary mass and the integral is determined by dimensional analysis, given that the free energy density has mass dimension $d = 4$. Renormalization group theory states that the dimensionless coupling g and the dimensioned scale μ are not independent. According to the β-function in Eq. (0.4), they must appear only in the combination

$$\frac{1}{g^2} - 2b_0 \ln\mu + \mathcal{O}\big(g^2\big), \tag{E.40}$$

where b_0 is the leading order coefficient of the β-function in Eq. (0.4). A general feature of Yang-Mills theories is asymptotic freedom, i.e. $b_0 > 0$. Using the leading order β-function of SU(N_c) YM theory given Eq. (1.25a), we arrive to

$$\frac{F(\theta)}{V} = -A\mathrm{e}^{-\frac{8\pi^2}{g^2}}(\cos\theta - 1)g^{-4N_\mathrm{c}}\int_0^\infty \frac{d\rho}{\rho^5}(\rho\mu)^{\frac{11N_\mathrm{c}}{3}}\big[1 + \mathcal{O}\big(g^2\big)\big], \tag{E.41}$$

where A is a constant independent of g, μ and ρ. Therefore in YM theory the integral is ultraviolet convergent but infrared divergent. This is good, because the integral is finite in a regime where the coupling g is small and we trust this semiclassical approximation, and it diverges in a regime where the coupling is big, and all small coupling expansions are wrong. Ultimately, there may be or not be an infrared cut-off, but the answer to this must come from a non-perturbative analysis. If there is no infrared cut-off, very large instantons exist. These interact strongly and the dilute gas approximation fails.

Assume that a cut-off exists and the dilute instanton gas model is valid. This implies that the θ-dependence of the free energy is given by Eq. (E.36) up to a multiplicative constant. This is not enough to predict the topological susceptibility. Indeed, from the definition in Eq. (3.15) we have

$$\chi_{\mathrm{t}} = \left.\frac{\partial^2}{\partial\theta^2} F(\theta)\right|_{\theta=0} = 2K\mathrm{e}^{-\frac{8\pi^2}{g^2}}. \tag{E.42}$$

However, once the topological susceptibility is specified, every cumulant of the distribution is fixed. According to Eq. (3.14), they are

$$C_n = -\frac{(-\mathrm{i})^{2n}}{V}\left.\frac{\partial^{2n}}{\partial\theta^{2n}} F(\theta)\right|_{\theta=0} = 2K\mathrm{e}^{-\frac{8\pi^2}{g^2}} = \chi_{\mathrm{t}}. \tag{E.43}$$

Therefore, in particular, the ratio R defined in Eq. (3.67) is

$$R = \frac{C_2}{C_1} = 1. \tag{E.44}$$

This is in contrast with Eq. (3.67) that predicts $R = \mathcal{O}(1/N_c^2)$ in the large-N_c expansion.

F. Runge–Kutta–Munthe-Kaas integrators

In this appendix we define the 4^{th}-order numerical integration method used in Chapter 5, that we introduced in Ref. [1]. Consider an ordinary differential equation

$$\dot{y} = f(y)y, \qquad y(0) = y_0, \tag{F.1}$$

where $y \in \mathcal{G}$ for some Lie group $\mathcal{G}$ and $f(y) : \mathcal{G} \to \mathfrak{g}$, with $\mathfrak{g}$ being the Lie algebra of $\mathcal{G}$. Runge–Kutta–Munthe-Kaas (RKMK) methods [157–159] are *structure-preserving* Runge–Kutta (RK) methods designed to integrate numerically these equations on the group manifold, for a general introduction see Ref. [205]. The starting point is to write the solution of Eq. (F.1) as

$$y(t) = \exp\{v(t)\}y(0), \tag{F.2}$$

and then solve the ordinary differential equation

$$\dot{v} = \mathrm{d}\exp_v^{-1}\{f(y)\}, \qquad v(0) = 0, \tag{F.3}$$

where $\mathrm{d}\exp_v^{-1}$ has the series expansion

$$\mathrm{d}\exp_v^{-1} = \sum_{k=0}^{\infty} \frac{B_k}{k!} \mathrm{ad}_v^n = 1 + \frac{1}{2}[v, \cdot] + \frac{1}{12}[v, [v, \cdot]] + \dots \tag{F.4}$$

with B_k being the Bernoulli numbers, and $\mathrm{ad}_v = [v, \cdot]$ the adjoint action. Since $v(t)$ takes values in the Lie algebra, the differential equation in Eq. (F.3) can be numerically integrated using an ordinary RK method. No extra conditions are needed, and any RK method of a given order can be used as a base for a RKMK method of the same order. The only complication is given by the operator $\mathrm{d}\exp_v^{-1}$, which can be substituted with its series expansion in Eq. (F.4) suitably truncated according to the order of the method. The RKMK method of q^{th} order with s stages is given by

$$\begin{aligned}
&\texttt{for } i = 1, 2, \dots, s : \\
&\qquad u_i = \sum_{j=1}^{s} a_{i,j}\tilde{k}_j \\
&\qquad k_i = hf\big\{\exp\big(u_i\big)y_0\big\} \\
&\qquad \tilde{k}_i = \mathrm{dexpinv}\big(u_i, k_i, q\big) \\
&v = \sum_{j=1}^{s} b_j\tilde{k}_j \\
&y_1 = \exp\{v\}y_0
\end{aligned} \tag{F.5}$$

where dexpinv(u, v, q) is the truncated series

$$\text{dexpinv}(u, v, q) = \sum_{k=0}^{q-1} \frac{B_k}{k!} \text{ad}_u^k \tag{F.6}$$

and $a_{i,j}$, b_i are the coefficients of q^{th} order s-stages RK method. The fourth order RKMK method that we implemented is obtained starting from the very common 4^{th} order 4-stages RK method with coefficients, arranged in a Butcher tableau,

$$\begin{array}{c|cccc} 0 & & & & \\ 1/2 & 1/2 & & & \\ 1/2 & 0 & 1/2 & & \\ 1 & 0 & 0 & 1 & \\ \hline & 1/6 & 1/3 & 1/3 & 1/6 \end{array} \tag{F.7}$$

introduced by Kutta himself. At a first sight this method entails the computation of six different commutators of k_i structures. However, it is possible to reduce the number of independent commutators needed to only two. As explained in Ref. [206], this is due to the fact that, whereas the k_i are in general $\mathcal{O}(h)$, some combinations of them are higher order in h and so the corresponding commutators can be neglected. The resulting integration algorithm is

$$\begin{aligned} u_1 &= 0, \quad k_i = hf\{\exp\big(u_i\big)y_0\} \\ u_2 &= \frac{1}{2}k_1, \\ u_3 &= \frac{1}{2}k_2 + \frac{1}{8}\big[k_1, k_2\big], \\ u_4 &= k_3, \\ v &= \frac{1}{6}k_1 + \frac{1}{3}k_2 + \frac{1}{3}k_3 + \frac{1}{6}k_4 - \frac{1}{12}\big[k_1, k_4\big], \\ y_1 &= \exp(v)y_0. \end{aligned} \tag{F.8}$$

