Application of contemporary renormalization group techniques to strongly-coupled field theories

by

Curtis Taylor Peterson

B.S., Arizona State University, 2019

M.S., University of Colorado Boulder, 2022

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Physics
2024

Committee Members:
Anna Hasenfratz, Chair
Ethan Neil
Thomas DeGrand
John Bohn
Markus Pflaum

Peterson, Curtis Taylor (Ph.D., Physics)

Application of contemporary renormalization group techniques to strongly-coupled field theories

Thesis directed by Professor Anna Hasenfratz

I explore the infrared properties of massless SU(3) gauge-fermion systems with $N_{\mathrm{f}} = 0$, 8, and 12 Dirac fermions in the fundamental representation of SU(3) using non-perturbative Wilsonian renormalization group (RG) techniques. From an infinite volume massless RG scheme based upon the gradient flow transformation, I calculate the non-perturbative RG β-function for all three systems. I verify my determination of the RG β-function by calculating the Λ-parameter of the $N_{\mathrm{f}} = 0$ system and the leading irrelevant critical exponent at the infrared fixed point of the $N_{\mathrm{f}} = 12$ system; both are in reasonable agreement with the literature. The $N_{\mathrm{f}} = 8$ β-function exhibits tantalizing signs of upward curvature, which could indicate that the $N_{\mathrm{f}} = 8$ system is either infrared conformal or slowly walking. Additionally, I develop a finite size scaling method based on radial basis function neural networks. This method is tested on the finite-temperature phase transition of various two-dimensional classical spin systems. It is then applied to a potential quantum (zero-temperature) phase transition that the $N_{\mathrm{f}} = 8$ appears to undergo in transiting from a weakly-coupled conformal phase to a strongly-coupled symmetric mass generation phase.

Dedication

In memory of Jeffrey Steward

Acknowledgements

I would not be in the position that I am now were it not for the back-breaking work that my parents and grandparents put in throughout their entire lives to make the life of their family better. I am especially grateful to my father, Curtis Peterson, for encouraging me to pursue what I love and to my mother, Wendy Steward, for teaching me how to make it a reality.

I am forever grateful to my lifelong friends Casey Morrison, Rigoberto Meza Lara, and Alejandro Martinez both for their friendship and for giving me the balance that I needed when I started to lose focus of what life has to offer outside of my aspirations. I would not have made it through my Ph.D. were it not for Chase & Amanda Brooks, Kenny Higginbotham, Evan Owen, Jesse & Emma Snelling, Sam & Becca Gross, Sally Rudolfs, Benjamin Marinoff, Kevin & Julia Winseck, and Tatum Ward; whether it was through skiing, playing Dungeons and Dragons, going to unhinged trivia nights, or simply talking, your friendship centered me when I needed it most.

I extend my heartfelt appreciation to my high school teachers/mentors Kristie Wiencek-Kilkelly, Andrew Stiger, and Mari Echols. The support of Richard Lebed, Wenbo Tang, Jeff Leiding, Christopher Ticknor, and Andrew Mugler made me both a better undergraduate researcher and human being. The incredible patience, insight, and humanity of my graduate mentors Anna Hasenfratz and Ethan Neil allowed me to both grow and flourish as a professional research scientist. No words can adequately convey my gratitude for the profound impact they've had on my life.

To my wife, Samantha Runnels, you have been my rock the entire way through. You bore much of the stress, self-doubt, and uncertainty that I've felt for years, yet you held me up with your love, laughter, and unyielding support. You are the light that guided me to the end.

Contents

Chapter

Appendix

Tables

Table

Figures

Figure

Chapter 1

Introduction

Renormalization is a practice which used to be widely regarded as distasteful, and so was largely done in the privacy of one's own home. That has all changed.

C. P. Burgess [59]

Our modern understanding of *quantum field theories* (QFTs), be it effective or fundamental, rests upon the bedrock ideas of *renormalization*. This is especially true for SU(N) gauge-fermion systems, which are often so strongly coupled at low energies that non-perturbative renormalization techniques are one of the only means by which one can hope to understand their low-energy properties. The primary objective of this thesis is to explore various aspects of SU(3) gauge-fermion systems with N_{f} fermions in the fundamental representation of SU(3) using non-perturbative *renormalization group* techniques that have been developed roughly over the past decade. This journey will take us through a variety of interesting topics, such as the physics of confinement, conformal systems, Standard Model physics (and beyond), quantum/classical phase transitions, symmetric mass generation, spin models, machine learning, optimization algorithms, Bayesian statistics/methods, Monte Carlo algorithms, high-performance computing, and much more, so sit back, relax, and enjoy the ride.

1.1 Gauge-fermion systems and renormalization

The classical Minkowski space action of a four-dimensional massless SU(N) gauge-fermion system with N_f fermions in the fundamental representation of SU(N) is

$$\mathcal{S}\big[\mathfrak{A}_\mu, \Psi, \overline{\Psi}\big] = \int \mathrm{d}^4x \Bigg[-\frac{1}{2g_0^2}\mathrm{Tr}_\mathrm{c}\big[\mathfrak{F}_{\mu\nu}(x)\mathfrak{F}^{\mu\nu}(x)\big] + \sum_{f=0}^{N_\mathrm{f}} \overline{\Psi}^{(f)}(x)\mathfrak{D}_f\Psi^{(f)}(x)\Bigg], \tag{1.1}$$

where

$$\mathfrak{D}_f \equiv i\gamma^\mu\mathfrak{D}_\mu - m_0^{(f)} \tag{1.2}$$

with Dirac matrix γ^μ and

$$\mathfrak{D}_\mu = \partial_\mu - i\mathfrak{A}_\mu(x) \tag{1.3}$$

the gauge covariant derivative. The non-Abelian vector potential $\mathfrak{A}_\mu(x) \in \mathfrak{su}(N)$ enters the field strength tensor $\mathfrak{F}_{\mu\nu}(x)$ as

$$\mathfrak{F}_{\mu\nu}(x) = i\big[\mathfrak{D}_\mu, \mathfrak{D}_\nu\big] \in \mathfrak{su}(N). \tag{1.4}$$

For the purpose of brevity, I shall refer to such systems as *gauge-fermion systems*. The $m_0^{(f)}$ parameters are the "bare mass". The fermion field $\Psi^{(f)}(x)$ of "flavor" f is Grassmann-valued and transforms under the fundamental representation of SU(N); likewise, $\overline{\Psi}^{(f)}(x) \equiv \Psi^{(f)}(x)^\dagger\gamma^0$. Furthermore, "$\mathrm{Tr}_\mathrm{c}$" denotes a trace over the "color indices" of either elements of SU(N) or its Lie algebra $\mathfrak{su}(N)$. For now, the bare gauge coupling g_0^2 is just a dimensionless parameter. In the quantized system on a (Euclidean) hypercubic lattice, g_0^2 controls the lattice spacing and hence the continuum limit. Quantization of the gauge-fermion system defined by Eqn. 1.1 proceeds schematically by defining the Minkowski space path integral

$$\mathcal{Z} \equiv \int \big[\mathrm{d}\mathfrak{A}_\mu \mathrm{d}\Psi \mathrm{d}\overline{\Psi}\cdots\big]\ \exp\big(i\mathcal{S}\big[\mathfrak{A}_\mu, \Psi, \overline{\Psi}, ...\big]\big), \tag{1.5}$$

where $\left[\mathrm{d}\mathfrak{A}_\mu \mathrm{d}\Psi \mathrm{d}\overline{\Psi}\cdots\right]$ is a formal "measure" (not actually a measure) that leaves room for additional unphysical "ghost" fields in $\mathcal{S}\left[\mathfrak{A}_\mu, \overline{\Psi}, \Psi, ...\right]$ that take care of ambiguities in the path integral arising from gauge freedom.

Essentially all of the difficulties pertaining to understanding the quantum properties of gauge-fermion systems arise from Eqn. 1.5. Perhaps the most egregious sin of Eqn. 1.5 is that the integral itself is not well-defined as it stands. In essence, the purpose of renormalization is to give expressions like Eqn. 1.5 meaning. In the early days of quantum field theory (roughly, 1930s-1950s), issues with the definition of quantum field theories (QFTs) like that of Eqn. 1.5, *quantum electrodynamics* (QED), revealed themselves through divergences that appear in perturbative expansions of observables at next-to-leading-order in the bare gauge coupling g_0^2. Due in large to the work of Richard Feynman, Julian Schwinger, Freeman Dyson, and Sin-Itiro Tomonaga in the 1950s, it became apparent that divergences appearing in perturbative calculations of QED observables could be removed by the procedure of *renormalization* as follows.

(1) Introduce a cutoff Λ in fluctuations of the QFT.

(2) Redefine (*renormalize*) the couplings $\left\{g_0^2, m_0^{(f)}\right\} \to \left\{g^2, m^{(f)}\right\}$ in Eqn. 1.1 by defining $\left\{g^2, m^{(f)}\right\}$ in terms physical observables $\{O_1, O_2, ...\}$.

(3) Take the $\Lambda \to \infty$ limit with $\left\{g^2, m^{(f)}\right\}$ fixed.

The latter three step procedure not only removed divergences from observables in QED, it also opened the door to performing calculations of the Lamb shift and anomalous magnetic moment of the electron with unprecedented accuracy. At about the same time, Murray Gell-Mann and Francis Low noted that the renormalized coupling $g^2(\mu)$ in a *massless ($m^{(f)} = 0$) renormalization scheme* can be expressed in terms of a *renormalization scale* μ^2 [143]. The *running* of $g^2(\mu)$ with μ^2 was later characterized in the 1970s by Curtis Callan and Kurt Symanzik in terms of the *renormalization group (RG) β-function* $\beta\left(g^2\right)$ as [65, 349]

$$\mu^2 \frac{\mathrm{d}g^2(\mu)}{\mathrm{d}\mu^2} = \beta\left(g^2\right). \tag{1.6}$$

Almost in parallel with the work of Callan and Symanzik, Kenneth Wilson reformulated the renormalization procedure in terms of Leo Kadanoff's *block spin transformation* [216], for which Wilson won the Nobel prize [372–374].

According to the perspective of *Wilsonian RG*, a *continuum quantum field theory* defined at a scale μ and parameterized by the renormalized coupling $g^2(\mu)$ is obtained from a sequence of *effective field theories* (EFT) that approach the continuum QFT in the $\Lambda \to \infty$ limit at fixed μ (equivalently, fixed $g^2(\mu)$). Each EFT comes with its own cutoff $\Lambda \geq \mu$ and is hence well-defined, at least in the eyes of the common physicist. In the parlance of high-energy physics, each EFT is said to be "defined at the scale Λ", meaning that the EFT is unable to describe physics that occurs at scales $\mu > \Lambda$. Each EFT can be made to better approximate its corresponding continuum QFT by integrating out fluctuations (*degrees of freedom*) between μ and Λ while imposing that the integration procedure does not disturb physical properties of the EFT at scales $\lesssim \mu$. This is achieved by performing a series of *renormalization group transformations* on the EFT, which are discussed at length in Chapter 2. Performing the same procedure for a continuous sequence of EFTs defined at Λ approaching infinity, the continuum QFT defined at μ is reached. Alternatively, one need not even take a continuum limit in the first place, as each EFT is a well-defined approximation of a continuum QFT that exists up in the high heavens. So long as a sufficiently low μ can be reached from the EFT, it can be used to probe the low-energy properties of the continuum QFT with a systematic error that goes as $(\mu/\Lambda)^p \log(\mu/\Lambda)^q$ for some $p, q \in \mathbb{Z}$ at leading order. The modern view of the Standard Model is that it is such an EFT; however, whether or not the Standard Model EFT (SMEFT) has a well-defined continuum limit is up for debate.

1.1.1 The zero-temperature phase diagram

Bringing ourselves back down to the task at hand, Wilsonian RG, and the many decades of research that lead up to it, tells us how to access the low-energy properties of the quantized gauge-fermion system defined by Eqns. 1.1-1.5 from a sequence of RG transformations and continuum $\Lambda \to \infty$ limits. Given an RG transformation, the renormalized coupling $g^2(\mu)$ defined by the RG transformation characterizes the scale-dependent behavior of the continuum gauge-fermion system. The β-function for $g^2(\mu)$ can be calculated in *renormalized perturbation theory*. Famously, the 1-loop β-function

$$\beta\big(g^2\big) \sim -b_0 \frac{g^4}{(4\pi)^2}$$

was first calculated in 1974 by David Gross and Frank Wilczek, along with David Politzer independently, which earned all three of them the Nobel prize [154, 155, 301].[1] The 1-loop coefficient

$$b_0 = \frac{1}{3}\Big(11N - 2N_\mathrm{f}\Big) \tag{1.7}$$

is greater than zero when $N_\mathrm{f} \leq 11N/2$, which implies that the renormalized coupling $g^2 \to 0$ when $\mu \to \infty$; in other words, the system is weakly interacting a high energies (short distances). The latter property is referred to as *asymptotic freedom*. The two-loop perturbative β-function for the gauge-fermion system defined by Eqn. 1.1 with $m_0^{(f)} = 0$ was calculated by William Caswell in 1974 [75], yielding

$$\beta\big(g^2\big) \sim -\frac{g^4}{(4\pi)^2}\left[b_0 + b_1 \frac{g^2}{(4\pi)^2}\right] \qquad \big(g^2/4\pi \to 0\big), \tag{1.8}$$

where

$$b_1 = \frac{2N(17N - 5N_\mathrm{f})}{3} - \frac{\big(N^2-1\big)N_\mathrm{f}}{N}. \tag{1.9}$$

In a *massless RG scheme*, the b_0 and b_1 coefficients are *universal*, meaning that they are independent of the RG transformation (or scheme) used to calculate them. RG scheme dependence in the perturbative RG β-function enters at 3-loops and beyond. In what follows, I am going to focus on the massless $N = 3$ case unless stated otherwise.

1.1.1.1 The quantum electrodynamics-like phase

According to the 2-loop β-function, massless SU(3) gauge-fermion systems lose asymptotic freedom at $N_\mathrm{f}^* \approx 16.5$. In other words, the system exists in a quantum electrodynamics (QED-like)

---

[1] Supposedly, Gerard ’t Hooft had discovered asymptotic freedom a year before, but never published the result [1].

phase with a positive β-function when $g^2/4\pi \ll 1$ for $N_\mathrm{f} \gtrsim N_\mathrm{f}^*$. In Chapter 3.4, we will learn that the existence of a repulsive *fixed point* ($\beta\big(g_\star^2\big) = 0$) in the RG flow dictated by the β-function implies that a continuum limit from a collection of EFTs can be defined. Such fixed points are referred to as a *ultraviolet fixed points* (UVFPs). Attractive fixed points are referred to as *infrared fixed points* (IRFPs), such as the IRFP of the QED phase at

$$g_\star^2 = 0 \qquad \text{(IRFP, } N_\mathrm{f} \gtrsim N_\mathrm{f}^*\text{).} \tag{1.10}$$

Should a UVFP exist in the QED phase, the value of the non-trivial fixed point coupling $g_\star^2$ at leading order in $1/N_\mathrm{f}$ is expected to go as

$$g_\star^2 \propto 1/N_\mathrm{f} \qquad \text{(hypothesized UVFP, } N_\mathrm{f} \gtrsim N_\mathrm{f}^*\text{)} \tag{1.11}$$

in any RG scheme [203]. However, it should be noted that evidence for presence of a UVFP in the QED phase is few and far between. Closer to N_f^*, it is possible that the fixed point is simply not accessible from perturbation theory and it instead arises non-perturbatively, as was hoped for QED and is still hoped for asymptotically safe approaches to quantum gravity [311].

1.1.1.2 The conformal phase

Below N_f^*, the two-loop β-function possesses both a UVFP at

$$g_\star^2 = 0 \qquad \text{(UVFP, } N_\mathrm{f} \lesssim N_\mathrm{f}^*\text{).} \tag{1.12}$$

It also possesses an IRFP for some range of $N_\mathrm{f}^\mathrm{c} \leq N_\mathrm{f} \lesssim 16$. The $N_\mathrm{f} = 16$ IRFP at

$$g_\star^2 \approx -(4\pi)^2 b_0/b_1 \approx 0.523 \qquad \text{(IRFP, } N_\mathrm{f} = 16\text{)} \tag{1.13}$$

is accessible from perturbation theory. This is the *Caswell–Banks–Zaks fixed point* [28, 75]. Though it had been noted before [75], Tom Banks and Alex Zaks were the first to explore it systematically [28]. The existence of an IRFP at some $g_\star^2 \neq 0$ has interesting consequences for the low-energy dynamics of the system. Most notably, deep within the infrared, the system is *conformal*; in other words, it is invariant under coordinate transformations $x^\mu \to x^\mu + \epsilon^\mu$ satisfying [138, 314]

$$\partial_\mu \epsilon_\nu + \partial_\nu \epsilon_\mu \approx \frac{1}{2}\big(\partial_\sigma \epsilon^\sigma\big) g_{\mu\nu} \qquad (\epsilon^\mu / x^\mu \ll 1), \tag{1.14}$$

where $g_{\mu\nu}$ is the metric tensor. Such a coordinate transformation changes the metric as

$$g_{\mu\nu} \to \Omega^2(x) g_{\mu\nu}, \tag{1.15}$$

where

$$\frac{1}{2}\Big(\partial_\mu \epsilon^\rho + \partial_\nu \epsilon^\sigma\Big) g^{\mu\nu} g_{\rho\sigma} \approx \Omega^2(x) - 1 \qquad (\epsilon^\mu / x^\mu \ll 1). \tag{1.16}$$

The transformations generated by infinitesimal ϵ^μ are described by the *conformal group*. According to Eqn. 1.15, the group of diltations (Ω^2 independent of x) is a subgroup of the conformal group. As such, conformally invariant field theories are also *scale invariant*. Conformal invariance implies that the correlation function of any operator $\mathcal{O}(x)$ transforms under $x \to \lambda x$ as

$$\langle \mathcal{O}(\lambda x) \mathcal{O}(0) \rangle = \lambda^{-2\Delta_{\mathcal{O}}} \langle \mathcal{O}(x) \mathcal{O}(0) \rangle \tag{1.17}$$

for $\Delta_{\mathcal{O}}$ a constant that is referred to as the *scaling dimension* of $\mathcal{O}$. This implies that the correlation function of a field Φ (e.g., a gauge or fermion field) decays as

$$\langle \Phi(x) \Phi(0) \rangle = |x|^{-2\Delta_\Phi}, \tag{1.18}$$

where Δ_Φ is the scaling dimension of Φ. Additionally, the expected trace of the stress tensor $T^{\mu\nu}$ vanishes

$$\left\langle T^{\mu}_{\mu} \right\rangle = 0. \tag{1.19}$$

Eqns. 1.17-1.18 imply that the spectrum is continuous; as such, the low-energy dynamics does not permit the existence of a particle spectrum. Furthermore, if conformal invariance is broken explicitly by introducing a fermion mass, the corresponding hadron spectrum will exhibit *conformal hyperscaling*. Deriving conformal hyperscaling follows along the same line of reasoning that leads to finite size scaling, which I describe in Chapter 2.5. In the language of RG, the fermion mass is a *relevant deformation* that repels any RG flow away from the critical surface. Therefore, the zero-momentum correlation function

$$G_{\mathcal{O}}\big(\hat{x}_4; \hat{m}\big) \equiv \int \mathrm{d}^3\mathbf{x} \; \langle \mathcal{O}(\mathbf{x}, x_4) \mathcal{O}(0) \rangle \tag{1.20}$$

of any operator $\mathcal{O}$ in an EFT of the mass-deformed conformal system is expected to scale as [100]

$$G_{\mathcal{O}}\big(x_4; \hat{m}\big) \propto \mathcal{F}_{\mathcal{O}}\Big(\hat{x}_4 \hat{m}^{1/(1+\gamma_m^\star)}\Big) \qquad \big(\hat{m} \ll 1\big) \tag{1.21}$$

where $\hat{x}_4 = \Lambda x_4$ is the source/sink separation along the time direction and $\hat{m} = m_f/\Lambda$ is the degenerate fermion mass. Both are expressed in units of the cutoff Λ. The constant $\gamma_m^\star$ is the mass anomalous dimension at the IRFP. The function $\mathcal{F}_{\mathcal{O}}$ is a *universal scaling function*. Eqn. 1.21 implies that any hadron mass of the mass-deformed EFT must scale as

$$M_{\mathcal{O}}/\Lambda \propto \hat{m}^{1/(1+\gamma_m^\star)} \qquad \big(\hat{m} \ll 1\big). \tag{1.22}$$

In other words, ratios of hadron masses from the mass-deformed EFT should be flat in $\hat{m}$. Such hyperscaling relations have been used extensively to test for conformality in lattice gauge theory simulations of many-flavor gauge-fermion systems. See, for example, references provided in overview of Chapter 5. The range in N_f over which massless $\mathrm{SU}(N)$ gauge-fermion systems are both asymptotically free and conformal is referred to as the *conformal window.* The lower boundary of the conformal window N_f^c in N_f is referred to as the *conformal sill.*

1.1.1.3 The confined phase

For $N_\mathrm{f} < N_\mathrm{f}^\mathrm{c}$, the low-energy dynamics is drastically different than it is for $N_\mathrm{f} > N_\mathrm{f}^\mathrm{c}$; however, it is much more familiar. The confined phase is characterized by the generation of an infrared scale, the *confinement scale*, that is associated with the phenomenon of *confinement*; i.e., the inability to separate fermions in a hadron without putting so much energy into the system that the act of pulling them apart simply creates more hadrons. The confined phase exists between $0 \leq N_\mathrm{f} < N_\mathrm{f}^\mathrm{c}$. For $N_\mathrm{f} \geq 1$, the confined phase is also characterized by *spontaneous chiral symmetry breaking*, which I shall now describe.

The classical action of Eqn. 1.1 in the $m_f = 0$ limit is symmetric under $U \in \mathrm{SU}(N_\mathrm{f})_\mathrm{R} \times \mathrm{SU}(N_\mathrm{f})_\mathrm{L}$ transformations of the form

$$U = \exp\left(i\boldsymbol{\alpha} \cdot \boldsymbol{T}\right) \text{ or } U = \exp\left(i\gamma^5 \boldsymbol{\alpha} \cdot \boldsymbol{T}\right), \tag{1.23}$$

where $\boldsymbol{T} = \left(T^1, T^2, ..., T^{N_\mathrm{f}^2\text{-}1}\right)$ is a collection of generators for $\mathfrak{su}(N_\mathrm{f})$, $\boldsymbol{\alpha} \in \mathbb{R}^{N_\mathrm{f}^2\text{-}1}$, and $\gamma^5 \equiv i\gamma^1\gamma^2\gamma^3\gamma^4$ is the standard chirality operator that separates any Dirac fermion into left-handed components (L)

$$\Psi_\mathrm{L}^{(f)} \equiv \frac{1}{2}\left(1 - \gamma^5\right)\Psi^{(f)} \tag{1.24}$$

and right-handed components (R)

$$\Psi_\mathrm{R}^{(f)} \equiv \frac{1}{2}\left(1 + \gamma^5\right)\Psi^{(f)} \tag{1.25}$$

in four dimensions. Elements of $\mathrm{SU}(N_\mathrm{f})_\mathrm{R} \times \mathrm{SU}(N_\mathrm{f})_\mathrm{L}$ act on an extended space of spinors involving all flavors $\boldsymbol{\Psi} \equiv \left(\Psi^{(0)}, \Psi^{(1)}, ..., \Psi^{(N_\mathrm{f}\text{-}1)}\right)$. Note that there are also two additional U(1) "axial" and "vector" symmetries; however, the U(1) axial symmetry is broken due to quantum effects. The transformations $U \in \mathrm{SU}(N_\mathrm{f})_\mathrm{R} \times \mathrm{SU}(N_\mathrm{f})_\mathrm{L}$ involving the chirality operator γ^5, otherwise known as *axial vector transformations*, mix both flavors $\Psi^{(f)}$ and chiral components $\Psi^{(f)}_\mathrm{L}/\Psi^{(f)}_\mathrm{R}$. In the confined phase, this *chiral symmetry* is spontaneously broken, taking

$$\mathrm{SU}(N_\mathrm{f})_\mathrm{R} \times \mathrm{SU}(N_\mathrm{f})_\mathrm{L} \to \mathrm{SU}\big(N_\mathrm{f}\big)_\mathrm{V}, \tag{1.26}$$

where $\mathrm{SU}\big(N_\mathrm{f}\big)_\mathrm{V}$ is the "vector subgroup" of $\mathrm{SU}(N_\mathrm{f})_\mathrm{R} \times \mathrm{SU}(N_\mathrm{f})_\mathrm{L}$; it is the *isospin symmetry* that is approximately realized in the light quark sector of QCD. This spontaneous *chiral symmetry breaking* (χSB) has profound consequences for the low-energy dynamics of gauge-fermion systems in the confined phase. For one, it produces *Goldstone bosons* $\boldsymbol{\pi} = \big(\pi^-, \pi^0, \pi^+\big)$, which are identified with the *pseudo-Goldstone pions* of QCD. Without pions, the nuclear force that keeps nuclei together would not be strong enough to keep them from falling apart. As Goldstone bosons, the pions are also the lightest states of QCD. Therefore, they have a dominant effect in just about every low-energy QCD process. χSB also has important consequences for the spectrum of baryons in QCD; for example, it is responsible for the non-degeneracy of nucleons with their parity partners, which are significantly heavier [142]. Quite frankly, it is hard to overstate the importance of χSB. In the isospin symmetric limit, the masses of the pions follow the famous Gell-Mann-Oakes-Renner (GMOR) relation [141, 144, 368]

$$M_\pi^2 \propto \Lambda_\chi m_f \qquad\qquad (m_f/\Lambda_\chi \ll 1), \tag{1.27}$$

where Λ_χ is the "*chiral symmetry breaking scale*". This is to be juxtaposed against the hyperscaling prediction of Eqn. 1.22 in the conformal phase.

1.1.1.4 Quantum phase transition at the conformal sill

The difference between the low-energy dynamics of the confined and conformal phase is quite astonishing. Below the conformal sill, the system is confining, chirally broken, and full of hadrons. Above the conformal sill, the system is interacting (and thereby non-trivial), chirally symmetric, and there are no particles in sight. Something quite dramatic must be happening at $N_{\mathrm{f}}^{\mathrm{c}}$. Such a transition is referred to as a *quantum phase transition* (QPT), as it is not a consequence of thermal fluctuations. Already, we expect two things to happen when crossing the conformal sill from the confined phase. For one, we of course expect that β-function in any RG scheme to pick up an infrared fixed point. Additionally, we expect the chiral condensate

$$\Sigma_f \equiv \left\langle \overline{\Psi}^{(f)}(x)\Psi^{(f)}(x)\right\rangle \tag{1.28}$$

to vanish. The position argument is suppressed because the chiral condensate expectation is translationally invariant. Because on the main focuses of this thesis is on the calculation non-perturbative β-functions, I shall not dwell too much on the relevance of the chiral condensate.

The β-function is expected to either "*jump*" or "*walk*" within the vicinity of $N_{\mathrm{f}}^{\mathrm{c}}$. In the *jumping* scenario, the β-function depends on g^2 and N_{f} like [323, 324]

$$\beta\left(g^2\right) \propto \frac{-g^4\left[1-k_1 g^2-\delta(N_{\mathrm{f}})\right]}{1-k_2 g^2} \qquad \text{(jumping)}, \tag{1.29}$$

where

$$\delta(N_{\mathrm{f}}) \to 0 \text{ as } N_{\mathrm{f}} \to N_{\mathrm{f}}^{\mathrm{c}} \tag{1.30}$$

and k_1, k_2 are free parameters. In the *walking* scenario, the β-function goes with g^2 and N_{f} as [323]

$$\beta\left(g^2\right) \propto -g^4\left[\left(g^2-k\right)^2-\delta(N_{\mathrm{f}})\right] \qquad \text{(walking)}, \tag{1.31}$$

where k is a free parameter and $\delta(N_{\mathrm{f}})$ is defined by Eqn. 1.30. In Figs. 1.1-1.2, I illustrate the shape

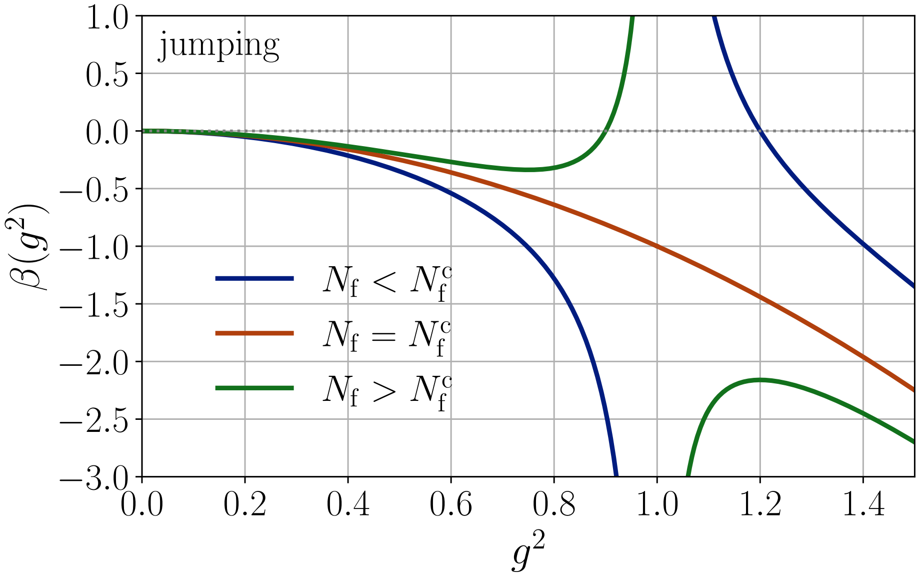

Figure 1.1: Illustration of the β-function in the jumping dynamics scenario just below the conformal sill ($N_\mathrm{f} < N_\mathrm{f}^\mathrm{c}$, blue line), at the conformal sill ($N_\mathrm{f} = N_\mathrm{f}^\mathrm{c}$, red line), and just below the conformal sill ($N_\mathrm{f} > N_\mathrm{f}^\mathrm{c}$, blue line).

of the β-function in a jumping scenario (Fig. 1.1, Eqn. 1.29) and walking scenario (Fig. 1.2, Eqn. 1.31) for N_f just below, at, and just above the conformal sill with $k = k_1 = k_2 = 1$ for the purposes of illustration.

β-functions of the form of Eqn. 1.29 have appeared in literature on supersymmetric Yang-Mills [279, 333] and non-supersymmetric Yang-Mills [77, 315]. As is illustrated in Fig. 1.1, the β-function runs into a pole at $g_\mathrm{p}^2 \approx 1/k_2$ below N_f^c (blue line). Past the pole, it runs into a UVFP. Above N_f^c (green line), the β-function exhibits an IRFP on the weak coupling side of the pole and does nothing exciting on the other side of the pole. Only at $N_\mathrm{f} = N_\mathrm{f}^\mathrm{c}$ (red line) is the β-function on one side of the pole connected to the other side of the pole. As jumping dynamics is associated with a first-order conformal QPT, the system on the weak coupling side of the pole is not connected to the strong coupling side, except at $N_\mathrm{f} = N_\mathrm{f}^\mathrm{c}$.

The walking scenario has received much more attention in the literature due to its relevance to beyond Standard Model (BSM) model building. Similar behavior has also been observed in a variety of systems [136, 221, 261]. When $N_\mathrm{f} > N_\mathrm{f}^\mathrm{c}$ (green line), the β-function runs from the UVFP

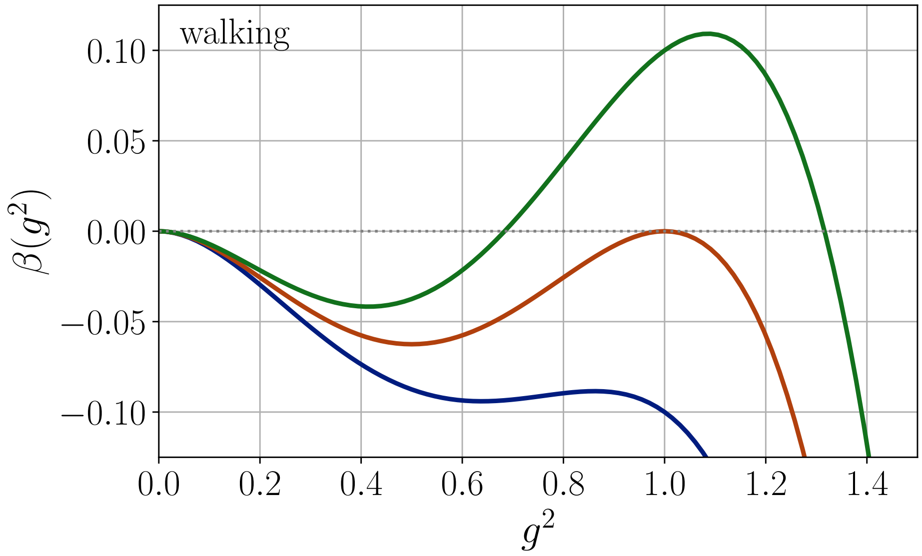

Figure 1.2: Illustration of the β-function in the walking dynamics scenario just below the conformal sill ($N_{\rm f} < N_{\rm f}^{\rm c}$, blue line), at the conformal sill ($N_{\rm f} = N_{\rm f}^{\rm c}$, red line), and just below the conformal sill ($N_{\rm f} > N_{\rm f}^{\rm c}$, blue line).

at $g_\star^2 = 0$ to an IRFP at some $g_{a\star}^2$. In addition to $g_{a\star}^2$, the RG flow also possesses another UVFP at some $g_{b\star}^2 > g_{a\star}^2$. As $N_{\rm f} \to N_{\rm f}^{\rm c}$ from the conformal phase, $g_{a\star}^2$ increases, while $g_{b\star}^2$ decreases. The conformal QPT occurs when $g_{a\star}^2$ and $g_{b\star}^2$ merge (red line). Below $N_{\rm f}^{\rm c}$, $g_{a\star}^2$ and $g_{b\star}^2$ move away the complex plane. Once the system is in the confined phase, chiral symmetry is broken and the system confines. From the confined phase and just below $N_{\rm f}^{\rm c}$, the Λ-parameter $\Lambda_{\rm X}$, which characterizes the scale of non-perturbative observables calculated within any particular RG scheme X from the confined phase (see Chapter 4.6), scales with $N_{\rm f}$ as [323]

$$\Lambda_X \propto \left(N_{\rm f}^{\rm c} - N_{\rm f}\right)^{-1} \exp\left[-\frac{\pi}{2}\left(N_{\rm f}^{\rm c} - N_{\rm f}\right)^{-1/2}\right]. \tag{1.32}$$

Eqn. 1.32 is the famous Miransky scaling [261]. Such scaling is reminiscent of the ∞-order phase transition of the two-dimensional XY model (explored in Chapter 6); as such, the picture posited by Miransky scaling is that the conformal QPT is ∞-order; i.e., due to an essential singularity in $\log \mathcal{Z}$. Corrections to Miransky scaling have been explored in Ref. [50]; see also Ref. [221] for more information on the relevance of Miransky scaling to the picture of merging fixed points.

1.2 Relevance to beyond Standard Model physics

1.2.1 The conformal window and beyond Standard Model physics

The pattern of chiral symmetry breaking (Eqn. 1.26) in the confined phase of a gauge-fermion system is similar in structure to the pattern of symmetry breaking that occurs in the electroweak sector. The symmetry breaking pattern for spontaneous electroweak symmetry breaking (EWSB) is

$$\mathrm{SU}(2)_{\mathrm{L}} \times \mathrm{SU}(2)_{\mathrm{R}} \to \mathrm{SU}(2)_{\mathrm{V}}. \tag{1.33}$$

More precisely, because the $\mathrm{U}(1)_Y$ subgroup of $\mathrm{SU}(2)_{\mathrm{R}}$ is gauged, the symmetry breaking pattern is

$$\mathrm{SU}(2)_{\mathrm{L}} \times \mathrm{U}(1)_Y \to \mathrm{U}(1)_Q, \tag{1.34}$$

where $\mathrm{U}(1)_Q$ is the Abelian electromagnetic gauge symmetry [287, 323]. As an aside, note that this standard realization of EWSB appears to be in conflict with *Elizur's theorem*, which states that a gauge symmetry cannot be spontaneously broken [116, 153]; as such, the interpretation of EWSB in the Standard Model must be more subtle than the standard approach that is expounded by most physicists, including the author. The formal similarity between the symmetry breaking pattern of χSB and that of EWSB led to the development of *technicolor models* (TC), which aimed to deliver EWSB from a form of χSB by extending the Standard Model with a new $\mathrm{SU}(N)$ gauge-fermion sector possessing $N_{\mathrm{f}} < N_{\mathrm{f}}^{\mathrm{c}}$ fermions in some representation of $\mathrm{SU}(N)$ (or some other Lie group) [119, 200, 236]. The technicolor Higgs arises from the formation of a non-zero chiral condensate (Eqn. 1.28) due to χSB, which generates of mass for the $W^{\pm}$ and Z bosons. *Extended technicolor* (ETC) models additionally aim to address the EWSB-induced generation of fermion masses in the Standard Model [119, 220]. While the foundations of early TC/ETC models were sound, they ultimately failed to meet the following experimental criteria [200, 236].

(1) TC/ETC models fail to pass electroweak precision tests.

(2) ETC models fail to suppress effects from flavor changing neutral currents (FCNCs) while also accommodating for the heaviness of the top quark[2] [236].

Interest in TC(ETC) models was revived with the advent of *walking technicolor* (WTC) [85, 121, 297], which is founded on the same χSB-induced EWSB principles as TC(ETC) models; however, the β-function of the strongly-coupled sector is of the walking type described in Sec. 1.1.1.4 and illustrated in Fig. 1.2. Defining $\mu_{\rm TC}$ to be the scale at which chiral symmetry breaks and $\mu_{\rm ETC}$ to be the scale at which the fermions acquire their mass (possibly from additional spontaneous symmetry breaking), the chiral condensate $\Sigma_f\left(\mu_{\rm TC}\right)$ at $\mu_{\rm TC}$ is related to $\Sigma_f\left(\mu_{\rm ETC}\right)$ in WTC models as [323]

$$\Sigma_f\left(\mu_{\rm ETC}\right) \sim \log\left(\frac{\mu_{\rm ETC}}{\mu_{\rm TC}}\right)^{\gamma_m^\star} \Sigma_f\left(\mu_{\rm TC}\right); \tag{1.35}$$

in other words, the condensate at the $\mu_{\rm ETC}$ scale is logarithmically enhanced. The fixed point mass anomalous dimension $\gamma_m^\star$ determines the size of the logarithmic enhancement. If the enhancement is large enough, then it is possible for WTC to evade the problems introduced by electroweak precision observables and FCNCs, so long as $\gamma_m^\star$ is at least O(1). WTC models built on top of conformal phases that are sufficiently close to the boundary of the conformal window, so that $\gamma_m^\star \sim$ O(1), could additionally evade the issue that are posed by needing to produce the top quark mass [64, 85]. Though, to the best of the author's knowledge, the latter statement has yet to be established conclusively; see, also, Refs. [64, 218] for workarounds involving *fermion partial compositeness.* Many of the modern BSM models that are WTC-adjacent are constructed from systems that are barely below the conformal sill. This allows them to utilize the approximate conformal invariance that is potentially realized by near-conformal systems to deliver EWSB from an approximate breakdown of conformal invariance. In such models, the Higgs is a pseudo-Nambu-Goldstone (pNGB) boson of spontaneous conformal symmetry breaking; i.e., it is a *dilaton.* Such models have been investigated

---

[2] The top quark is incredibly massive as far as Standard Model fermions are concerned, with a rest mass of $m_t \approx 173$ GeV. That is close to the mass of rhenium (≈ 173 GeV), which has an atomic number of 75.

extensively using dilaton effective chiral perturbation theory (dχPT) to describe potential candidates for such models using lattice simulations [14, 20, 149–151, 208]. I will discuss some of these attempts at describing the $(N, N_\mathrm{f}) = (3, 8)$ system using dχPT in Chapter 7. See Refs. [86, 287] for extensive reviews of many other pNGB models.

1.2.2 Symmetric mass generation

Before signing off on this section, let me briefly mention ideas from condensed matter physics that could have implications for the problem of dynamically generating fermion masses. Roughly over the past two decades, a mechanism for mass generation in interacting systems has been discovered and investigated extensively. Such *symmetric mass generation* (SMG) can occur if the system is free of quantum anomalies. It is also intimately connected to the physics of chiral edge modes [156, 366, 386]. For a review of SMG, see Ref. [367]. Recently, it has been realized that 4-dimensional systems of Kähler-Dirac fermions, or multiples of 4 Dirac fermions, could possess just the anomaly cancellation needed to realize SMG [62, 76]. The same result has also been interpreted in the context of chiral edge modes by realizing 4-dimensional Kähler-Dirac fermions as edge states of a 5-dimensional symmetry protected topological phase [156]. Lattice simulations of the $(N, N_\mathrm{f}) = (3, 8)$ gauge-fermion system show some signs of an SMG phase [170], though much more scrutiny is needed; this is explored further in Chapter 7. The existence of a strongly-coupled SMG phase that is continuously connected to some other weakly coupled phase (likely conformal) could open up exciting opportunities for BSM model building. It has been even put forth by Ref. [170] that the existence of such a continuous phase transition could signal the beginning of the conformal window. If not, it could at least be an interesting example of an SMG phase that is induced by strong dynamics, which, to the best of the author's knowledge, has yet to be realized.

1.3 Synopsis of key results

This thesis focuses upon two approaches to extracting information about properties of massless SU(3) gauge-fermion systems using tools from Wilsonian RG. The first approach utilizes non-

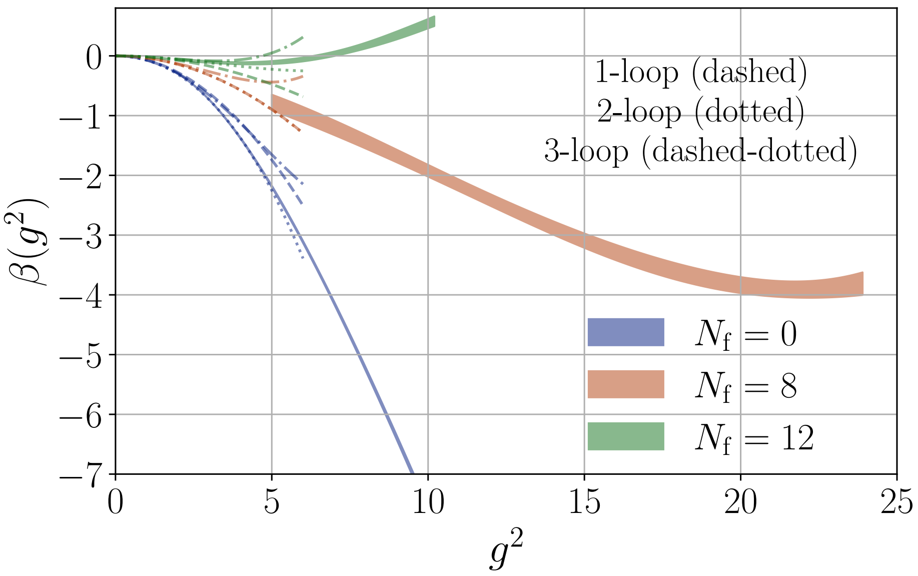

Figure 1.3: Gradient-flow-based continuum β-function for $N_\mathrm{f} = 0$ (blue), 8 (red), and 12 (green). Non-perturbative continuum β-function indicated by multicolored bands with the width of the band indicating the error. Perturbative 1- (dashed), 2- (dotted), and 3-loop β-function indicated by multicolored lines [161].

perturbative β-functions, which are calculated from Monte Carlo simulations of the $N_\mathrm{f} = 0, 8$, and 12 system in Chapters 4, 7, and 5, respectively. The second approach utilizes finite-size scaling (FSS). In Chapter 6, I describe a neural-network-based method for FSS and apply it to various spin systems. In Chapter 7, the same neural-network-based FSS method is applied to the zero-temperature phase of the $N_\mathrm{f} = 8$ system.

1.3.1 Non-perturbative β-functions

Fig. 1.3 shows the prediction for the continuum β-function $\beta\big(g^2\big)$ for $N_\mathrm{f} = 0$ (blue, Chapter 4), 8 (red, Chapter 7), and 12 (green, Chapter 5) as multicolored bands. The width of the band indicates the error. The continuum β-function is calculated from an infinite-volume/massless RG scheme that utilizes the continuous gradient flow smearing transformation of Ref. [252], as described in Chapter 4 [181, 189, 190, 293]. All simulations utilize the Hamiltonian (hybrid) Monte Carlo algorithm described in Chapter 8.1.1. The simulations of Chapters 5 and 7 additionally utilize Pauli-Villars improvement, which is described Chapter 5.2. The continuum β-function in Fig. 1.3 is juxtaposed against the universal 1- (dahsed) and 2-loop (dotted) perturbative β-functions, along with the 3-loop (dashed-dotted) perturbative β-function that is calculated from the same gradient flow-based RG scheme as the non-perturbative continuum β-functions [161]. The color of the perturbative β-function indicates the corresponding value of N_{f}.

The behavior of the β-function in the weak coupling region ($g^2/4\pi \ll 1$) for each N_{f} approaches its respective perturbative counterpart, as expected. On the other hand, the strong coupling behavior from different N_{f} is very different. For the $N_{\mathrm{f}} = 0$ system (blue band in Fig. 1.3), the β-function increases in magnitude without bound; eventually, it becomes asymptotically linear in g^2. Most of the $N_{\mathrm{f}} = 0$ β-function is cut off in Fig. 1.3 for the purposes of visualization. The full continuum β-function is shown in either Fig. 4.8 or Fig. 4.12. In Chapter 4, I verify the calculation of the $N_{\mathrm{f}} = 0$ β-function by calculating from it the $\overline{\mathrm{MS}}$ Λ-parameter. The calculation yields $\sqrt{8t_0}\Lambda_{\mathrm{MS}} = 0.622(10)$, which is consistent with the most recent determinations based on gradient flow [92, 380] (see Fig. 4.16). The scale t_0 is used in modern-day scale setting calculations[3] and is defined by $g^2(\mu_0) \approx 15.8$ at $\mu_0^2 \sim 1/8t_0$ in the aforementioned infinite-volume/massless gradient flow RG scheme [252].

The $N_{\mathrm{f}} = 8$ β-function (red band in Fig. 1.3) initially decreases, then begins to show signs of turning around around $15 \lesssim g^2 \lesssim 23$. If the $N_{\mathrm{f}} = 8$ system is below the conformal window, then such behavior could be indicative of walking (Fig. 1.2). If $N_{\mathrm{f}} = 8$ is instead inside the conformal window, then the turnaround is a sign of the β-function running into an IRFP. In any case, the β-function for the $N_{\mathrm{f}} = 8$ system is very different asymptotically from that of $N_{\mathrm{f}} = 0$ and even the $N_{\mathrm{f}} = 2$ β-function of Ref. [177] calculated from the same RG scheme.

The $N_{\mathrm{f}} = 12$ system (green band in Fig. 1.3) is likely in the conformal phase, which is supported by a majority of investigations probing its infrared behavior properties. Indeed, the $N_{\mathrm{f}} = 12$ RG β-function of this thesis exhibits an IRFP at $g_\star^2 = 6.60(62)$. To verify the calculation of the $N_{\mathrm{f}} = 12$ β-function in Chapter 5, I calculate from it the leading irrelevant critical exponent

---

[3] Slight correction: the w_0 scale is often preferred over t_0 for scale setting because $w_0^2 > t_0$; hence, cutoff effects in w_0 are suppressed relative to t_0.

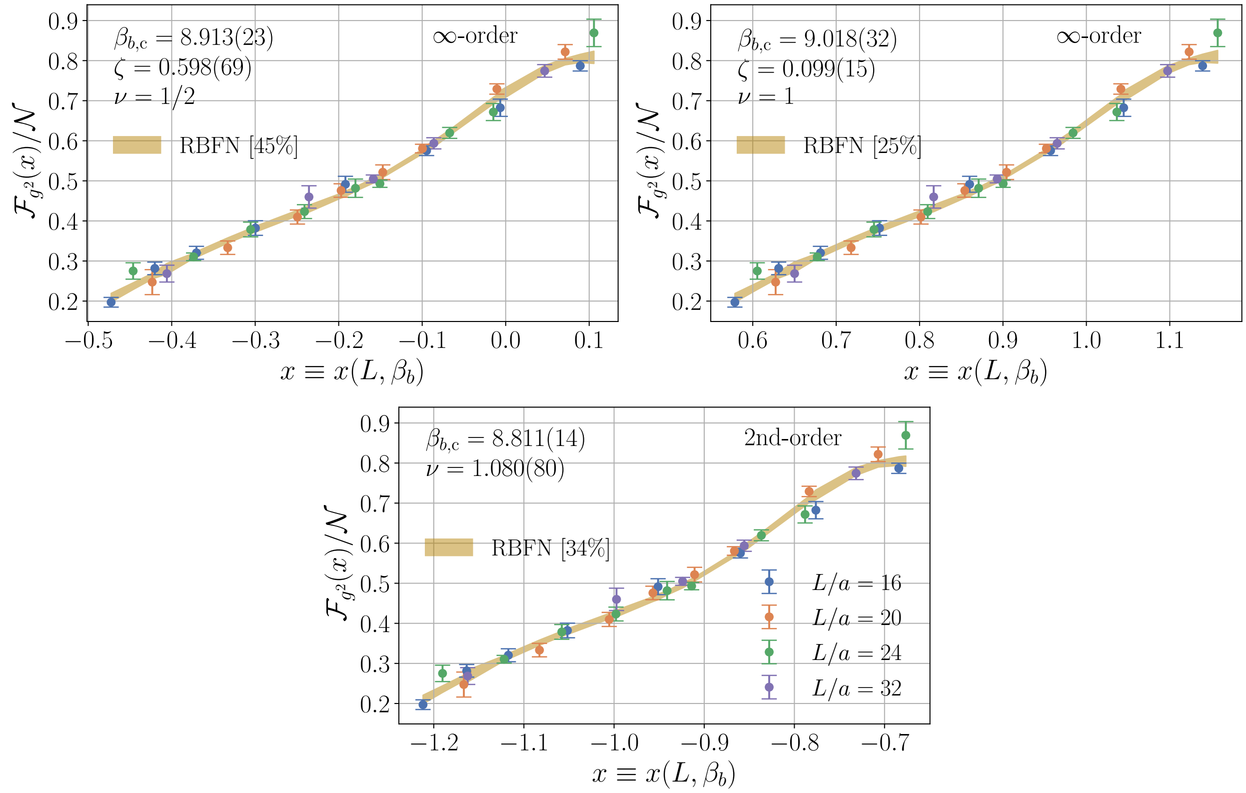

Figure 1.4: Result of finite size scaling analysis of the gradient flow renormalized coupling g^2 for the $N_{\rm f} = 8$ system using a radial basis function network. ∞-order scaling with $\nu = 1/2$ is shown in top left panel and $\nu = 1$ in the top right panel. 2nd-order scaling shown in bottom panel. The prediction for the scaling function $\mathcal{F}_{g^2}$ from the radial basis function (RBFN) is indicated by a gold band with the width of the band indicating the error. Scaling function is normalized by $\mathcal{N} = 128\pi^2/(3N^2 - 3)$ $(N = 3)$ for visualization purposes. Data entering the curve collapse analysis is indicated by multicolored error bars for $L/a = 16$ (blue), 20 (orange), 24 (green), and 32 (purple).

$\gamma_g^\star = 0.199(32)$. I find that the estimate for $\gamma_g^\star$ is consistent with the literature at the 1σ-2σ level [106, 186, 321] (see Fig. 5.8).

1.3.2 Finite size scaling with neural networks

The renormalized coupling g^2 is a scaling variable in the vicinity of an RG fixed point. Therefore, it can be used to probe the properties of phase transitions that occur as the bare gauge coupling $\beta_b \equiv 6/g_0^2 \to \beta_{bc}$ in infinite volume. In a finite volume, the RG coupling $g^2(L, \beta_b) \equiv g^2(\mu; L)$

with $\mu \equiv c/L$ for some $c \in (0, 1/2]$ is expected to scale with β_b and L in the vicinity of β_{bc} as

$$g^2(L, \beta_b) = \mathcal{F}_{g^2}(x) \qquad (\beta_b \approx \beta_{bc}),$$

where

$$x(L, \beta_b) = (\beta_b/\beta_{bc} - 1)L^{1/\nu} \qquad \text{(2nd-order)}$$

for a 2nd-order phase transition or

$$x(L, \beta_b) = L \exp\Big[-\zeta(\beta_b/\beta_{bc} - 1)^{-\nu}\Big] \qquad (\infty\text{-order})$$

for an ∞-order phase transition (like the BKT transition of the XY model). First-order phase transitions also scale as $x = (\beta_b/\beta_{bc} - 1)L^{1/\nu}$, but with $\nu = 1/d$ [96, 120]. The function $\mathcal{F}_{g^2}$ is a *universal scaling function*. It is similar to the scaling function of Eqn. 1.21 for the two-point correlation function of some operator in a mass-deformed conformal field theory. The parameters β_{bc}, ν and ζ of the phase transition can be extracted by requiring that $g^2(L, \beta_b) = F_{g^2}(x)$ forms a unique a one-dimensional curve in x. Once the correct critical parameters have been identified, the individual curves of $g^2(L, \beta_b)$ in β_b at fixed L *collapse* onto $\mathcal{F}_{g^2}(x)$; hence, this method is referred to as *curve collapse*. One of the particularly tricky road bumps encountered by curve collapse analyses is the need to estimate the scaling function $\mathcal{F}_{g^2}$, which is typically done by replacing it with some parametric ansatz; i.e., a polynomial. The parameters of the parametric ansatz for $\mathcal{F}_{g^2}$ are then estimated simultaneously with the critical parameters as part of the curve collapse. It is desirable to have on hand a collection of expressive functions that are capable of representing $\mathcal{F}_{g^2}$. I show in Chapter 6 that a type of single-layer artificial neural network known as a radial basis function network (RBFN) can be very useful for this purpose, as RBFNs are specially-designed for function approximation. I test the RBFN-based curve collapse on the finite-temperature phase transition of various two-dimensional spin systems in Chapter 6. I then apply the same curve collapse analysis to the zero-temperature phase transition of the $N_{\mathrm{f}} = 8$ system into an SMG-like phase in Chapter 7

In Fig. 1.4, I summarize the result of the $N_{\mathrm{f}} = 8$ curve collapse for ∞-order scaling with $\nu = 1/2, 1$ (top left and right panels, respectively) and 2nd-order scaling (bottom panel). In each panel of Fig. 1.4, the scaling function $\mathcal{F}_{g^2}$ is rescaled by $\mathcal{N} \equiv 128\pi^2/(3N^2 - 3) \approx 53.6$ for the purposes of visualization and set $c = 0.45$. The prediction for $\mathcal{F}_{g^2}$ from the radial basis function network is shown as a gold band. The width of the band indicates the statistical error. Individual volumes at different β_b are shown as colored markers with error bars. The prediction for β_{bc}, ν and ζ is shown in the top left corner of each panel. Though it is not shown in Fig. 1.4, 1st-order scaling is strongly disfavored compared to 2nd-order and ∞-order scaling. For reasons that I discuss in detail in Chapter 7, this does not necessarily imply that the transition is not 1st-order. Outside of 1st-order scaling, no one scaling is unambiguously preferred over another according to the simple statistical tests that are applied to the curve collapse analysis of each scaling scenario (χ^2 and marginal likelihood). If the correlation length, defined in Chapter 2.2.2, can be determined, then it will be possible determine whether the curve collapse analysis strongly disfavoring 1st-order scaling hold water. If so, then there is a real possibility that the $N_{\mathrm{f}} = 8$ system could possesses an SMG phase with a consistent continuum limit.[4]

1.4 Road map

The core chapters of this thesis are Chapters 4-7. Readers that are interested only in the main results of this thesis are advised to read just those chapters. Subleading developments are discussed in Chapter 8. Relevant background material is given in Chapters 2-3. Below, I provide a road map for navigating the material of this thesis.

(1) **Chapter 2 (Wilsonian renormalization group)**: I introduce the Wilsonian renormalization group tools that are used throughout this thesis, such RG transformations, RG β-functions, running anomalous dimensions, and finite size scaling. The content of this chapter is relevant to Chapters 4-8. It is also used in Chapter 3.4 to discuss the notion of continuum limits.

---

[4] Pending confirmation that the SMG-like phase observed by Ref. [170] is a legitimate SMG phase.

(2) **Chapter 3 (Lattice gauge theory)**: I introduce the the basic tools and language of lattice gauge theory with an emphasis on staggered fermions. The content of this chapter is relevant to Chapters 4, 5, 7, and 8.

(3) **Chapter 4 ($\boldsymbol{N_{\mathrm{f}} = 0}$)**: I first introduce the *continuous β-function method* (CBFM) for calculating the infinite-volume RG β-function from gradient flow. I then apply the CBFM to the $N_{\mathrm{f}} = 0$ system. The Λ-parameter of the $N_{\mathrm{f}} = 0$ system is calculated and a method for matching β-functions non-perturbatively is introduced. The CBFM is also deployed in Chapters 5, 7, and 8.3.

(4) **Chapter 5 ($\boldsymbol{N_{\mathrm{f}} = 12}$)**: I calculate the continuum gradient flow β-function for the $N_{\mathrm{f}} = 12$ system using the CBFM of Chapter 4. From the β-function, I calculate the leading irrelevant critical exponent $\gamma_g^{\star}$. The Bayesian model averaging procedure introduced in this chapter is deployed in Chapter 7.

(5) **Chapter 6 (Finite size scaling with neural networks)**: I both introduce the radial-basis-function-network (RBFN-based) curve collapse method and apply it to curve collapse analyses of the q-state Potts model and p-state clock model for $q = 2, 3$ and $p = 4, \infty$. I speculate on other applications of radial basis function networks to lattice field theory analyses and exemplify one such method by determining the critical temperature of the ∞-state clock (XY) model from a direct interpolation of helicity modulus using an RBFN. The RBFN-based curve collapse method is used in Chapter 7.

(6) **Chapter 7 ($\boldsymbol{N_{\mathrm{f}} = 8}$)**: I first analyze the zero-temperature phase transition of the $N_{\mathrm{f}} = 8$ system using the RBFN-based curve collapse method of Chapter 6. I then calculate the continuum gradient flow β-function using the CBFM of Chapter 4. Error estimates for the continuum β-function are improved by deploying the Bayesian model averaging procedure introduced in Chapter 5.

(7) **Chapter 8 (Other Developments)**: I first discuss the high-performance code development

that went into creating the `Quantum EXpressions` (`QEX`-based) `qex_staghmc` suite [284]. `qex_staghmc` is used for most of the large-volume simulations and gradient flow measurements of Chapters 5 and 7. I then discuss the code development that went into creating the `SwissFit` library, which implements many of the Bayesian-statistics-inspired analysis tools that are deployed in Chapters 5-7. I end by briefly discussing a calculation of operator anomalous dimensions in the $N_{\text{f}} = 10$ system using domain wall fermion simulations.

(8) **Chapter 8 (Conclusions)**: Summarizes the main technical results of this thesis and discusses future research directions.

Chapter 2

The Wilsonian renormalization group

> *In The Renormalization Group method you take a structure you don't understand and convert it to another structure you don't understand. You keep doing it until you finally understand.*
>
> Michael Berry [40]

Broadly speaking, the *Wilsonian renormalization group* (RG), or just "*renormalization group*", refers to a formal collection of ideas and techniques that aid in the investigation of scale-dependent properties of a physical system. Though the focus of this thesis is on gauge-fermion systems, this chapter aims to treat the renormalization group in as broad of terms as possible. In doing so, we can appreciate what the renormalization group tells us about field theories on a lattice in general and we can directly apply what we have learned to other systems throughout this thesis.

To this end, let us consider a generic collection of fields $\{\varphi_{\mathfrak{f}}(n)\}$ indexed by $\mathfrak{f}$ that are defined over a d-dimensional hypercubic lattice with lattice spacing a. For notational simplicity, we suppress any additional indices that each $\varphi_{\mathfrak{f}}$ may possess. The fields in $\{\varphi_{\mathfrak{f}}(n)\}$ interact according a classical action $\mathcal{S}[\varphi]$ that is a sum of terms

$$\mathcal{S}[\varphi] = \sum_i K_i \mathcal{O}_i[\varphi], \tag{2.1}$$

where each $\mathcal{O}_i$ respects the symmetries of $\mathcal{S}[\varphi]$ and each K_i is a *coupling constant* that renders the combination $K_i \mathcal{O}_i[\varphi]$ dimensionless. Quantum fluctuations of this system are captured by the partition function

$$\mathcal{Z}_\varphi = \int \prod_{\mathfrak{f}} [\mathrm{d}\varphi_{\mathfrak{f}}] \ \exp\big(\text{-}\mathcal{S}[\varphi]\big), \tag{2.2}$$

from which we define the vacuum expectation value of a generic observable $\mathcal{O}[\varphi]$ as

$$\langle \mathcal{O}[\varphi] \rangle = \mathcal{Z}_\varphi^{-1} \int \prod_{\mathfrak{f}} [\mathrm{d}\varphi_{\mathfrak{f}}] \ \mathcal{O}[\varphi] \exp\big(\text{-}\mathcal{S}[\varphi]\big). \tag{2.3}$$

The key theoretical tool of the renormalization group is the *renormalization group transformation*, which systematically removes irrelevant short-distance fluctuations while preserving physical long-distance properties of the system under investigation. One can learn quite a bit about the properties of field theories in general by understanding how the couplings K_i vary under repeated renormalization group transformations.

2.1 Renormalization group transformations

A generic renormalization group proceeds via the following two step procedure; see Refs. [222, 230, 255, 356] for more details and applications to specific systems.

(1) **Coarse grain:** It is common to define the coarse graining step of an RG transformation either in real space or in wave number (dual) space. Both achieve the task of eliminating short-distance fluctuations.

- **Real space:** Eliminate short-distance fluctuations by defining a new set of *coarse-grained* fields $\{\overline{\varphi}_{\mathfrak{f}}\}$ that are some local average over the original fields $\{\varphi_{\mathfrak{f}}\}$. If the average is performed over fields that fluctuate on scales less than or equal to ba, then the new coarse-grained fields live on a lattice with lattice spacing ba. The local average should be defined in such a way that the symmetries of the original action are preserved; this makes implementing a real space coarse-graining step in a gauge-fermion system quite difficult, but not impossible [99, 194, 204].
- **Dual space:** Eliminate short-distance fluctuations by integrating out wave number

fluctuations in the $\pi/ba \leq |k| \leq \pi/a$ shell. More precisely, let the Fourier-transformed fields be $\{\tilde{\varphi}_{\mathfrak{f}}(k)\}$. We can then eliminate modes in the $\pi/ba \leq |k| \leq \pi/a$ shell by integrating over them in the partition function. This defines a new action $\overline{\mathcal{S}}\big[\overline{\tilde{\varphi}}_{\mathfrak{f}}\big]$ for the low wave number fields $\big\{\overline{\tilde{\varphi}}_{\mathfrak{f}}(k)\big\}$ that is related to the original action $\mathcal{S}[\tilde{\varphi}_{\mathfrak{f}}]$ as

$$-\overline{\mathcal{S}}\big[\overline{\tilde{\varphi}}_{\mathfrak{f}}\big] = \log \int \prod_{\mathfrak{f}} \prod_{\pi/ba \leq |k| \leq \pi/a} \big[\mathrm{d}\tilde{\varphi}_{\mathfrak{f}}(k)\big] \;\exp\big(\text{-}\mathcal{S}[\tilde{\varphi}]\big). \tag{2.4}$$

The inverse Fourier transformed low wave number fields $\{\overline{\varphi}_{\mathfrak{f}}\}$ now live on a new lattice with lattice spacing ba. For complicated systems, coarse graining in wave number space can be more efficient computationally than coarse graining in real space.

(2) **Renormalization:** If necessary, restore the contrast of the fields of the original lattice by rescaling the coarse-grained fields as $\varphi'_{\mathfrak{f}} = Z_{\mathfrak{f}}\overline{\varphi}_{\mathfrak{f}}$.

Depending on your familiarity with RG, you may now be wondering why I have not included the *rescaling* step, whereby all dimensionful scales l are rescaled as $l \to l/b$. The latter step is usually included in treatments of RG that are geared toward condensed matter physics, where the lattice spacing is held fixed and the rescaling step is necessary to restore the original lattice spacing (resolution) [222, 230, 356]. However, the perspective in high-energy physics is quite different. The lattice spacing, or *ultraviolet cutoff*, changes under an RG transformation, but all dimensionful scales below the cutoff are held *fixed* [70]. As far as dimensionless quantities are concerned, this difference in perspectives does not matter. However, it is worth keeping in mind because it can lead to a whole array of confusions, as it had for me.

For the renormalization group transformation to preserve the long-distance properties of the original set of fields, it must be the case that it preserves the partition function; that is,

$$\mathcal{Z}_{\varphi} = \int \prod_{\mathfrak{f}} [\mathrm{d}\varphi_{\mathfrak{f}}] \;\exp\big(\text{-}\mathcal{S}[\varphi]\big) = \int \prod_{\mathfrak{f}} [\mathrm{d}\varphi'_{\mathfrak{f}}] \;\exp\big(\text{-}\mathcal{S}'[\varphi']\big) = \mathcal{Z}_{\varphi'}. \tag{2.5}$$

This condition imposes a set of constraints on the steps of the renormalization group transformation.

The action $\mathcal{S}'[\varphi']$ is composed of two sets of terms. The first set of terms are the same as the original action $\mathcal{S}[\varphi]$, but they are composed of the renormalized fields $\{\varphi'_{\mathfrak{f}}(n)\}$ and they are expressed in terms of a new set of *renormalized couplings* $K'_i(\{K_i\})$. The second set of terms are generated by the coarse-graining procedure of the renormalization group transformation (if they were not already present in $\mathcal{S}[\varphi]$). There are usually an infinite number of such terms. Due to the condition of Eqn. 2.5, iterating the renormalization group transformation induces a flow on the infinite-dimensional space of couplings $\boldsymbol{K}' \equiv (K'_1, K'_2, ...)$ defined by each $\mathcal{S}'[\varphi']$. This flow is the *renormalization group flow.*

2.2 Renormalization group flow

Consider now a map $\mathbf{R}_b$ that takes any initial set of couplings $\boldsymbol{K}_0$ to a new set of couplings $\boldsymbol{K}' = \mathbf{R}_b(\boldsymbol{K}_0)$ via an RG transformation. The coarse graining step of the RG transformation ensures that $\mathbf{R}_b$ does not have an inverse. It must also be the case that [230, 255]

$$\mathbf{R}_{b'}(\boldsymbol{K}') = \mathbf{R}_{b'b}(\boldsymbol{K}_0). \tag{2.6}$$

This where I must insert the obligatory statement that Eqn. 2.6 combined with $\mathbf{R}_b$ lacking an inverse implies that a renormalization group transformation does not actually form a group. It forms a semi-group. Hence, the name *renormalization group* is a misnomer; such is the way of science. We can think about the flow that is induced by $\mathbf{R}_b$ using the language of dynamical systems. Namely, the topology of the flow over the entire set of initial couplings $\boldsymbol{K}_0$ is completely specified by the action of $\mathbf{R}_b$ on $\boldsymbol{K}$ that are within the local vicinity of a *fixed point* $\boldsymbol{K}^\star$ of $\mathbf{R}_b$, which is defined by

$$\mathbf{R}_b\big(\boldsymbol{K}^\star\big) = \boldsymbol{K}^\star. \tag{2.7}$$

If we know what the local behavior of couplings $\boldsymbol{K}_0$ starting near every fixed point $\boldsymbol{K}^\star$ looks like under an RG transformation, then we can qualitatively determine what the global topology of the flow must look like.

2.2.1 The β-function and characterizing fixed points by their local topology

Let us take b to be a dimensionless quantity that can be varied continuously. This is most easily realized by a RG transformation in dual space, though it is possible to define RG transformations with continuous b in real space. We can cast the flow of the couplings K_i under an infinite number of infinitesimally small RG transformations in terms of a differential equation

$$-b^2 \frac{\mathrm{d}K_i(b)}{\mathrm{d}b^2} \equiv \beta_i(\boldsymbol{K}), \tag{2.8}$$

where $\beta_i(\boldsymbol{K})$ is referred to as the renormalization group β-function of coupling K_i. As it contains all of the information about the RG flow, it is *only* a function of the couplings $\boldsymbol{K}$. In Appendix A, I present a perspective of what is to follow in a form that may be more comfortable to folks in the condensed matter community. If you have never been introduced to the concepts of the renormalization group, I recommend reading Appendix A before continuing. Fixed points $\boldsymbol{K}^\star$ correspond to points in $\boldsymbol{K}$-space where the RG β-function is zero; that is,

$$\beta_i(\boldsymbol{K}^\star) = 0 \tag{2.9}$$

for all i. Let $\delta\boldsymbol{K} \equiv \boldsymbol{K} - \boldsymbol{K}^\star$. Then, for $\delta K_i / K_i^\star \ll 1$, we can approximate the RG flow by linearizing the β-function as

$$\beta_i(\boldsymbol{K}) = \mathbf{B}\delta\boldsymbol{K} + \mathrm{O}\big(\delta\boldsymbol{K}^2\big), \tag{2.10}$$

where $\mathbf{B}$ is a linear operator with components

$$\mathbf{B}_{ij} = \left.\frac{\partial\beta_i(\boldsymbol{K})}{\partial K_j}\right|_{\boldsymbol{K}=\boldsymbol{K}^\star}. \tag{2.11}$$

The principal axes of $\mathbf{B}$ completely characterize the local topology of the RG flow about $\boldsymbol{K}^\star$. Denoting $e_{\mathrm{L}}^{(\alpha)}$ as a normalized left eigenvector (principal axis) of $\mathbf{B}$, we have

$$\mathbf{e}_{\mathrm{L}}^{(\alpha)}\mathbf{B} = \mathbf{e}_{\mathrm{L}}^{(\alpha)}\lambda^{(\alpha)}. \tag{2.12}$$

Along the principal axes, the RG equation is decoupled with β-function

$$-b^2\frac{\mathrm{d}K_\alpha(b)}{\mathrm{d}b^2} = \beta_\alpha(K_\alpha) \approx \lambda^{(\alpha)}\delta K_\alpha + \mathrm{O}\big(\delta\boldsymbol{K}^2\big) \tag{2.13}$$

within the vicinity of $\boldsymbol{K}^\star$. The component K_α is obtained from $\boldsymbol{K}$ via the projection $K_\alpha = \mathbf{e}_{\mathrm{L}}^{(\alpha)}\boldsymbol{K}$; δK_α is referred to as a *scaling variable*. The sign of $\lambda^{(\alpha)}$ indicates whether K_α flows into the fixed point (*irrelevant*), out of the fixed point (*relevant*), or if higher-order in terms in $\delta\boldsymbol{K}$ are needed to determine if K_α flows into or out of the fixed point (*marginally irrelevant* or *marginally relevant*). The eigenvalues $\lambda^{(\alpha)}$ are *universal* in the sense that they do not depend on the RG transformation used to calculate them. This is our first instance of *universality* coming into play. Physical systems with the same RG eigenvalues $\lambda^{(\alpha)}$ are said to belong to the same *universality class*; here, equivalent "physical systems" can mean different lattice discretizations of a continuum field theory.

2.2.2 Characterizing fixed points by the correlation length

Aside from the local topology the RG flow about a fixed point $\boldsymbol{K}^\star$, fixed points are also characterized by the minimum length $\tilde{\xi}(\boldsymbol{K})$ (in units of the lattice spacing) over which the fields $\varphi_{\mathfrak{f}}(n)$ at lattice site n are correlated with another lattice site m that is separated from n by a vector of length $\tilde{\xi}(\boldsymbol{K})$. After an RG transformation, the correlation length in units of the lattice spacing $\tilde{\xi}(\boldsymbol{K})$ at $\boldsymbol{K} = \boldsymbol{R}_b(\boldsymbol{K}_0)$ is related to $\tilde{\xi}(\boldsymbol{K}_0)$ at $\boldsymbol{K}_0$ before the RG transformation as

$$\tilde{\xi}(\boldsymbol{K}) = \tilde{\xi}(\boldsymbol{K}_0)/b. \tag{2.14}$$

Therefore, once the fixed point $\boldsymbol{K}^\star$ is “reached”, the correlation length transforms as

$$\tilde{\xi}(\boldsymbol{K}^\star) = \tilde{\xi}(\boldsymbol{K}^\star)/b \tag{2.15}$$

under an RG transformation. The only way that Eqn. 2.15 can be true is if $\tilde{\xi}(\boldsymbol{K}^\star)$ is either zero or infinite. A fixed point is called a *critical fixed point* if $\tilde{\xi}(\boldsymbol{K}^\star)$ is infinite; otherwise, it is called a *trivial fixed point*. The collection of couplings that flow *into* a critical fixed point form what is called the *critical manifold* or *critical surface*. Since the correlation lengths up to the fixed point are related to one another by Eqn. 2.14, it must be the case that each of the correlation lengths on the critical surface are infinite. In Appendix B, I use the Ising model to describe how the global topology of an RG flow can be characterized by the local topology and correlation length at all fixed points of an RG transformation.

2.3 Connection to β-functions in high energy physics

At this point, you may be wondering how the notion of a renormalized coupling in this chapter is related to the renormalized coupling that you may be familiar with from the Review for Particle Physics, the Flavor Lattice Averaging Group report, or whatever your favorite source of information about high energy physics is [6, 381]. In the high energy physics literature on quantum chromodynamics (QCD), it is common to *define* a renormalized coupling in terms of some observable $O(\mu)$ at an energy scale μ that is related to the renormalized coupling in the $\overline{\mathrm{MS}}$ scheme $g^2_{\overline{\mathrm{MS}}}(\mu) \equiv 4\pi\alpha_{\mathrm{s}}(\mu)$ in perturbation theory as

$$O(\mu) \sim a_1\alpha_{\mathrm{s}}(\mu) + a_2\alpha_{\mathrm{s}}(\mu)^2 + \mathrm{O}\big(\alpha_{\mathrm{s}}^4\big) \tag{2.16}$$

for constants a_1, a_2, and so forth. For the perturbative calculations that take up much of the particle physics literature, such a practical definition is perfectly reasonable and allows for important Standard Model calculations to be carried out at energy scales that are relevant to the Large Hadron

Collider (LHC). I delve into calculations of α_{s} in slightly more detail in Chapter 4.6. The connection between the procedure of defining a renormalization coupling by relations like Eqn. 2.16 and the RG-based definition that I have discussed already in this chapter is subtle and, from the author's perspective, not well-understood. Nonetheless, I shall make an argument that will convince some, but not all. Given that renormalization is an infamously difficult subject to understand and different sub-disciplines have varying notions of RG that are tailored to fit their idea of what RG is useful for, both reactions are quite frankly within reason.

Suppose that I have defined a renormalization group transformation that preserves the partition function $\mathcal{Z}_{\varphi}$ of some physical system as in Sec. 2.1. The expectation value of an RG-transformed observable $\mathcal{O}'(b)$ in the "bare system" is related to the expectation value of $\mathcal{O}$ before the RG transformation in the "renormalized system" as [74, 256]

$$\langle \mathcal{O}'(b) \rangle_{\boldsymbol{K}_0} = Z_{\mathcal{O}}(b) \langle \mathcal{O} \rangle_{\boldsymbol{K}(b)}, \tag{2.17}$$

where $\langle \cdots \rangle_{\boldsymbol{K}_0}$ is an expectation value in the bare system, $\langle \cdots \rangle_{\boldsymbol{K}(b)}$ is an expectation value in the renormalized system, $Z_{\mathcal{O}}(b)$ is the wave function renormalization of $\mathcal{O}$, and I have neglected any potential operator mixing. Note that Eqn. 2.17 states that expectation values of RG blocked observables $\overline{O}(b)$ in the bare system are equivalent to expectation values of unblocked observables in the renormalized system; i.e.,

$$\left\langle \overline{\mathcal{O}}(b) \right\rangle_{\boldsymbol{K}_0} = \langle \mathcal{O} \rangle_{\boldsymbol{K}(b)}, \tag{2.18}$$

as $\mathcal{O}'(b) = Z_{\mathcal{O}}(b)\overline{\mathcal{O}}(b)$ by definition [73, 157, 347].

Suppose now that the RG transformation that we are performing has a fixed point. Additionally, suppose that there exists some set of relevant observables $\{\mathcal{O}_\alpha\}$ with $Z_{\mathcal{O}_\alpha}(b) = 1$ about the fixed point; in other words,

$$\langle \mathcal{O}'(b) \rangle_{\boldsymbol{K}_0} = \langle \mathcal{O} \rangle_{\boldsymbol{K}(b)}. \tag{2.19}$$

If the number of $\mathcal{O}_\alpha$ is equal to the number of relevant couplings of the fixed point, then Eqn. 2.19

implies that there is a one-to-one relation between the set $\{O_\alpha\}$ and the relevant couplings $\{K_\alpha\}$ as far as mapping out the relevant directions of the RG flow about the fixed point is concerned [258]. As such, the $\{\mathcal{O}_\alpha\}$ can be used to define a set of renormalized couplings

$$g_\alpha^2(b) \sim \langle \mathcal{O}'_\alpha(b) \rangle_{\boldsymbol{K}_0} \tag{2.20}$$

and corresponding β-functions

$$\beta_\alpha\left(g_\alpha^2\right) \equiv -b^2 \frac{\mathrm{d}g_\alpha^2(b)}{\mathrm{d}b^2} \tag{2.21}$$

that characterize the relevant directions of the RG flow. To the best of the author's knowledge, there is no reason to expect $Z_{\mathcal{O}}(b)$ to diverge away from unity as the flow evolves away from the local vicinity of the fixed point; hence, the $\left\{g_\alpha^2(b)\right\}$ should map out the entire RG flow that emerges from the fixed point. Moreover, Eqn. 2.19 ensures that the couplings $\left\{g_\alpha^2(b)\right\}$ are also regular if the couplings of the Wilsonian effective action are regular; this is always the case. As such, if the RG flow emerging out of one fixed point runs into another fixed point, the relevant couplings $\left\{g_\alpha^2(b)\right\}$ of the former fixed point must evolve into irrelevant couplings of the latter fixed point. There are often many different choices for the $\{\mathcal{O}_\alpha\}$ from which one can define $\{g_\alpha^2(b)\}$ to map out the RG flow. As far as calculations for QCD are concerned, defining a running coupling in terms of some observable $O(\mu)$ about the ultraviolet fixed point of QCD (see Sec. 3.4), where asymptotic freedom reigns supreme, is similar to the procedure of picking out a coupling $g^2(\mu)$ that maps out the RG flow of QCD. Note, however, that perturbatively-defined RG couplings like those of Eqn. 2.16 can lead to certain pathologies where the coupling appears to be well-defined pertubatively, but it does not actually track the RG flow outside of the local vicinity of the fixed point. In such cases, one is advised to refer to the non-perturbative definition motivated by Eqn. 2.19. Even so, because there are no proofs of any of the statements that I have made about Wilsonian RG in this chapter, it is possible that even the couplings based on Eqn. 2.19 break down. To this end, high energy physics is really in need of a constructive approach to non-perturbative quantum field theory and renormalization. Nonetheless, one has to start somewhere and check that what they're doing along the way is sane.