Alternative RK methods for integrating Eq. (F.1) are given by the Crouch–Grossman integrators [207, 208]. They are a special case of so-called *commutator-free* Lie group methods [209]. The third order algorithm described in Ref. [29] belongs to this class. The conditions which the coefficients need to satisfy, order by order, are computable up to arbitrary order [210]. They are given by the order conditions for a classical RK method, plus specific extra conditions. At fourth order, however, we did not find a coefficient scheme with the useful properties of Lüscher's integrator in terms of exponential reusing.

F.1. Application to the Yang–Mills gradient flow

The Yang–Mills gradient flow equation in Eq. (4.18) can be written as an ordinary first-order autonomous differential equation

$$\dot{V}(t) = Z[V(t)]V(t), \qquad V(0) = V_0, \tag{F.9}$$

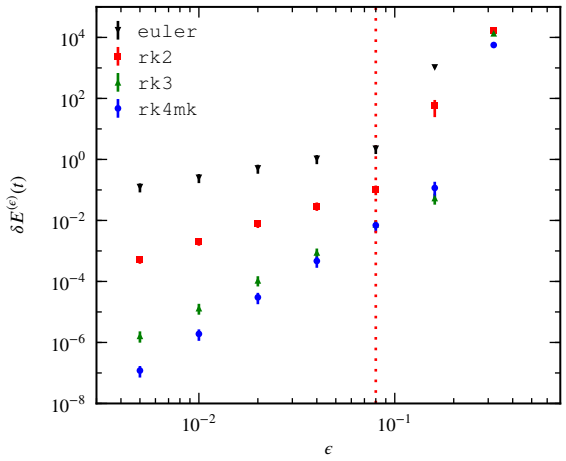


Figure F.1.: Comparison of the numerical integration methods. The systematic error $\delta E^{(\epsilon)}(t)$, where $t = 3.2a^2$, is estimated from 100 configurations of the lattice B_1 by taking the difference between $E^{(\epsilon)}(t)$, evolved with step size ϵ, and $E^{(\epsilon/2)}(t)$, evolved with $\epsilon/2$.

where

$$Z[V(t)] = -g_0^2\{\partial_{x,\mu} S[V(t)]\}, \tag{F.10}$$

and the link differential operators are defined in Eq. (A.16). The fourth order RKMK method in Eq. (F.8) reads

$$\begin{aligned}
W_1 &= V(t), \qquad Z_i = \epsilon Z[W_i], \\
W_2 &= \exp\Big\{\frac{1}{2}Z_1\Big\}V(t), \\
W_3 &= \exp\Big\{\frac{1}{2}Z_2 + \frac{1}{8}\big[Z_1, Z_2\big]\Big\}V(t), \\
W_4 &= \exp\big\{Z_3\big\}V(t), \\
V(t + a^2\epsilon) &= \exp\Big\{\frac{1}{6}Z_1 + \frac{1}{3}Z_2 + \frac{1}{3}Z_3 + \frac{1}{6}Z_4 - \frac{1}{12}\big[Z_1, Z_4\big]\Big\}V(t).
\end{aligned} \tag{F.11}$$

This method computes four times the force field $Z[W_i]$ and four times the Lie group exponential. The commutators are economically implemented exploiting structure constants of $\mathfrak{g}$. Each iteration needs space in memory for one auxiliary gauge field and three Z_i fields. Gauge fields are stored in memory with a full 3×3 complex matrix, which has 18 real components, for each link. A Z_i field is an element of $\mathfrak{su}(3)$, which is a 8 dimensional linear space, for each link. Thus, the method in Eq. (F.11) requires space for $(18 + 3 \cdot 8) \cdot 4V$ floating point numbers. Each exponential of a $\mathfrak{g}$-valued combination of $Z[W_i]$ reduce to $4V$ $\mathfrak{su}(3)$ matrices exponentials, which can be computed economically exploiting the Cayley–Hamilton theorem as described in Ref. [211].

In the left plot of Figure F.1 the RKMK method is compared to lower-order Runge–Kutta methods, such as the third order method `rk3` found in Ref. [29]. The comparison is done averaging over 100 configurations at $\beta = 5.96$ on a 12^4 lattice evolved at $t = 3.2a^2$. The `rk4mk` algorithm scales correctly as a fourth-order method. However, the pre-factor appears to be larger, thus the new method is more precise with respect to `rk3` in Ref. [29] for $\epsilon \lesssim 0.1$.

Bibliography

[1] M. Cè, C. Consonni, G. P. Engel and L. Giusti, 'Non-Gaussianities in the topological charge distribution of the SU(3) Yang-Mills theory', Phys. Rev. D **92**, 074502 (2015), arXiv:1506.06052 [hep-lat].

[2] M. Cè, L. Giusti and S. Schaefer, 'Domain decomposition, multilevel integration, and exponential noise reduction in lattice QCD', Phys. Rev. D **93**, 094507 (2016), arXiv:1601.04587 [hep-lat].

[3] M. Cè, M. García Vera, L. Giusti and S. Schaefer, 'The topological susceptibility in the large-N limit of SU(N) Yang-Mills theory', Phys. Lett. B **762**, 232–236 (2016), arXiv:1607.05939 [hep-lat].

[4] M. Cè, L. Giusti and S. Schaefer, 'A local factorization of the fermion determinant in lattice QCD', (2016), arXiv:1609.02419 [hep-lat].

[5] M. Cè, C. Consonni, G. P. Engel and L. Giusti, 'Testing the Witten-Veneziano mechanism with the Yang-Mills gradient flow on the lattice', PoS **LATTICE2014**, 353 (2014), arXiv:1410.8358 [hep-lat].

[6] M. Cè, 'Non-Gaussianity of the topological charge distribution in SU(3) Yang-Mills theory', PoS **LATTICE 2015**, 318 (2015), arXiv:1510.08826 [hep-lat].

[7] M. Cè, M. García Vera, L. Giusti and S. Schaefer, 'The large-N limit of the topological susceptibility of Yang-Mills gauge theory', PoS **LATTICE2016**, 350 (2016), arXiv:1610.08797 [hep-lat].