2.4 Running operator anomalous dimensions

The correlation function of an operator $\mathcal{O}(n)$ in the bare system $\boldsymbol{K}_0$ is related to the expectation value of the same observable $\mathcal{O}(n)$ in the renormalized (*blocked*) system $\boldsymbol{K}$ as [74]

$$\langle \mathcal{O}(n)\mathcal{O}(0)\rangle_{\boldsymbol{K}_0} = Z^2_{\mathcal{O}}(b)\langle \mathcal{O}(n_b)\mathcal{O}(0)\rangle_{\boldsymbol{K}(b)}, \tag{2.22}$$

where $n_b \equiv n/b$. Eqn. 2.22 is the Wilsonian RG version of the Callan-Symanzik equation [65, 349, 350]. Combining Eqns. 2.17 and Eqn. 2.22 yields the following string of identities ($n_b \gg 1$)

$$\langle \mathcal{O}(n)\mathcal{O}(0)\rangle_{\boldsymbol{K}_0} = \langle \mathcal{O}'(n_b;b)\mathcal{O}'(0;b)\rangle_{\boldsymbol{K}_0} = Z^2_{\mathcal{O}}(b)\big\langle \overline{\mathcal{O}}(n_b;b)\overline{\mathcal{O}}(0;b)\big\rangle_{\boldsymbol{K}_0}, \tag{2.23}$$

where $\mathcal{O}'(n_b;b)$ and $\overline{\mathcal{O}}(n_b;b)$ denote the renormalized and blocked $\mathcal{O}(n)$, respectively. The first equality states that correlations in $\mathcal{O}$ are identical to those of $\mathcal{O}'$ as far as fluctuations in the bare system at scales $n_b \gg 1$ are concerned. In other words, renormalized observables preserve the long-distance properties of their unrenormalized counterparts. Eqn. 2.23 also hints at a method for determining $Z_{\mathcal{O}}(b)$ in terms of the expectation value of the RG blocked observables $\overline{\mathcal{O}}(n_b;b)$ in the bare system

$$Z^{-2}_{\mathcal{O}}(b) = \frac{\big\langle \overline{\mathcal{O}}(n_b;b)\overline{\mathcal{O}}(0;b)\big\rangle_{\boldsymbol{K}_0}}{\langle \mathcal{O}(n)\mathcal{O}(0)\rangle_{\boldsymbol{K}_0}}. \tag{2.24}$$

Note again that Eqn. 2.24 requires $n_b \gg 1$. Eqn. 2.24 can be taken as a non-perturbative definition of the wave function renormalization for $\mathcal{O}$ for the RG transformation (scheme) that takes $\boldsymbol{K}_0 \to \boldsymbol{K}(b)$.

The wave function renormalization $Z_{\mathcal{O}}$ is not physical in the sense that its running depends upon the RG transformation that it is defined from. However, like the renormalized coupling, its behavior in the vicinity of a RG fixed point $\boldsymbol{K}^\star$ is universal. Supposing then that the abstract RG flow we are working with exhibits an RG fixed point at $\boldsymbol{K}^\star$, the wave function renormalization runs with b in the vicinity of $\boldsymbol{K}^\star$ as

$$Z_{\mathcal{O}}(b) \sim b^{-\Delta_{\mathcal{O}}} \qquad (\boldsymbol{K} \to \boldsymbol{K}^\star), \tag{2.25}$$

where

$$\Delta_{\mathcal{O}} \equiv d_{\mathcal{O}} + \gamma^\star_{\mathcal{O}} \tag{2.26}$$

is the *scaling dimension*, $d_{\mathcal{O}}$ is the *canonical dimension*, and $\gamma^\star_{\mathcal{O}}$ is the *anomalous dimension* of $\mathcal{O}$. The anomalous dimension $\gamma^\star_{\mathcal{O}}$ is universal and characterizes various properties of the system close to or at $\boldsymbol{K}^\star$; it is a critical exponent. If $\mathcal{O}$ is dimensionless, as is the case for any observable that is measured from lattice simulations before scale setting, and the RG flow is nonlinear, then Eqns. 2.24-2.26 can be used to define a *running operator anomalous dimension* in terms of $Z_{\mathcal{O}}$ as

$$b\frac{\mathrm{d}}{\mathrm{d}b}\log Z_{\mathcal{O}}(b) \equiv \gamma_{\mathcal{O}}\big(\boldsymbol{K}\big). \tag{2.27}$$

Furthermore, if one defines [74, 177]

$$R_{\mathcal{O}}(b) \equiv \frac{\big\langle \overline{\mathcal{O}}(n_b; b)\overline{\mathcal{O}}(0; b)\big\rangle_{\boldsymbol{K}_0}}{\langle \mathcal{O}(n)\mathcal{O}(0)\rangle_{\boldsymbol{K}_0}}, \tag{2.28}$$

then Eqn. 2.24 implies that

$$\gamma_{\mathcal{O}}\big(\boldsymbol{K}\big) = -2b\frac{\mathrm{d}}{\mathrm{d}b}\log R_{\mathcal{O}}(b). \tag{2.29}$$

In any case,

$$\gamma_{\mathcal{O}}\big(\boldsymbol{K}\big) \to \gamma^\star_{\mathcal{O}}. \tag{2.30}$$

If the RG flow is linear, then $\gamma_{\mathcal{O}}$ can be calculated similarly by extending the definition of $R_{\mathcal{O}}(b)$. I will use the ideas of this section, specifically Eqn. 2.29, in Chapter 8.3.

2.5 Phase transitions and finite size scaling

A *phase transition* for a system living on a lattice is generally characterized by a non-analyticity in the *reduced* Helmholtz free energy per site, defined as

$$A_{\varphi}(\boldsymbol{K}) \equiv -N^{-d} \log \mathcal{Z}_{\varphi}, \tag{2.31}$$

at some critical value $\boldsymbol{K} = \boldsymbol{K}_{\mathrm{c}}$ of the couplings $\boldsymbol{K}$. For notational convenience, I have defined $N \equiv L/a$, such that the volume in units of the lattice spacing is N^d. Such non-analyticities exist only in the infinite volume limit. The phenomenology of any particular phase transition depends strongly on its *order*. Discontinuities in any of the first-order partial derivatives of $A_{\varphi}(\boldsymbol{K})$ are associated with *first-order* transitions, which exhibit phenomena such as phase coexistence, hysteresis, and the release of latent heat. *Second*-order phase transitions are associated with discontinuities in any of the second-order derivatives of $A_{\varphi}(\boldsymbol{K})$ and most notably exhibit a diverging correlation length, leading to phenomena such as scale invariance and power law scaling. The continuum limit of the gauge-fermion systems that we have been exploring in this thesis is a second-order phase transition. Non-analyticities in $A_{\varphi}(\boldsymbol{K})$ that are due to an *essential singularity* are categorized as ∞-*order* phase transitions. The correlation length at an infinite-order phase transition also diverges; however, the manner in which it diverges is not characterized by a power law like it is for a second-order phase transition. First-order phase transitions are fascinating in their own right; however, they are extremely difficult to simulate using conventional canonical Monte Carlo techniques due to hysteresis.[1] As such, I will only explore 2nd- and infinite-order phase transitions in this chapter, which I collectively refer to as *continuous phase transitions*. All statements that follow hold for continuous phase transitions unless stated otherwise.

According to the definition for an RG transformation that I gave in Sec. 2.1, $A_{\varphi}(\boldsymbol{K})$ transforms

---

[1] One way to get around the hysteresis-induced problems that canonical algorithms experience when simulating first-order phase transitions is to use the beautiful *multicanonical* sampling technique; see Refs. [33, 38, 39].

under an RG transformation ($\boldsymbol{K} \to \boldsymbol{K'}$) as

$$A_\varphi(\boldsymbol{K}) = F_\varphi(\boldsymbol{K}) + b^{-d} A_\varphi(\boldsymbol{K'}) \qquad \text{(infinite volume)}, \tag{2.32}$$

where $F_\varphi(\boldsymbol{K})$ is a "constant" that originates from averaging over the short-distance ($a \leq l \leq ba$) degrees of freedom in the coarse graining step; it does not contribute any singularities to $A_\varphi(\boldsymbol{K})$ [70]. Therefore, it is common to define *singular* part of $A_\varphi(\boldsymbol{K})$ by its behavior under an RG transformation:

$$A_\varphi^{(\mathrm{s})}(\boldsymbol{K}) = b^{-d} A_\varphi^{(\mathrm{s})}(\boldsymbol{K'}) \qquad \text{(infinite volume)}. \tag{2.33}$$

Eqn. 2.33 is the famous *scaling hypothesis*, first posited on phenomenological grounds by Widom in 1965 and put to work soon after for the Ising system by Kadanoff [216, 371]. In a finite volume, the infinite volume non-analyticities in $A_\varphi^{(\mathrm{s})}(\boldsymbol{K})$ are smoothed out. From the perspective of RG, this is because $N^{-1} = a/L$ acts as a relevant variable that pushes the system off of any critical surface. The transformation for the singular part of the free energy is modified in a finite volume as [69, 70]

$$A_\varphi^{(\mathrm{s})}\left(\boldsymbol{K}, N^{-1}\right) = b^{-d} A_\varphi^{(\mathrm{s})}\left(\boldsymbol{K'}, bN^{-1}\right) \qquad \text{(finite volume)}. \tag{2.34}$$

From Eqn. 2.34, all of the well-known FSS relations for a 2nd-order phase transition follow. Infinite-order phase transitions have to be treated with care, though the overall structure of the FSS relations are similar. Both for the purposes of simplicity and concreteness, let us now focus on a system with one relevant parameter (with respect to some fixed point). Moreover, let us focus specifically on the one-dimensional subspace of the full set of couplings, so the $\boldsymbol{K} \equiv (K, 0, 0, ...)$ with K the single relevant coupling. For example, we could be studying the Ising model with no external magnetic field and no nearest-neighbor coupling; the RG flow in this case is demonstrated in Fig. B.1 for $K_2 = K_3 = 0$ (as defined in that appendix).

Define $k \equiv K/K_\mathrm{c} - 1$ for K_c the critical coupling of the infinite volume phase transition. The singular part of the free energy transforms under an RG transformation in a finite volume as

$$A_{\varphi}^{(\mathrm{s})}\big(k, N^{-1}\big) = b^{-d} A_{\varphi}^{(\mathrm{s})}\big(b^{1/\nu}k, bN^{-1}\big) \qquad \text{(2nd-order)} \tag{2.35}$$

at leading order (no corrections to scaling) for a 2nd-order phase transition. The $1/\nu$ exponent is the conventional renormalization group eigenvalue of K; it is also the critical exponent of the correlation length (in units of the lattice spacing)

$$\tilde{\xi}(K) \sim |k|^{-\nu} \qquad \text{(2nd-order)}. \tag{2.36}$$

Eqn. 2.35 implies that the singular part of the free energy is a *homogeneous* function of k and N^{-1}. The homogeneity of $A_{\varphi}^{(\mathrm{s})}\big(k, N^{-1}\big)$ around $K \approx K_{\mathrm{c}}$ implies that it can be rewritten as [148, 230]

$$A_{\varphi}^{(\mathrm{s})}\big(k, N^{-1}\big) = N^{d} \Phi_{\pm}\big(|k| N^{1/\nu}\big) \qquad \text{(2nd-order)}, \tag{2.37}$$

where $\Phi_{\pm}$ is a *universal scaling function* that could differ for $K^{+} \to K_{\mathrm{c}}$ (above, Φ_{+}) or $K^{-} \to K_{\mathrm{c}}$ (below, Φ_{-}). The scaling behavior of *any* observable that is derived from $A_{\varphi}^{(\mathrm{s})}\big(k, N^{-1}\big)$ will depend upon the dimensionless combination $|k|N^{1/\nu}$ within the vicinity of $K \approx K_{\mathrm{c}}$ due to the presence of the scaling function in Eqn. 2.37. Interestingly enough, the scaling of Eqn. 2.37 also holds for a first-order phase transition, but with $\nu = 1/d$ [96, 120]. Assuming that the correlation length is known, this makes Eqn. 2.37 a good test for distinguishing a 2nd-order phase transition from a 1st-order phase transition. However, it is absolutely crucial to stress that this test only holds any weight if the correlation length in units of the lattice spacing is known at each K; otherwise, it is possible to *misidentify* a phase transition as 2nd-order due to the pseudocritical behavior that is observed for first-order phase transitions with large (but not infinite) correlation lengths.

Similar behavior is observed for ∞-order phase transitions; however, the argumentation is more challenging. A heuristic way to arrive at the scaling form of the free energy for an infinite-order phase transition is to first note that the argument of the scaling function for a 2nd-order phase transition is related to the correlation length $\tilde{\xi}$ of the infinite volume system as [70]

$$|k|N^{1/\nu} \sim \big(\tilde{\xi}(K)/N\big)^{-1/\nu} \qquad \text{(2nd-order)}. \tag{2.38}$$

As such, I could have also written Eqn. 2.37 as

$$A_{\varphi}^{(\mathrm{s})}\big(k, N^{-1}\big) = N^{d}\Phi_{\pm}\big(\tilde{\xi}(K)/N\big), \tag{2.39}$$

which holds for any continuous phase transition of a single variable. Note that I am sloppily still referring to the scaling function as $\Phi_{\pm}$ despite changing its argument; for this I have no remorse and I will continue this practice for the sake of notational brevity. The infinite-volume correlation length of an ∞-order phase transition diverges as

$$\tilde{\xi}(K) \sim \exp\big(\zeta|k|^{-\nu}\big) \qquad (\infty\text{-order}), \tag{2.40}$$

where ν is a universal critical exponent and ζ is a non-universal constant. Therefore, $A_{\varphi}^{(\mathrm{s})}\big(k, N^{-1}\big)$ scales in the vicinity of an ∞-order phase transition as

$$A_{\varphi}^{(\mathrm{s})}\big(k, N^{-1}\big) = N^{d}\Phi_{\pm}\big(N\exp\big(\text{-}\zeta|k|^{-\nu}\big)\big) \qquad (\infty\text{-order}) \tag{2.41}$$

after some rearranging of the arguments. As is the case for a 2nd-order phase transition, all observables derived from Eqn. 2.41 will scale with the combination $N\exp\big(\text{-}\zeta|k|^{-\nu}\big)$ around $K \approx K_{\mathrm{c}}$ in some manner. I will use the ideas of this section in Chapters 6 and 7.

Chapter 3

Gauge-fermion systems on a lattice

> *[Lattice gauge theory] is merely a corner of quantum field theory, and the techniques lattice theorists use are simply decorated versions of techniques used across the board by physicists studying problems with many degrees of freedom in particle, condensed matter, and nuclear physics.*
>
> Thomas DeGrand and Carleton DeTar [98]

In this chapter, I introduce the lattice discretization of SU(N) gauge-fermion systems on a hypercubic Euclidean spacetime lattice. In Secs. 3.1-3.2, I discuss the classical aspects of discretizing gauge and fermion fields. I then discuss the quantization of lattice-discretized gauge-fermion systems in Sec. 3.3 and end in Sec. 3.4 by treating the notion of continuum limits with the tools of the renormalization group that I introduced in Chapter 2. For further reading, see Refs. [88, 142, 227, 265]. For mathematically-rigorous treatments of gauge-fermion systems in the continuum and on a lattice, see Refs. [105, 159, 270] and Refs. [26, 107–109, 329], respectively.

3.1 Classical pure Yang-Mills on a lattice

The classical SU(N) Yang-Mills system in the continuum is described by a vector potential $\mathfrak{A}_\mu(x) \in \mathfrak{su}(N)$, from which the field strength tensor

$$\mathfrak{F}_{\mu\nu}(x) = \partial_\mu \mathfrak{A}_\nu(x) - \partial_\nu \mathfrak{A}_\mu(x) + i[\mathfrak{A}_\mu(x), \mathfrak{A}_\nu(x)] \in \mathfrak{su}(N). \tag{3.1}$$

is specified. Under a local gauge transformation $\Lambda(x) \in \mathrm{SU}(N)$, $\mathfrak{F}_{\mu\nu}(x)$ transforms as

$$\mathfrak{F}_{\mu\nu}(x) \rightarrow \Lambda(x)\mathfrak{F}_{\mu\nu}(x)\Lambda(x)^{\dagger}.$$

As such, the simplest local action that we can construct from $\mathfrak{A}_{\mu}(x)$ that satisfies both (Euclidean) Poincaré invariance and local gauge invariance is

$$\mathcal{S}_{\mathrm{YM}}[\mathfrak{A}_{\mu}] = -\frac{1}{2g_0^2}\int \mathrm{d}^d x \ \mathrm{Tr}_{\mathrm{c}}\big[\mathfrak{F}_{\mu\nu}(x)\mathfrak{F}^{\mu\nu}(x)\big], \tag{3.2}$$

where Tr_{c} denotes a trace over color indices and g_0^2 is a dimensionless parameter that we shall identify with the *bare gauge coupling* once we quantize the classical Yang-Mills system. Eqn. 3.2 is referred to as the Yang-Mills action.

The key mathematical object that will allow us to transcribe the continuum Yang-Mills action onto a lattice is the *parallel transporter*. Given a curve $\mathcal{C}$ in $\mathbb{R}^d$, the parallel transporter associated to $\mathcal{C}$ is

$$\mathcal{U}(\mathcal{C}) = \mathcal{P}\exp\left[i\int_{\mathcal{C}} \mathrm{d}x^{\mu}\mathfrak{A}_{\mu}(x)\right], \tag{3.3}$$

where $\mathcal{P}$ is the path ordering operator [265]. Taking $\mathcal{C}$ to form a closed boundary of a 2-dimensional surface Ω ($\mathcal{C} = \partial\Omega$), a generalization of Stokes' theorem implies [142, 265]

$$\mathcal{U}(\mathcal{C}) = \mathcal{P}\exp\left[i\int_{\Omega} \mathrm{d}x^{\mu}\mathrm{d}y^{\mu}\mathfrak{F}_{\mu\nu}(x)\right]. \tag{3.4}$$

Eqn. 3.4 forms the basis of Wilson's discretization of Eqn. 3.2 [375].

Pass now from the continuum to a d-dimensional hypercubic lattice $\mathbb{Z}^d$ with uniform lattice spacing a. We take the discretized parallel transporter of a curve $\mathcal{C}$ connecting any lattice site $n \equiv x/a$ to the next lattice site $n + \hat{\mu}$ to be

$$\log \mathcal{U}_{\mu}(n) \equiv ia\mathfrak{A}_{\mu}(an). \tag{3.5}$$

We refer to $\mathcal{U}_\mu(n)$ as a *gauge link*; the gauge links play the role of the continuum vector potential. Under a local gauge transformation, gauge links transform as

$$\mathcal{U}_\mu(n) \to \Lambda(an)\mathcal{U}_\mu(n)\Lambda(an + a\hat{\mu})^\dagger. \tag{3.6}$$

Now defining a curve $\mathcal{C}$ that starts at n and traverses the smallest possible square loop in the μ-ν plane, the lattice discretization of Eqn. 3.4 is

$$\log \mathcal{U}_{\mu\nu}(n) \equiv \log \mathcal{U}_\mu(n)\mathcal{U}_\nu(n+\hat{\mu})\mathcal{U}_\mu(n+\hat{\nu})^\dagger\mathcal{U}_\nu(n)^\dagger \approx ia^2\mathfrak{F}_{\mu\nu}(an) + \mathcal{O}\big(a^4\big), \tag{3.7}$$

which is referred to as a *plaquette*. The simplest discretization of Eqn. 3.2 that we can construct from the gauge links $\mathcal{U}_\mu(n)$ that is invariant under gauge transformations is

$$\mathcal{S}_{\mathrm{W}}[\mathcal{U}_\mu] \equiv \frac{\beta_b}{N}\sum_n \sum_{\mu<\nu} \Re\mathrm{Tr}_{\mathrm{c}}\big[\mathbb{1} - \mathcal{U}_{\mu\nu}(n)\big], \tag{3.8}$$

where $\beta_b \equiv 2N/g_0^2$. The action S_{W} is referred to as the *Wilson action*, as it was first written down by Kenneth G. Wilson in 1974 [375]. The classical continuum limit of Eqn. 3.8, taken by driving $a \to 0$ directly, yields

$$\begin{aligned}\mathcal{S}_{\mathrm{W}}[\mathcal{U}_\mu] &= -\frac{a^4}{2g_0^2}\left[\sum_n \sum_{\mu\nu} \mathrm{Tr}_{\mathrm{c}}\big[\mathfrak{F}_{\mu\nu}(an)\mathfrak{F}_{\mu\nu}(an)\big] + \mathcal{O}\big(a^4\big)\right] \\ &\sim -\frac{1}{2g_0^2}\int \mathrm{d}^d x \ \mathrm{Tr}_{\mathrm{c}}\big[\mathfrak{F}_{\mu\nu}(an)\mathfrak{F}^{\mu\nu}(an)\big] \quad (a \to 0),\end{aligned}$$

which follows from Eqns. 3.7-3.8.

3.2 Classical fermions on a lattice

Free classical Dirac fermions in continuum Euclidean space are described by Grassmann-valued spinor fields $\Psi^{(f)}(x)$ that are endowed with the action

$$\mathcal{S}_{\mathrm{F}}[\Psi] \equiv \int \mathrm{d}^d x \sum_{f=1}^{N_f} \overline{\Psi}^{(f)}(x) \mathfrak{D}_f \Psi^{(f)}(x), \tag{3.9}$$

where $\mathfrak{D}_f \equiv \gamma^\mu \partial_\mu + m_f$ is the Euclidean Dirac operator with Euclidean Dirac matrices satisfying

$$\{\gamma^\mu, \gamma^\nu\} = 2\delta^{\mu\nu}.$$

N_f is the number of fermion flavors, and $\overline{\Psi}^{(f)}(x) \equiv \Psi^{(f)}(x)^\dagger \gamma^0$. Discretizing Eqn. 3.9 is a formidible task due to the infamous Nielsen–Ninomiya no-go theorem. For simplicity, consider just one of the flavors in Eqn. 3.9, with action

$$\mathcal{S}_{\mathrm{F}}^{(f)}[\Psi] \equiv \int \mathrm{d}^d x \; \overline{\Psi}^{(f)}(x) \mathfrak{D}_f \Psi^{(f)}(x). \tag{3.10}$$

Discretizing Eqn. 3.10 amounts to discretizing the Dirac operator $\mathfrak{D}_f$. Consider the Fourier-transformed lattice Dirac operator $\widetilde{\mathfrak{D}}_f(p)$ with Fourier transformed action

$$\mathcal{S}_{\mathrm{F}}^{(f)}[\tilde{\Psi}] = \int \frac{\mathrm{d}^d p}{(2\pi)^d} \; \overline{\tilde{\Psi}}(p) \widetilde{\mathfrak{D}}_f(p) \tilde{\Psi}(p)$$

and take the following conditions on $\mathfrak{D}_f$ to hold [227].

- *Translational invariance*: for a generic Dirac spinor $\tilde{u}(p)$, $\mathfrak{D}_f e^{ipx}\tilde{u}(p) = \widetilde{\mathfrak{D}}_f(p) e^{ipx}\tilde{u}(p)$.
- *Locality*: $\widetilde{\mathfrak{D}}_f(p)$ is both an analytic and periodic function of p.
- *Proper continuum limit*: $\widetilde{\mathfrak{D}}_f(p) = i\gamma^\mu p_\mu + \mathcal{O}\big(a|p|\big)$.

The Nielsen–Ninomiya theorem then states that the following cannot hold simultaneously in four dimensions for a discretized massless Dirac operator [227].

- *No species doubling*: $\widetilde{\mathfrak{D}}_f(p)$ is invertible for all non-zero p.
- *Strict continuum chiral symmetry*: for $\gamma_5 \equiv i\gamma^1\gamma^2\gamma^3\gamma^4$, $\{\mathfrak{D}_f, \gamma_5\} = 0$.

In other words, we either have more fermions in the continuum than we intended or we have strict chiral symmetry. Any lattice fermion formulation in four dimensions must contend with the Nielsen-Ninomiya theorem. I discuss in detail only the lattice fermion formulations that are utilized in the present thesis. From here on, I shall bring back the flavor index f only when necessary. Moreover, I distinguish *lattice fermions* ψ from continuum fermions Ψ notationally by

$$\psi(n) = \Psi(an). \tag{3.11}$$

3.2.1 Staggered lattice fermions

Let us begin with the simplest discretization of $\mathfrak{D}$ that fully preserves chiral symmetry. Define a forward difference operator ∂_μ as

$$a\partial_\mu\psi(n) \equiv \psi(n+\hat{\mu}) - \psi(n) \tag{3.12}$$

and a backward difference operator ∂^*_μ as

$$a\partial^*_\mu\psi(n) \equiv \psi(n) - \psi(n-\hat{\mu}). \tag{3.13}$$

In tems of ∂_μ and ∂^*_μ, we can write down a discretization of the Dirac operator that preserves chiral symmetry when $m = 0$ as

$$\mathfrak{D}_{\mathrm{N}} = \frac{1}{2}\sum_\mu \gamma^\mu\left(\partial_\mu + \partial^*_\mu\right) + m. \tag{3.14}$$

The discretized Dirac operator $\mathfrak{D}_{\mathrm{N}}$ is often referred to as the *naïve* Dirac operator; as is the case for many names in the physical sciences, this is quite unnecessarily disparaging. In any case, $\widetilde{\mathfrak{D}}_{\mathrm{N}}(p)$ is not invertible when all $p_\mu = 0$ or π/a. There are sixteen such momenta. Therefore, $\mathfrak{D}_{\mathrm{N}}$ yields sixteen degenerate continuum fermions. This is despite the fact that we discretized the $\mathfrak{D}$ for only a single flavor. That is the price we pay for enforcing chiral symmetry.

Staggered fermions reduce the number of doublers from sixteen to four while retaining a residual U(1) chiral symmetry by mixing lattice indices n with spinor indices α. Consider the local transformation

$$\psi(n) \to \Omega(n)\psi(n), \tag{3.15}$$

where

$$\Omega(n) = \gamma_1^{n_1}\gamma_2^{n_2}\gamma_3^{n_3}\gamma_4^{n_4} \tag{3.16}$$

for $n = (n_1, n_2, n_3, n_4)$ [98, 142]. Performing the *staggered transformation* of Eqn. 3.15 on the naïve lattice fermion action with naïve lattice Dirac operator $\mathfrak{D}_{\mathrm{N}}$ produces a new action that, when written in terms of the transformed lattice fermions fields, has a new lattice Dirac operator

$$\mathfrak{D}_{\mathrm{S}} = \frac{1}{2}\sum_{\mu}\alpha_\mu(n)\big(\partial_\mu + \partial_\mu^*\big) + m, \tag{3.17}$$

where

$$\alpha_1(n) = 1 \text{ and } \alpha_\mu(n) = (-1)^{n_{\mu-1}}\alpha_{\mu-1} \text{ for } \mu > 1. \tag{3.18}$$

The new action with staggered Dirac operator $\mathfrak{D}_{\mathrm{S}}$ is diagonal in spin; in other words, the components of the Dirac spinor have been decoupled. We call these decoupled components *staggered fermions* χ. The action for a single staggered fermion is

$$\mathcal{S}_{\mathrm{SF}}[\chi] = a^4\sum_{n}\overline{\chi}(n)\mathfrak{D}_{\mathrm{S}}\chi(n). \tag{3.19}$$

Staggered fermions were first introduced in a seminal paper by Leonard Susskind and John Kogut in 1975 [228]. Therefore, they are sometimes referred to as *Kogut-Susskind* fermions. Note that there also exists a beautiful differential geometric formulation of staggered fermions that starts in the continuum with Kähler-Dirac fermions [35]. The staggered fermion action of Eqn. 3.19 with $m = 0$

is invariant under the U(1) transformation [98]

$$\chi(n) \to \exp\big(i\Gamma_5(n)\theta\big)\psi(n), \tag{3.20}$$

where $\Gamma_5(n) = 1$ for n even and -1 for n odd. This U(1) transformation is a remnant of chiral symmetry.

One staggered fermion is equivalent to four Dirac fermions. The four degenerate Dirac fermions are referred to as *tastes* because physicists think they're funny. To see how one gets four Dirac fermions out of one staggered fermion, consider the unitary change of basis [98, 265]

$$\psi^{(f)}_\alpha(n) \equiv \frac{1}{8}\sum_\eta \Omega^f_\alpha(\eta)\chi(2n+\eta), \tag{3.21}$$

where $\eta_\mu = 0$ or 1. Each $\psi^{(f)}(n)$ is a four-component Dirac fermion. The matrix Ω is the same as Eqn. 3.16, but one index is for spin and the other is for flavor. With the change of basis in Eqn. 3.16, the staggered action of Eqn. 3.19 reads

$$\mathcal{S}_{\text{SF}}[\psi] \propto a^4 \sum_{n,\mu} \overline{\psi}(n)\left[\frac{1}{2}(\gamma_\mu \otimes \mathbb{1})\big(\partial_\mu + \partial^*_\mu\big) - a\big(\gamma_5 \otimes \gamma^*_\mu\gamma_5\big)\Box_\mu\right]\psi(n) + a^4 m \sum_n \overline{\psi}(n)\psi(n), \tag{3.22}$$

where the "$\otimes$" denotes a tensor product "spin" $\otimes$ "taste" and

$$4a^2\Box_\mu\psi(n) \equiv 4a^2\partial^*_\mu\partial_\mu\psi(n) \equiv \psi(n+\hat{\mu}) + \psi(n-\hat{\mu}) - 2\psi(n). \tag{3.23}$$

The γ-matrices on the flavor side are understood to act on the flavor components of the Dirac spinor in Eqn. 3.21. The term

$$a\overline{\psi}(n)\big(\gamma_5 \otimes \gamma^*_\mu\gamma_5\big)\Box_\mu\psi(n)$$

in Eqn. 3.22 breaks taste symmetry at non-zero lattice spacing. The massless Dirac operator in this new “spin-taste” basis

$$\mathfrak{D}_{\text{SST}} \propto \frac{1}{2}(\gamma_\mu \otimes \mathbb{I})\big(\partial_\mu + \partial^*_\mu\big) - a\big(\gamma_5 \otimes \gamma^*_\mu\gamma_5\big)\Box_\mu \tag{3.24}$$

has a Fourier transform that is not invertible only at four momenta due to the appearance of the taste-breaking term. The four momenta at which $\mathfrak{D}^{-1}_{\text{SST}}$ has a pole correspond to four “tastes” of Dirac fermion in the continuum. This reduction in the number of doublers comes at the cost of explicitly breaking taste symmetry. However, there is still a residual U(1) chiral symmetry, and when this U(1) chiral symmetry is spontaneously broken, it produces a Goldstone boson [98].

3.2.2 Other lattice fermions

Staggered fermions are only one of a whole zoo of lattice fermion contenders. I have focused on staggered fermions because they are the primary formulation that is utilized in this thesis. In Sec. 8.3, I use another formulation of lattice fermion known as *domain wall* (DW) fermions. For the sake of completeness, let me briefly discuss DW fermions by starting with *Wilson fermions.*

Wilson fermions reduce the number of fermion doublers in four dimensions from sixteen to one by explicitly breaking chiral symmetry [375]. The Wilson Dirac operator is

$$\mathfrak{D}_{\text{W}} = \frac{1}{2}\sum_\mu \Big[\gamma_\mu\big(\partial_\mu + \partial^*_\mu\big) - a\Box_\mu\Big] + m. \tag{3.25}$$

The $-a\Box_\mu$ term is responsible for explicitly breaking chiral symmetry in Eqn. 3.25. Notice that the chiral symmetry breaking term for the Wilson Dirac operator is similar to the taste breaking term in the staggered Dirac operator in the spin-taste basis. *Domain wall fermions* are five-dimensional Wilson fermions with a topological defect in the fifth-dimension that separates left-handed modes from right-handed modes on the four-dimensional boundary [217, 219, 331, 363]. Denote the size of the fifth dimension as N_5. When $N_5 \to \infty$, DW fermions are equivalent to a formulation of lattice fermion that obeys a lattice analogue of chiral symmetry that is expressed by the famous

Ginsparg-Wilson equation [146, 195, 196, 275]

$$\{\mathfrak{D}_{\text{DW}}, \gamma_5\} \sim a\mathfrak{D}_{\text{DW}}\gamma_5\mathfrak{D}_{\text{DW}} \quad (N_5 \to \infty), \tag{3.26}$$

where $\mathfrak{D}_{\text{DW}}$ denotes the domain wall discretization of the Dirac operator. The $N_5 \to \infty$ limit of a DW fermion is known as an *overlap* fermion [276]. Both overlap fermions and DW fermions possess no doublers at the cost of explicitly breaking chiral symmetry according to Eqn. 3.26. When N_5 is finite, there are corrections to Eqn. 3.26. For an explicit form of the DW Dirac operator $\mathfrak{D}_{\text{DW}}$, see Ref. [142]. Note that modern simulations with DW fermions, including those in Sec. 8.3, typically utilize a generalization of DW fermions known as *Möbius domain wall fermions* [54].

3.2.3 Coupling lattice fermions to lattice gauge fields

When one says that a continuum fermion field Ψ "transforms under the fundamental representation of SU(N)", they are asserting that Ψ is smooth map $x \mapsto \Psi(x) \in V_x$, where $x \in \mathbb{R}^4$ and V_x is a vector space that furnishes the fundamental representation of SU(N).[1] Because the vector spaces V_x and V_y for $x, y \in \mathbb{R}^4$ are different vector spaces, there is no meaning to comparing $\Psi(x) \in V_x$ to $\Psi(y) \in V_y$ without some mathematical device that transports vectors in V_y to vectors in V_x (and vice-versa). That device is the parallel transporter that I briefly introduced in Sec. 3.1.

Take $\mathcal{C}$ to be a curve that connects any two points $x, y \in \mathbb{R}^4$. The action of the parallel transporter $\mathcal{U}(\mathcal{C})$ on $\Psi(y)$ is to transport it to V_x; e.g., $\mathcal{U}(\mathcal{C})\Psi(y) \in V_x$. By transporting $\Psi(y) \in V_y$ to V_x, one can now compare $\Psi(y)$ to $\Psi(x)$.[2] In Sec. 3.1, I introduced the gauge link $\mathcal{U}_\mu(n)$ as a discretization of the parallel transporter that links $n \to n + \hat{\mu}$. To make the forward and backward difference operators in Eqns. 3.12-3.13 well-defined when the lattice field ψ transforms under the fundamental representation of SU(N) at each lattice site, one uses the gauge link to connect

---

[1] Many of the following statements readily generalize to any other spacetime manifold and Lie group representation.

[2] It is worth noting that one can also easily define the gauge covariant derivative without reverting to arguments from gauge invariance, which I always found to be unsatisfying.

neighboring lattice sites as

$$a\partial_\mu \psi(n) \equiv \mathcal{U}_\mu(n)\psi(n+\hat{\mu}) - \psi(n) \tag{3.27}$$

and

$$a\partial^*_\mu \psi(n) \equiv \psi(n) - \mathcal{U}_\mu(n)^\dagger \psi(n-\hat{\mu}). \tag{3.28}$$

By redefining the forward- and backward-difference operators in Eqns. 3.12-3.13 as 3.27-3.28, one has coupled the lattice gauge field $\mathcal{U}_\mu(n)$ to a generic lattice fermion field ψ. That is all there is to it. All expressions involving ∂_μ and ∂^*_μ remain the same.

3.3 Quantization of gauge-fermion systems on a lattice

Quantum fluctuations of a gauge-fermion system on a Euclidean space-time lattice $\mathbb{Z}^4$ are captured by the partition function

$$\mathcal{Z} = \int \left[\mathrm{d}\mathcal{U}_\mu\right]\left[\mathrm{d}\overline{\psi}\mathrm{d}\psi\right] \exp\left(\text{-}\mathcal{S}_{\text{lat.}}[\mathcal{U}_\mu, \psi, \overline{\psi}]\right), \tag{3.29}$$

where $\mathcal{S}_{\text{lat.}}$ a lattice discretization of the continuum gauge-fermion action. No Faddeev-Popov ghost fields are needed to fix up the gauge integration measure. The integration measure for the gauge fields

$$\left[\mathrm{d}\mathcal{U}_\mu\right] \equiv \prod_{n\in\mathbb{Z}^4}\prod_{\mu=1}^{4} \mathrm{d}\mathcal{U}_\mu(n) \tag{3.30}$$

is a *Haar* measure. The integration “measure” for the fermion fields

$$\left[\mathrm{d}\overline{\psi}\mathrm{d}\psi\right] \propto \prod_{n\in\mathbb{Z}^4}\prod_{\alpha} \mathrm{d}\overline{\psi}_\alpha(n)\mathrm{d}\psi_\alpha(n) \tag{3.31}$$

indicates notationally that we are performing a Berezin integral. It is not a true Lebesgue measure. The integration measure for the fermions is normalized such that

$$\int \left[\mathrm{d}\overline{\psi}\mathrm{d}\psi\right] \exp\left(\text{-}\sum_{n,\alpha}\overline{\psi}_\alpha(n)\psi_\alpha(n)\right) = 1. \tag{3.32}$$

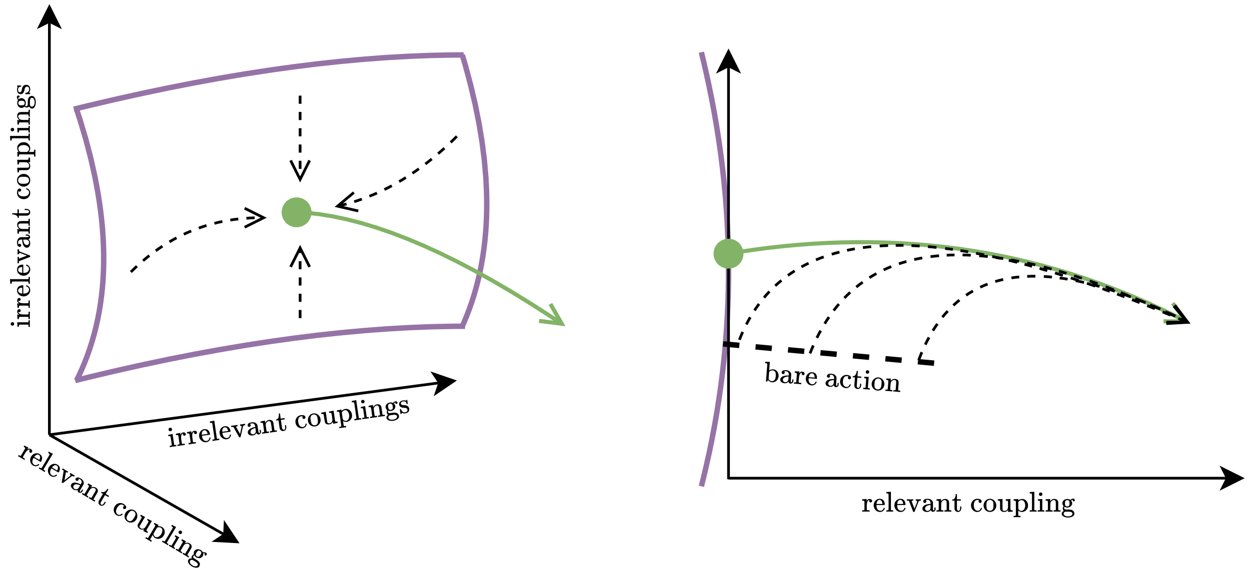

Figure 3.1: Illustration of an RG flow for a gauge-fermion system with a single relevant coupling that emerges from a critical fixed point. The left panel shows the critical surface of the critical fixed point as a bounded surface with purple lines. The flows starting on the critical surface are indicated by dashed arrows. The renormalized trajectory emerging from the critical fixed point is indicated by a solid green arrow. The right panel shows the RG flow of systems starting off of the critical surface, which are indicated by a thick dashed line marked as “bare action”. The RG flows of such systems are indicated by dashed arrows. As in the left panel, the renormalized trajectory is indicated by a green arrow.

For naïve, Wilson, and domain wall fermions, $\alpha \in \{1,2,3,4\}$. For staggered fermions, I drop the α index. Let us assume that *physical* observables derived from $\mathcal{Z}$ have a well-defined continuum limit. I’ll discuss this in the next section. I also assume that the continuum limit satisfies the Osterwalder-Schrader axioms [285], which implies that the Wick-rotated continuum limit defines a Minkowski space quantum field theory in the sense of the Wightmann axioms [344]. Rather infamously, this has yet to be proven, but you can win one million dollars from the Clay Mathematics Institute if you’re the lucky one to figure it out; hence, this is left as an exercise to the reader.

3.4 Renormalization and the continuum limit

In Chapter 2, I introduced the renormalization group. In this section, I shall apply the ideas and language that we learned in that chapter to understand the continuum limit in gauge-fermion systems with $N_{\mathrm{f}} \lesssim 11N/2$. To start, recall that the gauge-fermion system defined by the partition function of Eqn. 3.29 is formally a classical statistical mechanical system with reduced Hamiltonian $\mathcal{S}_{\mathrm{lat.}}[\mathcal{U}_\mu, \psi, \overline{\psi}]$. In Sec. 2.2.2, I briefly discussed the notion of a correlation length $\tilde{\xi}\left(g_0^2\right)$, which is the shortest distance over which the fields of such systems are correlated with one another. Specifically, we learned that the correlation length in units of the lattice spacing diverges at a critical fixed point. If we are to keep all scales that are greater than the lattice spacing fixed along the continuum limit, then the correlation length $\tilde{\xi}\left(g_0^2\right)$ must also diverge in the continuum limit. Therefore, the existence of a continuum limit in a gauge-fermion system is tied to the existence of a critical fixed point in the renormalization group flow of the underlying system. This is also true in general. Even more, the RG trajectory that emerges from the critical fixed point furnishes a definition of the continuum gauge-fermion system at any length scale [193, 194]. This special trajectory is referred to as the *renormalized trajectory* (RT).

In Fig. 3.1, I illustrate the RG flow for a system with a critical fixed point and a single relevant coupling. Such a critical fixed point is referred to as an *ultraviolet fixed point* (UVFP). The RT is shown as a green arrow that emerges from the critical fixed point. Consider that any point along the RT is connected to the fixed point by an infinite number of infinitesimally small RG transformations. Within the vicinity of the fixed point, the system is completely free of any cutoff effects; this includes any and all long-distance properties of that system. Since RG transformations preserve the long-distance properties of the system, it must be the case that any point along the RT is completely free of cutoff effects [193]. Each point of the RT therefore furnishes a definition of the continuum system at any length scale that we may observe it at. Returning to the flow diagram of Fig. 3.1, the surface that is bounded by purple lines is the critical surface. Any RG trajectory starting on the critical surface, but not at the fixed point, flows to the fixed point, as indicated by the dashed lines in the left panel of Fig. 3.1. A massless gauge-fermion system with a particular lattice action and bare gauge coupling g_0^2 defines a set of starting points in couplings space, as indicated by the thick dashed line that is marked with “bare action” in the right panel of Fig. 3.1. Each RG trajectory of such a system, indicated by thin dashed lines in the right panel of Fig. 3.1, converges to the renormalized trajectory. This is merely reflecting the fact that the gauge-fermion system defined over a spacetime lattice with lattice spacing a approximates the continuum gauge-fermion system well over distance scales $l/a \gg 1$. The idea that the RG trajectory of any lattice action at any g_0^2 converges onto the RT is a consequence of *universality.* Each and every action defines its own line of starting points in the space of couplings illustrated in the right panel of Fig. 3.1. A *classically improved* lattice action defines a line of starting points that is close to the RT within the vicinity of the critical surface (see Appendix F).

The flow diagram in the right panel of Fig. 3.1 tells us how to take the continuum limit. Each point along the RT is parameterized by a renormalized coupling $g^2(b)$ with a β-function defined as in Eqn. 2.8. As the bare gauge coupling is tuned closer to the critical surface, the RG flow converges onto the RT earlier because it is a better approximation of the continuum system at distances scales of increasing length. By fixing $g^2(b)$, one is picking a point along the renormalized trajectory. Once the RG flows that start off of the critical surface reach the scale b, one can collapse them onto the RT by tuning g_0^2 to the critical surface [205]. Hence, one takes the continuum limit by fixing $g^2(b)$ and tuning g_0^2 to the critical surface. In principle, it is not even necessary to that the continuum limit. If one is far enough along the the RG flow in the sense that one is close to the RT, then cutoff effects are suppressed and they are able to access long-distance observables of the continuum system. As I describe in Appendix C, this is achieved by tuning the bare gauge coupling in a gauge fermion system to zero ($g_0^2 \to 0$).

3.4.1 Connection to quantum chromodynamics

The notion of a continuum limit that I have described may not be familiar to many lattice field theory practitioners, especially those that work on quantum chromodynamics (QCD) at the physical point. Namely, where is the fixed renormalized coupling in a calculation of the hadronic contribution to the muon's anomalous magnetic moment or hadronic spectroscopy calculations (e.g., Ref. [115]), for example [12, 122]? The answer is that it is hidden in the *scale setting* component of many lattice calculations, whereby some observable $O(l)$ is fixed is fixed to some value $O(l_0) = c$

to estimate a reference scale l_0/a that all dimensionful quantities, such as hadron masses, can be expressed in terms of [6, 31, 43, 336].[3] As has been described in Chapter 2.3, the observable $O(l)$ can be used to define a renormalized coupling that runs with l and has the same global properties as the couplings $g^2(b)$ of the RG transformation if it obeys certain conditions under a renormalization group transformation. Hence, by fixing $O(l_0) = c$ along the continuum limit, one is indeed fixing a renormalized coupling. The bare masses $\{m_i\}$ in such calculations are typically, but not always, fixed by requiring that the same number of hadron masses $\{M_i\}$ are fixed to their physical values. This is particularly evident in spectroscopic calculations.[4] For example, the massess of the up m_u and down m_d quarks in the isospin-symmetric limit ($m_u = m_d$) are often fixed by requiring that the pion mass M_π is equal to the mass of the neutral pion $M_{\pi^0} \approx 135$ MeV [381].[5]

---

[3] As of the writing of this thesis, the most recent Flavor Lattice Averaging Group (FLAG) review contains up-to-date information on scale setting calculations in lattice gauge theory [6].

[4] See Fig. 3 of Ref. [115] and note which meson masses are fixed.

[5] See Refs. [32, 42, 93] for an example in the context of calculations of the hadronic vacuum polarization contribution to the muon's anomalous magnetic moment, especially Table 1 of Ref. [93].

Chapter 4

The β-function of the pure Yang-Mills system

In this chapter, we get our feet wet with calculating RG β-functions on the lattice using the pure SU(3) Yang-Mills system. Gradient flow and its connection to the renomalization group is discussed in Secs. 4.1-4.2. The *continuous β-function method* (CBFM) for extracting the continuum β-function from gradient flow is outlined in Sec. 4.3. I calculate the continuum RG β-function for the $N_\mathrm{f} = 0$ system using the CBFM in Sec. 4.5. From the β-function, I calculate the pure Yang-Mills Λ-parameter in Sec. 4.6. I end this chapter with a proposal for matching β-functions from different RG schemes non-perturbatively in Sec. 4.7. The main results of this chapter are based on the works of Refs. [181, 294].

4.1 Gradient flow

The β-function in this chapter relies upon the gradient flow transformation [252, 253, 272] and its connection to renormalization group transformations [72–74]. Using the notation of Chapter 2, the gradient flow transformation is a continuous smearing operation on the elementary fields $\{\varphi_\mathfrak{f}\}$ along the gradient of some *flow action* $\mathcal{S}^\mathrm{f}\left[\varphi_\mathfrak{f}\right]$.[1]

The gradient flow equation for a pure Yang-Mills system is [253, 254]

$$\frac{\mathrm{d}\mathfrak{A}_\mu(x,t)}{\mathrm{d}t} = -g_0^2 \frac{\delta \mathcal{S}_\mathrm{YM}[\mathfrak{A}_\mu]}{\delta \mathfrak{A}_\mu(x,t)}, \tag{4.1}$$

---

[1] Sometimes, it is preferable to include extra terms in the gradient flow equation that force the gradient flow to preserve some symmetries of $\mathcal{S}^\mathrm{f}\left[\varphi_\mathfrak{f}\right]$; as a simple example, I derive a gradient flow equation for the XY model in Appendix E.

where $\mathfrak{A}_\mu(x,t)\big|_{t=0} = \mathfrak{A}_\mu(x)$, $\mathcal{S}_{\mathrm{YM}}[\mathfrak{A}_\mu]$ is defined in Eqn. 3.2, and $\delta\mathcal{S}_{\mathrm{YM}}[\mathfrak{A}_\mu]/\delta\mathfrak{A}_\mu(x,t)$ is a somewhat sloppy notation for a distributional derivative of $\mathcal{S}_{\mathrm{YM}}$ with respect to $\mathfrak{A}_\mu(x,t)$. A lattice discretization of the continuum Yang-Mills gradient flow is [253, 254]

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{U}_\mu(n,t) = -g_0^2\left(\partial_{x,\mu}\mathcal{S}^{\mathrm{f}}[\mathcal{U}_\mu]\right)\mathcal{U}_\mu(n,t), \tag{4.2}$$

with $\mathcal{U}_\mu(n,t)\big|_{t=0} = \mathcal{U}_\mu(n)$ and the action $\mathcal{S}^{\mathrm{f}}[\mathcal{U}_\mu]$ some lattice discretization of $\mathcal{S}_{\mathrm{YM}}[\mathfrak{A}_\mu]$ (such as the Wilson action of Eqn. 3.8). I call $\mathcal{S}^{\mathrm{f}}[\mathcal{U}_\mu]$ the *flow action.* The differential operator

$$\partial_{n,\mu} = T^a\partial^a_{n,\mu} \tag{4.3}$$

is defined by the action of $\partial^a_{n,\mu}$ on a differentiable function F of the SU(N) links $\mathcal{U}_\mu(n,...)$ as [252]

$$\partial^a_{n,\mu}f(\mathcal{U}_\mu(n,...)) = \frac{\mathrm{d}}{\mathrm{d}s}F\left(e^{sX^a(m,\nu)}\mathcal{U}_\mu(n,...)\right)\bigg|_{s=0}, \tag{4.4}$$

where

$$X^a(m,\nu) = T^a \text{ if } (m,\nu) = (n,\mu) \text{ else } 0. \tag{4.5}$$

I have intentionally not specified the co-domain of F. Note that the basis for the Lie algebra that $\partial^a_{n,\mu}$ is defined with respect to is arbitrary. For this thesis, I follow Ref. [253] and choose the conventional basis $\{T^a\}$ normalized as

$$\mathrm{Tr}_{\mathrm{c}}\left[T^aT^b\right] = -\frac{1}{2}\delta^{ab} \tag{4.6}$$

and satisfying the structure equation

$$\left[T^a, T^b\right] = f^{abc}T^c, \tag{4.7}$$

along with various other relations that follow from the completeness of the $\{T^a\}$ basis. See Chapter 8.1.2 for a description of numerically integrating the gradient flow equation of Eqn. 4.2.

4.2 Renormalization and gradient flow

The gradient flow equation is a dissipative Langevin equation without noise [72, 389]. The latter realization bears a fruitful connection between gradient flow and stochastic renormalization that was worked out in Refs. [72, 73]; however, there is still more work that could be done along this direction. Though the construction of Refs. [72, 73] is more rigorous, it is conceptually simpler to parse out the connection between gradient flow and renormalization by thinking in terms of the real-space RG transformation that I briefly touched upon in Sec. 2.1 [74]. Note that there are also many wonderful works that discuss the connection between gradient flow and the renormalization group at length in terms of *exact* renormalization group transformations [263, 264, 337–339], along with various other perspectives [2, 68, 238, 256, 257, 345].

The gradient flow transformation suppresses wave number fluctuations in the $\pi/a \leq |k| \lesssim \pi/b_t a$ shell for $b_t \propto \sqrt{8t/t_0}$ with t_0 some dimensionful constant that makes b dimensionless [250]. Returning to the notation of Chapter 2, define the RG blocked fields $\{\overline{\varphi}_{\mathfrak{f}}^{(b_t)}(n_{b_t})\}$ $(n_{b_t} = n/b_t)$ in terms of the gradient flowed fields $\{\varphi_{\mathfrak{f}}(n,t)\}$ as [74]

$$\overline{\varphi}_{\mathfrak{f}}^{(b_t)}(n_{b_t}) = \varphi_{\mathfrak{f}}(n,t). \tag{4.8}$$

From the RG blocked fields, define the rescaled (renormalized) fields as

$$\varphi'^{(b_t)}_{\mathfrak{f}}(n_{b_t}) = Z_{\mathfrak{f}}(b_t)\overline{\varphi}_{\mathfrak{f}}^{(b_t)}(n_{b_t}), \tag{4.9}$$

where $Z_{\mathfrak{f}}(b_t)$ is the wave function renormalization of $\varphi_{\mathfrak{f}}$. The wave function renormalization $Z_{\mathfrak{f}}(b_t)$ is fixed by requiring that the correlation functions of $\{\varphi_{\mathfrak{f}}\}$ are preserved at long distances (compared to the lattice spacing of the unblocked lattice). Assuming that the definitions in Eqns. 4.8-4.9 *describe* an RG transformation for local operators at asymptotically large b_t, one can define a renormalized coupling $g^2(b_t)$ in terms of the expectation value of a local operator $\mathcal{O}(b_t)$ that does not renormalize;

i.e.,

$$g^2(b_t) \sim \langle \mathcal{O}(b_t) \rangle. \tag{4.10}$$

The corresponding β-function is

$$\beta_\alpha\left(g_\alpha^2\right) \equiv -b^2 \frac{\mathrm{d}g_\alpha^2(b)}{\mathrm{d}b^2}. \tag{4.11}$$

Note that I have been careful to state that gradient flow *describes* an RG transformation, as the definitions of Eqns. 4.8-4.9 cannot fully specify a true RG transformation on their own. This is evident from the fact that the GF equations are reversible; gradient flow *suppresses* short-distance fluctuations, but it does not remove them completely. Nonetheless, gradient flow is capable of *describing* an RG transformation for (expectation values of) local operators at asymptotic b_t [72, 73].

4.3 Introduction to the continuous β-function method

Consider the Yang-Mills energy density

$$E(x,t) = -\frac{1}{2} \Re \mathrm{Tr}_{\mathrm{c}}\left[\mathfrak{F}_{\mu\nu}(x,t)\mathfrak{F}_{\mu\nu}(x,t)\right]. \tag{4.12}$$

at GF flow time t. The observable $t^2\langle E(x,t)\rangle$ has no wave function renormalization [250]. Hence, I can use it to define a renormalized coupling in infinite volume as

$$g_{\mathrm{GF}}^2(t) \equiv \mathcal{N}\langle t^2 E(t)\rangle \tag{4.13}$$

and corresponding renormalization group β-function as

$$\beta_{\mathrm{GF}}\left(g_{\mathrm{GF}}^2\right) \equiv -t\frac{\mathrm{d}}{\mathrm{d}t} g_{\mathrm{GF}}^2(t). \tag{4.14}$$

The normalization $\mathcal{N} = 128\pi^2/(3N^2-3)$ is chosen such that g_{GF}^2 matches $g_{\overline{\mathrm{MS}}}^2$ at tree-level [252]. The renormalized coupling of Eqn. 4.13 tracks the renormalized trajectory that emerges from the UVFP of the gauge-fermion system, should it exist [256]. Note that I have not once referred to perturbation theory in defining g^2_{GF} outside of the choice for $\mathcal{N}$. That is not meant to say that perturbation theory will not be useful for either improving or modifying the definition of g^2_{GF}.

4.3.1 Defining the gradient flow renormlized coupling in a finite volume

I wish to calculate the RG β-function of Eqn. 4.14 from finite-volume lattice simulations. To this end, one must start out by defining the renormalized coupling in a finite volume. This is more tricky than it may sound. The low-energy dynamics of an interacting gauge-fermion system in a finite volume with periodic boundary conditions is dominated by gauge zero modes [247]. The effect of the gauge zero-modes on the finite-volume Yang-Mills energy density $\langle E(t;L)\rangle$ must be treated exactly when defining the renormalized coupling. The details of properly treating the gauge zero modes were worked out in Ref. [130]. At leading order in the renormalized coupling of the $\overline{\text{MS}}$ scheme $g^2_{\overline{\text{MS}}}$, the Yang-Mills energy density is related to $g^2_{\overline{\text{MS}}}$ as [130]

$$\langle t^2 E(t;L)\rangle = \mathcal{N}^{-1} g^2_{\overline{\text{MS}}}(\mu)\big(1+\delta(t,L)\big), \tag{4.15}$$

where

$$\delta(t,L) \equiv -\frac{1}{3}\left(\frac{8\pi t}{L^2}\right)^2 + \vartheta^4\Big(\exp\Big(\text{-}8t/L^2\Big)\Big) - 1 \tag{4.16}$$

and $\vartheta(\cdot)$ is the Jacobi elliptic function

$$\vartheta(x) \equiv \sum_{n=-\infty}^{\infty} x^{n^2}. \tag{4.17}$$

As

$$g^2_{\overline{\text{MS}}} \approx g^2_{\text{GF}} + O\big(g^4_{\text{GF}}\big) \qquad \big(g^2_{\overline{\text{MS}}}/4\pi \ll 1\big) \tag{4.18}$$

at tree level, a suitable definition of the gradient flow coupling in finite volume is

$$g^2_{\text{GF}}(t;L) \equiv \frac{\mathcal{N}}{1+\delta(t,L)}\langle t^2 E(t;L)\rangle. \tag{4.19}$$

Crucially, as $t/L^2 \to 0$, so too does $\delta(t, L) \to 0$. Therefore, the definition of Eqn. 4.19 approaches the infinite volume definition of the renormalized coupling in Eqn. 4.13 as $t/L^2 \to 0$. Note, also, that Eqns. 4.15-4.19 imply that the leading finite-volume effects in $g^2_{\mathrm{GF}}(t; L)/4\pi \lesssim 1$ are

$$g^2_{\mathrm{GF}}(t; L) \approx g^2_{\mathrm{GF}}(t) + \kappa\big(t/L^2\big)^2 \qquad \big(g^2_{\mathrm{GF}}/4\pi \ll 1\big), \tag{4.20}$$

where κ is a constant whose dependence on t is revealed through non-perturbative simulations. One could have guessed the latter scaling of $g^2_{\mathrm{GF}}(t; L)$ in t/L from the mass dimension of $E(t, L)$; such scaling is likely to hold outside of the weakly-coupled perturbative regime.

4.3.2 Discretization of the Yang-Mills energy density and tree-level improvement

In this chapter, I discretize $E(t)$ with the Wilson discretization (W), the Symanzik discretization (S), and the clover discretization (C). I have described the Wilson and Symanzik discretization of the Yang-Mills energy density in Chapter 3.1 and Appendix F, respectively. The clover discretization of the Yang-Mills energy density is obtained from a direct discretization of $\mathfrak{F}_{\mu\nu}(x)$ as [142]

$$\mathcal{F}^{(\text{clov.})}_{\mu\nu}(n) \sim \big(Q_{\mu\nu}(n) - Q_{\nu\mu}(n)\big) \sim a^2 \mathfrak{F}_{\mu\nu}(an) \qquad (a \to 0), \tag{4.21}$$

where

$$Q_{\mu\nu}(n) \equiv \mathcal{U}_{\mu\nu}(n) + \mathcal{U}_{\nu,-\mu}(n) + \mathcal{U}_{-\mu,-\nu}(n) + \mathcal{U}_{-\nu,\mu}(n) \tag{4.22}$$

and the plaquette $\mathcal{U}_{\mu\nu}(n)$ is defined in Chapter 3. The negative indices are to be interpreted as reversing the direction of the gauge links that make up a particular plaquette via the prescription

$$\mathcal{U}_{-\mu}(n) \equiv \mathcal{U}_{\mu}(n - \hat{\mu})^{\dagger}. \tag{4.23}$$

To this end, I calculate the lattice-discretized renormalized coupling in finite volume as

$$g^2_{\mathrm{GF}}\big(t; L, g_0^2\big) \equiv \frac{\mathcal{N}}{1+\delta(t,L)} \langle t^2 E\big(t;L\big)\rangle_{g_0^2}, \tag{4.24}$$

where $E\big(t;L\big)$ is discretized with the Wilson, Symanzik or Clover discretization. I refer to these discretizations as different "*operators*" to follow the language used in the literature. In fact, one can do better than Eqn. 4.24. By redefining $\delta(t,L)$ in Eqn. 4.16, one can correct tree-level discretization effects in $g^2_{\mathrm{GF}}\big(t;L,g_0^2\big)$. The manner in which this is done is similar to the tree-level improvement that I briefly discuss in Appendix F and it is outlined clearly in Ref. [123]. Defining

$$a\hat{p}_\mu = 2\sin\big(ap_\mu/2\big) \text{ and}$$

$$a\tilde{p}_\mu = \sin(ap_\mu),$$

one can express the kernel $\mathcal{S}_{\mu\nu}$ for an improved lattice action as [123, 139, 248, 369, 370]

$$\mathcal{S}_{\mu\nu} = \delta_{\mu\nu}\Big(\hat{p}^2 - a^2 c_{\mathrm{I}} \sum_\rho \hat{p}_\rho^4 - a^2 c_{\mathrm{I}} \hat{p}_\mu^2 \hat{p}^2\Big) - \hat{p}_\mu \hat{p}_\nu \big(1 - a^2 c_{\mathrm{I}} \hat{p}_\mu^2 - a^2 c_{\mathrm{I}} \hat{p}_\nu^2\big) \qquad \text{(Symanzik)} \tag{4.25}$$

$$\text{or } \big(\delta_{\mu\nu}\tilde{p}^2 - \tilde{p}_\mu \tilde{p}_\nu\big) \frac{\partial \hat{p}_\mu}{\partial p_\mu} \frac{\partial \hat{p}_\mu}{\partial p_\mu} \qquad \text{(clover)}, \tag{4.26}$$

where c_{I} is an improvement coefficient similar to $c_{\mathrm{p}}/c_{\mathrm{r}}$ in Appendix F. In terms of $\mathcal{S}_{\mu\nu}$ for the flow $\mathcal{S}^{\mathrm{f}}_{\mu\nu}$, action $\mathcal{S}^{\mathrm{a}}_{\mu\nu}$, and $E(t)$ operator $\mathcal{S}^{\mathrm{e}}_{\mu\nu}$, the tree-level improved $\delta(t,L)$ is [123]

$$\delta(t,L) = \frac{1}{3}\left(\frac{8\pi t}{L^2}\right)^2 \left(2 + \mathrm{Tr} \sum_{n_\mu = 0, n^2 \neq 0}^{L/a-1} \exp\Big[\text{-}t\big(\mathcal{S}^{\mathrm{f}} + \mathcal{G}\big)\Big] (\mathcal{S}^{\mathrm{a}} + \mathcal{G})^{-1} \exp\Big[\text{-}t\big(\mathcal{S}^{\mathrm{f}} + \mathcal{G}\big)\Big] \mathcal{S}^{\mathrm{e}}\right) - 1, \tag{4.27}$$

where

$$\mathcal{G}_{\mu\nu} = \frac{1}{\alpha} \hat{p}_\mu \hat{p}_\nu \tag{4.28}$$

is a gauge fixing factor. Despite the Eqn. 4.27 involving $\mathcal{G}$, the correction $\delta(t,L)$ is gauge-invariant. The tree-level correction $\delta(t,L)$ from Eqn. 4.27 is included in the definition of $g^2_{\mathrm{GF}}\big(t;L,g_0^2\big)$ by calculating $\delta(t, L)$ numerically with $\alpha = 1$.

4.3.3 Extracting the continuum β-function from finite-volume simulations

I wish to calculate the infinite-volume gradient-based RG β-function in the continuum from the finite-volume gradient flow coupling defined in Eqn. 4.19 with $\delta(t, L)$ given by either Eqn. 4.16 or Eqn. 4.27. One way to do this is by the method of *step-scaling* [130, 254, 307], whereby one defines a renormalized coupling that runs with the volume L as [130]

$$g_c^2\big(L, g_0^2\big) = g_{\mathrm{GF}}^2\big(t; L, g_0^2\big)\Big|_{8t=(cL)^2}. \tag{4.29}$$

As such, each value of c defines a renormalization scheme. From $g_c^2\big(L, g_0^2\big)$, one typically either calculates the discrete β-function

$$\beta_{c,s}\big(g_c^2; L, g_0^2\big) \equiv \frac{g_c^2\big(sL, g_0^2\big) - g_c^2\big(L, g_0^2\big)}{\log s^2} \tag{4.30}$$

or its closely-related cousin, the step-scaling function

$$\Sigma_{c,s}\big(u; L, g_0^2\big) = g_c^2\big(sL; g_0^2\big)\Big|_{g_c^2(L, g_0^2)=u}. \tag{4.31}$$

Note that the difference between the discrete β-function and step-scaling function is superficial. Once one has either of the two, a continuum extrapolation to the $a/L \to 0$ limit at fixed g_c^2 is performed. Step-scaling has been applied with great success to a variety of important gauge-fermion systems; see, for example, Refs. [56, 66, 91, 92, 124, 125, 130, 132–134, 183–187, 244, 245, 269, 306]. However, there is one issue that step-scaling cannot overcome: it requires that the volume L is the only dimensionful scale that is available to the system [131, 294]. As such, step-scaling is not applicable to large-volume confined regime of a lattice gauge-fermion system, where confinement introduces another infrared scale; namely, the confinement scale. In the lattice gauge theory literature, the "large-volume confined regime" is referred to the *p-regime*; usually, "large volume" means $M_{\pi^0} L \gg 1$, where M_{π^0} is the mass of the neutral pion. This is to be juxtaposed against the small-volume deconfined regime; i.e., the ϵ-*regime* ($M_{\pi^0}L \ll 1$). To get around the latter shortcoming of step-scaling, a method for extracting the infinite-volume gradient flow β-function was proposed independently by the authors of Ref. [131] and the authors of Refs. [189, 190].

The method by which the infinite volume β-function of Refs. [131, 189, 190] is calculated is referred to in the literature as either the *continuous* β*-function method* (CBFM) or *infinite volume* β*-function*. In this thesis, I follow the convention of Refs. [189, 190] and refer to it as the CBFM. For a massless gauge-fermion system, the CBFM proceeds conceptually in three steps.

(1) **Finite-volume coupling and β-function:** Calculate $g^2_{\mathrm{GF}}(t; L, g_0^2)$ using Eqn. 4.24 and $\beta_{\mathrm{GF}}(t; L, g_0^2) = -t\mathrm{d}g^2_{\mathrm{GF}}(t; L, g_0^2)/\mathrm{d}t$, where $\delta(t, L)$ is given by either Eqn. 4.16 or Eqn. 4.27.

(2) **Infinite volume limit:** There are many ways that this can be done. The two that have been explored for massless systems thus far are as follows.

- Extrapolate $\beta_{\mathrm{GF}}(t; L, g_0^2)$ linearly in a^4/L^4 at fixed t/a^2 and $g^2_{\mathrm{GF}}(t; L, g_0^2)$ [189, 190].
- Extrapolate *both* $g^2_{\mathrm{GF}}(t; L, g_0^2)$ and $\beta_{\mathrm{GF}}(t; L, g_0^2)$ linearly in a^4/L^4 at fixed t/a^2 and g_0^2. This was first explored in Ref. [294] and subsequently deployed in Refs. [180, 181, 235, 294, 380]. This approach avoids potential systematic uncertainties that are associated with finite-volume effects in $g^2_{\mathrm{GF}}(t; L, g_0^2)$.

In Ref. [131], another method for extracting the infinite volume β-function for massive gauge-fermion systems has been explored. This method corrects for finite-volume effects using a generic ansatz for Goldstone-boson-induced *round-the-world* effects. The Goldstone boson masses are then extrapolated to the chiral limit using information from chiral perturbation theory applied to gradient flow observables [29]. See Ref. [177] for recent developments regarding the CBFM for massive systems. If finite-volume effects are small, then extrapolating to the infinite volume limit may not be strictly necessary, depending on the desired precision; see, for example, Refs. [178, 179].

(3) **Continuum limit:** In Sec. 3.4, I stated that the continuum limit of a gauge-fermion system is taken by fixing the renormalized coupling and tuning g_0^2 to a critical surface. This is accomplished by extrapolating infinite-volume-extrapolated β-function $\beta_{\mathrm{GF}}\big(t; g_0^2\big)$ linearly in a^2/t at fixed $g_{\mathrm{GF}}^2 = g_{\mathrm{GF}}^2\big(t\big)$. Extrapolating in $a^2/t \to 0$ automatically tunes $g_0^2 \to 0$. Each $g_{\mathrm{GF}}^2(t)$ specifies a point on the renormalized trajectory in the right panel of Fig. 3.1. Hence, the continuum extrapolation can be thought of as projecting the RG flows that start off of the critical surface onto the renormalized trajectory.

Since its inception is Refs. [131, 189, 190], the CBFM has been applied to a variety of massless SU(3) gauge-fermion systems with $N_{\mathrm{f}} = 0$ [181, 294, 380], $N_{\mathrm{f}} = 2$ [177, 189, 190], $N_{\mathrm{f}} = 10$ [179, 235], and $N_{\mathrm{f}} = 12$ [180, 189] fermions in the fundamental representation of SU(3). In Chapters 5 and 7, I shall apply it to the $N_{\mathrm{f}} = 12$ and 8 systems, respectively. The CBFM has also been applied to the massless SU(4) gauge-fermion system with four fermions in the fundamental respresentation of SU(4) and another four fermions in the two-index antisymmetric representation of SU(4) [178].

4.4 Simulation details

The continuum β-function in this chapter is extracted from Hamiltonian (hybrid) Monte Carlo (HMC) simulations that utilize a tree-level improved Symanzik (Lüscher-Weisz) gauge action [112, 248, 251], as implemented in the `GRID` `C++` mathematical object library [47]. If the reader is unfamiliar with the HMC algorithm, consider reading Chapter 8.1.1. For a description of tree-level improved gauge actions, see Appendix F. All lattices used in this chapter are symmetric and possess periodic boundary conditions in all four directions. In Table 4.1, I list the full set of volumes $20 \leq L/a \leq 48$ (five total) and bare gauge couplings $4.3 \leq \beta_b \equiv 6/g_0^2 \leq 9.5$ (nineteen total), along with the total number of thermalized samples “No.” and the acceptance rate “Acc.” for each statistical ensemble $(L/a,\, \beta_b)$. The molecular dynamics trajectory length (see Chapter 8.1.1) is set to $\tau = 2$ and each statistical sample is separated by a total of 20 molecular dynamics time units (MDTU). Integration of the gradient flow equations (see Chapter 8.1.2), along with the measurement of gradient flow

	L/a									
	20		24		28		32		48	
β_b	Acc.	No.	Acc.	No.	Acc.	No.	Acc.	No.	Acc.	No.
4.30	87.9%	451	86.6%	467	85.3%	297	80.7%	165	···	···
4.35	86.5%	451	84.8%	458	82.0%	277	78.8%	171	···	···
4.40	86.6%	451	80.8%	460	83.7%	272	82.3%	167	···	···
4.50	85.1%	451	84.2%	501	86.2%	1391	81.0%	250	···	···
4.60	86.1%	451	85.2%	490	83.3%	1040	84.9%	202	···	···
4.70	84.2%	451	84.1%	490	80.5%	681	82.1%	201	···	···
4.80	86.5%	451	88.0%	469	80.5%	681	78.9%	140	···	···
4.90	85.0%	451	85.3%	491	82.7%	701	83.4%	163	···	···
5.00	82.6%	451	85.5%	456	81.0%	772	77.3%	211	80.8%	124
5.30	*84.4%*	*451*	*88.3%*	*534*	*82.9%*	*911*	*78.4%*	*656*	*81.8%*	*139*
5.50	83.6%	451	87.6%	456	81.8%	701	77.8%	608	78.2%	149
6.00	84.4%	451	84.6%	476	84.6%	661	79.2%	472	76.8%	227
6.50	81.1%	451	80.7%	486	82.8%	661	85.0%	563	77.4%	233
7.00	81.3%	451	79.2%	461	81.7%	701	84.6%	527	74.7%	241
7.50	82.6%	451	81.3%	466	80.5%	661	83.7%	489	73.6%	224
8.00	81.3%	451	78.3%	456	76.1%	701	85.0%	487	73.3%	211
8.50	78.8%	451	77.4%	461	79.5%	661	81.6%	462	74.6%	211
9.00	78.2%	451	76.8%	581	78.0%	524	81.6%	531	71.7%	208
9.50	77.4%	621	77.5%	481	77.7%	547	81.7%	541	69.2%	208

Table 4.1: From Ref. [181]. Total number of configurations “No.” and acceptance rate “Acc.” for each volume $(L/a)^4$ and bare gauge coupling $\beta_b \equiv 6/g_0^2$ that is used to extract the continuum renormalization group β-function for the pure Yang-Mills system. Each statistical sample (configuration) is separated by 20 molecular dynamics time units (see Chapter 8.1.1).

observables is performed using the `QLUA` software library [300].

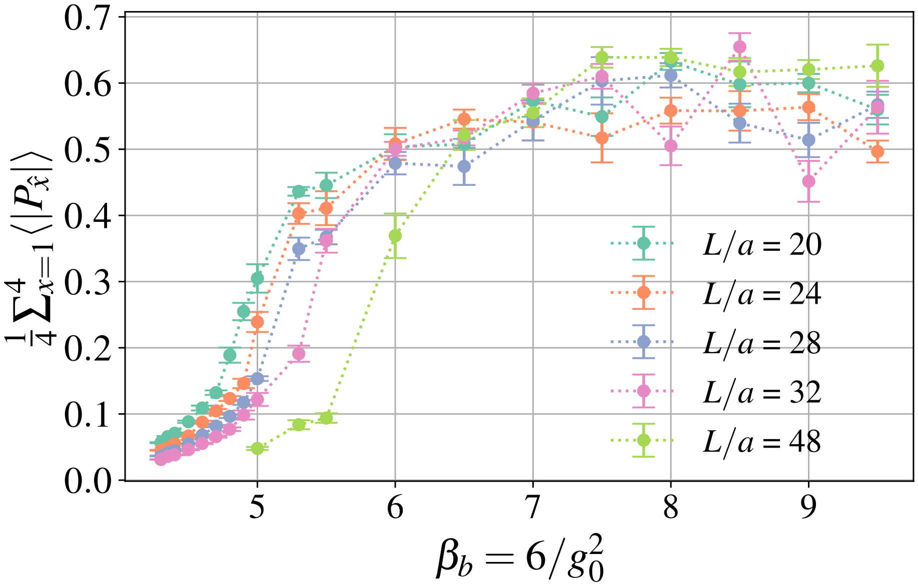

Figure 4.1: From Ref. [181]. The (direction-averaged) gradient-flowed Polyakov loop magnitude at flow time $8t = (L/2)^2$ against the bare gauge coupling on different volumes (indicated by different colors; see legend).

4.4.1 Physical regimes of the pure Yang-Mills system in a finite box

As the coupling β_b increases at fixed L/a, the pure Yang-Mills system transitions from exhibiting signs of confinement to being deconfined. It is important to note, however, that there is no zero-temperature deconfinement phase transition in the pure Yang-Mills system. Instead, the smooth transition between the two regimes reflects a difference in the physics that occurs at short distances (small physical volumes, ϵ-regime), where the system is asymptotically free, and the physics that occurs at long distances (large physical volumes, p-regime), where the system is confining. As a consequence of *center symmetry breaking*, the temporal *Polyakov loop* (a.k.a. *thermal Wilson line*) [36]

$$P_{\hat{4}} = \frac{1}{N(L/a)^3} \sum_{\mathbf{n}} \mathrm{Tr}_{\mathrm{c}} \prod_{n_4=0}^{L/a-1} \Big[\mathcal{U}_4(n)\Big]_{n=(\mathbf{n},n_4)} \tag{4.32}$$

is an order parameter for the first-order deconfinement phase transition that the pure Yang-Mills system exhibits *at finite-temperature* [55, 153, 310]. One can define the Polyakov loop similarly in any direction; therefore, I will refer to it as $P_{\hat{x}}$ for $x = 1, 2, 3, 4$ henceforth. At zero-temperature, the expectation value of the Polyakov loop magnitude $\langle |P_{\hat{x}}| \rangle$ is *exactly* zero in the infinite volume limit. In a finite volume, the zero temperature $\langle |P_{\hat{x}}| \rangle$ approaches zero as the size of the volume increases. Because small physical volumes[2] probe only short-distance properties of the Yang-Mills system, they are governed by asymptotic freedom. The physics of confinement should be manifest in the large physical volume regime; hence, the expectation value of the Polyakov loop magnitude should approach zero in this regime. Because the Polyakov loop expectation is merely reflecting the infrared properties of the zero-temperature pure Yang-Mills system that are accessible from a particular physical volume, the transition between the two regimes must be smooth. Moreover, the infinite volume limit of observables that are measured in either regime must be consistent over the physical scales that they are capable of capturing. Put another way, there is no phase transition that separates one finite-volume zero-temperature regime from another in the infinite volume limit; therefore, the phase of their zero-temperature thermodynamic limit must be consistent.

Estimating the Polyakov loop from lattice simulations can be quite difficult, as it exhibits considerable statistical noise. Luckily, gradient flow can be used to dramatically reduce the statistical noise in the Polyakov loop [295]. In Fig. 4.1, I plot the magnitude of the gradient-flowed Polyakov loop averaged over all four directions at gradient flow time $8t = (L/2)^2$. As expected, the Polyakov loop increases with β_b at fixed L/a and varies inversely with L/a at fixed $\beta_b \lesssim 4.0$, where all volumes are safely in the confined regime. Moreover, because the simulations in this chapter are performed in a finite box, the Polyakov loop is never exactly zero. There is a sudden, yet smooth, jump in the Polyakov loop on all volumes around $5.0 \lesssim \beta_b \lesssim 6.0$. The region about which the jump occurs corresponds to the range in g_0^2 over which the volumes that have been simulated in this chapter traverse the confinement scale.

The physics of confinement could be intimately tied to the presence of instantons in the vacuum of gauge-fermion systems[3] [326]. Lending support to this idea is the observation of an increase in

---

[2] By "physical", I am referring to the size of the volume expressed in units of the scale at which confinement kicks in.