[8] M. Cè, L. Giusti and S. Schaefer, 'Domain decomposition and multilevel integration for fermions', PoS **LATTICE2016**, 263 (2016), arXiv:1612.06424 [hep-lat].

[9] S. L. Glashow, 'Partial-symmetries of weak interactions', Nucl. Phys. **22**, 579–588 (1961).

[10] S. Weinberg, 'A Model of Leptons', Phys. Rev. Lett. **19**, 1264–1266 (1967).

[11] A. Salam, 'Weak and Electromagnetic Interactions', in 8th Nobel Symposium, Lerum, Sweden, May 19–25, 1968 (1968), pp. 367–377.

[12] H. Fritzsch and M. Gell-Mann, 'Current algebra: quarks and what else?', in Proceedings, 16th International Conference on High-Energy Physics, ICHEP, Batavia, Illinois, 6-13 Sep 1972 (1973), pp. 135–165, arXiv:hep-ph/0208010 [hep-ph].

[13] H. Fritzsch, M. Gell-Mann and H. Leutwyler, 'Advantages of the color octet gluon picture', Phys. Lett. B **47**, 365–368 (1973).

[14] D. J. Gross and F. Wilczek, 'Ultraviolet behavior of non-Abelian gauge theories', Phys. Rev. Lett. **30**, 1343–1346 (1973).

[15] H. D. Politzer, 'Reliable perturbative results for strong interactions?', Phys. Rev. Lett. **30**, 1346–1349 (1973).

[16] C.-N. Yang and R. L. Mills, 'Conservation of isotopic spin and isotopic gauge invariance', Phys. Rev. **96**, 191–195 (1954).

[17] K. A. Olive et al., 'Review of Particle Physics', Chin. Phys. C **40**, 100001 (2016).

[18] A. M. Jaffe and E. Witten, *Quantum Yang-Mills theory*, (2000) `http://www.claymath.org/sites/default/files/yangmills.pdf`.

[19] K. G. Wilson, 'Confinement of quarks', Phys. Rev. D **10**, 2445–2459 (1974).

[20] M. Gell-Mann, 'Symmetries of baryons and mesons', Phys. Rev. **125**, 1067–1084 (1962).

[21] M. Gell-Mann and Y. Ne'emam, *The Eightfold way, A review with a collection of reprints* (W.A. Benjamin, 1964), 317 pp.

[22] G. 't Hooft, 'A planar diagram theory for strong interactions', Nucl. Phys. B **72**, 461–473 (1974).

[23] E. Witten, 'Current algebra theorems for the U(1) "Goldstone boson"', Nucl. Phys. B **156**, 269–283 (1979).

[24] G. Veneziano, 'U(1) without instantons', Nucl. Phys. B **159**, 213–224 (1979).

[25] G. Veneziano, 'Goldstone mechanism from gluon dynamics', Phys. Lett. B **95**, 90–92 (1980).

[26] H. B. Nielsen and M. Ninomiya, 'A no-go theorem for regularizing chiral fermions', Phys. Lett. B **105**, 219–223 (1981).

[27] L. Giusti, G. C. Rossi, M. Testa and G. Veneziano, 'The $U_A(1)$ problem on the lattice with Ginsparg-Wilson fermions', Nucl. Phys. B **628**, 234–252 (2002), arXiv:`hep-lat/0108009 [hep-lat]`.

[28] P. H. Ginsparg and K. G. Wilson, 'A remnant of chiral symmetry on the lattice', Phys. Rev. D **25**, 2649–2657 (1982).

[29] M. Lüscher, 'Properties and uses of the Wilson flow in lattice QCD', JHEP **1008**, 071 (2010), arXiv:`1006.4518 [hep-lat]`, [Erratum: JHEP **1403** 092 (2014)].

[30] M. Albanese et al., 'Glueball masses and string tension in lattice QCD', Phys. Lett. B **192**, 163–169 (1987).

[31] M. Lüscher and P. Weisz, 'Locality and exponential error reduction in numerical lattice gauge theory', JHEP **0109**, 010 (2001), arXiv:`hep-lat/0108014 [hep-lat]`.

[32] H. B. Meyer, 'Locality and statistical error reduction on correlation functions', JHEP **0301**, 048–048 (2003), arXiv:`hep-lat/0209145 [hep-lat]`.

[33] K. Osterwalder and R. Schrader, 'Axioms for Euclidean Green's functions', Commun. Math. Phys. **31**, 83–112 (1973).

[34] K. Osterwalder and R. Schrader, 'Axioms for Euclidean Green's functions II', Commun. Math. Phys. **42**, 281–305 (1975).

[35] S. Weinberg, *Quantum theory of fields: volume 2, modern applications* (Cambridge University Press, 1996), 512 pp.

[36] C. G. Bollini and J. J. Giambiagi, 'Dimensional renormalization: the number of dimensions as a regularizing parameter', Nuovo Cimento B **12**, 20–26 (1972).

[37] G. 't Hooft, 'Dimensional regularization and the renormalization group', Nucl. Phys. B **61**, 455–468 (1973).

[38] S. Weinberg, 'New approach to the renormalization group', Phys. Rev. D **8**, 3497–3509 (1973).

[39] K. G. Wilson and J. B. Kogut, 'The renormalization group and the ϵ expansion', Phys. Rep. **12**, 75–199 (1974).

[40] E. C. G. Stueckelberg and A. Petermann, 'La normalisation des constantes dans la théorie des quanta', Helv. Phys. Acta **26**, 499–520 (1953).

[41] M. Gell-Mann and F. E. Low, 'Quantum electrodynamics at small distances', Phys. Rev. **95**, 1300–1312 (1954).

[42] C. G. Callan, 'Broken scale invariance in scalar field theory', Phys. Rev. D **2**, 1541–1547 (1970).

[43] K. Symanzik, 'Small distance behaviour in field theory and power counting', Commun. Math. Phys. **18**, 227–246 (1970).

[44] K. Symanzik, 'Small-distance-behaviour analysis and Wilson expansions', Commun. Math. Phys. **23**, 49–86 (1971).

[45] I. B. Khriplovich, 'Green's functions in theories with non-Abelian gauge group', Sov. J. Nucl. Phys. **10**, 235–242 (1969), [Yad. Fiz. **10**, 409–424 (1969)].

[46] J. Gasser and H. Leutwyler, 'Quark masses', Phys. Rep. **87**, 77–169 (1982).