[3] To a physicist, instantons are particle-like local optima of the classical action. See Ref. [326] for a classic overview of instanton physics and Ref. [138] for more modern treatment. For a more mathematically precise treatment of instantons, see Refs. [41, 270].

statistical fluctuations of the topological charge once the physical volume crosses the confinement scale. According to the Atiyah-Singer index theorem [24], the continuum topological charge

$$Q_{\text{top.}}^{(\text{cont.})} = \frac{1}{32\pi^2}\epsilon_{\mu\nu\rho\sigma}\int \mathrm{d}^4x\ \mathrm{Tr}_{\mathrm{c}}\left[\mathfrak{F}_{\mu\nu}(x)\mathfrak{F}_{\rho\sigma}(x)\right] \tag{4.33}$$

is related to the number of instantons n_+ and anti-instantons n_- as

$$Q_{\text{top.}}^{(\text{cont.})} = n_+ - n_- \in \mathbb{Z}. \tag{4.34}$$

Different integral values for Q are referred to as *topological sectors*. An analogue of the topological charge on a hypercubic spacetime lattice can be defined in terms of the clover discretization of $\mathfrak{F}_{\mu\nu}(x)$ in Eqn. 4.21 as

$$Q_{\text{top.}}^{(\text{lat.})} = \frac{1}{32\pi^2}\epsilon_{\mu\nu\rho\sigma}\sum_{n}\mathrm{Tr}_{\mathrm{c}}\left[\mathcal{F}_{\mu\nu}^{(\text{clov.})}(n)\mathcal{F}_{\rho\sigma}^{(\text{clov.})}(n)\right] \tag{4.35}$$

or any other lattice discretization of $\mathfrak{F}_{\mu\nu}(x)$ [4]. As was the case for the Polaykov loop, measuring the topological charge from lattice simulations is extremely difficult without the use of noise reduction techniques. Moreover, since the Atiyah-Singer index theorem applies to *continuum* gauge fields, the lattice discretization of the topological charge is not necessarily an integer. Once again, gradient flow comes to the rescue, as it acts to suppress the short-distance fluctuations that give rise to the statistical noise in $Q_{\text{top.}}^{(\text{lat.})}$ and which prevent individual gauge configurations from at least approximately realizing the index theorem [252]. Note that this is not because gradient flow is somehow reducing the lattice spacing. Rather, the gradient flow takes the topological charge closer to the renormalized trajectory (“longer distances”), where cutoff effects are suppressed and the lattice discretization of the topological charge converges to its integral continuum counterpart.

Corroborating the onset of confinement in the simulations of this chapter is the presence of fluctuations in gradient-flowed topological charge at $8t = (L/2)^2$ for $\beta_b \lesssim 5.0$. As β_b decreases, the frequency and magnitude of tunneling events between different topological sectors increases. For the purposes of illustration, I have plotted Monte Carlo time history of the topological charge on a

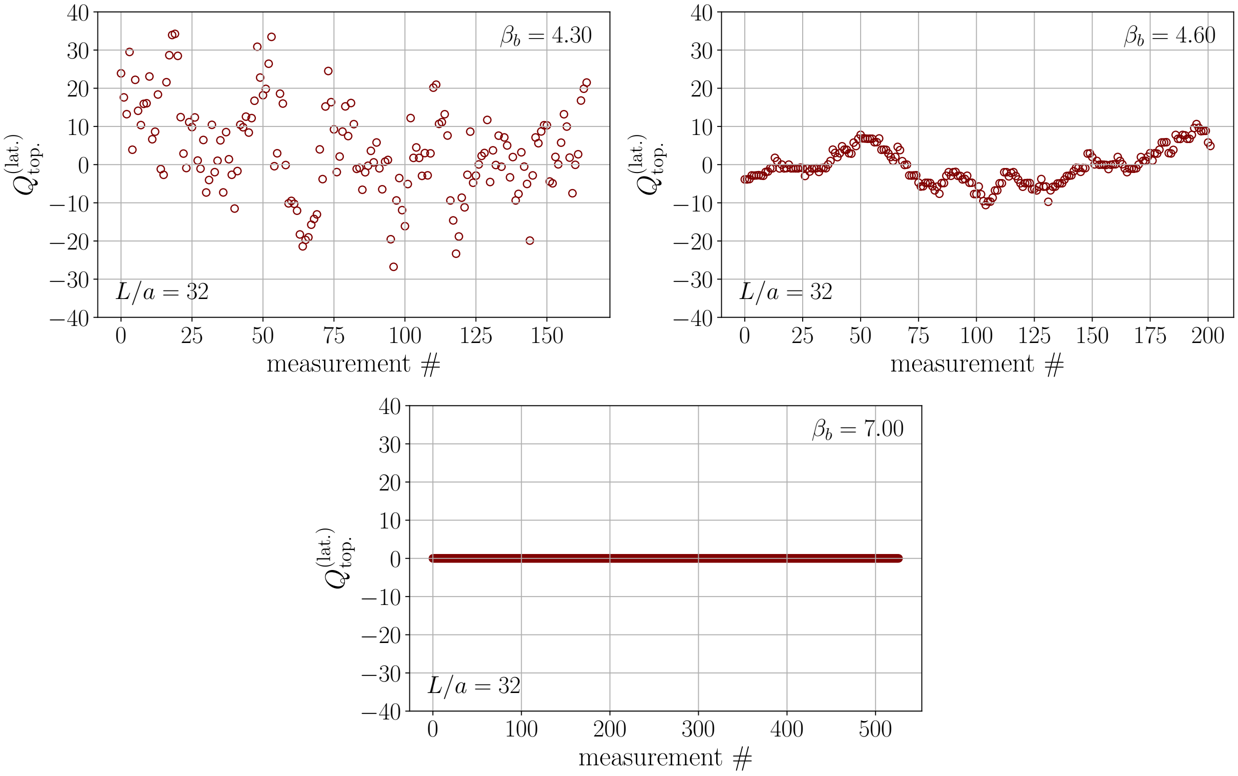

Figure 4.2: The Monte Carlo time history of the gradient-flowed topological charge at flow time $t = (L/2)^2$ on a box of size $(L/a)^4 = 32^4$ at $\beta_b = 4.3$ (top left panel), 4.6 (top right panel), and 7.0 (bottom panel).

$(L/a)^4 = 32^4$ volume at $\beta_b = 4.3, 4.6$ and 7.0 in Fig. 4.2. At the strongest coupling $\beta_b = 4.3$, the topological charge fluctuates in the range $-40 \lesssim Q^{(\mathrm{lat.})}_{\mathrm{top.}} \lesssim 40$. Closer to the region where $L/a = 32$ traverses the confinement scale at $\beta_b = 4.6$, the topological charge only fluctuates in the range $-10 \lesssim Q^{(\mathrm{lat.})}_{\mathrm{top.}} \lesssim 10$ and it takes longer for the topological charge to tunnel between sectors. The increase in the Monte Carlo time that it takes for the system tunnel between topological sectors as β_b increases is a consequence of *critical slowing down*. Well within the deconfined regime at $\beta_b = 7.0$, the topological charge stays at $Q^{(\mathrm{lat.})}_{\mathrm{top.}} \approx 0$. Due to the fast running of the pure Yang-Mills system, the impact of a non-zero topological charge is not resolved statistically. This is quite different from what is observed in slow-running systems, where tunneling between topological sectors can have a significant impact on $g^2_{\mathrm{GF}}(t; L, g_0^2)$ [191]. Furthermore, following the recommendation of the step-scaling study of Ref. [92], I have estimated $g^2_{\mathrm{GF}}(t; L, g_0^2)$ from filtered configurations with

$Q_{\text{top.}}^{(\text{lat.})} \approx 0$ and no significant impact on the resulting estimate for $g_{\text{GF}}^2(t;L,g_0^2)$ is observed, aside from a significant increase in the statistical error.

4.5 Calculation of the continuum β-function

Following the first step of the CBFM laid out in Sec. 4.3, I calculate the renormalized coupling $g_{\text{GF}}^2(t;L,g_0^2)$ using Eqn. 4.24 with $\delta(t,L)$ given by either Eqn. 4.16 or Eqn. 4.27. The operator $\mathcal{S}^{\text{e}}$ from which I estimate the Yang-Mills energy density $E(t;L)$ is either the Wilson operator of Eqn. 3.8 ($\mathcal{S}^{\text{e}} = \text{W}$), the Symanzik operator defined in Appendix F ($\mathcal{S}^{\text{e}} = \text{S}$), or the clover operator defined in terms of Eqn. 4.21 ($\mathcal{S}^{\text{e}} = \text{C}$).The flow action $\mathcal{S}^{\text{f}}$ is either Wilson flow $\mathcal{S}^{\text{f}} = \text{W}$ or Zeuthen flow $\mathcal{S}^{\text{f}} = \text{Z}$, which is defined in Chapter 8.1. I refer to a specific combination of flow $\mathcal{S}^{\text{f}}$ and $E(t,L)$ operator $\mathcal{S}^{\text{e}}$ as "$\mathcal{S}^{\text{f}}\mathcal{S}^{\text{e}}$"; e.g., the Zeuthen flow action for $\mathcal{S}^{\text{e}}$ with Symanzik action for $\mathcal{S}^{\text{e}}$ is "ZS". If tree-level corrections are included in $\delta(t,L)$ (Eqn. 4.24), then an "n" is prepended to "$\mathcal{S}^{\text{f}}\mathcal{S}^{\text{e}}$"; e.g., "nZS". From $g_{\text{GF}}^2(t;L,g_0^2)$, I calculate the RG β-function in a finite volume as

$$\beta_{\text{GF}}(t;L,g_0^2) = -t\frac{\mathrm{d}}{\mathrm{d}t}g_{\text{GF}}^2(t;L,g_0^2), \tag{4.36}$$

where the derivative $\mathrm{d}/\mathrm{d}t$ is discretized using a 5-point stencil [302]; explicitly,

$$\begin{aligned} &12\delta t\frac{\mathrm{d}}{\mathrm{d}t}g_{\text{GF}}^2(t;L,g_0^2) \\ &\quad\approx -g_{\text{GF}}^2(t+2\delta t;L,g_0^2) + 8g_{\text{GF}}^2(t+\delta t;L,g_0^2) - 8g_{\text{GF}}^2(t-\delta t;L,g_0^2) + g_{\text{GF}}^2(t-2\delta t;L,g_0^2). \end{aligned} \tag{4.37}$$

I set value for $\delta t = 0.04$ by the time step "ϵ" that is used to integrate the gradient flow equations (see Chapter 8.1.2). I have checked to ensure that the estimate of $\beta_{\text{GF}}(t;L,g_0^2)$ from Eqns. 4.36-4.37 does not change significantly if I use a high-order stencil. I have also checked that lower-order stencils converge comfortably to the 5-point stencil of Eqn. 4.37. A time step of $\delta t = 0.01$ has also been run on select ensembles as a crosscheck of the choice for ϵ used in this chapter and no significant change in $E(t,L)$ is observed. Correlated statistical uncertainties are estimated and kept track of using the automatic error propagation tools of the `gvar` library [241]. All fits in the rest of this chapter are performed using the `lsqfit` library [240], which utilizes the robust *trust region reflective* algorithm implemented in the `SciPy` library to estimate fit parameters via *maximum a posteriori* (MAP) estimation [49, 361]. Correlated uncertainties in fit parameters are estimated via Laplace approximation of the posterior distribution over the fit parameters about their MAP estimate; see Appendix D and Chapter 8.2 for further details.

4.5.1 Infinite volume extrapolation

As the Yang-Mills energy density is a dimension-4 operator, finite-volume effects in $g^2_{\rm GF}\big(t; L, g_0^2\big)$ and $\beta_{\rm GF}\big(t; L, g_0^2\big)$ are expected to be $\mathrm{O}\big(t^2/L^4\big)$. By fixing both $\beta_b = 6/g_0^2$ (i.e., the lattice spacing) and the flow time in lattice units t/a^2, both $g^2_{\rm GF}\big(t; L, g_0^2\big)$ and $\beta_{\rm GF}\big(t; L, g_0^2\big)$ scale with the linear extent of the lattice L/a according the generic ansatz

$$\mathrm{FV}\big(t; L, g_0^2\big) = k_1\big(t; g_0^2\big) + k_2\big(t; g_0^2\big)(a/L)^4 \qquad \text{(fixed } t/a^2 \text{ and } \beta_b\text{)}, \tag{4.38}$$

at leading order in t/L. The "constants" $k_1\big(t; g_0^2\big)$ and $k_2\big(t; g_0^2\big)$ are fixed for any specific $(t/a^2, \beta_b)$ pair. I extrapolate $g^2_{\rm GF}\big(t; L, g_0^2\big)$ and $\beta_{\rm GF}\big(t; L, g_0^2\big)$ to the $a/L \to 0$ limit by independently fitting both of them to the ansatz of Eqn. 4.38. In Fig. 4.3, I illustrate typical infinite volume extrapolations for $g^2_{\rm GF}\big(t; L, g_0^2\big)$ (left panels) and $\beta_{\rm GF}\big(t; L, g_0^2\big)$ (right panels) at one "weak coupling" $\beta_b = 6.00$, two "intermediate couplings" $\beta_b = 5.5, 4.9$, and one "strong coupling" $\beta_b = 4.35$ (top panel to bottom panel). The renormalized couplings and correspond β-functions in Fig. 4.3 are derived from the "nZS" combination. Each panel shows the infinite volume extrapolation at five fixed flow times $t/a^2 = 2.0$ (yellow), 2.52 (green), 3.0 (blue), 3.52 (purple), and 4.0 (red).

In the strong coupling regime, confinement introduces an additional infrared scale. As such, finite volume effects in the strong coupling regime are suppressed. Therefore, I utilize the full set of volumes $20 \leq L/a \leq 32$ over $4.3 \leq \beta_b \leq 4.9$ (see Table. 4.1), as illustrated in the bottom two panels of Fig. 4.3. In the weak coupling regime, where the volume furnishes the only infrared

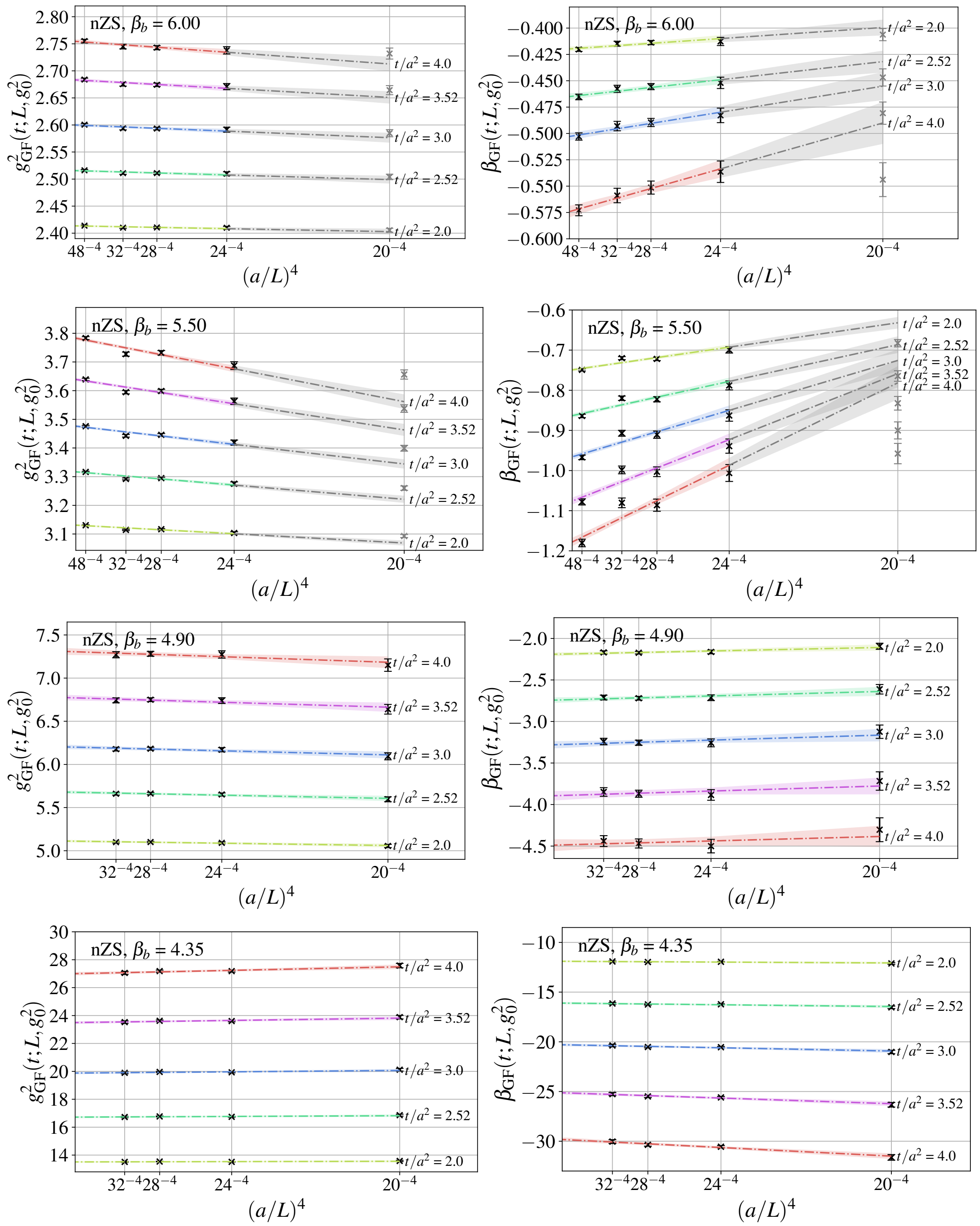

Figure 4.3: From Ref. [181]. Illustration of the infinite volume extrapolation of $g^2_{\mathrm{GF}}(t; L, g_0^2)$ (left panels) and $\beta_{\mathrm{GF}}(t; L, g_0^2)$ (right panels) at $\beta_b = 6.00, 5.50, 4.90, 4.35$ (top to bottom). Colored bands indicate the statistical error of the extrapolating curve and the central value is indicates the central value. Black markers (with error bars) contribute to the fit, whereas grey markers are not included in the fit and are shown for the purposes of illustration.

scale, finite-volume effects are more pronounced. This can be seen in the top panel of Fig. 4.3, where it is clear that the $L/a = 20$ simulations deviate significantly from the linear trend in a^4/L^4. Therefore, I drop the $L/a = 20$ volumes and utilize an additional set of $L/a = 48$ volumes in the weak coupling regime ($6.0 \lesssim \beta_b \leq 9.5$, with the exception of $\beta_b = 6.0$ on $L/a = 20$). In the transition region ($4.6 \lesssim \beta_b \lesssim 6.0$ for $20 \lesssim L/a \lesssim 40$), the integrated autocorrelation time for $g^2_{\mathrm{GF}}\left(t; L, g_0^2\right)$ and $\beta_{\mathrm{GF}}\left(t; L, g_0^2\right)$ increases significantly from a maximum for 80 MDTU outside of the transition region to a maximum of 300 MDTU within the transition region. Such long integrated autocorrelation times likely lead to underestimated statistical errors, even after they have been explicitly taken into account via binning.

Overall, most fits have acceptable p-values ($p \gtrsim 10\%$). See Chapter 8.2 for a refresher on goodness-of-fit; namely, the notion of a p-value. Notable exceptions occur at $\beta_b = 5.0, 5.3$ and 5.5, which are well within the transition region. The p-values for extrapolations at $\beta_b = 5.0$ are in the $1.4\% \lesssim p \lesssim 3.1\%$ range. The infinite volume extrapolations at $\beta_b = 5.0$ are the only extrapolations that utilize all five volumes; however, the extrapolations at β_b are dominated by the largest volumes. Infinite volume extrapolations at $\beta_b = 5.3$ possess nearly vanishing p-values. Significantly increasing the number of statistical samples for each ensemble at $\beta_b = 5.3$, especially those that are solidly within the transition region, does not improve the p-value for the extrapolations at $\beta_b = 5.3$. Therefore, $\beta_b = 5.3$ likely suffers from sitting on top of the transition region and is dropped from the central analysis. The p-values for infinite volume extrapolations at $\beta_b = 5.5$ are in the $0.14\% \lesssim p \lesssim 0.4\%$ range. The extrapolation at $\beta_b = 5.5$ is illustrated in the second-to-top panel of Fig. 4.3. Such small p-values are possibly attributable to underestimated errors on $L/a = 32$, as it generally deviates from the linear trend in a^4/L^4 at the $\approx 2\sigma$ level.

4.5.2 The effect of tree-level improvement

Including corrections for tree-level cutoff effects in $\delta(t, L)$ (Eqn. 4.27) has a significant impact on the consistency of estimates for $g^2_{\mathrm{GF}}\left(t; L, g_0^2\right)$ from different $\mathcal{S}^{\mathrm{f}}\mathcal{S}^{\mathrm{e}}$ combinations in the weak coupling regime, and, surprisingly, in the strong coupling regime, though to a lesser extent. Fig.

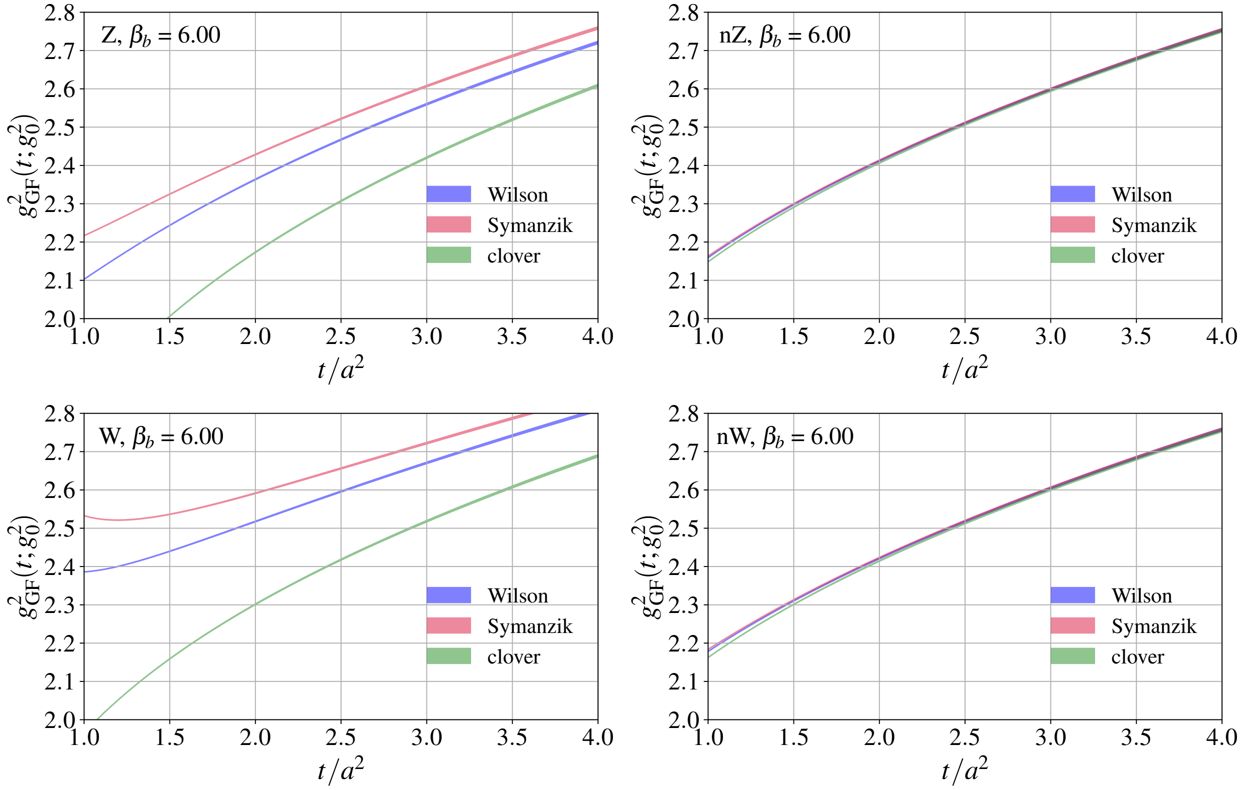

Figure 4.4: From Ref. [181]. Infinite-volume-extrapolated gradient flow coupling $g^2_{\rm GF}(t; g_0^2)$ at $\beta_b = 6.0$ without tree-level corrections (left panels) and with tree-level corrections (right panels) against t/a^2. Wilson operator indicated by blue back, Symanzik by red, and clover by green. Zeuthen flow shown in top panels and Wilson flow shown in bottom panels.

4.4 illustrates the effect of tree-level corrections on the infinite-volume-extrapolated GF coupling $g^2_{\rm GF}(t; g_0^2)$ at weak coupling ($\beta_b = 6.0$) for $\mathcal{S}^{\rm f} = {\rm Z}$ (top panels) and $\mathcal{S}^{\rm f} = {\rm W}$ (bottom panels). Different colors indicate different operators $\mathcal{S}^{\rm e}$, with $\mathcal{S}^{\rm e} = {\rm W}$ (Wilson, blue), $\mathcal{S}^{\rm e} = {\rm S}$ (Symanzik, red), and $\mathcal{S}^{\rm e} = {\rm C}$ (clover, green). Similarly, Fig. 4.5 illustrates the effect of tree-level corrections on $g^2_{\rm GF}(t; g_0^2)$ at strong coupling ($\beta_b = 4.35$) for the same set of flow/operator combinations $\mathcal{S}^{\rm f}\mathcal{S}^{\rm e}$ as in Fig. 4.4.

At weak coupling, where the perturbatively calculated tree-level corrections should have the most significant impact on the estimate of $g^2_{\rm GF}(t; g_0^2)$, there is a substantial gain in consistency between different $\mathcal{S}^{\rm f}\mathcal{S}^{\rm e}$ (right panels of Fig. 4.4) in comparison to unimproved estimates of $g^2_{\rm GF}(t; g_0^2)$ (left panels of Fig. 4.4). This is holds for the entire range of t/a^2 that I use in the continuum

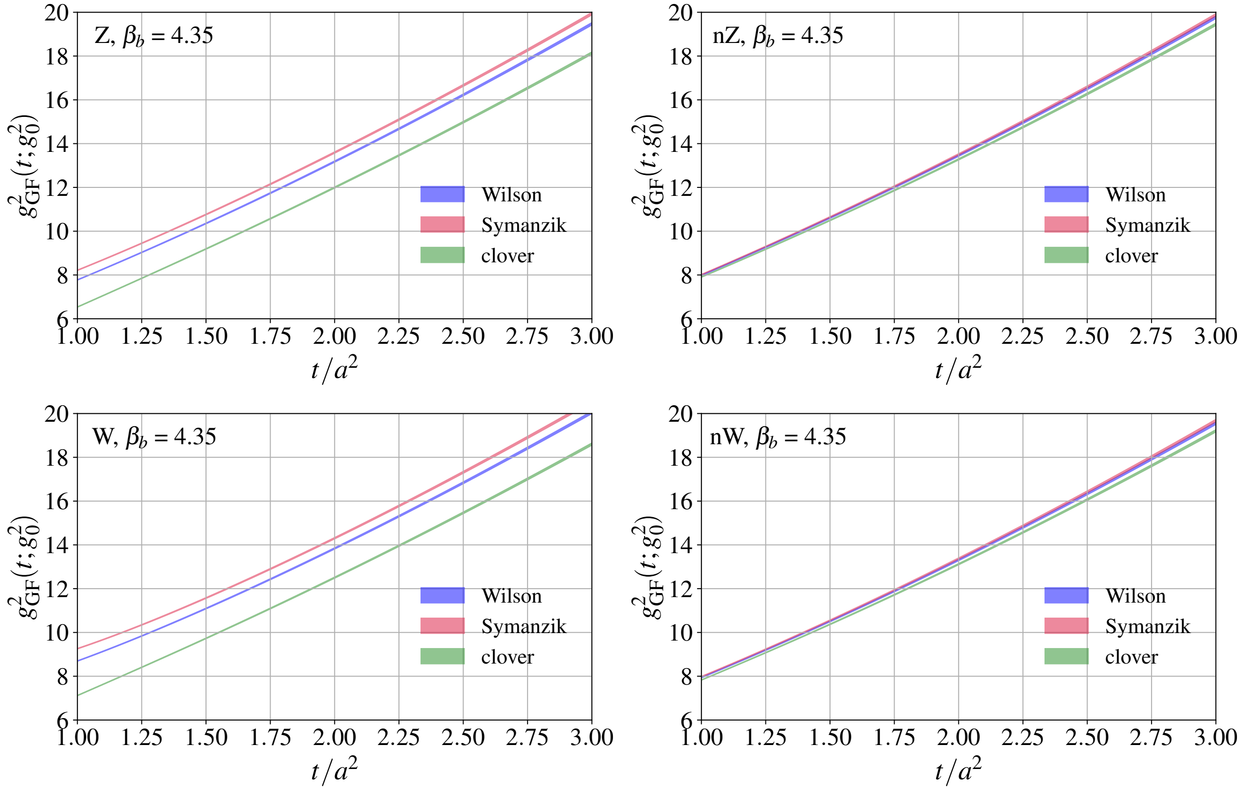

Figure 4.5: From Ref. [181]. Infinite-volume-extrapolated gradient flow coupling $g^2_{\mathrm{GF}}(t; g_0^2)$ at $\beta_b = 4.35$ without tree-level corrections (left panels) and with tree-level corrections (right panels) against t/a^2. Wilson operator indicated by blue back, Symanzik by red, and clover by green. Zeuthen flow shown in top panels and Wilson flow shown in bottom panels.

extrapolations of Sec. 4.5.3, though there is a slight gain in consistency as t/a^2 increases. Tree level corrections also improve the consistency of different $\mathcal{S}^{\mathrm{f}}\mathcal{S}^{\mathrm{e}}$ combinations at strong coupling, though the improvement is less dramatic than it is at weak coupling. It also degrades with increasing t/a^2. That tree-level corrections provide any considerable level of improvement at such strong couplings ($6 \lesssim g^2_{\mathrm{GF}}(t; g_0^2) \lesssim 20$) is surprising, though the observation a degradation of the improvement as t/a^2 increases is less so. In any case, the significant effect that tree-level corrections on the estimate of $g^2_{\mathrm{GF}}(t; g_0^2)$ over the entire range of renormalized couplings that I explore in this chapter justifies their exclusive use throughout the rest of this chapter. Henceforth, all estimates of $g^2_{\mathrm{GF}}(t; g_0^2)$ will utilize tree-level corrections. It has been checked that unimproved $g^2_{\mathrm{GF}}(t; g_0^2)$ yield consistent results for the continuum β-function (though with a considerable increase in error). Furthermore, based

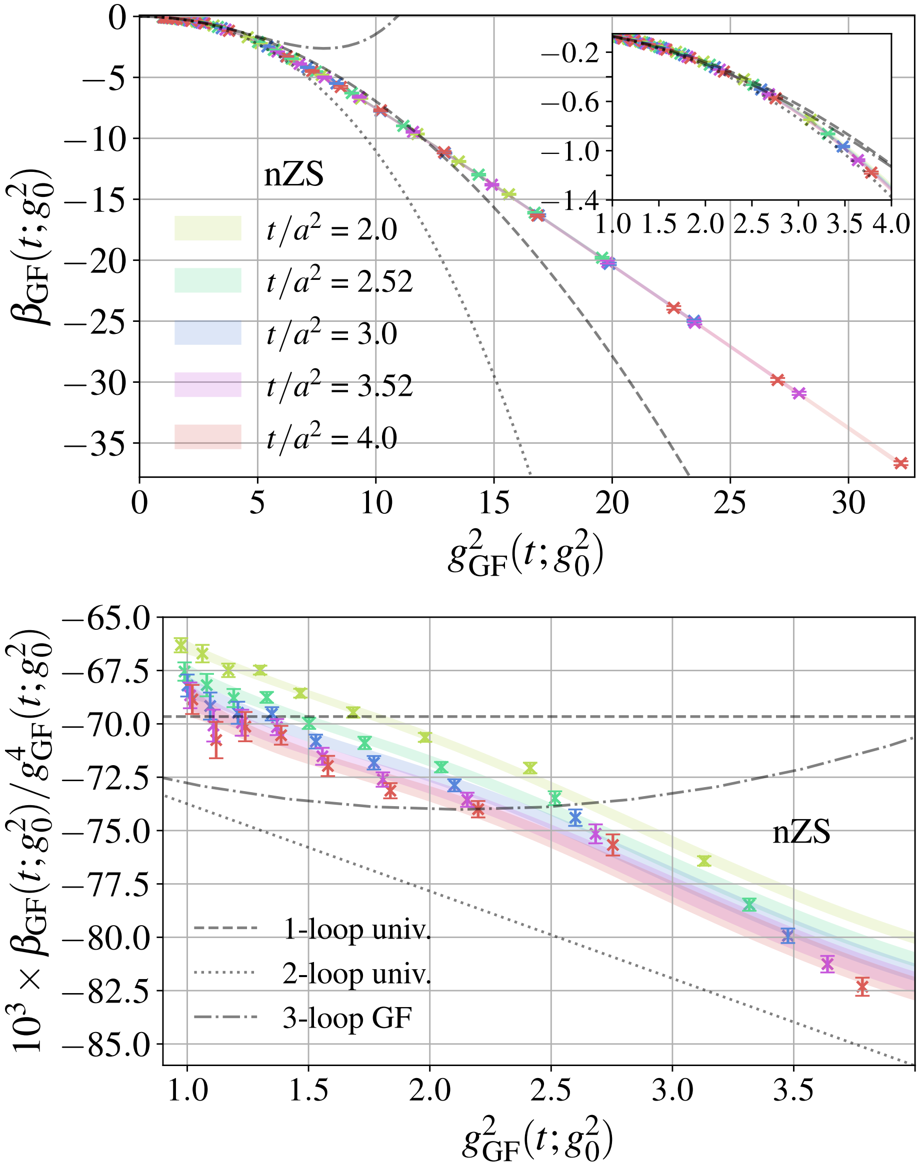

Figure 4.6: From Ref. [181]. Interpolation of $\beta_{\rm GF}(t;g_0^2)$ in $g^2_{\rm GF}(t;g_0^2)$ at fixed $t/a^2 = 2.0$ (yellow), 2.52 (green), 3.0 (blue), 3.52 (purple), and 4.0 (red). Different colored symbols correspond to $(g^2_{\rm GF}(t;g_0^2), \beta_{\rm GF}(t;g_0^2))$ pairs at $2.0 \leq t/a^2 \leq 4.0$. Colored bands indicate the interpolated β-function from Eqn. 4.42. In the lower panel, I plot $\beta_{\rm GF}(t;g_0^2)/g^4_{\rm GF}(t;g_0^2)$ to enhance the weak coupling regime.

on the improvement from tree-level corrections observed in Figs. 4.5-4.5, the central results of this chapter will be based on the nZS combination; however, I will return to the systematic uncertainty that is associated with choosing a particular $\mathcal{S}^{\rm f}\mathcal{S}^{\rm e}$ combination in Sec. 4.5.4.

4.5.3 Continuum extrapolation

The continuum extrapolation at a particular g^2_{GF} is taken by fixing $g^2_{\mathrm{GF}} = g^2_{\mathrm{GF}}(t)$ and extrapolating $\beta_{\mathrm{GF}}\big(t; g_0^2\big)$ to the $a^2/t \to 0$ limit. As the leading discretization effects in $\beta_{\mathrm{GF}}\big(t; g_0^2\big)$ are $\mathrm{O}\big(a^2/t\big)$, achieving the latter goal requires obtaining a collection of pairs $\big(\beta_{\mathrm{GF}}\big(t; g_0^2\big), t/a^2\big)$ at each fixed g^2_{GF} over which the continuum extrapolation is performed. For each fixed g^2_{GF}, the β-function $\beta_{\mathrm{GF}}\big(t; g_0^2\big)$ is extrapolated to the continuum $a^2/t \to 0$ limit by fitting the $\big(\beta_{\mathrm{GF}}\big(t; g_0^2\big), t/a^2\big)$ at that g^2_{GF} to the ansatz

$$\beta_{\mathrm{GF}}\big(t; g_0^2\big) = \beta_{\mathrm{GF}}\big(g^2_{\mathrm{GF}}\big) + \kappa(t)(a^2/t) \qquad (\text{fixed } g^2_{\mathrm{GF}}), \tag{4.39}$$

where $\kappa(t)$ is a t-dependent "constant" and $\beta_{\mathrm{GF}}\big(g^2_{\mathrm{GF}}\big)$ is the continuum β-function at g^2_{GF}. Note that there is a hidden assumption here, being that the leading discretization effects in $g^2_{\mathrm{GF}}\big(t; g_0^2\big)$ are suppressed compared to those in $\beta_{\mathrm{GF}}\big(t; g_0^2\big)$ or, at least, that they can be absorbed into $\kappa(t)$ when extrapolating $\beta_{\mathrm{GF}}\big(t; g_0^2\big)$ to the continuum limit with Eqn. 4.39. If this is not the case, then there is an ambiguity in the $a^2/t \to 0$ limit due to t not necessarily being fixed along the continuum limit. In practice, this potential issue does not appear to affect the continuum extrapolation, so long as there is sufficient control over cutoff effects; though it should be kept in mind in any application of the CBFM.

4.5.3.1 Intermediate interpolation

To obtain the $\big(\beta_{\mathrm{GF}}\big(t; g_0^2\big), t/a^2\big)$ pairs that are required to take the continuum limit at *any* fixed g^2_{GF}, I must interpolate $\beta_{\mathrm{GF}}\big(t; g_0^2\big)$ in $g^2_{\mathrm{GF}}\big(t; g_0^2\big)$ over available t/a^2 entering the continuum limit. Obtaining the necessary $\big(\beta_{\mathrm{GF}}\big(t; g_0^2\big), t/a^2\big)$ pairs then follows by choosing a set of t/a^2 over which one wishes to perform the continuum extrapolation, fixing g^2_{GF}, then obtaining $\beta_{\mathrm{GF}}\big(t; g_0^2\big)$ at that fixed $g^2_{\mathrm{GF}} = g^2_{\mathrm{GF}}\big(t; g_0^2\big)$ from each fixed t/a^2 interpolation. The interpolation at each fixed t/a^2 must be capable of accommodating the differing curvatures of $\beta_{\mathrm{GF}}\big(t; g_0^2\big)$ in $g^2_{\mathrm{GF}}\big(t; g_0^2\big)$ in the weak coupling regime ($g^2_{\mathrm{GF}}\big(t; g_0^2\big) \lesssim 4\pi$) and the strong coupling regime ($g^2_{\mathrm{GF}}\big(t; g_0^2\big) \gtrsim 4\pi$).

Once t/a^2 is large enough to have reached the vicinity of the renormalized trajectory, the dependence of the weak coupling β-function $\beta_{\mathrm{GF}}\left(t;g_0^2\right)$ on $g_{\mathrm{GF}}^2\left(t;g_0^2\right)$ should converge asymptotically to its one-loop counterpart (see Chapter 1) as

$$\beta_{\mathrm{GF}}\left(t;g_0^2\right) \sim -b_0\left(g_{\mathrm{GF}}^2\left(t;g_0^2\right)/4\pi\right)^2 \qquad \left(g_{\mathrm{GF}}^2\left(t;g_0^2\right)\to 0\right) \tag{4.40}$$

up to discretization effects. See Eqn. 1.9 for a definition of the one-loop universal constant b_0. In the strong coupling regime, the β-function is observed empirically to be linear in $g_{\mathrm{GF}}^2\left(t;g_0^2\right)$; that is,

$$\beta_{\mathrm{GF}}\left(t;g_0^2\right) \propto g_{\mathrm{GF}}^2\left(t;g_0^2\right) \qquad \left(g_{\mathrm{GF}}^2\left(t;g_0^2\right) \gtrsim 4\pi\right). \tag{4.41}$$

See, for example, the top panel of Fig. 4.6, which I will discuss in detail soon. This linearity in the β-function survives the continuum limit and I will discuss it further in Sec. 4.5.5. A simple polynomial is not capable of describing such desperate regimes in the curvature without also introducing a number of systematic effects that are difficult to control for. A better approach is to construct an interpolating function that matches onto Eqn. 4.40 at weak coupling, Eqn. 4.41 at strong coupling, and everything else in-between. This is achieved with an N-order ratio of polynomials $\mathcal{I}_N$ of the form

$$\mathcal{I}_N\left(g_{\mathrm{GF}}^2\right) \equiv \frac{-p_0 g_{\mathrm{GF}}^4\left(1+\sum_{i=1}^{N} p_i g_{\mathrm{GF}}^{2i}\right)}{1+\sum_{j=1}^{N+1} q_j g_{\mathrm{GF}}^{2j}}. \tag{4.42}$$

The leading power of g_{GF}^2 in the numerator of Eqn. 4.42 explicitly forces $\mathcal{I}_N\left(g_{\mathrm{GF}}^2\right)\to -p_0 g_{\mathrm{GF}}^4$ as $g_{\mathrm{GF}}^2\to 0$. The order of the polynomial in the denominator forces $\mathcal{I}_N\left(g_{\mathrm{GF}}^2\right)\propto g_{\mathrm{GF}}^2$ when $g_{\mathrm{GF}}^2/4\pi \gg 1$. Because discretization effects affect the asymptotic behavior of the β-function, the ansatz of Eqn. 4.42 may have the effect of cutting off the smallest values of t/a^2 that could viably enter the continuum extrapolation. Nonetheless, the lowest order N that reasonably fits the data is $N=4$, which yields p-values in the $17\% - 32\%$ range for $2.0 \le t/a^2 \le 4.0$.

In the top panel of Fig. 4.6, I show the result of the intermediate interpolation over the

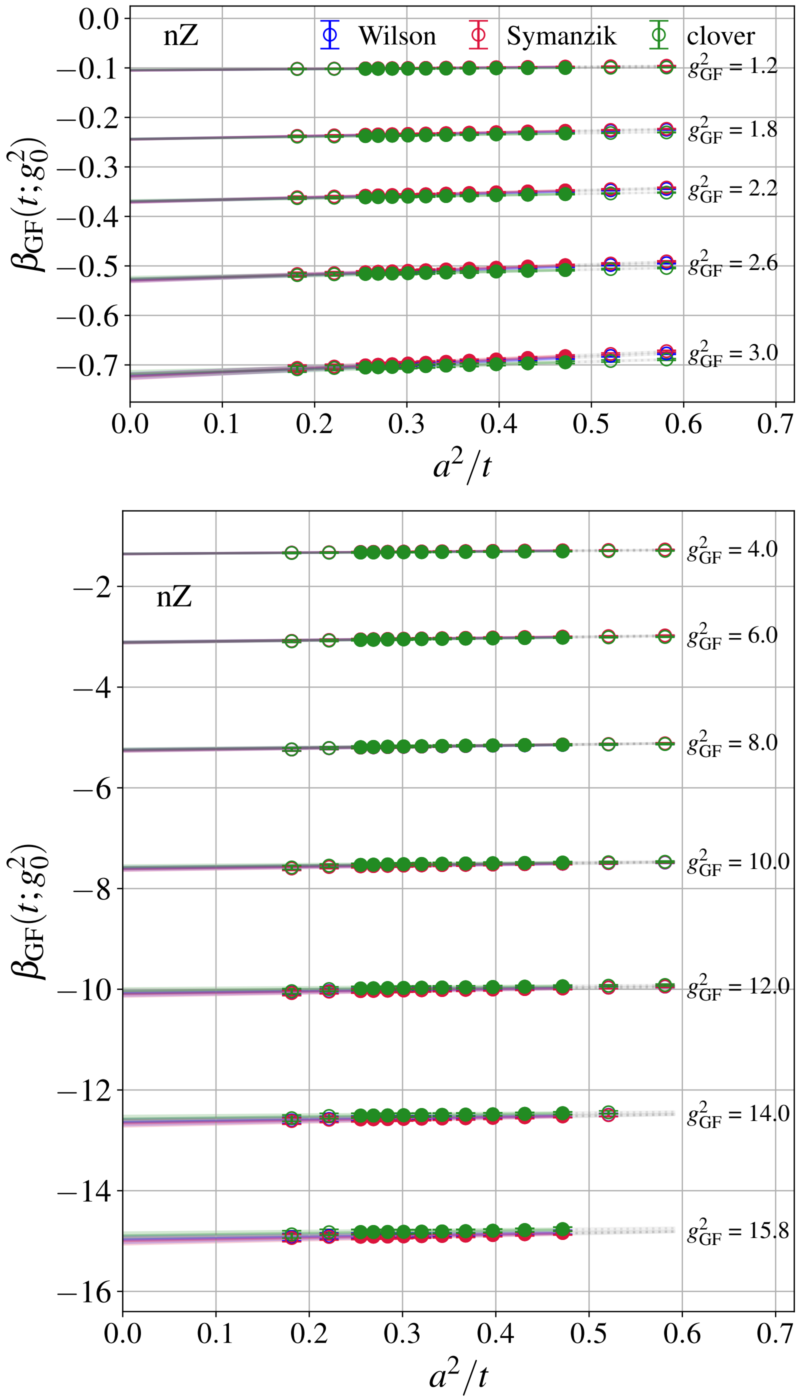

Figure 4.7: From Ref. [181]. Continuum $a^2/t \to 0$ extrapolation between $g^2_{\rm GF} = 1.2$ to 15.8. The results are shown in two separate panels to accommodate the increasingly faster running of the coupling. In all cases, we show all three operators with Zeuthen flow after tln improvement, though the different operators overlap and are barely distinguishable in the plot. The open symbols are not included in the extrapolation fit. They are shown to illustrate the linear behavior of the data even outside the region used in the fit.

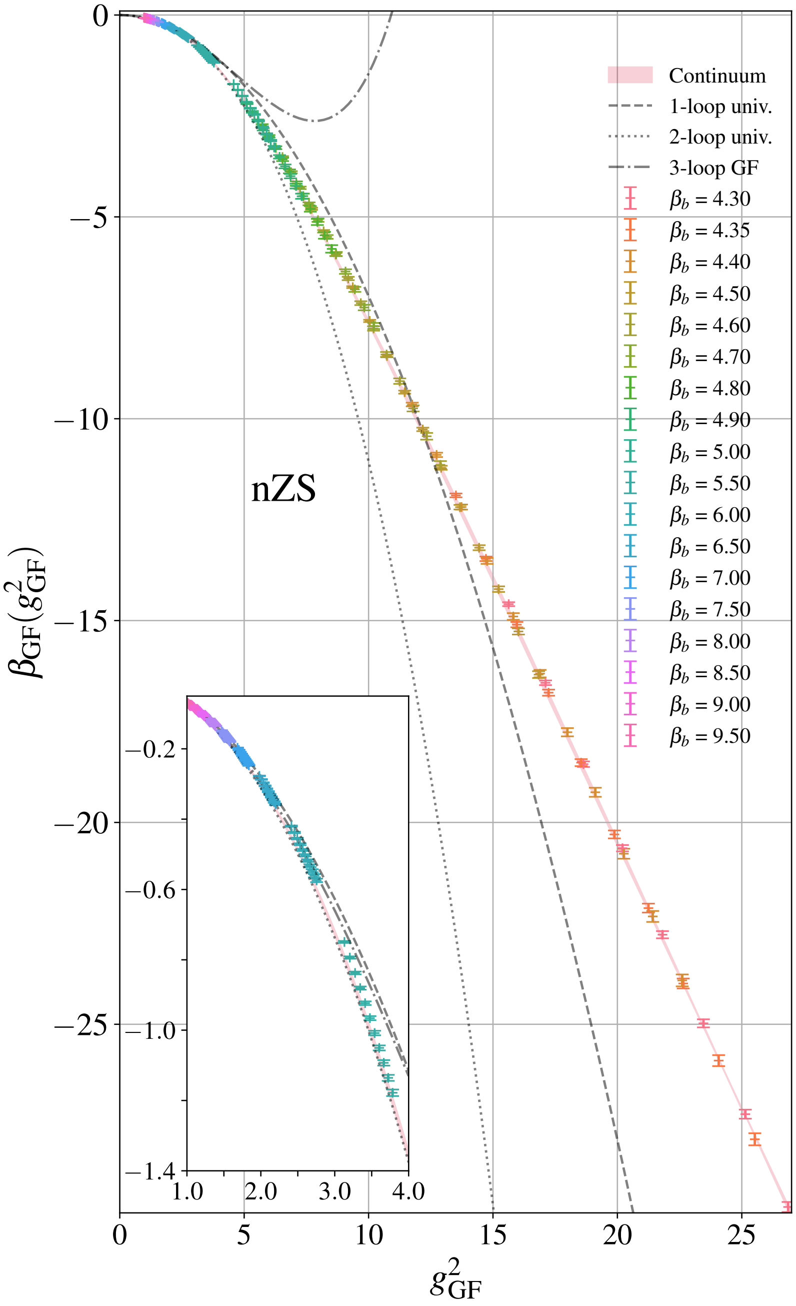

Figure 4.8: From Ref. [181]. The predicted β-function (salmon colored band) overlayed with the infinite volume extrapolated data at different bare coupling β_b (colored data points) for our main analysis based on nZS for flow times $t/a^2 \in [2.0, 4.0]$, separated by $\Delta t/a^2 = 0.2$. The insert magnifies the weak coupling region. The nZS combination shows very little cutoff dependence and the raw lattice data sit on top of the continuum extrapolated value.

entire range of $g^2_{\mathrm{GF}}(t; g^2_0)$ that the simulations of this chapter are capable of covering. I focus in on the weak coupling region in the bottom panel of Fig. 4.6. Following the suggestion of Ref. [92], I normalize $\beta_{\mathrm{GF}}\left(t; g_0^2\right)$ by $g_{\mathrm{GF}}^4\left(t; g_0^2\right)$. Focusing first on the top panel of Fig. 4.6, I show the result for the interpolation as multicolored bands with the same color scheme for the flow time as in Fig. 4.3. Data points entering any interpolation at a particular t/a^2 are marked with "×" markers with an error bar. The subpanel shows the weak coupling β-function. Grey lines correspond to the perturbative 1- (dashed) and 2-loop (dotted) universal β-function, along with the 3-loop (dashed-dotted) GF β-function from Ref. [161]. Qualitatively, the β-function appears to converge onto its perturbative counterpart as $g_{\mathrm{GF}}^2\left(t; g_0^2\right)$ reaches deeper into the weak coupling regime. Any mismatch between $\beta_{\mathrm{GF}}\left(t; g_0^2\right)$ and its perturbative counterpart in the weak coupling regime are better visualized by inspecting $\beta_{\mathrm{GF}}\left(t; g_0^2\right)/g_{\mathrm{GF}}^4\left(t; g_0^2\right)$, as is done in the bottom panel of Fig. 4.6. Qualitatively, $\beta_{\mathrm{GF}}\left(t; g_0^2\right)/g_{\mathrm{GF}}^4\left(t; g_0^2\right)$ appears to approach the 3-loop GF curve as $a^2/t \to 0$ at fixed $g_{\mathrm{GF}}^2\left(t; g_0^2\right) \lesssim 2.0$. In the strong coupling regime, the interpolated β-function becomes increasingly linear in $g_{\mathrm{GF}}^2\left(t; g_0^2\right)$, as I have already alluded to. Over the entire range of $g_{\mathrm{GF}}^2\left(t; g_0^2\right)$, the scale of the cutoff effects are much smaller than the absolute scale of $\beta_{\mathrm{GF}}\left(t; g_0^2\right)$. Hence, one may expect a fairly mild continuum extrapolation throughout the range of investigated renormalized couplings $g_{\mathrm{GF}}^2\left(t; g_0^2\right)$. We will see in Sec. 4.5.3 that the continuum extrapolations indeed support the latter observation.

4.5.3.2 Continuum extrapolation

Now that I have interpolated $\beta_{\mathrm{GF}}\left(t; g_0^2\right)$ in $g_{\mathrm{GF}}^2\left(t; g_0^2\right)$ over a set of fixed t/a^2, I can take the continuum extrapolation at any g_{GF}^2 that is covered by the fixed-t/a^2 interpolations using Eqn. 4.39. The minimum flow time $t_{\min}/a^2$ and maximum flow time $t_{\max}/a^2$ over which the extrapolation is performed must be chosen with care. If $t_{\min}/a^2$ is too small, then the flow may not be close enough to the renormalized trajectory for the linear scaling in Eqn. 4.39 to be realized. If $t_{\max}/a^2$ is too large, then the continuum extrapolation will pick up residual finite-volume effects due to the presence of higher-order corrections to the finite-volume scaling of $g_{\mathrm{GF}}^2\left(t; L, g_0^2\right)$ and $\beta_{\mathrm{GF}}\left(t; L, g_0^2\right)$ with t^2/L^4. Because the β-function obtained from different flow/operator combinations must be consistent in the

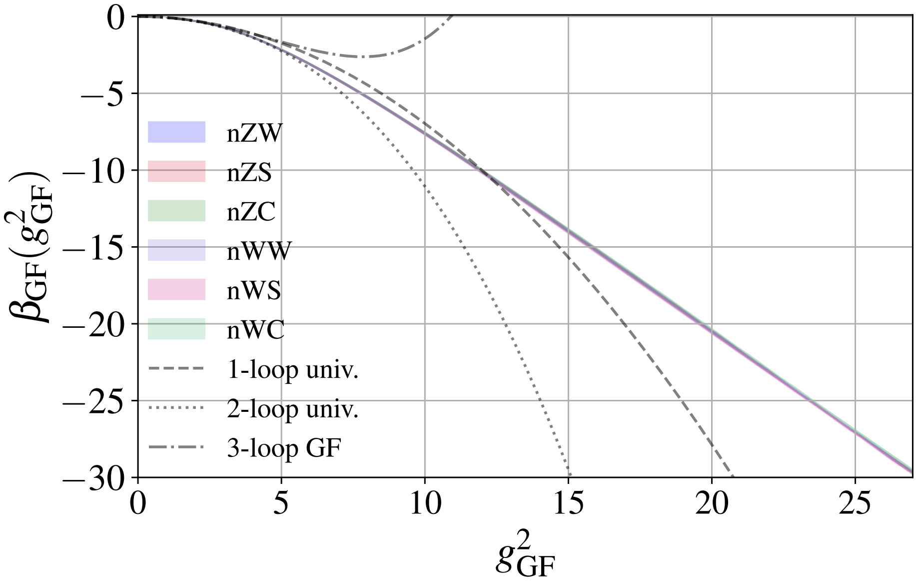

Figure 4.9: From Ref. [181]. Comparison of different continuum limit results obtained with tln improved flow operator combinations. All results overlap and form a single band.

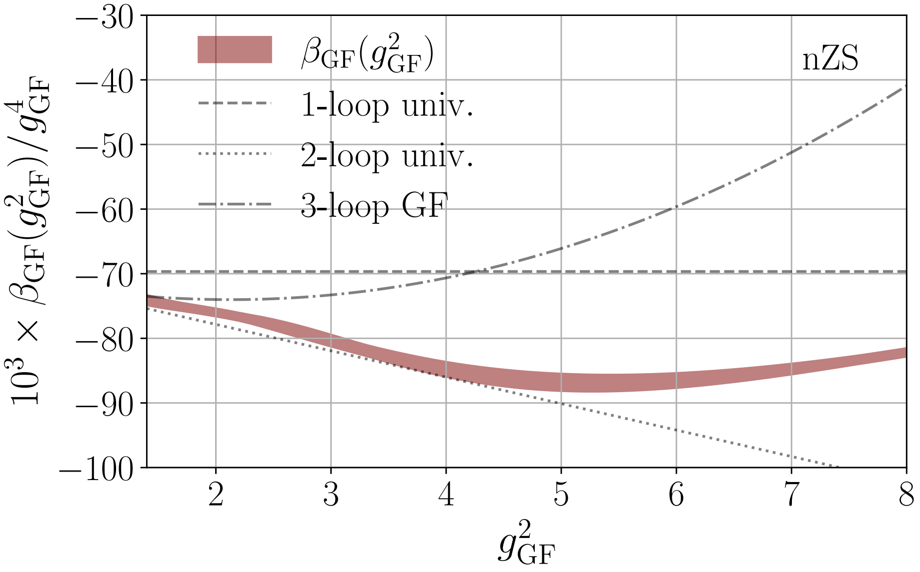

Figure 4.10: Continuum $\beta_{\mathrm{GF}}\left(g^2_{\mathrm{GF}}\right)/g^4_{\mathrm{GF}}$ over region where $\beta_{\mathrm{GF}}\left(g^2_{\mathrm{GF}}\right)/g^4_{\mathrm{GF}}$ changes curvature. Colored maroon band indicated continuum prediction for $\beta_{\mathrm{GF}}\left(g^2_{\mathrm{GF}}\right)/g^4_{\mathrm{GF}}$ from nZS combination. Grey lines indicate 1- (dashed), 2- (dotted) and 3-loop (dashed-dotted) perturbative β-function [161].

continuum limit, I use the consistency of the continuum extrapolation of $g^2_{\mathrm{GF}}(t; g^2_0)$ from operators $\mathcal{S}^{\mathrm{e}}$ to set the $t_{\mathrm{min}}/a^2, t_{\mathrm{max}}/a^2$ over which the continuum extrapolation is performed. The range

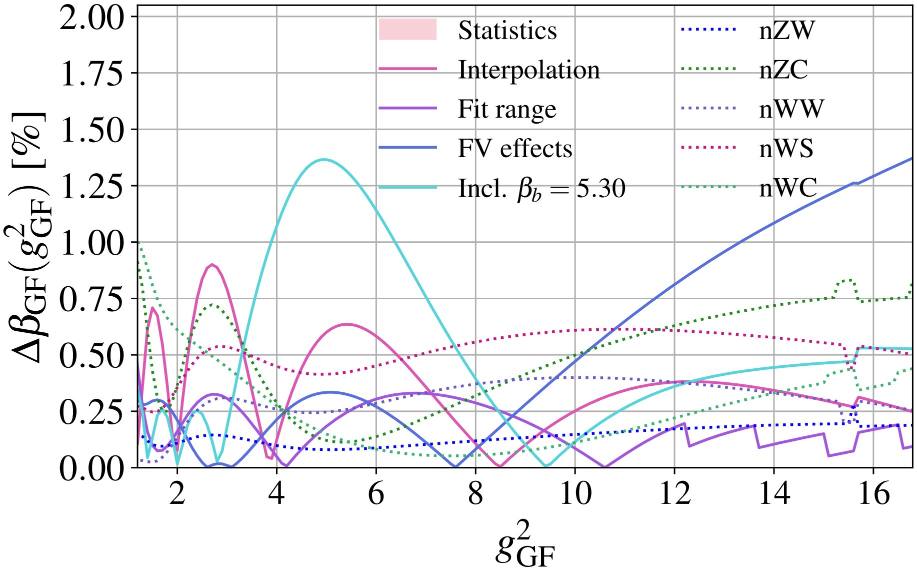

Figure 4.11: From Ref. [181]. Systematic uncertainties with respect to our main analysis based on nZS. By varying different parts of our analysis one after the other, we calculate the relative changes of the central value and compare the size of the different systematic uncertainties (colored lines) to our statistical uncertainty (salmon-colored band).

$[t_{\mathrm{min}}, t_{\mathrm{max}}]/a^2 = [2.0, 4.0]$ achieves a reasonable degree of consistency over $1.2 \lesssim g^2_{\mathrm{GF}} \lesssim 20$; hence, this is the range that is chosen for the central analysis. In Sec. 4.5.4, I estimate the systematic uncertainty that is associated with different choices for $t_{\mathrm{min}}/a^2, t_{\mathrm{max}}/a^2$.

In Fig. 4.7, I illustrate the continuum extrapolation at fixed $1.2 \leq g^2_{\mathrm{GF}} \leq 0.3\mathcal{N} \approx 15.8$ over $[t_{\mathrm{min}}, t_{\mathrm{max}}]/a^2 = [2.0, 4.0]$ for $\mathcal{S}^{\mathrm{e}} = \mathrm{W}$ (Wilson, blue), S (Symanzik, red), and C (clover, green). The multicolored bands correspond to the extrapolation and circular markers (with error bars) are the data points used in the fit (filled) and left out of the fit (open). Note two interesting observations:

(1) relative to the absolute scale of the continuum limit, the $\beta_{\mathrm{GF}}(t; g_0^2)$ from different $\mathcal{S}^{\mathrm{e}}$ are fairly close to one another and

(2) the continuum extrapolation for all $\mathcal{S}^{\mathrm{e}}$ are indeed quite mild in comparison to the absolute scale of $\beta_{\mathrm{GF}}(t; g_0^2)$.

Both of the latter observations are attributed primarily to the tree-level improvement discussed in Sec. 4.5.2. The closeness of $\beta_{\mathrm{GF}}(t; g_0^2)$ from different $\mathcal{S}^{\mathrm{e}}$ was already noted in Sec. 4.5.2, as it was the primary motivation for choosing only tree-level-corrected $g^2_{\mathrm{GF}}\big(t; L, g^2_0\big)$ for the central analysis. Without tree-level corrections, the slope for the continuum extrapolation from different $\mathcal{S}^{\mathrm{e}}$ is generally larger than those with tree-level corrections, which increases the statistical error in the continuum extrapolation. To emphasize the latter point, I juxtapose the $\beta\big(t; g^2_0\big)$ at $4.3 \leq \beta_b \leq 9.5$ (colored markers with error bars) and $2.0 \leq t/a^2 \leq 4.0$ from the nZS combination against its continuum prediction $\beta\big(g^2_{\mathrm{GF}}\big)$ (maroon band) in Fig. 4.8. For better visibility, I spread out fixed β_b data points in steps of $\delta t/a^2 = 0.2$. For the vast majority of $1.2 \lesssim g^2_{\mathrm{GF}} \lesssim 27$, the values for $\beta\big(t; g^2_0\big)$ at fixed β_b either overlap with or are within the local vicinity of their continuum counterpart.

Note that the flow time is actually a continuous variable, as the gradient flow is a continuous stochastic process. The discrete t/a^2 that enter the continuum extrapolation are highly correlated with one another, which makes the covariance matrix for the data points that enter the continuum extrapolation poorly conditioned (see Appendix D). One way out is to approximate the covariance matrix by its diagonal components; however, the resulting χ^2 and p-value lose their meaning. Even worse, this “fix” can lead to underestimating the statistical error. To obtain $\beta_{\mathrm{GF}}\big(g^2_{\mathrm{GF}}\big)$ with a reliable statistical error, I first estimate the central value for $\beta_{\mathrm{GF}}\big(g^2_{\mathrm{GF}}\big)$ from a continuum extrapolation with a diagonal covariance matrix. I then repeat the latter fit with data shifted by $\pm 1\sigma$. The statistical error in $\beta_{\mathrm{GF}}\big(g^2_{\mathrm{GF}}\big)$ is estimated from the half difference of the central value for $\beta_{\mathrm{GF}}\big(g^2_{\mathrm{GF}}\big)$ from the $\beta_{\mathrm{GF}}\big(t; g^2_0\big) \pm 1\sigma$ fits. This ensures that I am able to perform the continuum extrapolation without running into issues with the poorly conditioned covariance matrix and/or underestimated statistical errors, as it essentially assumes that the data are 100% correlated.

The authors of Ref. [57] recently (as of the writing of this thesis) devised a method for estimating the χ^2 and p-value of “uncorrelated fits” (those that must approximate the covariance matrix by its diagonal components) in such a way that both goodness-of-fit measures retain their original meaning. I was not aware of this method at the time that I performed the analysis in this chapter; however, I intend to include it in future analyses of the continuous β-function. Doing so could open the door to using Bayesian model averaging as a means to estimate the various systematic uncertainties that are present in the continuum extrapolation step of the continuous β-function method directly [210, 273, 274], assuming that the notion of subset select for continuous Gaussian variables (i.e., a Gaussian process) can be defined unambiguously. I explore the use of Bayesian model averaging for the infinite volume extrapolation step of the $N_{\mathrm{f}} = 12$ continuous β-function method in Chapter 5. It is also applied to the infinite volume extrapolation step of the $N_{\mathrm{f}} = 8$ system in Chapter 7.

4.5.4 Systematic errors in the β-function

I estimate the combined statistical (Fig. 4.8) and systematic error in the continuum $\beta_{\mathrm{GF}}\left(g^2_{\mathrm{GF}}\right)$ from the nZS combination by first considering variations in the analysis that I have presented in Secs. 4.5.1-4.5.3. The percent shift $\Delta\beta_{\mathrm{GF}}\left(g^2_{\mathrm{GF}}\right)$ in the central value for continuum prediction for $\beta_{\mathrm{GF}}\left(g^2_{\mathrm{GF}}\right)$ resulting from each variation is summarized in Fig. 4.11. Note that, as g^2_{GF} increases, different $\mathcal{S}^{\mathrm{f}}\mathcal{S}^{\mathrm{e}}$ begin to lose a^2/t values over which the continuum extrapolation discussed in Sec. 4.5.3 can be performed. This results is small discontinuities in the central value for the continuum β-function as g^2_{GF} crosses such thresholds. The values $\beta_{\mathrm{GF}}\left(g^2_{\mathrm{GF}}\right)$ on either side of the discontinuity are consistent within their respective statistical errors; however, they result in small discontinuities in the estimate for the systematic error, as is observed in Fig. 4.11. The variations considered in this chapter are as follows.

- **Including $\boldsymbol{\beta_{\mathrm{b}}}$= 5.3:** Data at $\beta_b = 5.3$ sits in the middle of the transition region. As such, the autocorrelation time is quite large and obtaining a reliable estimate for the statistical error is challenging. This results in a nearly vanishing p-value for the infinite volume extrapolation at $\beta_b = 5.3$ and it is hence dropped from the central analysis. Unfortunately, it also lies in a sensitive region of the β-function, where $\beta_{\mathrm{GF}}\left(g^2_{\mathrm{GF}}\right)/g^4_{\mathrm{GF}}$ is observed to change curvature. I illustrate the change in curvature in Fig. 4.10. $\beta_b = 5.3$ is therefore added back into the analyses to estimate the impact that it has on the β-function in that region. Unsurprisingly, it is the largest systematic effect (O(1.3%)) around $g^2_{\mathrm{GF}} \sim 5.0$.

- **Infinite volume extrapolation:** Two variations are considered.

(1) Drop the smallest volume and perform a linear fit to all three volumes.

(2) Repeat the analysis with only the largest volume.

Both effects yield similar shifts in $\beta_{\mathrm{GF}}\big(g^2_{\mathrm{GF}}\big)$, but using just the largest volume has the largest effect. Hence, it is what is used to estimate the systematic uncertainty that is associated with the infinite volume extrapolation. According to Fig. 4.11, it is the dominant source of systematic uncertainty for $g^2_{\mathrm{GF}} \gtrsim 11$.

- **Intermediate interpolation:** The value for the order N of the interpolating function in Eqn. 4.42 is chosen based on the stability of the β-function against increasing N. The intermediate interpolation is varied by choosing $N = 2$ and 6. The largest effect is observed in going from $N = 4$ to 2; hence, it is taken as an estimate for the systematic error that is associated to the intermediate interpolation.

- **Continuum fit range:** The continuum extrapolation is varied by changing the t_{min}/a^2 and t_{max}/a^2 over which the continuum extrapolation is performed as follows.

  (1) Fix $t_{\mathrm{min}}/a^2 = 2.0$ and vary $4.0 \leq t_{\mathrm{max}}/a^2 \leq 5.0$.

  (2) Fix $t_{\mathrm{max}}/a^2 = 4.0$ and vary $1.52 \leq t_{\mathrm{min}}/a^2 \leq 2.0$.

  The systematic error is estimated from the maximum of the latter variations at each g^2_{GF}. The maximum systematic effect from varying the flow time range is O(0.3%).

- **Flow/operator combination:** In Fig. 4.9, I show the continuum $\beta_{\mathrm{GF}}\big(g^2_{\mathrm{GF}}\big)$ from each $\mathcal{S}^{\mathrm{f}}\mathcal{S}^{\mathrm{e}}$ combination. Taken together, the continuum curves nearly form a uniform curve. In Fig. 4.11, one observes ever-so-slight variations in the central result for $\beta_{\mathrm{GF}}\big(g^2_{\mathrm{GF}}\big)$, most of which are either well within 1σ of the central nZS statistical error or just outside of 1σ. The largest variations are observed from the clover operator, which are O(0.8%-1.0%). Because varying the flow/operator combination is the dominant systematic effect for $7.5 \lesssim g^2_{\mathrm{GF}} \lesssim 11$, the largest flow/operator deviation at each g^2_{GF} is taken as an estimate for the systematic error that is associated with choosing a particular $\mathcal{S}^{\mathrm{f}}\mathcal{S}^{\mathrm{e}}$ in the central analysis.

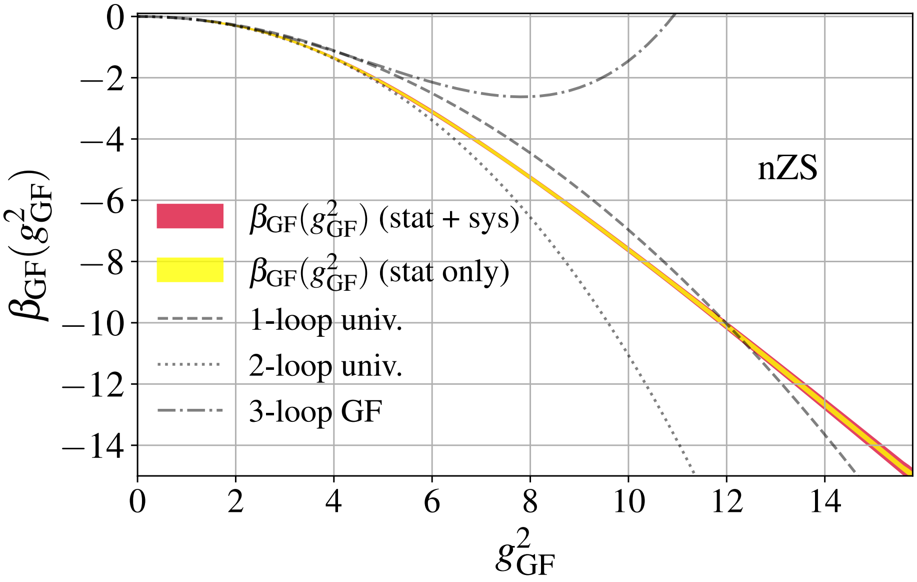

Figure 4.12: From Ref. [181]. Final result for $\beta_{\mathrm{GF}}\big(g^2_{\mathrm{GF}}\big)$ as a function of g^2_{GF} for the coupling range relevant to determine the Λ parameter. The yellow inner band shows only the statistical uncertainty, whereas the red outer band shows the combined statistical and systematic uncertainties.

The latter systematic errors are included in the combined statitsical/systematic error by adding them to the statistical error in quadrature. In Fig. 4.12, I compare the continuum $\beta_{\mathrm{GF}}\big(g^2_{\mathrm{GF}}\big)$ with just statistical errors (yellow) against $\beta_{\mathrm{GF}}\big(g^2_{\mathrm{GF}}\big)$ with both statistical and systematic errors (red) over the range of g^2_{GF} that will enter the Λ-parameter calculation of Sec. 4.6.

4.5.5 The continuum β-function

The continuum prediction for the β-function illustrated in Figs. 4.8 and Fig. 4.12 exhibits a number of expected and unexpected features. In the weak coupling region, it appears to converge to the 1-, 2- and 3-loop perturbative β-function [161], as one should expect. This will be discussed further in Sec. 4.6, where matching onto the perturbative regime will be crucial for estimating the Λ-parameter. I have also alluded many times already to the linearity of the β-function in the strong coupling regime. The slope of $\beta_{\mathrm{GF}}\big(g^2_{\mathrm{GF}}\big)$ in g^2_{GF} in the linear region predicts directly how the continuum gradient-flowed Yang-Mills energy depends on the gradient flow time t. First, take

$$\beta_{\mathrm{GF}}\left(g_{\mathrm{GF}}^2\right) \approx c_0 + c_1 g_{\mathrm{GF}}^2 \qquad \left(g_{\mathrm{GF}}^2/4\pi \gg 1\right) \tag{4.43}$$

as an ansatz. Then the solution to the RG equation for the continuum g_{GF}^2 (Eqn. 4.14) is

$$c_1 g_{\mathrm{GF}}^2(t) \approx c_0 + (\Lambda t)^{-c_1} \propto t^2 \langle E(t) \rangle \qquad \left(g_{\mathrm{GF}}^2/4\pi \gg 1\right), \tag{4.44}$$

where Λ is an integration constant that fixes the dimension of Λt. Early literature on gradient flow observed that $-c_1$ is O(1) [252]. More recent literature posits that $-c_1 = 1$ *exactly* [271, 328], albeit from simulations that have been performed on a single bare gauge coupling and gradient flow times that extend well beyond $8t = (L/2)^2$ on their largest volume ($L/a = 32$). At such large flow times, statements based on gradient flow regarding infinite volume properties of Yang-Mills systems lose their meaning because the smearing radius now wraps around the lattice and finite-volume effects are manifest. The dependence of c_1 on g_{GF}^2 is illustrated in Fig. 4.13.

Over $20 \lesssim g_{\mathrm{GF}}^2 \lesssim 27$, the slope (derivative) of the continuum β-function from the simulations of this chapter levels off at $c_1 \approx -1.32$. The spikes in c_1 that are occasionally observed for $g_{\mathrm{GF}}^2 \gtrsim 15$ in Fig. 4.13 are due to occasional discontinuities in the continuum β-function that occur when the continuum extrapolation loses a^2/t values over which to extrapolate $\beta_{\mathrm{GF}}\left(t; g_0^2\right)$ to $a^2/t \to 0$ at larger fixed g_{GF}^2. The discontinuities have a small effect on the $\beta_{\mathrm{GF}}\left(g_{\mathrm{GF}}^2\right)$, but they lead to occasional numerical spikes in the derivative that do not reflect the slope in the strong coupling region. Their effect is also observed as discontinuities is the systematic error estimates reported in Fig. 4.11. They also do not affect the estimate for $c_1 = -1.32(1)$ reported in Ref. [181], which is shown as a grey line in Fig. 4.13. Note that value for c_1 in Ref. [181] is obtained from a cubic spline interpolation of $\beta_{\mathrm{GF}}\left(g_{\mathrm{GF}}^2\right)$ in g_{GF}^2 using the *Steffen* algorithm provided by `gvar` [241, 342].[4] The spline-based estimate was crosschecked against a simple linear fit of $\beta_{\mathrm{GF}}\left(g_{\mathrm{GF}}^2\right)$ in g_{GF}^2 over the region where $\mathrm{d}\beta_{\mathrm{GF}}/\mathrm{d}g_{\mathrm{GF}}^2$ levels off. The slope in Fig. 4.13 is obtained from a naive estimate of the slope over a

---

[4] Using a spline is justified by the high degree of statistical correlation in $\beta_{\mathrm{GF}}\left(g_{\mathrm{GF}}^2\right)$ amongst neighboring g_{GF}^2.

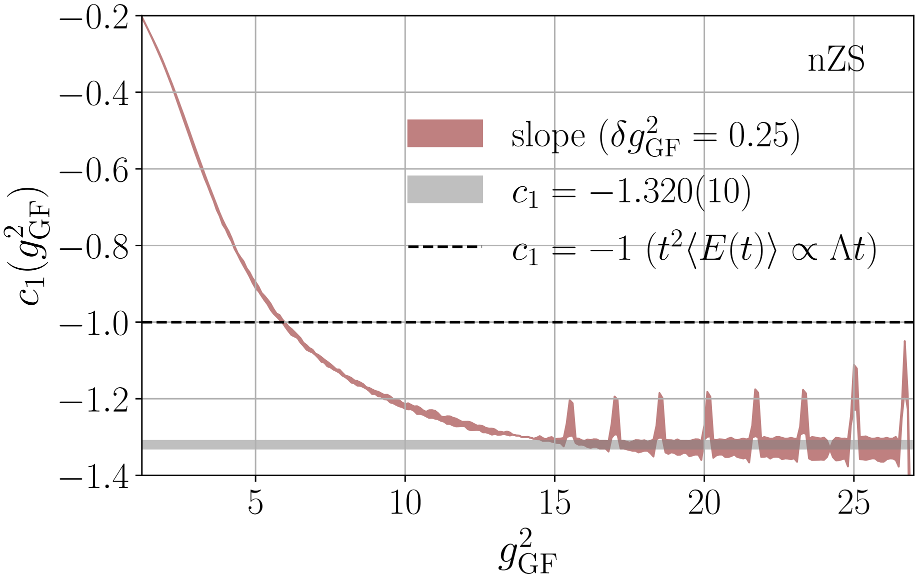

Figure 4.13: Slope of continuum $\beta_{\mathrm{GF}}(g^2_{\mathrm{GF}})$ predicted from $c_1\delta g^2_{\mathrm{GF}} \equiv \beta_{\mathrm{GF}}(g^2_{\mathrm{GF}} + \delta g^2_{\mathrm{GF}}) - \beta_{\mathrm{GF}}(g^2_{\mathrm{GF}})$ for $\delta g^2_{\mathrm{GF}} = 0.25$ (maroon band). See text in Sec. 4.5.5 for a description of the numerical spikes that occasionally occur when $g^2_{\mathrm{GF}} \gtrsim 15$. These spikes do not affect the central estimate $c_1 = -1.32(1)$ from Ref. [181] is shown as a grey line. The value for $c_1 = -1$ is shown as a black dashed line for visualization purposes.

coarser set of β-function values and is shown for the purposes of visualization. As such, the estimate from Ref. [181] is slightly lower than the slope in Fig. 4.13, but it is consistent nonetheless. In any case, the slope obtained from the numerical simulations of this chapter is not consistent with unity, which implies that $t^2\langle E(t)\rangle$ is not linear in Λt over the largest range of investigated flow times in this chapter. As the linearity of $t^2\langle E(t)\rangle$ in Λt discussed in Refs. [252, 271, 328] is merely an observation and not a rigorous result, there is no good reason to believe that the slope reported in this chapter is incorrect based on the commonly-held notation that it should be consistent with unity. It is also possible that the slope has a mild dependence on g^2_{GF} at strong coupling and slowly converges to unity. If this is the case, then the simulations of this chapter are simply not capable reaching large enough g^2_{GF} to detect such dependence of c_1 on g^2_{GF}.