[47] J. Gasser and H. Leutwyler, 'Chiral perturbation theory to one loop', Ann. Phys. **158**, 142–210 (1984).

[48] J. Gasser and H. Leutwyler, 'Chiral perturbation theory: expansions in the mass of the strange quark', Nucl. Phys. B **250**, 465–516 (1985).

[49] K. G. Wilson and W. Zimmermann, 'Operator product expansions and composite field operators in the general framework of quantum field theory', Commun. Math. Phys. **24**, 87–106 (1972).

[50] S. Durr et al., 'Ab initio determination of light hadron masses', Science **322**, 1224–1227 (2008), arXiv:0906.3599 [hep-lat].

[51] S. Coleman and S. L. Glashow, 'Electrodynamic properties of baryons in the unitary symmetry scheme', Phys. Rev. Lett. **6**, 423–425 (1961).

[52] J. Beringer et al., 'Review of Particle Physics', Phys. Rev. D **86**, 010001 (2012).

[53] S. Borsanyi et al., 'Ab initio calculation of the neutron-proton mass difference', Science **347**, 1452–1455 (2015), arXiv:1406.4088 [hep-lat].

[54] S. Capitani, M. Lüscher, R. Sommer and H. Wittig, 'Non-perturbative quark mass renormalization in quenched lattice QCD', Nucl. Phys. B **544**, 669–698 (1999), arXiv:hep-lat/9810063 [hep-lat], [Erratum: Nucl. Phys. B **582**, 762 (2000)].

[55] M. Della Morte et al., 'Computation of the strong coupling in QCD with two dynamical flavors', Nucl. Phys. B **713**, 378–406 (2005), arXiv:hep-lat/0411025 [hep-lat].

[56] M. Lüscher, P. Weisz and U. Wolff, 'A numerical method to compute the running coupling in asymptotically free theories', Nucl. Phys. B **359**, 221–243 (1991).

[57] M. Lüscher, R. Narayanan, P. Weisz and U. Wolff, 'The Schrödinger functional — a renormalizable probe for non-abelian gauge theories', Nucl. Phys. B **384**, 168–228 (1992), arXiv:hep-lat/9207009 [hep-lat].

[58] P. Fritzsch et al., 'The strange quark mass and lambda parameter of two flavor QCD', Nucl. Phys. B **865**, 397–429 (2012), arXiv:1205.5380 [hep-lat].

[59] J. C. Ward, 'An identity in quantum electrodynamics', Phys. Rev. **78**, 182–182 (1950).

[60] Y. Takahashi, 'On the generalized Ward identity', Nuovo Cimento **6**, 371–375 (1957).

[61] M. Gell-Mann, R. J. Oakes and B. Renner, 'Behavior of current divergences under SU(3) × SU(3)', Phys. Rev. **175**, 2195–2199 (1968).

[62] G. P. Engel, L. Giusti, S. Lottini and R. Sommer, 'Chiral symmetry breaking in QCD with two light flavors', Phys. Rev. Lett. **114**, 112001 (2015), arXiv:1406.4987 [hep-ph].

[63] T. Banks and A. Casher, 'Chiral symmetry breaking in confining theories', Nucl. Phys. B **169**, 103–125 (1980).

[64] L. Del Debbio, L. Giusti, M. Luscher, R. Petronzio and N. Tantalo, 'Stability of lattice QCD simulations and the thermodynamic limit', JHEP **0602**, 011 (2006), arXiv:hep-lat/0512021 [hep-lat].

[65] L. Giusti and M. Lüscher, 'Chiral symmetry breaking and the Banks–Casher relation in lattice QCD with Wilson quarks', JHEP **0903**, 013 (2009), arXiv:0812.3638 [hep-lat].

[66] G. P. Engel, L. Giusti, S. Lottini and R. Sommer, 'Spectral density of the Dirac operator in two-flavor QCD', Phys. Rev. D **91**, 054505 (2015), arXiv:1411.6386 [hep-lat].

[67] K. G. Wilson, 'Quarks and Strings on a Lattice', in 13th International School of Subnuclear Physics: New Phenomena in Subnuclear Physics Erice, Italy, July 11-August 1, 1975 (1977).

[68] L. Susskind, 'Lattice fermions', Phys. Rev. D **16**, 3031–3039 (1977).

[69] H. J. Rothe, *Lattice gauge theories: an introduction*, 4th Edition, World Scientific Lecture Notes in Physics 82 (World Scientific Publishing Company, 2012), 628 pp.

[70] C. Gattringer and C. B. Lang, *Quantum chromodynamics on the lattice: an introductory presentation*, Lecture Notes in Physics 788 (Springer, 2009), 343 pp.

[71] L. Lellouch, R. Sommer, B. Svetitsky, A. Vladikas and L. F. Cugliandolo, eds., *Modern perspectives in lattice QCD: Quantum field theory and high performance computing. Proceedings, International School, 93rd Session, Les Houches, France, August 3-28, 2009* (Oxford University Press, 2011), 760 pp.

[72] H. B. Nielsen and M. Ninomiya, 'Absence of neutrinos on a lattice: (i). proof by homotopy theory', Nucl. Phys. B **185**, 20–40 (1981).

[73] H. B. Nielsen and M. Ninomiya, 'Absence of neutrinos on a lattice: (ii). intuitive topological proof', Nucl. Phys. B **193**, 173–194 (1981).

[74] D. Friedan, 'A proof of the Nielsen-Ninomiya theorem', Commun. Math. Phys. **85**, 481–490 (1982).

[75] D. B. Kaplan, 'A method for simulating chiral fermions on the lattice', Phys. Lett. B **288**, 342–347 (1992), arXiv:hep-lat/9206013 [hep-lat].

[76] Y. Shamir, 'Chiral fermions from lattice boundaries', Nucl. Phys. B **406**, 90–106 (1993), arXiv:hep-lat/9303005 [hep-lat].

[77] P. Hasenfratz and F. Niedermayer, 'Perfect lattice action for asymptotically free theories', Nucl. Phys. B **414**, 785–814 (1994), arXiv:hep-lat/9308004 [hep-lat].

[78] H. Neuberger, 'Exactly massless quarks on the lattice', Phys. Lett. B **417**, 141–144 (1998), arXiv:hep-lat/9707022 [hep-lat].

[79] P. Hasenfratz, 'Prospects for perfect actions', Nucl. Phys. B Proc. Suppl. **63**, 53–58 (1998), arXiv:hep-lat/9709110 [hep-lat].

[80] H. Neuberger, 'More about exactly massless quarks on the lattice', Phys. Lett. B **427**, 353–355 (1998), arXiv:hep-lat/9801031 [hep-lat].

[81] M. Lüscher, 'Exact chiral symmetry on the lattice and the Ginsparg–Wilson relation', Phys. Lett. B **428**, 342–345 (1998), arXiv:hep-lat/9802011 [hep-lat].

[82] P. Hernández, K. Jansen and M. Lüscher, 'Locality properties of Neuberger's lattice Dirac operator', Nucl. Phys. B **552**, 363–378 (1999), arXiv:hep-lat/9808010 [hep-lat].

[83] M. Creutz, 'Gauge fixing, the transfer matrix, and confinement on a lattice', Phys. Rev. D **15**, 1128–1136 (1977).

[84] M. Lüscher, 'Construction of a selfadjoint, strictly positive transfer matrix for Euclidean lattice gauge theories', Commun. Math. Phys. **54**, 283–292 (1977).

[85] S. Sint, 'On the Schrödinger functional in QCD', Nucl. Phys. B **421**, 135–156 (1994), arXiv:hep-lat/9312079 [hep-lat].

[86] K. Osterwalder and E. Seiler, 'Gauge field theories on a lattice', Ann. Phys. **110**, 440–471 (1978).

[87] P. Menotti and A. Pelissetto, 'General proof of Osterwalder-Schrader positivity for the Wilson action', Commun. Math. Phys. **113**, 369–373 (1987).

[88] M. Bochicchio, L. Maiani, G. Martinelli, G. Rossi and M. Testa, 'Chiral symmetry on the lattice with Wilson fermions', Nucl. Phys. B **262**, 331–355 (1985).

[89] M. Lüscher, 'Topology and the axial anomaly in abelian lattice gauge theories', Nucl. Phys. B **538**, 515–529 (1999), arXiv:hep-lat/9808021 [hep-lat].

[90] Y. Kikukawa and A. Yamada, 'Axial vector current of exact chiral symmetry on the lattice', Nucl. Phys. B **547**, 413–423 (1999), arXiv:hep-lat/9808026 [hep-lat].

[91] T. Reisz, 'A convergence theorem for lattice Feynman integrals with massless propagators', Commun. Math. Phys. **116**, 573–606 (1988).

[92] T. Reisz, 'Renormalization of Feynman integrals on the lattice', Commun. Math. Phys. **117**, 79–108 (1988).

[93] T. Reisz, 'Lattice gauge theory: renormalization to all orders in the loop expansion', Nucl. Phys. B **318**, 417–463 (1989).

[94] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, 'Equation of state calculations by fast computing machines', J. Chem. Phys. **21**, 1087–1092 (1953).

[95] M. Creutz, 'Monte Carlo study of quantized SU(2) gauge theory', Phys. Rev. D **21**, 2308–2315 (1980).

[96] N. Cabibbo and E. Marinari, 'A new method for updating SU(N) matrices in computer simulations of gauge theories', Phys. Lett. B **119**, 387–390 (1982).

[97] S. L. Adler, 'Over-relaxation method for the Monte Carlo evaluation of the partition function for multiquadratic actions', Phys. Rev. D **23**, 2901–2904 (1981).

[98] S. Duane, A. Kennedy, B. J. Pendleton and D. Roweth, 'Hybrid Monte Carlo', Phys. Lett. B **195**, 216–222 (1987).

[99] A. Ukawa, 'Computational cost of full QCD simulations experienced by CP-PACS and JLQCD collaborations', Nucl. Phys. B Proc. Suppl. **106**, 195–196 (2002).

[100] C. Bernard et al., 'Panel discussion on the cost of dynamical quark simulations', Nucl. Phys. B Proc. Suppl. **106**, 199–205 (2002).

[101] J. C. Sexton and D. H. Weingarten, 'Hamiltonian evolution for the hybrid Monte Carlo algorithm', Nucl. Phys. B **380**, 665–677 (1992).

[102] M. Hasenbusch, 'Speeding up the hybrid Monte Carlo algorithm for dynamical fermions', Phys. Lett. B **519**, 177–182 (2001), arXiv:hep-lat/0107019 [hep-lat].

[103] C. Urbach, K. Jansen, A. Shindler and U. Wenger, 'HMC algorithm with multiple time scale integration and mass preconditioning', Comput. Phys. Commun. **174**, 87–98 (2006), arXiv:hep-lat/0506011 [hep-lat].

[104] M. Lüscher, 'Solution of the Dirac equation in lattice QCD using a domain decomposition method', Comput. Phys. Commun. **156**, 209–220 (2004), arXiv:hep-lat/0310048 [hep-lat].

[105] M. Lüscher, 'Schwarz-preconditioned HMC algorithm for two-flavor lattice QCD', Comput. Phys. Commun. **165**, 199–220 (2005), arXiv:hep-lat/0409106 [hep-lat].

[106] M. Lüscher, 'Local coherence and deflation of the low quark modes in lattice QCD', JHEP **0707**, 081–081 (2007), arXiv:0706.2298 [hep-lat].

[107] M. Lüscher, 'Deflation acceleration of lattice QCD simulations', JHEP **0712**, 011 (2007), arXiv:0710.5417 [hep-lat].

[108] M. A. Clark and A. D. Kennedy, 'Accelerating dynamical-fermion computations using the rational hybrid Monte Carlo algorithm with multiple pseudofermion fields', Phys. Rev. Lett. **98**, 051601 (2007), arXiv:hep-lat/0608015 [hep-lat].

[109] M. Lüscher and S. Schaefer, 'Lattice QCD without topology barriers', JHEP **1107**, 036 (2011), arXiv:1105.4749 [hep-lat].

[110] L. Giusti, 'Light dynamical fermions on the lattice: toward the chiral regime of QCD', PoS **LAT2006**, 009 (2006), arXiv:hep-lat/0702014 [hep-lat].

[111] S. Schaefer, 'Status and challenges of simulations with dynamical fermions', PoS **Lattice 2012**, 001 (2012), arXiv:1211.5069 [hep-lat].

[112] R. Sommer, 'A new way to set the energy scale in lattice gauge theories and its application to the static force and σ_s in SU(2) Yang-Mills theory', Nucl. Phys. B **411**, 839–854 (1994), arXiv:hep-lat/9310022 [hep-lat].

[113] K. Symanzik, 'Continuum limit and improved action in lattice theories', Nucl. Phys. B **226**, 205–227 (1983).

[114] K. Symanzik, 'Continuum limit and improved action in lattice theories', Nucl. Phys. B **226**, 187–204 (1983).