As of the writing of this thesis, a rigorous explanation for the linearity of the continuum β-function strong coupling regime does not exist. Moreover, it is not evident that such linear strong coupling behavior should be universal in confining systems. Nonetheless, there exists a small literature on the behavior of non-perturbative β-functions in the confining regime [77, 271, 315, 328], all of which exhibit some form of linearity in the confined regime. Even the textbook β-function that is defined in terms of the static $q\bar{q}$ "quark"[5] potential $V(R)$ as

$$g_V^2(R) \propto R^2 V(R) \implies \beta_V\left(g_V^2\right) \equiv -R^2 \frac{\mathrm{d}g_V^2(R)}{\mathrm{d}R^2} \tag{4.45}$$

is linear in its respective renormalized coupling $g_V^2(R)$ in the confining regime, where $V(R) \approx \sigma R$ with σ the string tension [313].[6] The slope of the β-function is such a scheme is $c_1^{(V)} = -3/2$, which is slightly higher in magnitude, but not statistically consistent with, the slope obtained in the gradient-flow-based scheme of this chapter. There is no reason to expect that such a slope should be universal without more theoretical input; moreover, the coupling g_V^2 may not properly furnish a proper definition of the renormalized coupling in the pure Yang-Mills system. The proposal in Ref. [315] for the pure Yang-Mills system is based on the Novikov-Shifman-Vainshtein-Zakharov (NSVZ) β-function of $\mathcal{N} = 1$ supersymmetric Yang-Mills systems [279, 333]. A similar β-function is obtained in Ref. [77]. However, the β-function obtained from both methods has a pole singularity that it must cross before the β-function becomes linear. Moreover, the slope is positive in both Refs. [77, 315], which is the opposite of what is predicted from the GF β-function. Refs. [271, 328] claim that the linearity of GF-based β-functions is a consequence of confinement; as such, its linearity is attributed to the same source of the linearity of the β-function defined in Eqn. 4.45. It is also plausible that the linearity of the β-function can be attributed to Yang-Mills instantons, as they are all that is left after gradient flow suppresses all other high-wavenumber fluctuations. Unfortunately, the source of the observed asymptotic linearity in the β-function is unknown. The same can be said for whether or not they asymptotic linearity has some universal component to it in a confined system. However, there are plenty of directions for research that could be explored along the same vein.

---

[5] More precisely, $V(R)$ is the potential between *any* source/anti-source of color flux; hence, it is well-defined for the pure Yang-Mills system [265].

[6] Note using $V(R)$ to defined a β-function is valid *only* in the pure Yang-Mills system, where there is no string breaking due to the presence of fermions.

4.6 The Λ-parameter and strong coupling constant

The continuum β-function $\beta_{\mathrm{GF}}\left(g^2_{\mathrm{GF}}\right)$ that I calculated in Sec. 4.5 characterizes the running of $g^2_{\mathrm{GF}}(t)$ with $8t \propto 1/\mu^2$ according to Eqn. 4.14. More generally, in any renormalization scheme X,

$$\mu^2 \frac{\mathrm{d}g^2_X(\mu)}{\mathrm{d}\mu^2} \equiv \beta_X\left(g^2_X\right) \tag{4.46}$$

for $g^2_X(\mu)$ the renormalized coupling of renormalization scheme X. It is assumed that $g^2_X(\mu)$ is either a genuine renormalized coupling appearing in some Wilsonian effective action or, equivalently, defined in terms of an observable O_X that does not renormalize under some RG transformation. There are two equivalent ways that one could solve Eqn. 4.46 assuming that $\beta_X\left(g^2_X\right)$ is known over the desired range of $g^2_X = g^2_X(\mu)$.

4.6.1 Initial value problem: the strong coupling constant

Assuming that one knows $g^2_X(\mu_0) = 4\pi\alpha_X$ for some constant α_X, they can use Eqn. 4.46 to evolve $g^2_X(\mu_0)$ at μ_0 to $g^2_X(\mu)$ at some other scale μ. This is what is done in many studies of the Standard Model that need input from quantum chromodynamics (QCD) at scales that are well above the pole mass of the Z boson, where QCD is pertubative. In such studies, α_X in the $X = \overline{\mathrm{MS}}$ scheme is defined in terms of the $\overline{\mathrm{MS}}$ running coupling $g^2_{\overline{\mathrm{MS}}}(\mu_0)$ at $\mu_0 = M_Z$, where M_Z is the pole mass of the Z boson. The value of $\alpha_{\overline{\mathrm{MS}}}$ is often called the "strong coupling constant" α_s, though there is nothing special about α_{s} outside of being deep enough in the perturbative regime of QCD that $g^2_{\overline{\mathrm{MS}}}(\mu) \equiv 4\pi\alpha_s(\mu)$ for any $\mu \geq M_Z$ can be determined from the perturbative β-function in the $\overline{\mathrm{MS}}$ scheme.[7] The value of α_s is therefore considered to be an important Standard Model parameter, entering a variety of perturbative QCD processes; calculations of the t quark mass and decay constants; Higgs production processes; hadronic Z widths; and much more [89, 90, 95]. See the most recent Flavor Lattice Averaging Group (FLAG) report for more information on contemporary α_s determinations based on lattice field theory. Currently, lattice-based determinations

---

[7] The $\overline{\mathrm{MS}}$ β-function is known to a whopping 5-loop order [198]

of $\alpha_s = 0.1184(8)$ ($N_\mathrm{f} = 4$) are the most precise.[8]

4.6.2 Direct integration: the Λ-parameter

One can also just integrate Eqn. 4.46 directly using separation of variables. In the "$\log b_0$" convention, the solution is

$$\frac{\Lambda_X^2}{\mu^2} = \left(b_0 g_X^2(\mu)\right)^{-\frac{b_1}{b_0^2}} \exp\left(-\frac{1}{b_0 g_X(\mu)^2}\right) \times \exp\left[-\int_0^{g_X^2(\mu)} \mathrm{d}x \left(\frac{1}{\beta(x)} + \frac{1}{b_0 x^2} - \frac{b_1}{b_0^2 x}\right)\right], \tag{4.47}$$

where b_0, b_1 are universal 1- and 2-loop coefficients of the β-function defined in Chapter 1 [173, 227, 335]. Strictly speaking, the Λ-parameter is an integration constant that fixes the dimension of Λ_X/μ. As an integration constant, knowing Λ_X is equivalent to knowing any initial value of $g_X^2(\mu_0)$ as far as solving Eqn. 4.46 is concerned. The Λ-parameter possesses a number of theoretical properties that tend to make it easier to calculate than α_X. For one, it is a RG invariant, meaning

$$\frac{\mathrm{d}\Lambda_X}{\mathrm{d}\mu^2} = 0. \tag{4.48}$$

This is already evident from the fact that it is an integration constant. Conveniently, the Λ-parameter in any other scheme Y is related to Λ_X by an exact one-loop relation

$$\Lambda_Y^2/\Lambda_X^2 = \exp(d_1/b_0), \tag{4.49}$$

where d_1 is the one-loop coefficient relating g_Y^2 to g_X^2 as

$$g_Y^2 = g_X^2 + d_1 g_X^4 + \mathrm{O}\big(g_X^6\big) \qquad (g_X^2, g_Y^2 \ll 4\pi) \tag{4.50}$$

and b_0 is the one-loop universal β-function coefficient [92]. Eqn. 4.49 follows directly from Eqns. 4.47 and 4.50. That it is exact appears to be a consequence of asymptotic freedom, though it is

---

[8] Compare to the current global average $\alpha_s = 0.1179(9)$ from the most recent Particle Data Group (PDG) Review of Particle Physics [381].

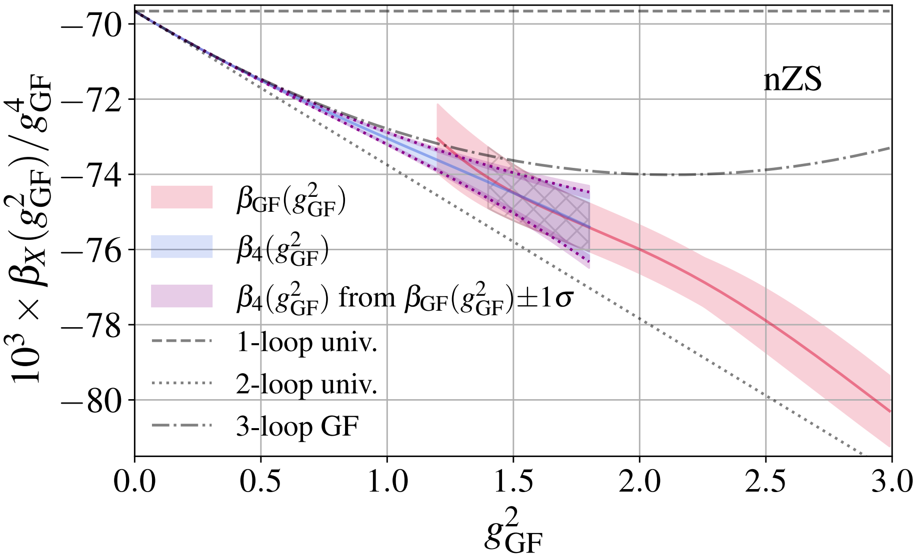

Figure 4.14: From Ref. [181]. $\beta_{\rm GF}(g^2_{\rm GF})/g^4_{\rm GF}$ in the weak coupling region. The salmon-colored band shows our nonperturbatively determined $\beta_{\rm GF}$ with combined statistical and systematic uncertainties. We match to the 3-loop GF function using Eq. (4.52) in the range $g^2_{\rm GF}\in[1.4,1.8]$ indicated by the grey hatched area. Shifting our nonperturbative values by $\pm1\sigma$, we obtain the magenta bands providing the upper and lower limits of the resulting matched function shown in blue.

plausible that there is a better and/or more fundamental explanation for it being exact. In principle, if the Λ-parameter is known in one scheme, then is known in any other scheme, so long as d_1 is calculable. For example, $\Lambda_{\overline{\rm MS}}$ is related numerically to the Λ-parameter of the infinite volume gradient-flow-based scheme of this chapter $\Lambda_{\rm GF}$ as [92, 161]

$$\Lambda_{\overline{\rm MS}}/\Lambda_{\rm GF}=0.534162960405763... \tag{4.51}$$

In principle, if one can calculate $\tilde{t}\Lambda^2_{\rm GF}$ in terms of some *known* hadronic scale $\tilde{t}$, then $\Lambda_{\overline{\rm MS}}$, and hence α_s, is determined. Calculating $\tilde{t}\Lambda^2_{\rm GF}$ at the $\tilde{t}=8t_0$ scale defined by $g^2_{\rm GF}(t_0)\equiv0.3\mathcal{N}$ is the objective of the next section [252]. Note that t_0 (and its counterpart from the β-function w_0) is used most modern scale setting studies [6, 31, 43, 336], though w_0 has largely superseded it.

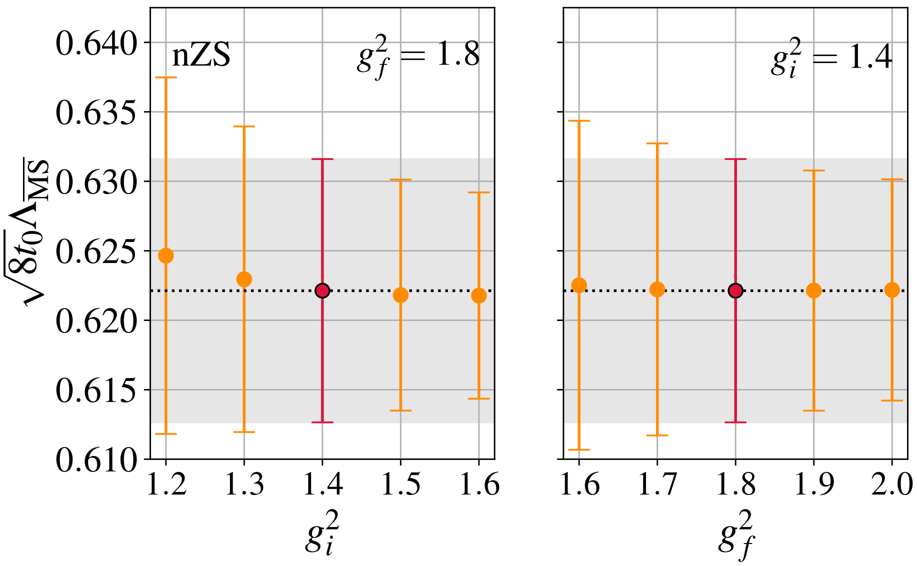

Figure 4.15: From Ref. [181]. Systematic uncertainty in our perturbative matching procedure due to choosing g_i^2 or g_f^2. The value for our preferred choices for g_i^2 and g_f^2 is highlighted in red.

4.6.3 Calculation of the pure Yang-Mills Λ-parameter

The definition of the Λ-parameter in Eqn. 4.47 hints at a method for determining $t\Lambda_{\mathrm{GF}}$ in terms of the β-function that I calculated in Sec. 4.5: simply throw β_{GF} in for β_X and integrate up to $g^2_{\mathrm{GF}}(t)$. However, this requires that I am able to determine the continuum β-function for $g^2_{\mathrm{GF}} \lesssim 1.2$, where the non-perturbative simulations in this chapter no longer cover the β-function. Deep within the weak coupling regime, one can use the 3-loop perturbative β-function from Ref. [161]; however, in-between it is necessary to match the non-perturbative β-function to the 3-loop perturbative β-function. To this end, first define the following ansatz for the β-function in the weak coupling regime

$$\beta_4(g^2_{\mathrm{GF}}) \equiv -\frac{g^4_{\mathrm{GF}}}{(4\pi)^2}\left(b_0 + b_1\frac{g^2_{\mathrm{GF}}}{(4\pi)^2} + b_2\frac{g^4_{\mathrm{GF}}}{(4\pi)^3} + b_p g^6_{\mathrm{GF}}\right), \tag{4.52}$$

where b_0, b_1, and b_2 are the one-, two-, and 3-loop coefficients for the perturbative β-function [161] and b_p is a free parameter [161]. I determine b_p by requiring that

$$\int_{g_i^2}^{g_f^2} \mathrm{d}x\ 1/\beta(x)$$

is the same for $\beta = \beta_4$ and $\beta = \beta_{\rm GF}$ over $g^2_{\rm GF} \in \left[g^2_i, g^2_f\right]$ for some choice of g^2_i, g^2_f. Because different $\beta_{\rm GF}\left(g^2_{\rm GF}\right)$ are correlated with one another and I no longer have access to the covariance matrix, I estimate the error in $\beta_4(g^2_{\rm GF})$ from this procedure by repeating this exercise with $\beta_{\rm GF} \pm 1\sigma$ and estimating the error in b_p from the half difference. The error in b_p is then propagated directly into $\beta_4\left(g^2_{\rm GF}\right)$ at any $g^2_{\rm GF}$ using the automatic error propagation tools provided by the `gvar` library. I choose $g^2_i, g^2_f = 1.4, 1.8$, respectively, and show the result of the matching procedure in Fig. 4.14. The maroon band in Fig. 4.14 is the non-perturbative $\beta_{\rm GF}\left(g^2_{\rm GF}\right)$ that I calculated in Sec. 4.5. The blue band indicates the estimate for β_4 from the matching procedure described above, with errors bounded by the matching procedure performed at $\beta_{\rm GF} \pm 1\sigma$ (purple bands). The region over which the matching is performed ($g^2_i \leq g^2_{\rm GF} \leq g^2_f$) is shown as a crosshatched band in Fig. 4.14.

With $\beta_4\left(g^2_{\rm GF}\right)$ determined, I am able to extend the non-perturbative β-function down to $g^2_{\rm GF} \to 0$. To calculate $8t_0\Lambda^2_{\rm GF}$, I use $\beta_4\left(g^2_{\rm GF}\right)$ in place of β_X in Eqn. 4.47 for $g^2_{\rm GF} \leq 1.4$ and $\beta_{\rm GF}\left(g^2_{\rm GF}\right)$ for $1.4 < g^2_{\rm GF} \leq 0.3\mathcal{N}$. To determine the integral in Eqn. 4.49, I interpolate over $\beta_{\rm GF}\left(g^2_{\rm GF}\right)$ using a fine grid of $g^2_{\rm GF}$ so that I can represent $\beta_{\rm GF}\left(g^2_{\rm GF}\right)$ as a continuous variable, then perform the integral

$$\int_0^{g^2_{\rm GF}(t_0)} \mathrm{d}x \left(\frac{1}{\beta_{\rm GF}(x)} + \frac{1}{b_0 x^2} - \frac{b_1}{b_0^2 x}\right)$$

numerically via an adaptive 4th-order Runge-Kutta integration algorithm [302, 342]. Note that the error in $\beta_{\rm GF}\left(g^2_{\rm GF}\right)$ is the combined statistical/systematic error from Sec. 4.5.4. Both the spline algorithm and the Runge-Kutta integration are performed using the numerical analysis tools provided by `gvar`, which ensures that all errors are properly propagated back into $8t_0\Lambda^2_{\rm GF}$ [241]. I have checked to ensure that the central value and error of $8t_0\Lambda^2_{\rm GF}$ does not change if I make the grid in $g^2_{\rm GF}$ finer for the spline interpolation over $\beta_{\rm GF}\left(g^2_{\rm GF}\right)$.

In Fig. 4.15, I demonstrate the effect varying the g^2_i, g^2_f in the matching procedure used to obtain β_4 has on the estimate of $8t_0\Lambda^2_{\overline{\rm MS}}$. Keeping $g^2_f = 1.8$ fixed (left panel), the central value for $8t_0\Lambda^2_{\overline{\rm MS}}$ settles around $g^2_i = 1.4$. Keeping $g_i = 1.4$ fixed (right panel), the central value for settles around $g^2_f = 1.8$, though the central value changes much less when varying $g^2_f < 1.8$ at fixed $g^2_i = 1.4$

than it does when varying $g_i^2 < 1.4$ at fixed $g_f^2 = 1.8$. Increasing g_i^2 beyond 1.4 at fixed $g_f^2 = 1.8$ or g_f^2 beyond 1.8 at fixed $g_i^2 = 1.4$ only decreases the error on $8t_0\Lambda^2_{\overline{\text{MS}}}$; therefore, $g_i^2, g_f^2 = 1.4, 1.8$ is chosen, so as to be as conservative with the error as possible. Taken together, I get the following estimate for the Λ-parameter

$$\sqrt{8t_0}\Lambda_{\text{GF}} = 1.164(19), \tag{4.53}$$

which yields

$$\sqrt{8t_0}\Lambda_{\overline{\text{MS}}} = 0.622(10). \tag{4.54}$$

using the relation between Λ_{GF} and $\Lambda_{\overline{\text{MS}}}$ in Eqn. 4.51.

4.6.4 Comparison of the Λ-parameter against the literature

Calculations of the Λ-parameter for the pure Yang-Mills system have a long history going all the way back to 1980 with derivation/standardization of Eqn. 4.49 [173]. A variety of methods have been both proposed and deployed to calculate the Λ-parameter from lattice simulations. In Fig. 4.16, I juxtapose the determinations of the Λ-parameter that have met the Flavor Lattice Averaging Group (FLAG) criteria against the value for $t_0\Lambda^2_{\overline{\text{MS}}}$ that I obtain from the β-function in this chapter [6]. Such studies utilize Schrödinger functional methods [67, 209], Wilson loops [147, 226], and the short-distance static quark potential $V(R)$ [48], and gradient-flow-based step-scaling [92]. Note that the gradient flow result of Ref. [92] was reanalyzed with a better method for controlling systematic errors in Ref. [269], from which a consistent estimate of $t_0\Lambda^2_{\overline{\text{MS}}}$ was obtained. Just a few weeks before the the result for $8t_0\Lambda^2_{\overline{\text{MS}}}$ presented in this chapter was announced, the group of Ref. [380] announced their value for $8t_0\Lambda^2_{\overline{\text{MS}}}$ using the same CBFM method discussed in this chapter. Values for the Λ-parameter that are not based on gradient flow are expressed in terms of the Sommer scale r_0, defined as $F(r_0)r_0^2 = 1.65$ in terms of the force $F(R)$ that is derived from the static quark

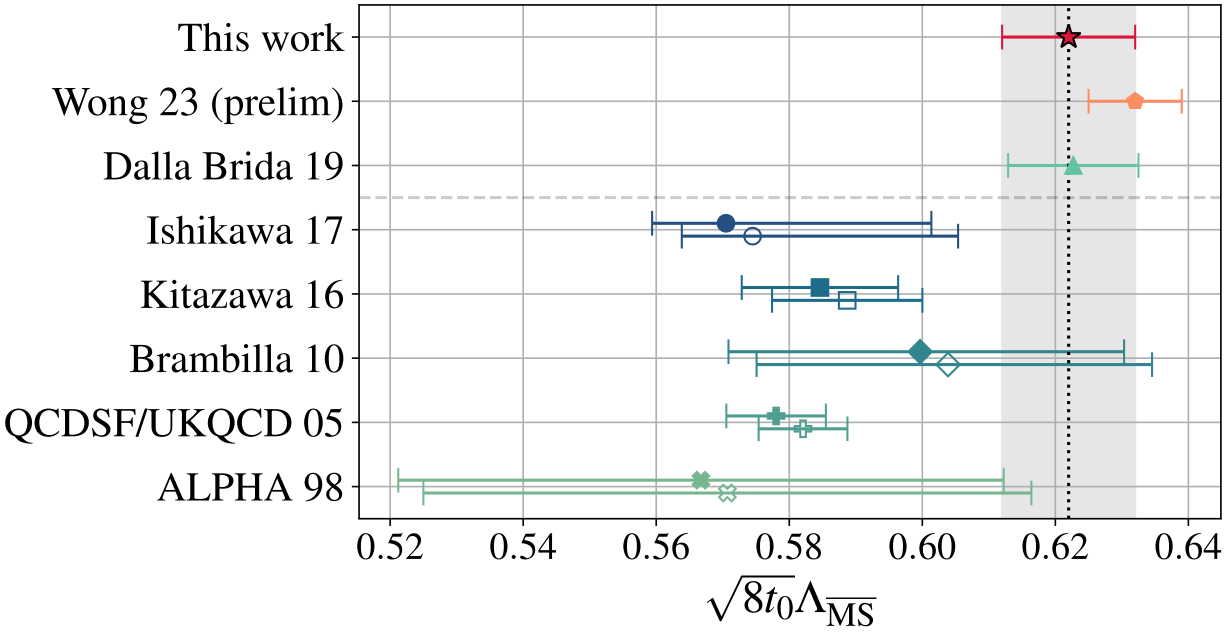

Figure 4.16: From Ref. [181]. Comparison of our result for $\sqrt{8t_0}\Lambda_{\mathrm{GF}}$ (maroon star) to the preliminary result by Wong et al. [380] (orange pentagon) and Dalla Brida/Ramos [92] (green triangle). In addition we show values for $r_0\Lambda_{\overline{\mathrm{MS}}}$ which enter the FLAG 2021 averages: ALPHA 98 [67], QCDSF/UKQCD 05 [147], Brambilla 10 [48], Kitazawa 16 [226], and Ishikawa 17 [209]. These values are converted to $\sqrt{8t_0}\Lambda_{\mathrm{GF}}$ using $\sqrt{8t_0}/r_0$ from [252] (open symbols) or Ref. [92] (filled symbols).

potential $V(R)$ 4.5.5. Therefore, such results for $r_0\Lambda_{\overline{\mathrm{MS}}}$ must be converted to $8t_0\Lambda^2_{\overline{\mathrm{MS}}}$ using the value for $\sqrt{8t_0}/r_0 = 0.948(7)$ for Ref. [252] (open symbols in Fig. 4.16) or $\sqrt{8t_0}/r_0 = 0.9414(90)$ from Ref. [92] (closed symbols in Fig. 4.16).

The estimate for $8t_0\Lambda^2_{\overline{\mathrm{MS}}}$ compares well with the other gradient-flow-based estimates from the literature [91, 380]. However, there is a gap between estimates based on gradient flow and estimates that utilize the Sommer scale r_0. It is plausible that the discrepancy between the two is simply due to the estimate for $\sqrt{8t_0}/r_0$ itself, as r_0 can be a difficult quantity to calculate precisely. Regardless, given the spread in $8t_0\Lambda^2_{\overline{\mathrm{MS}}}$ observed in Fig. 4.16, much scrutiny is needed. This especially rings true given that the pure Yang-Mills system is supposed to be fairly well-understood. The authors of Refs. [92, 269] have even called for a revision of the FLAG criteria for the Λ-parameter altogether, given the precision of their gradient-based-result. Hopefully, the resolution to this discrepancy will lead to a better understanding of lattice-based determinations of $8t_0\Lambda^2_{\overline{\mathrm{MS}}}$ and their associate systematics.

4.7 Scheme transformations and non-perturbative matching

I wrap this chapter up with a brief discussion of renormalization scheme transformations. Additionally, I propose a method for matching the β-function of one RG scheme to another *non-perturbatively.* One of the largest sources of systematic error in the calculation of $8t_0\Lambda^2_{\overline{\text{MS}}}$ in this chapter is the matching procedure discussed in Sec. 4.6.3. This is due in part to the size of the 3-loop correction to the β-function, which causes the 3-loop GF β-function to exhibit poor convergence compared to other RG schemes [91, 161]. This was already pointed out in the gradient-flow-based step-scaling study of $8t_0\Lambda^2_{\overline{\text{MS}}}$ from Ref. [92] (green triangle in Fig. 4.16). An alternative to matching the non-perturbative continuum β-function to its continuum perturbative counterpart (in the same RG scheme) could be to match the GF β-function of this chapter to a β-function from another RG scheme that has either better convergence properties at weak coupling. Consider that, for any two renormalized couplings g^2_X, g^2_Y defined as in Sec. 4.2, there is some bijection $G \in \mathcal{C}^1$ such that [322]

$$g^2_Y = G_{XY}\left(g^2_X\right). \tag{4.55}$$

The Jacobian of G_{XY} (just the derivative with respect to g^2_X in this case) and its inverse cannot be singular. To see why, note that the corresponding β-functions β_X, β_Y are related as

$$\beta_Y\left(g^2_Y\right) = \frac{\partial G_{XY}}{\partial g^2_X}\beta_X\left(g^2_X\right). \tag{4.56}$$

For the RG β-functions to have the same number of fixed points and the same slope at those fixed points, Eqn. 4.56 requires that $\partial G_{XY}/\partial g^2_X$ (and its inverse) is not singular; otherwise, the existence of a fixed point could be scheme-dependent. Eqn. 4.56 also hints at a method for matching the RG β-function in one scheme to another *non-perturbatively.* Taking, for example, the ansatz

$$G_{XY}(x) = x + x^2\sum_{i=0}^{N_{\text{p}}\text{-}1} c_n x^n, \tag{4.57}$$

it is plausible that $G_{XY}(x)$ could be determined *approximately* over some range of g_X^2 by requiring

$$\beta_X\big(g_X^2\big) = \beta_Y\big(G_{XY}(x)\big)/G'_{XY}(x)\Big|_{x=g_X^2}, \tag{4.58}$$

where

$$G'_{XY}(x) = \frac{\mathrm{d}G_{XY}(x)}{\mathrm{d}x}. \tag{4.59}$$

If such a procedure could be performed in a controlled manner, it would allow for lattice calculations to utilize different schemes for different parts of the calculation by patching them together where the overlap. This could be especially useful for the calculation of $8t_0\Lambda^2_{\overline{\mathrm{MS}}}$ in this chapter, where being able to utilize calculations based on Schrödinger functional methods could significantly improve the control that I have over the weak coupling region. See Refs. [317, 318, 322, 334] for more information on the properties of G_{XY}, which are currently not well-understood in the high-energy physics community as a whole.

Chapter 5

The β-function of the massless twelve flavor system

In the last chapter, I calculated the renormalization group β-function within the $N_{\mathrm{f}} = 0$ extreme of the $N_{\mathrm{f}} < N_{\mathrm{f}}^{\mathrm{c}}$ side of the conformal window. In this chapter, I am going to take you to the $N_{\mathrm{f}} = 12$ extreme on the $N_{\mathrm{f}} > N_{\mathrm{f}}^{\mathrm{c}}$ side of the conformal window. Nevertheless, the techniques that I used to calculate the β-function of the pure Yang-Mills system shall carry over directly to the twelve-flavor system. This chapter is laid out as follows. In Sec. 5.1, I briefly review the last several decades of research into the infrared properties of the $N_{\mathrm{f}} = 12$ system. I discuss an improvement technique in Sec. 5.2 that allows for the simulations in this chapter to reach deep into the infrared regime of the twelve flavor system. I summarize the simulations that are used to extract the β-function for this system in Sec. 5.3. In Sec. 5.4, I calculate the continuum renormalization group β-function using the methods that I introduced in Chapter 4. The continuumm β-function in Sec. 4 strongly suggests that the twelve-flavor system exhibits an infrared fixed point. I wrap up by calculating the leading irrelevant critical exponent at the infrared fixed point in Sec. 5.5. The content of this chapter is based on Ref. [293].

5.1 Overview of the twelve-flavor system

The massless twelve-flavor system was one of the first systems to be targeted by investigations of the SU(3) conformal window. As such, a variety of analytical and numerical techniques have been deployed to understand both the properties of this system and where it lies with respect to the conformal window. Along the vein of analytical techniques, groups have utilized perturbation theory [106, 298, 316, 319–321], the gap equation [22, 30], functional renormalization group methods [50, 51], conformal expansion [237], conformal bootstrap [243], the background field method [152], perturbative non-relativistic quantum chromodynamics [81], and large-N expansion [312]. As a whole, analytical investigations tend to predict that the twelve flavor system is infrared conformal, though a sizeable chunk of such investigations are taken up by perturbative approaches that are more or less likely to agree with one another in the first place. Taking perturbative calculations out of the picture, which also includes investigations that combine perturbation theory with the gap equation, analytical studies still tend to lie on the side of infrared conformality, though the margin is quite a bit smaller. Numerical studies deploying techniques that are based on non-perturbative lattice simulations also tend to lean toward the twelve-flavor system being infrared conformal. Such studies have utilized finite-volume step-scaling [18, 19, 126, 132, 134, 167, 182, 186, 244, 245] (see Sec. 4.3), Monte Carlo renormalization group methods [168, 169], hadron mass and decay constant spectroscopy [9, 11, 16, 78, 97, 102, 128, 129, 246], and the Dirac eigenmode spectrum [79, 127]. Of these studies, some of the most convincing evidence has come from the large-scale studies of the Lattice Kobayashi–Maskawa Instutute (LatKMI) collaboration, which collectively observe evidence for conformal hyperscaling and the presence of a light scalar boson at non-zero fermion mass [9, 11]. The latter studies are to be juxtaposed against the detailed study from Ref. [128], which concluded that the twelve-flavor is consistent with being chirally broken based multiple observables, such as the chiral condensate and the pseudoscalar spectrum. However, various claims made in Ref. [128] have been refuted using the same dataset [16, 97]. The debate surrounding infrared conformality in the twelve-flavor system morphed from being centered around spectroscopy and chiral symmetry breaking to non-perturbative determinations of the RG β-function, of which Refs. [126, 167, 182, 186, 244] claim to find evidence of an infrared fixed point and Refs. [126, 132, 134, 245] do not. Taken as a whole, the majority of investigations that probe the infrared properties of the twelve flavor system conclude that it is infrared conformal,[1] while a minority conclude that it is confining,

---

[1] See Refs. [9, 11, 16, 18, 19, 22, 30, 78, 79, 97, 102, 106, 167–169, 182, 186, 237, 243, 244, 246, 298, 316, 319, 321]

chirally broken, or are inconclusive.[2]

5.2 Bulk phase transitions and Pauli-Villars improvement

The debate surrounding infrared conformality in the twelve flavor system, as approached from the perspective of non-perturbative β-functions, is largely due to the presence of unphysical first-order bulk phase transitions that prevent simulations from reaching far enough into the infrared regime to observe to observe an infrared fixed point [80, 103, 280, 327, 340]. Bulk first-order phase transitions are triggered by unphysical ultraviolet fluctuations [188]. Such fluctuations get worse as N_f increases. Recently, it has been proposed to utilize heavy Pauli-Villars (PV) fields to cancel off the ultraviolet fluctuations that trigger bulk first-order phase transitions [188].

Heavy staggered Pauli-Villars fields $\phi^{(s)}$ with species index s and action

$$\mathcal{S}_\mathrm{PV}\left[\phi, \phi^\dagger\right] = \frac{1}{2}\sum_{n,s} \phi^{(s)\dagger}(n)\mathfrak{D}_\mathrm{S}^{(s)\dagger}\mathfrak{D}_\mathrm{S}^{(s)}\phi^{(s)}(n) \tag{5.1}$$

induce an effective action of the form

$$\mathcal{S}_\mathrm{ind}\left[\mathcal{U}_\mu\right] = N_\mathrm{sPV}\sum_l \frac{(-1)^{l/2}}{l(am)^l}\sum_n\sum_{\mathcal{C}_l}\mathcal{E}_{\mathcal{C}_l}\mathrm{Tr}_\mathrm{c}\left[\mathcal{U}_{\mathcal{C}_l}(n)\right], \tag{5.2}$$

where $N_\mathrm{sPV} \equiv 8N_\mathrm{PV}$ is the number of staggered species of Pauli-Villars fields and N_PV is the total number of Pauli-Villars fields, am_PV is the mass of the degenerate Pauli-Villars fields, $\mathcal{C}_l$ is closed a loop of size l (in units of the lattice spacing), $\mathcal{U}_{\mathcal{C}_l}(n)$ is the lattice-discretized parallel transporter starting at n and traversing $\mathcal{C}_l$ (defined in terms of a produce of links $\mathcal{U}_\mu$ as in Eqn. 3.7), and $\mathcal{E}_{\mathcal{C}_l} = \frac{1}{4}\mathrm{Tr}\left[\gamma^{\mu_1}\cdots\gamma^{\mu_l}\right]$ is a pure sign factor that is determined by the geometry of $\mathcal{C}_l$ [171, 172, 174, 188]. Note that N_sPV can be halved by using "half fields", which are described in Chapter 8.1.1. The staggered Dirac operator $\mathfrak{D}_\mathrm{S}^{(s)}$ was defined in Chapter 3. Note that the form of Eqn. 5.2 with smeared gauge links (see Appendix G) could be slightly more complicated. At leading order in $1/am$, observables calculated from fluctuations in the induced action of Eqn. 5.2 are equal to observables

---

[2] See Refs. [126, 128, 129, 135, 152, 245, 312]

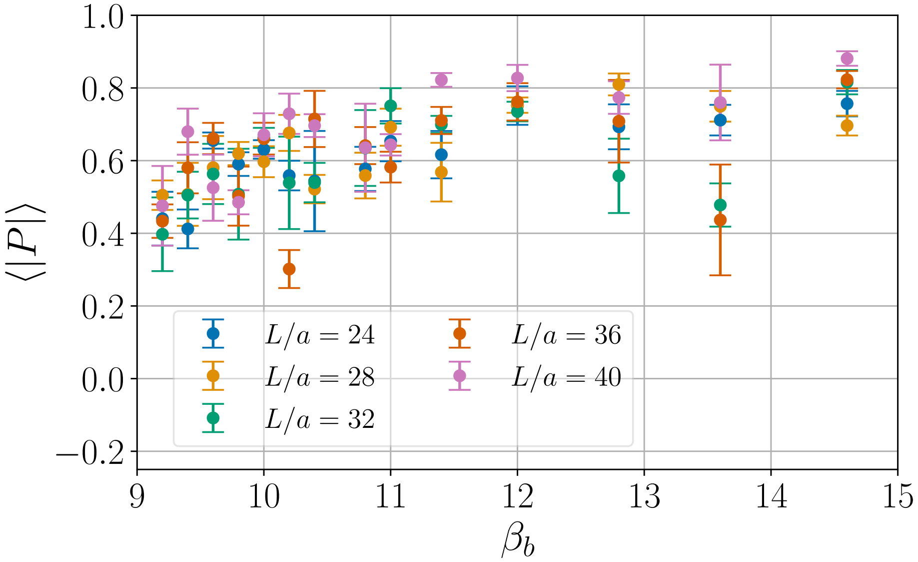

Figure 5.1: From Ref. [293]. The gradient-flowed Polyakov loop expectation value at flow time $8t/a^2 \approx (L/2a)^2$ versus the bare gauge coupling β_b on each volume in Table 4.1. The absolute value of the Polyakov loop is shown by colored error bars: $L/a = 24$ (blue), 28 (yellow), 32 (green), 36 (orange), and 40 (pink).

generated by fluctuations in the Wilson action (Eqn. 3.8), but with a coupling [188]

$$\beta_{\text{ind}} = -\frac{N_{\text{sPV}}}{(2am_{\text{PV}})^4} \tag{5.3}$$

In other words, heavy Pauli-Villars felds have an anti-screening effect. The anti-screening of the PV fields can be used to cancel off the ultraviolet fluctuations that trigger first-order bulk phase transitions. Moreover, by keeping the mass of the PV fields constant in units of the lattice spacing along the continuum limit, the PV fields decouple. As such, the leave the infrared properties of the target system untouched in the continuum limit. PV-improvement was first tested in the four-, eight-, and twelve-flavor system in Ref. [188], which yielded promising results. Pauli-Villars fields have since been deployed in simulations using Wilson fermions (briefly discussed in Sec. 3.2.2) with great success [178, 179]. I have implemented PV improvment in the `Quantum EXpressions` (`QEX`-based) `qex_staghmc` Monte Carlo suite.[3] I describe `qex_staghmc` in detail in Chapter 8.

---

[3] My fork of the `QEX` library is publicly available at `https://github.com/ctpeterson/qex`.

β_b	L/a				
	24	28	32	36	40
9.20	340	253	188	188	133
9.40	347	262	215	273	186
9.60	244	233	251	203	166
9.80	275	329	250	297	280
10.0	271	246	312	151	134
10.2	184	209	217	221	133
10.4	283	241	299	221	142
10.8	246	220	288	208	306
11.0	236	288	156	151	156
11.4	188	194	223	193	183
12.0	182	248	200	254	167
12.8	180	179	204	254	209
13.6	251	183	168	254	228
14.6	253	191	178	251	226

Table 5.1: The number of thermalized configurations analyzed at each bare coupling β_b and volume L/a. The configurations are separated by 10 MDTUs.

5.3 Simulation details

The simulations in this chapter utilize an adjoint-plaquette gauge action (see Appendix F), a massless ($am_f = 0$) nHYP-smeared staggered fermion action (see Appendix G for details on nHYP smearing), and an nHYP-smeared Pauli-Villars action with $N_{\text{PV}} = 48$ PV fields and $am_{\text{PV}} = 0.5$. Both the PV and fermion files have completely anti-periodic boundary conditions in all four directions. All simulations are performed either using a modified version of the `MILC` library (`KS_nHYP_FS`)[4] or `qex_staghmc` (see Sec. 8.1 for details). The same is true for all gauge flow measurements. Details of how gauge flow measurements are performed are discussed in Chapter 8.1.2. In Table 5.1, I list the total number of thermalized configurations for each $(L/a, \beta_b)$ pair that I use to extract the

---

[4] The modified `MILC` library can be found at `https://github.com/daschaich/KS_nHYP_FA`

β-function in this chapter.

In Fig. 5.1, I plot the gradient flowed temporal Polyakov loop magnitude (Eqn. 4.32) $P_{\hat{4}} \equiv P$ at $8t = (L/2)^2$ for all ensembles in Table 5.1 against β_b. The statistical error in the Polyakov loop is estimated using the Γ-method implemented in the `pyerrors` library, which automatically accounts for autocorrelation [214, 215, 308, 325, 378]. However, the statistical errors in Fig. 5.1 are still likely underestimated. As I discussed in Chapter 4.4.1, the Polyakov loop indicates the onset of confinement, as was evident from the sharp drop in the Polaykov loop observed in the $N_{\mathrm{f}} = 0$ system in Fig. 4.1. For the $N_{\mathrm{f}} = 12$ simulations of this chapter, the Polyakov loop is showing no such signs over the investigated β_b. In fact, if the simulations were to show any signs of confinement (or even fluctuations in the topological charge), then it would not be possible to perform them with $am_f = 0$. The reason for this is that the Dirac operator develops a zero-mode at $am_f = 0$ in a confined system; such a zero more prevents the Dirac operator from being invertible and hence the fermion sector from being simulations using pseudofermions. Hence, the simulations in this chapter show no signs of confinement over the range of investigated volumes and couplings. That, or they somehow only cover the ϵ-regime, which is unlikely.