[115] M. Lüscher and P. Weisz, 'On-shell improved lattice gauge theories', Commun. Math. Phys. **97**, 59–77 (1985), [Erratum: Commun. Math. Phys. **98**, 433 (1985)].

[116] G. Curci, P. Menotti and G. Paffuti, 'Symanzik's improved lagrangian for lattice gauge theory', Phys. Lett. B **130**, 205–208 (1983), [Erratum: Phys. Lett. B **135**, 516 (1984)].

[117] S. Caracciolo, P. Menotti and A. Pelissetto, 'One-loop analytic computation of the energy-momentum tensor for lattice gauge theories', Nucl. Phys. B **375**, 195–239 (1992).

[118] M. Lüscher, S. Sint, R. Sommer and P. Weisz, 'Chiral symmetry and $O(a)$ improvement in lattice QCD', Nucl. Phys. B **478**, 365–397 (1996), arXiv:hep-lat/9605038 [hep-lat].

[119] B. Sheikholeslami and R. Wohlert, 'Improved continuum limit lattice action for QCD with Wilson fermions', Nucl. Phys. B **259**, 572–596 (1985).

[120] R. Wohlert, 'Improved continuum limit action for quarks', (1987).

[121] K. Jansen and R. Sommer, '$O(\alpha)$ improvement of lattice QCD with two flavors of Wilson quarks', Nucl. Phys. B **530**, 185–203 (1998), arXiv:hep-lat/9803017 [hep-lat], [Erratum: Nucl. Phys. B **643**, 517–518 (2002)].

[122] N. Yamada et al., 'Nonperturbative $\mathcal{O}(a)$ improvement of Wilson quark action in three-flavor QCD with plaquette gauge action', Phys. Rev. D **71**, 054505 (2005), arXiv:hep-lat/0406028 [hep-lat], [Erratum: Phys. Rev. D **71**, 079902 (2005)].

[123] L. Giusti, G. Rossi and M. Testa, 'Topological susceptibility in full QCD with Ginsparg–Wilson fermions', Phys. Lett. B **587**, 157–166 (2004), arXiv:hep-lat/0402027 [hep-lat].

[124] M. Lüscher, 'Topological effects in QCD and the problem of short-distance singularities', Phys. Lett. B **593**, 296–301 (2004), arXiv:hep-th/0404034 [hep-th].

[125] E. Seiler, 'Some more remarks on the Witten–Veneziano formula for the η' mass', Phys. Lett. B **525**, 355–359 (2002), arXiv:hep-th/0111125 [hep-th].

[126] J. M. Pendlebury et al., 'Revised experimental upper limit on the electric dipole moment of the neutron', Phys. Rev. D **92**, 092003 (2015), arXiv:1509.04411 [hep-ex].

[127] S. Dar, 'The Neutron EDM in the SM: a review', (2000), arXiv:hep-ph/0008248 [hep-ph].

[128] M. F. Atiyah and I. M. Singer, 'The index of elliptic operators on compact manifolds', Bull. Amer. Math. Soc. **69**, 422–434 (1963).

[129] E. Witten, 'Theta dependence in the large N limit of four-dimensional gauge theories', Phys. Rev. Lett. **81**, 2862–2865 (1998), arXiv:hep-th/9807109 [hep-th].

[130] H. Leutwyler and A. Smilga, 'Spectrum of Dirac operator and role of winding number in QCD', Phys. Rev. D **46**, 5607–5632 (1992).

[131] S. Aoki, H. Fukaya, S. Hashimoto and T. Onogi, 'Finite volume QCD at fixed topological charge', Phys. Rev. D **76**, 054508 (2007), arXiv:0707.0396 [hep-lat].

[132] D. Espriu and R. Tarrach, 'Renormalization of the axial anomaly operators', Z. Phys. C Part. Fields **16**, 77–82 (1982).

[133] G. Shore and G. Veneziano, 'The U(1) Goldberger-Treiman relation and the proton 'spin': a renormalisation group analysis', Nucl. Phys. B **381**, 23–65 (1992).

[134] S. L. Adler, 'Axial-vector vertex in spinor electrodynamics', Phys. Rev. **177**, 2426–2438 (1969).

[135] J. S. Bell and R. Jackiw, 'A PCAC puzzle: $\pi^0 \to \gamma\gamma$ in the σ-model', Nuovo Cimento A **60**, 47–61 (1969).

[136] R. A. Bertlmann, *Anomalies in quantum field theory*, Revised edition, International Series of Monographs on Physics 91 (Clarendon Press, 2001), 584 pp.

[137] K. Fujikawa, 'Path-integral measure for gauge-invariant fermion theories', Phys. Rev. Lett. **42**, 1195–1198 (1979).

[138] P. Hasenfratz, V. Laliena and F. Niedermayer, 'The index theorem in QCD with a finite cut-off', Phys. Lett. B **427**, 125–131 (1998), arXiv:hep-lat/9801021 [hep-lat].

[139] Y. Kikukawa and A. Yamada, 'Weak coupling expansion of massless QCD with a Ginsparg-Wilson fermion and axial U(1) anomaly', Phys. Lett. B **448**, 265–274 (1999), arXiv:hep-lat/9806013 [hep-lat].

[140] K. Fujikawa, 'A continuum limit of the chiral Jacobian in lattice gauge theory', Nucl. Phys. B **546**, 480–494 (1999), arXiv:hep-th/9811235 [hep-th].

[141] E. Seiler and I. Stamatescu, 'Some remarks on the Witten–Veneziano formula for the η' mass', (1987).

[142] S. Narison, G. M. Shore and G. Veneziano, 'Target independence of the EMC-SMC effect', Nucl. Phys. B **433**, 209–233 (1995), arXiv:hep-ph/9404277 [hep-ph].

[143] S. Narison, G. M. Shore and G. Veneziano, 'Topological charge screening and the 'proton spin' beyond the chiral limit', Nucl. Phys. B **546**, 235–278 (1999), arXiv:hep-ph/9812333 [hep-ph].

[144] G. M. Shore, 'Pseudoscalar meson decay constants and couplings, the Witten–Veneziano formula beyond large N_c, and the topological susceptibility', Nucl. Phys. B **744**, 34–58 (2006), arXiv:hep-ph/0601051 [hep-ph].

[145] J. Ashman et al., 'A measurement of the spin asymmetry and determination of the structure function g_1 in deep inelastic muon-proton scattering', Phys. Lett. B **206**, 364–370 (1988).

[146] R. Kaiser and H. Leutwyler, 'Pseudoscalar decay constants at large N_c', in Nonperturbative methods in quantum field theory. Proceedings, Workshop, Adelaide, Australia, February 2-13, 1998 (1998), pp. 15–29, arXiv:hep-ph/9806336 [hep-ph].

[147] R. Kaiser and H. Leutwyler, 'Large N_c in chiral perturbation theory', Eur. Phys. J. C **17**, 623–649 (2000), arXiv:hep-ph/0007101 [hep-ph].

[148] M. Lüscher and P. Weisz, 'Perturbative analysis of the gradient flow in non-abelian gauge theories', JHEP **1102**, 051 (2011), arXiv:1101.0963 [hep-th].

[149] A. M. Polyakov, *Gauge fields and strings*, Vol. 3 (CRC Press, 1987), 312 pp.

[150] M. Lüscher, 'Lattice regularization of chiral gauge theories to all orders of perturbation theory', JHEP **0006**, 028 (2000), arXiv:hep-lat/0006014 [hep-lat].

[151] M. Lüscher, 'Chiral gauge theories revisited', in Theory and Experiment Heading for New Physics, Erice, Italy, August 27 – September 5, 2000, Vol. 38 (Jan. 2002), pp. 41–89, arXiv:hep-th/0102028 [hep-th].

[152] B. Allés, M. D'Elia and A. D. Giacomo, 'Topological susceptibility at zero and finite T in SU(3) Yang-Mills theory', Nucl. Phys. B **494**, 281–292 (1997), arXiv:hep-lat/9605013 [hep-lat], [Erratum: Nucl. Phys.B679,397(2004)].

[153] M. Bruno, S. Schaefer and R. Sommer, 'Topological susceptibility and the sampling of field space in $N_f = 2$ lattice QCD simulations', JHEP **1408**, 150 (2014), arXiv:1406.5363 [hep-lat].

[154] L. Del Debbio, L. Giusti and C. Pica, 'Topological susceptibility in the SU(3) gauge theory', Phys. Rev. Lett. **94**, 032003 (2005), arXiv:hep-th/0407052 [hep-th].

[155] G. de Divitiis et al., 'Universality and the approach to the continuum limit in lattice gauge theory', Nucl. Phys. B **437**, 447–470 (1995), arXiv:hep-lat/9411017 [hep-lat].

[156] L. Giusti, S. Petrarca and B. Taglienti, 'θ dependence of the vacuum energy in SU(3) gauge theory from the lattice', Phys. Rev. D **76**, 094510 (2007), arXiv:0705.2352 [hep-th].

[157] H. Munthe-Kaas, 'Lie-Butcher theory for Runge-Kutta methods', BIT **35**, 572–587 (1995).

[158] H. Munthe-Kaas, 'Runge-Kutta methods on Lie groups', BIT **38**, 92–111 (1998).

[159] H. Munthe-Kaas, 'High order Runge-Kutta methods on manifolds', Appl. Numer. Math. **29**, 115–127 (1999).

[160] U. Wolff, 'Monte Carlo errors with less errors', Comput. Phys. Commun. **156**, 143–153 (2004), arXiv:hep-lat/0306017 [hep-lat], [Erratum: Comput. Phys. Commun. **176**, 383 (2007)].

[161] L. Del Debbio, H. Panagopoulos and E. Vicari, 'θ dependence of SU(N) gauge theories', JHEP **0208**, 044 (2002), arXiv:hep-th/0204125 [hep-th].

[162] M. Lüscher, *From instantons to the Yang-Mills gradient flow: history & resolution of a conceptual puzzle in non-perturbative QCD*, https://luscher.web.cern.ch/luscher/talks/TopCharge15.pdf.

[163] M. Guagnelli, R. Sommer and H. Wittig, 'Precision computation of a low-energy reference scale in quenched lattice QCD', Nucl. Phys. B **535**, 389–402 (1998), arXiv:hep-lat/9806005 [hep-lat].

[164] J. Garden, J. Heitger, R. Sommer and H. Wittig, 'Precision computation of the strange quark's mass in quenched QCD', Nucl. Phys. B **571**, 237–256 (2000), arXiv:hep-lat/9906013 [hep-lat].

[165] M. Lüscher and F. Palombi, 'Universality of the topological susceptibility in the SU(3) gauge theory', JHEP **1009**, 110 (2010), arXiv:1008.0732 [hep-lat].

[166] K. Cichy, E. Garcia-Ramos, K. Jansen, K. Ottnad and C. Urbach, 'Non-perturbative test of the Witten-Veneziano formula from lattice QCD', JHEP **1509**, 020 (2015), arXiv:1504.07954 [hep-lat].

[167] S. Borsányi et al., 'High-precision scale setting in lattice QCD', JHEP **1209**, 010 (2012), arXiv:1203.4469 [hep-lat].

[168] M. Bruno and R. Sommer, 'On the N_f-dependence of gluonic observables', PoS **LATTICE 2013**, 321 (2013), arXiv:1311.5585 [hep-lat].

[169] B. Lucini and M. Teper, 'SU(N) gauge theories in four dimensions: exploring the approach to $N = \infty$', JHEP **0106**, 050 (2001), arXiv:hep-lat/0103027 [hep-lat].

[170] B. Lucini, M. Teper and U. Wenger, 'Topology of SU(N) gauge theories at $T = 0$ and $T = T_c$', Nucl. Phys. B **715**, 461–482 (2005), arXiv:hep-lat/0401028 [hep-lat].

[171] N. Cundy, M. Teper and U. Wenger, 'Topology and chiral symmetry breaking in SU(N_c) gauge theories', Phys. Rev. D **66**, 094505 (2002), arXiv:hep-lat/0203030 [hep-lat].

[172] S. Schaefer, R. Sommer and F. Virotta, 'Critical slowing down and error analysis in lattice QCD simulations', Nucl. Phys. B **845**, 93–119 (2011), arXiv:1009.5228 [hep-lat].

[173] G. Parisi, 'The strategy for computing the hadronic mass spectrum', Phys. Rep. **103**, 203–211 (1984).

[174] G. P. Lepage, 'The analysis of algorithms for lattice field theory', in Boulder ASI 1989 (1989), pp. 97–120.

[175] G. Parisi, R. Petronzio and F. Rapuano, 'A measurement of the string tension near the continuum limit', Phys. Lett. B **128**, 418–420 (1983).

[176] M. Della Morte and L. Giusti, 'Exploiting symmetries for exponential error reduction in path integral Monte Carlo', Comput. Phys. Commun. **180**, 813–818 (2009).

[177] M. Della Morte and L. Giusti, 'Symmetries and exponential error reduction in Yang–Mills theories on the lattice', Comput. Phys. Commun. **180**, 819–826 (2009), arXiv:0806.2601 [hep-lat].

[178] M. Della Morte and L. Giusti, 'A novel approach for computing glueball masses and matrix elements in Yang-Mills theories on the lattice', JHEP **1105**, 056 (2011), arXiv:1012.2562 [hep-lat].

[179] Y. Saad, *Iterative methods for sparse linear systems* (Cambridge University Pr., Jan. 2003), 528 pp.

[180] M. Lüscher and S. Schaefer, 'Lattice QCD with open boundary conditions and twisted-mass reweighting', Comput. Phys. Commun. **184**, 519–528 (2013), arXiv:1206.2809 [hep-lat].

[181] C. R. Allton, V. Giménez, L. Giusti and F. Rapuano, 'Light quenched hadron spectrum and decay constants on different lattices', Nucl. Phys. B **489**, 427–452 (1997), arXiv:hep-lat/9611021 [hep-lat].

[182] M. Lüscher and S. Schaefer, *openQCD – simulation program for lattice QCD*, https://luscher.web.cern.ch/luscher/openQCD/.

[183] C. Thron, S. J. Dong, K. F. Liu and H. P. Ying, 'Padé–Z_2 estimator of determinants', Phys. Rev. D **57**, 1642–1653 (1998), arXiv:hep-lat/9707001 [hep-lat].

[184] C. McNeile and C. Michael, 'Mixing of scalar glueballs and flavor-singlet scalar mesons', Phys. Rev. D **63**, 114503 (2001), arXiv:hep-lat/0010019 [hep-lat].

[185] G. S. Bali, S. Collins and A. Schäfer, 'Effective noise reduction techniques for disconnected loops in lattice QCD', Comput. Phys. Commun. **181**, 1570–1583 (2010), arXiv:0910.3970 [hep-lat].

[186] R. Sommer, 'Leptonic decays of *b*- and *d*-mesons', Nucl. Phys. B Proc. Suppl. **42**, 186–193 (1995), arXiv:hep-lat/9411024 [hep-lat].

[187] M. Foster and C. Michael, 'Quark mass dependence of hadron masses from lattice QCD', Phys. Rev. D **59**, 074503 (1999), arXiv:hep-lat/9810021 [hep-lat].

[188] T. DeGrand and S. Schaefer, 'Improving meson two-point functions in lattice QCD', Comput. Phys. Commun. **159**, 185–191 (2004), arXiv:hep-lat/0401011 [hep-lat].

[189] L. Giusti, P. Hernandez, M. Laine, P. Weisz and H. Wittig, 'Low-energy couplings of QCD from current correlators near the chiral limit', JHEP **0404**, 013 (2004), arXiv:hep-lat/0402002 [hep-lat].

[190] J. Foley et al., 'Practical all-to-all propagators for lattice QCD', Comput. Phys. Commun. **172**, 145–162 (2005), arXiv:hep-lat/0505023 [hep-lat].

[191] T. Blum, T. Izubuchi and E. Shintani, 'New class of variance-reduction techniques using lattice symmetries', Phys. Rev. D **88**, 094503 (2013), arXiv:1208.4349 [hep-lat].

[192] S. Weinberg, 'Nonlinear realizations of chiral symmetry', Phys. Rev. **166**, 1568–1577 (1968).

[193] S. Weinberg, 'Phenomenological lagrangians', Physica A **96**, 327–340 (1979).

[194] S. Aoki et al., 'Review of lattice results concerning low-energy particle physics', (2016), arXiv:1607.00299 [hep-lat].

[195] L. Giusti, 'Transfer matrix on the lattice: basics facts and baryonic projectors', Unpublished notes.

[196] S. Coleman, '1/N', in 17th International School of Subnuclear Physics: Pointlike Structures Inside and Outside Hadrons Erice, Italy, July 31-August 10, 1979 (1980), p. 0011.

[197] S. Coleman, *Aspects of Symmetry: selected Erice lectures of Sidney Coleman* (Cambridge University Press, 1988), 420 pp.

[198] E. Witten, 'Baryons in the 1/N expansion', Nucl. Phys. B **160**, 57–115 (1979).

[199] S. Coleman, 'The Uses of Instantons', in 15th Erice School of Subnuclear Physics: The Why's of Subnuclear Physics Erice, Italy, July 23-August 10, 1977, Vol. 15 (1979), p. 805.

[200] D. J. Gross, R. D. Pisarski and L. G. Yaffe, 'QCD and instantons at finite temperature', Rev. Mod. Phys. **53**, 43–80 (1981).

[201] A. Belavin, A. Polyakov, A. Schwartz and Y. Tyupkin, 'Pseudoparticle solutions of the Yang-Mills equations', Phys. Lett. B **59**, 85–87 (1975).

[202] G. 't Hooft, 'Computation of the quantum effects due to a four-dimensional pseudoparticle', Phys. Rev. D **14**, 3432–3450 (1976), [Erratum: Phys. Rev. D **18**, 2199 (1978)].

[203] M. F. Atiyah and R. S. Ward, 'Instantons and algebraic geometry', Commun. Math. Phys. **55**, 117–124 (1977).

[204] C. W. Bernard, 'Gauge zero modes, instanton determinants, and quantum-chromodynamic calculations', Phys. Rev. D **19**, 3013–3019 (1979).

[205] E. Hairer, C. Lubich and G. Wanner, *Geometric numerical integration: structure-preserving algorithms for ordinary differential equations*, Springer Series in Computational Mathematics 31 (Springer, 2006), 644 pp.

[206] H. Munthe-Kaas and B. Owren, 'Computations in a free Lie algebra', Philos. T. Roy. Soc. A **357**, 957–981 (1999).

[207] P. E. Crouch and R. Grossman, 'Numerical integration of ordinary differential equations on manifolds', J. Nonlinear Sci. **3**, 1–33 (1993).

[208] B. Owren and A. Marthinsen, 'Runge–Kutta methods adapted to manifolds and based on rigid frames', BIT **39**, 116–142 (1999).

[209] E. Celledoni, A. Marthinsen and B. Owren, 'Commutator-free Lie group methods', Future Gener. Comp. Sy. **19**, 341–352 (2003).

[210] B. Owren, 'Order conditions for commutator-free Lie group methods', J. Phys. A - Math. Gen. **39**, 5585–5599 (2006).

[211] M. Lüscher, SU(3) *matrix functions*, (2009) `https://luscher.web.cern.ch/luscher/notes/su3fcts.pdf`.