5.4 Calculation of the β-function

As in Chapter 4, I calculate the continuum RG β-function $\beta_{\mathrm{GF}}\left(g^2_{\mathrm{GF}}\right)$ using gradient flow and the continuous β-function method (CBFM). The steps of the calculation are the same; however, some details differ by improving upon the analysis presented in Chapter 4. The present study utilizes only Wilson flow ($\mathcal{S}^{\mathrm{f}} = \mathrm{W}$) and either the Wilson or clover operator ($\mathcal{S}^{\mathrm{e}} = \mathrm{W}, \mathrm{C}$) to discretize the Yang-Mills energy density $E(t, L)$. The finite-volume renormalized coupling $g^2_{\mathrm{GF}}\left(t; L, g^2_0\right)$ is calculated from the Yang-Mills energy density using Eqn. 4.19 with $\delta(t, L)$ defined by Eqn. 4.16; i.e., I do not consider any tree-level corrections in $\delta(t, L)$. Moreover, I denote the calculation of $g^2_{\mathrm{GF}}\left(t; L, g^2_0\right)$ from a specific flow/operator combination as "$\mathcal{S}^{\mathrm{f}}\mathcal{S}^{\mathrm{e}}$"; e.g., "WC" for Wilson flow and clover operator. The finite-volume β-function $\beta_{\mathrm{GF}}\left(t; L, g^2_0\right)$ is calculated from the renormalized coupling numerically using Eqn. 4.36 and the 5-point stencil of Eqn. 4.37.

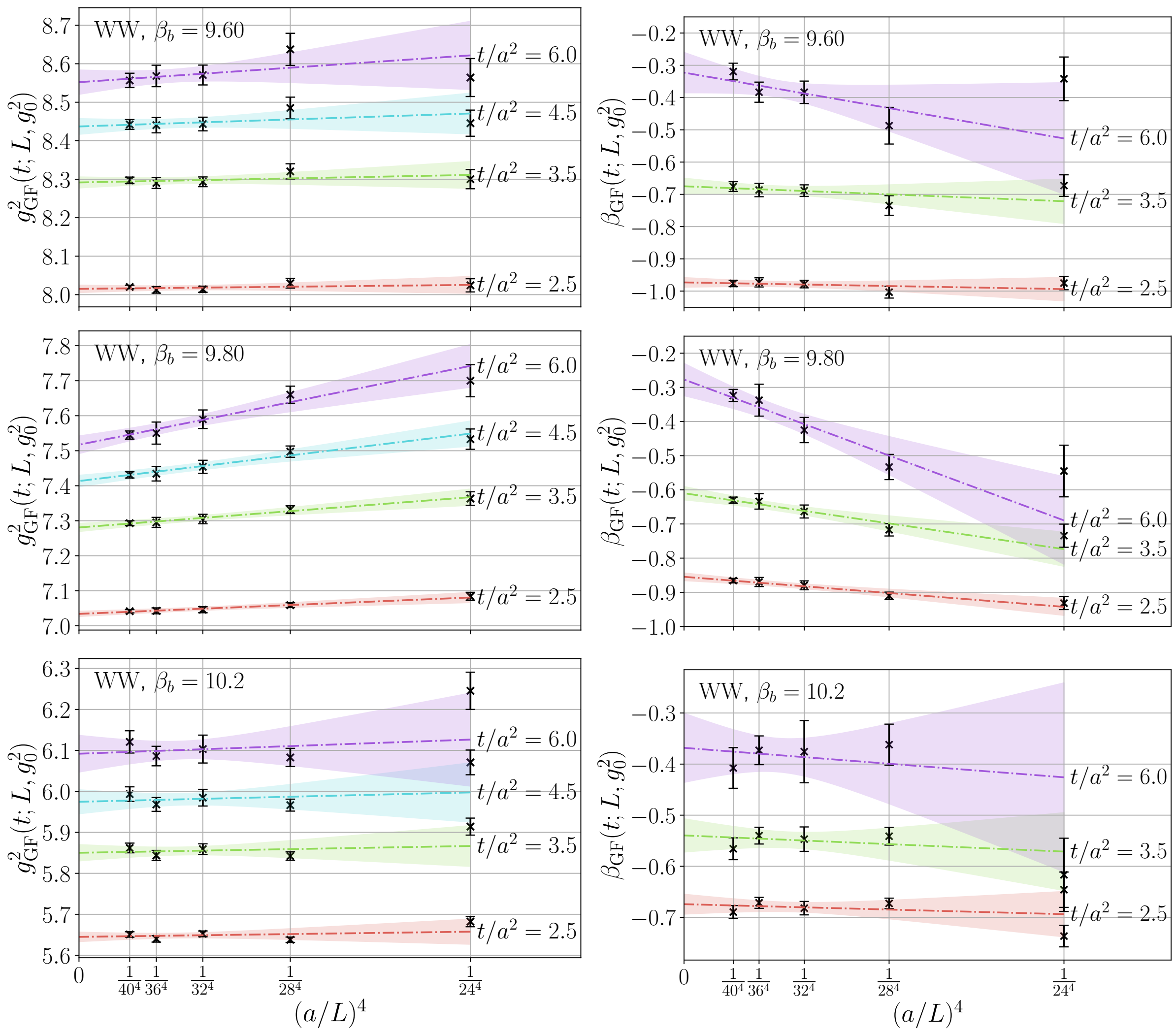

Figure 5.2: From Ref. [293]. Result of our infinite volume extrapolation of $g^2_{\rm GF}\big(t; L, g_0^2\big)$ (left panels) and $\beta_{\rm GF}\big(t; L, g_0^2\big)$ (right panels) for the Wilson (W) operator at $\beta_b = 9.60$ (top panels), 9.80 (middle panels) and 10.2 (bottom panels). Black (×) markers with error bars are the data included in our extrapolation. Extrapolations with errors that are predicted from Bayesian model averaging are indicated by multi-colored bands at $t/a^2 = 2.5$ (red), 3.5 (light green), 4.5 (cyan), and 6.0 (light purple). We do not show the infinite volume extrapolation of $\beta_{\rm GF}\big(t; L, g_0^2\big)$ at $t/a^2 = 4.5$ for visualization purposes.

Correlated errors are calculated and kept track of using the automatic error propagation tools provided by the `gvar` library [241]. Autocorrelation is accounted for by binning. Moreover, any fits to data utilize either `lsqfit` or my `Python`-based fitting library, `SwissFit` [240, 290]. Both `lsqfit` and `SwissFit` are integrated with `gvar` so as to ensure that any statistical correlations are

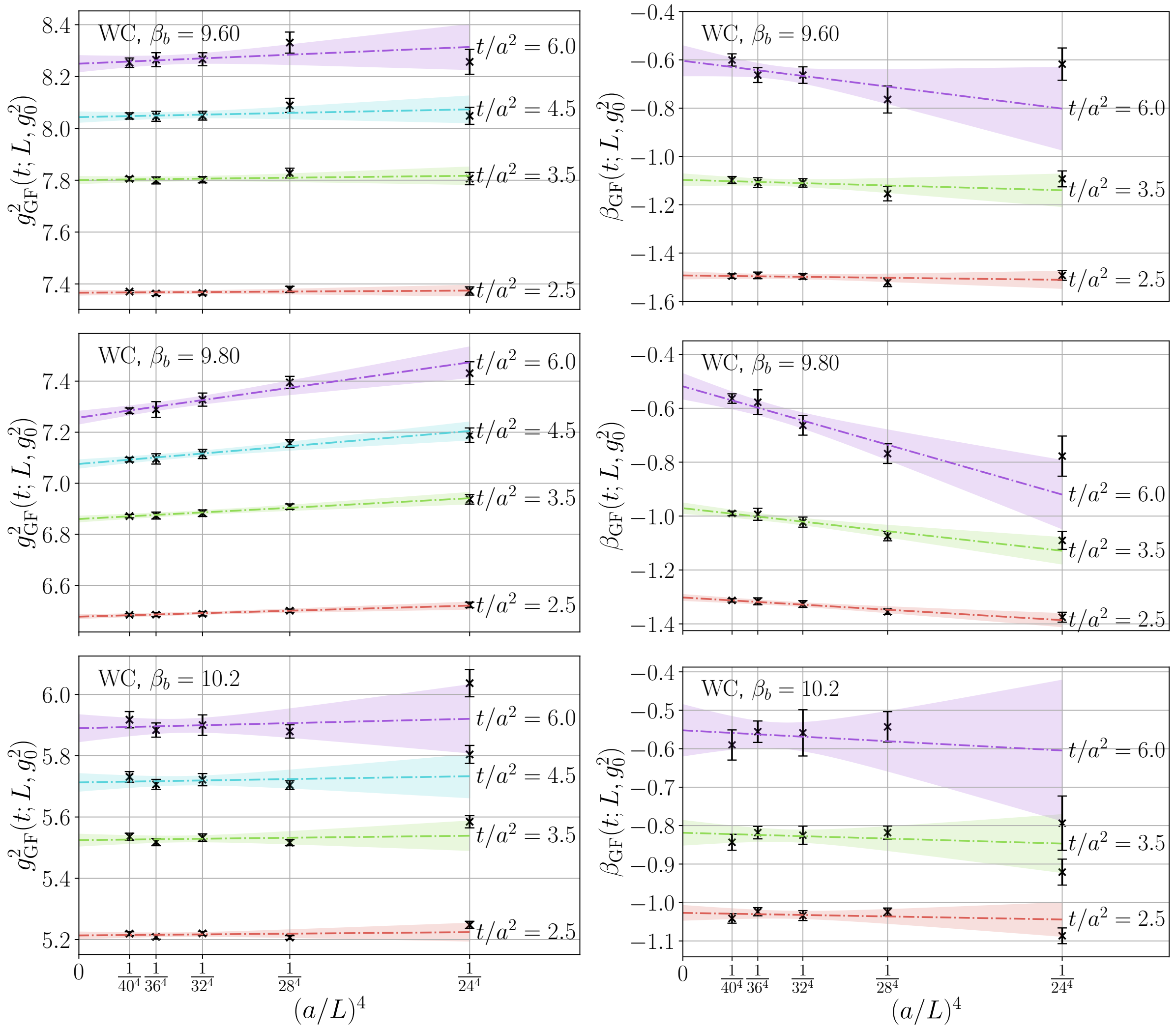

Figure 5.3: From Ref. [293]. Result of our infinite volume extrapolation of $g^2_{\rm GF}(t; L, g_0^2)$ (left panels) and $\beta_{\rm GF}(t; L, g_0^2)$ (right panels) for the clover (C) operator at $\beta_b = 9.60$ (top panels), 9.80 (middle panels) and 10.2 (bottom panels). Black (×) markers with error bars are the data included in our extrapolation. Extrapolations with errors that are predicted from Bayesian model averaging are indicated by multi-colored bands at $t/a^2 = 2.5$ (red), 3.5 (light green), 4.5 (cyan), and 6.0 (light purple). We do not show the infinite volume extrapolation of $\beta_{\rm GF}(t; L, g_0^2)$ at $t/a^2 = 4.5$ for visualization purposes.

estimated and kept track of throughout the fitting process. See Chapter 8.2 for more information about `SwissFit`.

5.4.1 Infinite volume extrapolation with Bayesian model averaging

The first step of the CBFM is to take the infinite volume limit. In this chapter, I extrapolate both $g^2_{\rm GF}(t;L,g^2_0)$ and $\beta_{\rm GF}(t;L,g^2_0)$ to $a/L\to 0$ at fixed β_b and t/a^2 using the ansatz of Eqn. 4.38, as was done in Chapter 4. However, I improve upon the error estimation by leveraging Bayesian model averaging (BMA) [210, 273, 274] to include the systematic error that is associated with fitting over subsets of volumes into the combined statistical/systematic error automatically. In Appendix D, I discuss the background of BMA in detail. In practice, I implement the BMA procedure for the infinite volume extrapolation at each fixed $\beta_b, t/a^2$ as follows.

(1) **Fit over subsets:** For each subset η of the full set of available volumes $L/a\in\{24,28,32,36,40\}$ with at least three volumes, I perform a fit over the subset η using Eqn. 4.38. I further denote the estimate for $k_i(t;g^2_0)$ in Eqn. 4.38 from fit η as $k_i^{(\eta)}(t;g^2_0)$. From the χ^2 of fit η (see Chapter 8.2 for a refresher on the χ^2), I associate a *model weight* w_η to subset η as

$$w_\eta \propto \exp\left[\text{-}\frac{1}{2}\left(\chi^2_\eta + 2d_\eta\right)\right], \tag{5.4}$$

where χ^2_η is the χ^2 of fit η and d_η is the number of volumes *not* included in fit η from the full set of available volumes. The extra $2d_\eta$ term acts as a penalty in the model weight w_η for not including data [210, 274]. The model weights w_η are normalized such that $\sum_\eta w_\eta = 1$.

(2) **Estimate of the mean:** Denote the posterior mode of $k_i(t;g^2_0)$ from fit η as $\overline{k}_i^{(\eta)}(t;g^2_0)$. Then the Bayesian-model-averaged mean $\overline{k}_i(t;g^2_0)$ is estimated from the weights w_η as

$$\overline{k}_i(t;g^2_0) = \sum_\eta \overline{k}_i^{(\eta)}(t;g^2_0)\,w_\eta. \tag{5.5}$$

In other words, the mean of $\overline{k}_i(t;g^2_0)$ is estimated from a weighted sum over the posterior mode of $\overline{k}_i^{(\eta)}(t;g^2_0)$ from each fit η.

(3) **Estimate of the covariance:** Denote covariance of $\{k_i^{(\eta)}(t;g^2_0)\}$ from fit η as $C_{ij}^{(\eta)}(t;g^2_0)$.

Then the covariance suggested by BMA is estimated from $C_{ij}^{(\eta)}\big(t;g_0^2\big)$ as

$$C_{ij}\big(t;g_0^2\big)=\sum_{\eta}C_{ij}^{(\eta)}\big(t;g_0^2\big)w_\eta+\sum_{\eta}\overline{k}_i^{(\eta)}\big(t;g_0^2\big)\overline{k}_j^{(\eta)}\big(t;g_0^2\big)w_\eta-\overline{k}_i\big(t;g_0^2\big)\overline{k}_j\big(t;g_0^2\big). \tag{5.6}$$

The total error for k_i is C_{ii}. It includes both the statistical error $C_{ii}^{(\eta)}$ from each fit η via a weighted sum and an additional two terms that collectively represent the contribution from the *model uncertainty.*

The BMA procedure above is repeated over each $(\beta_b, t/a^2)$ pair for both $g_{\mathrm{GF}}^2\big(t;L,g_0^2\big)$ and $\beta_{\mathrm{GF}}\big(t;L,g_0^2\big)$.

In Fig. 5.2, I illustrate the infinite volume extrapolation of $g_{\mathrm{GF}}^2\big(t;L,g_0^2\big)$ (left panels) and $\beta_{\mathrm{GF}}\big(t;L,g_0^2\big)$ (right panels) at $\beta_b=9.6$ (top panels), 9.8 (middle panels), and 10.2 (bottom panels) for the $\mathcal{S}^{\mathrm{f}}\mathcal{S}^{\mathrm{e}}=\mathrm{WW}$ combination. I show the same information in Fig. 5.2, but for the $\mathcal{S}^{\mathrm{f}}\mathcal{S}^{\mathrm{e}}=\mathrm{WC}$ combination. The bare gauge couplings shown in both figures are chosen such the the $g_{\mathrm{GF}}^2\big(t;g_0^2\big)$ from the infinite volume extrapolation are within the vicinity of the continuum predictions for $g_{\mathrm{GF}\star}^2$. For all three β_b, the $L/a=24$ volume deviates from the linear trend in a^4/L^4. As such, its contribution to the model average is negligible. The four largest volumes generally fit the linear trend in a^4/L^4 well and any minor deviations result in a larger model uncertainty.

5.4.2 Continuum extrapolation

The continuum extrapolation at each fixed g_{GF}^2 is performed by fitting $\big(\beta_{\mathrm{GF}}\big(t;g_0^2\big),t/a^2\big)$ to Eqn. 4.39. As was the case in Chapter 4, doing so at *any* fixed g_{GF}^2 requires first interpolating $\beta_{\mathrm{GF}}\big(t;g_0^2\big)$ in $g_{\mathrm{GF}}^2\big(t;g_0^2\big)$ at fixed t/a^2.

5.4.2.1 Intermediate interpolation

The β-function for the twelve flavor system is slowly running and its curvature in g_{GF}^2 does not change as dramatically as it had for the pure Yang-Mills system. Therefore, the interpolating

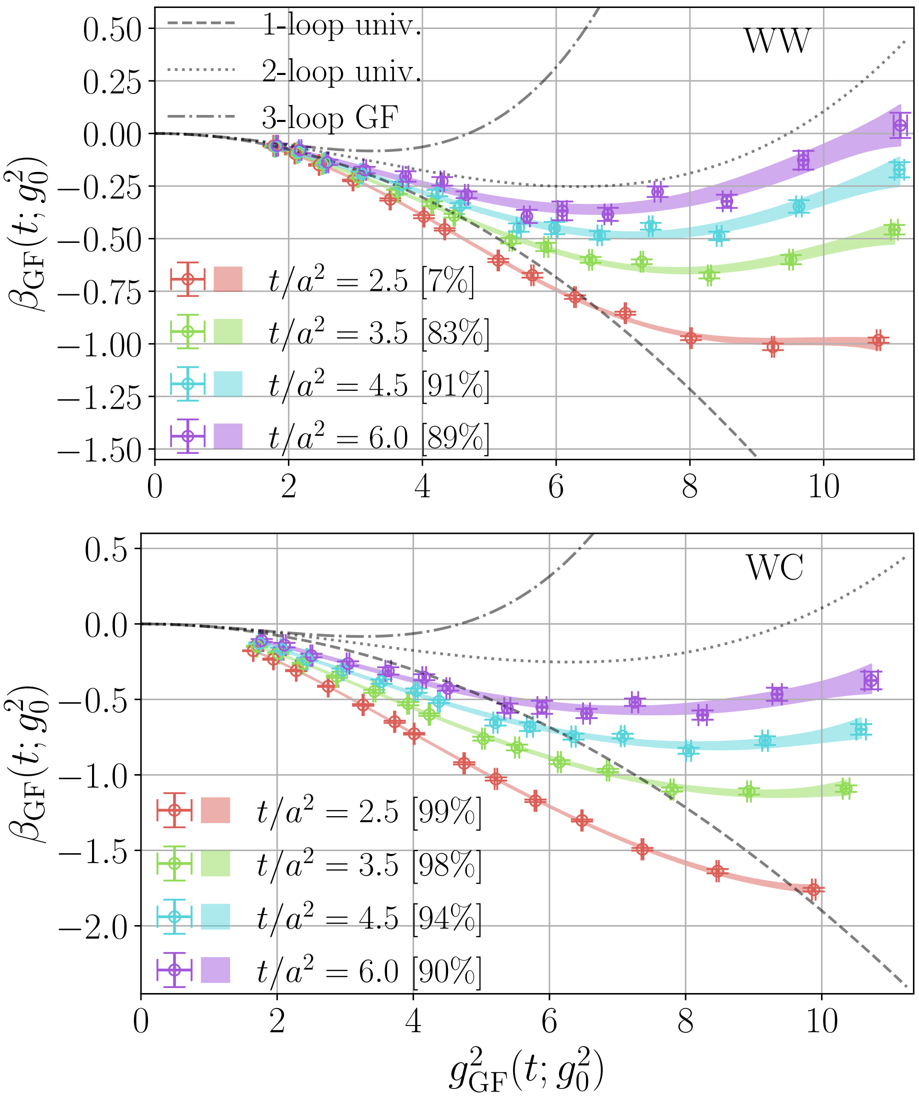

Figure 5.4: From Ref. [293]. Illustration of our interpolation of $\beta_{\mathrm{GF}}(t; g_0^2)$ in $\beta_{\mathrm{GF}}(t; g_0^2)$ for the Wilson operator (top panel) and clover operator (bottom panel). Interpolations at fixed t/a^2 are indicated by colored bands, with $t/a^2 = 2.5$ (red), 3.5 (light green), 4.5 (cyan), and 6.0 (light purple). The width of the band indicates the error. The data contributing to each interpolation is indicated by an open circular marker with both x- and y-errors. We compare our interpolation against the continuum 1- (dashed), 2- (dotted), and 3-loop (dashed-dotted) gradient flow β function from perturbation theory [161].

function need not be overly complicated and a simply polynomial ansatz of the form

$$\mathcal{I}_N\left(g^2_{\mathrm{GF}}\right) = g^4_{\mathrm{GF}} \sum_{i=0}^{N-1} p_n g^{2n}_{\mathrm{GF}} \tag{5.7}$$

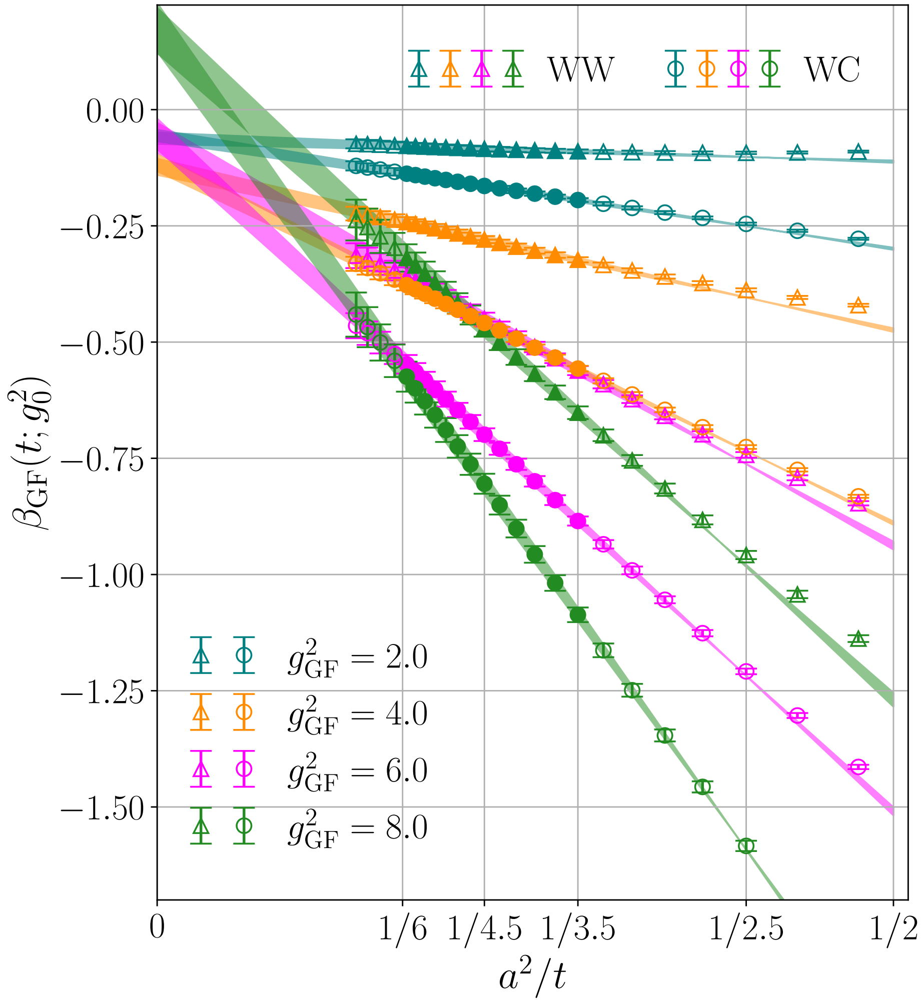

Figure 5.5: From Ref. [293]. Illustration of our continuum extrapolation of $\beta_{\rm GF}(t;g_0^2)$ at fixed $g^2_{\rm GF} = 2.0$ (teal), 4.0 (dark orange), 6.0 (magenta), and 8.0 (forest green). Data contributing to our extrapolation with the W operator are shown as error bars with triangular markers and the C operator are shown as error bars with circular markers. Our extrapolations are shown as colored bands, where the error is indicated by the width of the band.

is sufficient to capture the curvature for all intents and purposes. The leading $\mathrm{O}\left(g^4_{\rm GF}\right)$ term in Eqn. 5.7 forces the known weak coupling behavior of the β-function up to discretization effects. As was the case for the pure Yang-Mills system, forcing such behavior at weak coupling may cut off otherwise viable small t/a^2 for use in the continuum extrapolation. Unlike the analysis presented in Chapter 4, I include the statistical error in $g^2_{\rm GF}(t;g_0^2)$ by treating it as a Gaussian prior (see Appendix D). Moreover, I put a prior on each p_n with a mean of zero and a width of unity to stabilize each fit.

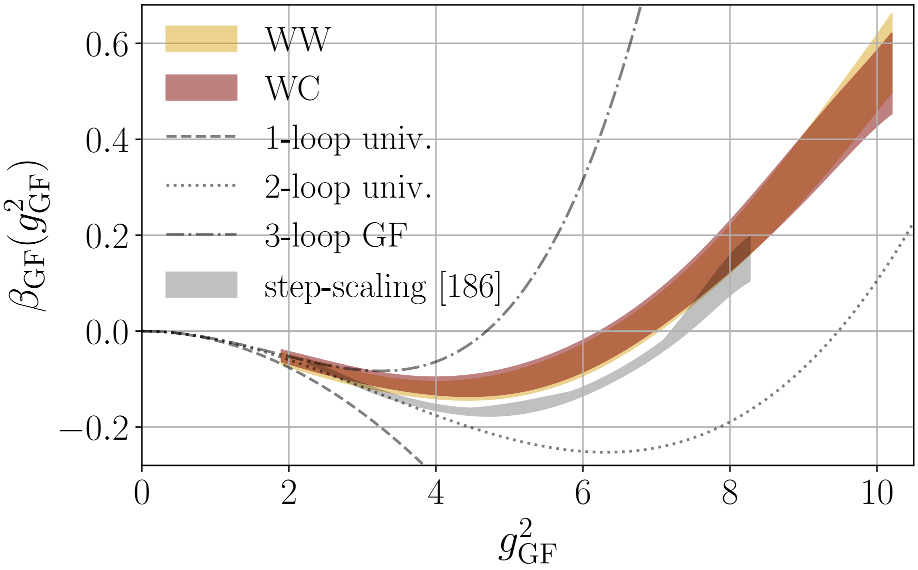

Figure 5.6: From Ref. [293]. Our continuum prediction for $\beta_{\mathrm{GF}}(g^2_{\mathrm{GF}})$ as a function of g^2_{GF} for the W operator (gold band) and C operator (maroon band). The width of the band indicates the error. The nonperturbative results are juxtaposed against the 1- (dashed), 2- (dotted), and 3-loop (dashed-dotted) gradient flow β function from perturbation theory [161]. Also shown is the step-scaling β function in the $c = 0.25$ scheme from Ref. [186] as a grey band.

The order N of the polynomial is chosen to be the smallest value that reasonably fits the data. This turns out to be $N = 4$, as $N \leq 3$ are unable to accommodate for slight variations in the curvature and hence possess poor p-values. In Sec. 5.5, I explore the systematic effect associated with the order N of the interpolating polynomial in my estimate of the fixed point coupling $g^2_{\mathrm{GF}\star}$ and leading irrelevant critical exponent $\gamma^\star_g$.

If Fig. 5.4, I illustrate the intermediate interpolation of $\beta_{\mathrm{GF}}(t; g_0^2)$ in $g^2_{\mathrm{GF}}(t; g_0^2)$ using the ansatz of Eqn. 5.7 at fixed $t/a^2 = 2.5$ (red), 3.5 (green), 4.5 (blue), and 6.0 (purple) for the WW combination (top panel) and the WC combination (bottom panel). The interpolation is juxtaposed against the 1-, 2-, and 3-loop continuum β-function from perturbation theory [161]. Over the t/a^2 that will enter the continuum extrapolation, the p-values tend to be quite high. This indicates that I could be overfitting. However, it may also be the case that the estimate for the error from the BMA is an overestimate. As I have already noted, reducing the order makes the intermediate interpolation significantly worse. As a remedy for potentially overfitting, I could decrease the width of the prior on the p_n; however, this has little effect on the outcome of the continuum extrapolation, so I leave the prior on p_n as is.

5.4.2.2 Continuum extrapolation

With the intermediate interpolation taken care of, we are ready to take the continuum limit. No improvements over the analysis in Chapter 4 are made to the continuum extrapolation. The β-function $\beta_{\mathrm{GF}}\big(t; g_0^2\big)$ is again extrapolated to the continuum limit over a set of fixed g_{GF}^2 and the issue of strong correlations in each fit persist. Issues with such such correlations are again remedied by first performing an uncorrelated fit of $\beta_{\mathrm{GF}}\big(t; g_0^2\big)$ in a^2/t to estimate the central value. Then uncorrelated fits to $\beta_{\mathrm{GF}}\big(t; g_0^2\big) \pm 1\sigma$ are used to estimate the statistical uncertainty from the half difference of the central value for $\beta_{\mathrm{GF}}\big(g_{\mathrm{GF}}^2\big)$ from the shifted fits.

In Fig. 5.5, I illustrate the result of the continuum extrapolation for both WW and WC at $g_{\mathrm{GF}}^2 = 2.0$ (blue), 4.0 (yellow), 6.0 (pink), and 8.0 (green). For $t/a^2 \gtrsim 6.0$, the data begins to deviate from the linear trend, indicating the the infinite volume extrapolation is no longer reliable beyond $t/a^2 \approx 6.0$. The same is true for $t/a^2 \lesssim 3.5$, which indicates that cutoff effects are not sufficiently suppressed until $t/a^2 \approx 3.5$. The choice $[t_{\mathrm{min}}, t_{\mathrm{min}}]/a^2 = [3.5, 6.0]$ lies between these two regimes; hence, it is chosen as the preferred flow time range for the central analysis of this chapter. In Sec. 5.5, I will explore the effect of varying $t_{\mathrm{min}}/a^2, t_{\mathrm{min}}/a^2$ on the estimate of the fixed point coupling $g_{\mathrm{GF}\star}^2$ and leading irrelevant critical exponent $\gamma_g^\star$.

5.4.3 The continuum β-function

The continuum prediction for $\beta_{\mathrm{GF}}\big(g_{\mathrm{GF}}^2\big)$ from the WW and WC combinations is shown in Fig. 5.6. I once again juxtapose the continuum prediction for $\beta_{\mathrm{GF}}\big(g_{\mathrm{GF}}^2\big)$ against its 1-, 2-, and 3-loop continuum counterparts from perturbation theory [161]. Also shown for the purposes of comparison is the continuum step-scaling β-function from Ref. [186]. The continuum prediction for $\beta_{\mathrm{GF}}\big(g_{\mathrm{GF}}^2\big)$ from both flow/operator combinations is consistent, overlapping well within error. At weak coupling ($g_{\mathrm{GF}}^2 \lesssim 4\pi$), the β-function shows signs of converging to the 1-, 2-, and 3-loop perturbative β-function.

The β-function crosses the origin in g^2_{GF} around $g^2_{\mathrm{GF}\star} \approx 6.6$ for both operators. The prediction for $g^2_{\mathrm{GF}\star}$ from the step-scaling β-function in Ref. [186] differs from the latter prediction; however, the step-scaling β-function is obtained from a different RG scheme. Therefore, the prediction for the fixed point coupling need not agree with the step-scaling prediction.

5.5 Leading irrelevant critical exponent

The renormalized coupling g^2_{GF} at the IRFP is *irrelevant*. The renormalized coupling g^2_{GF} is also a scaling variable of the IRFP. According to the discussion in Chapter 2, the β-function $\beta_{\mathrm{GF}}\left(g^2_{\mathrm{GF}}\right)$ depends on g^2_{GF} within the vicinity of $g^2_{\mathrm{GF}\star}$ as

$$\beta_{\mathrm{GF}}\left(g^2_{\mathrm{GF}}\right) \approx \frac{\gamma^\star_g}{2}\left(g^2_{\mathrm{GF}} - g^2_{\mathrm{GF}\star}\right) \qquad \left(g^2_{\mathrm{GF}} \approx g^2_{\mathrm{GF}\star}\right), \tag{5.8}$$

where $\gamma^\star_g$ is the *leading irrelevant critical exponent.* The factor of 1/2 in Eqn. 5.8 is conventional. In Chapter 2, I stated that the critical exponent of scaling variables are *universal.* That is, they are a *property* of the physical system that must be independent of the RG scheme used to extract them. The solution to the RG equation within the vicinity of the fixed point is

$$g^2_{\mathrm{GF}}(t) = g^2_{\mathrm{GF}\star} + \left(\Lambda t\right)^{-\gamma^\star_g/2} \qquad \left(g^2_{\mathrm{GF}} \approx g^2_{\mathrm{GF}\star}\right), \tag{5.9}$$

where Λ is an integration constant that sets the dimension of t. Eqn. 5.9 shows explicitly that the β-function for positive $\gamma^\star_g$ describes the running of an *irrelevant* coupling, as the solution to the RG equation within the vicinity of the fixed point indicates that $g^2_{\mathrm{GF}}(t) \to g^2_{\mathrm{GF}\star}$ as $t \to \infty$. The rate at which $g^2_{\mathrm{GF}}(t)$ decays to $g^2_{\mathrm{GF}\star}$ is characterized by the size of $\gamma^\star_g$, which enters as a potentially non-integral exponent in Eqn. 5.9. Note that $\gamma^\star_g$ is also often referred to as the *anomalous dimension* of $\langle \mathrm{Tr_c}\left[\mathfrak{F}_{\mu\nu}\mathfrak{F}_{\mu\nu}\right]\rangle$, which is somewhat coincidentally quite clear already from the definition of g^2_{GF} in terms of the gradient-flowed Yang-Mills energy $E(t)$ density in Eqn. 4.13.

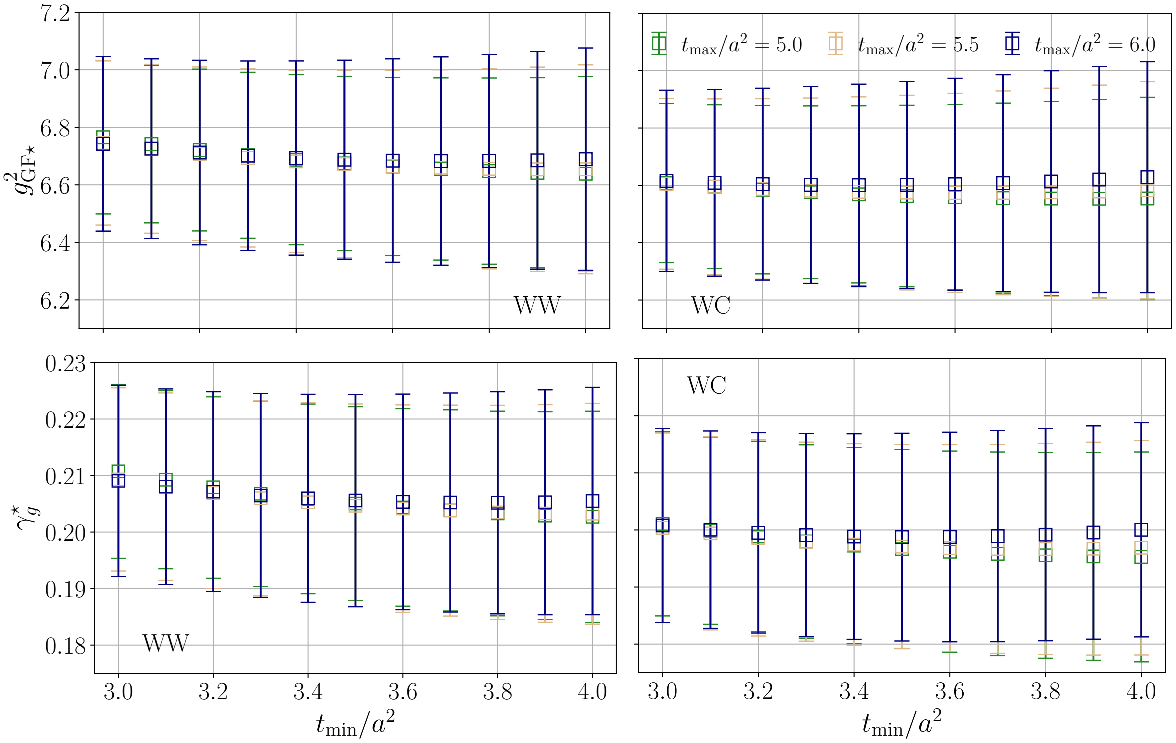

Figure 5.7: From Ref. [293]. Comparison of our estimated $g^2_{\mathrm{GF}\star}$ and $\gamma_g^\star$ for different t_{min}/a^2 (x-axes) and $t_{\mathrm{max}}/a^2 = 5.0$ (green), 5.5 (gold), 6.0 (navy) from the continuum extrapolation.

5.5.1 Calculation of the leading irrelevant critical exponent

Because the continuum extrapolation in Sec. 5.4.2.2 approximates the covariance matrix of $\beta_{\mathrm{GF}}\left(t; g_0^2\right)$ by its diagonal component, I no longer have access to the correlation matrix for $\beta_{\mathrm{GF}}\left(g^2_{\mathrm{GF}}\right)$ among the chosen g^2_{GF}. To ensure that the statistical error is not underestimated, I take $\beta_{\mathrm{GF}}\left(g^2_{\mathrm{GF}}\right)$ among each g^2_{GF} to be 100% correlated and estimate both $g^2_{\mathrm{GF}\star}$ and $\gamma_g^\star$ via the following procedure.

(1) **Interpolation:** I first interpolate the central value for $\beta_{\mathrm{GF}}\left(g^2_{\mathrm{GF}}\right)$ among a grid of g^2_{GF} using the "*Steffen*" monotonic spline algorithm provided by the `gvar` library [241, 342]. The latter spline interpolation is also repeated for the central value of $\beta_{\mathrm{GF}}\left(g^2_{\mathrm{GF}}\right) \pm 1\sigma$. The grid of g^2_{GF} for all three spline interpolations has $\delta g^2_{\mathrm{GF}} = 0.1$. The resulting estimate for $g^2_{\mathrm{GF}\star}$ and $\gamma_g^\star$ from Steps 2-4 is not sensitive to the choice of δg^2_{GF} within their statistical uncertainty.

(2) **Central value of $\boldsymbol{g^2_{\mathrm{GF}\star}}$:** I estimate the central value for $g^2_{\mathrm{GF}\star}$ by applying a derivative-free

root finding algorithm on the spline interpolation of $\beta_{\mathrm{GF}}\big(g^2_{\mathrm{GF}}\big)$ from Step 1 [52]. The root finding algorithm is provided by the `gvar` library [241].

(3) **Central value of $\gamma^{\star}_g$:** The central value of $\gamma^{\star}_g$ is obtained from the prediction for $g^2_{\mathrm{GF}\star}$ in Step 1 by evaluating the natural derivative of the monotonic spline at $g^2_{\mathrm{GF}\star}$ [342].

(4) **Statistical error of $g^2_{\mathrm{GF}\star}$ and $\gamma^{\star}_g$:** I estimate the statistical error for both $g^2_{\mathrm{GF}\star}$ and $\gamma^{\star}_g$ by repeating Steps 2-3 for the spline interpolation over the central value of $\beta_{\mathrm{GF}}\big(g^2_{\mathrm{GF}}\big) \pm 1\sigma$ from Step 1 and taking the statistical error to be the half difference in the central value obtained from both, respectively.

The latter steps yield a prediction of $g^2_{\mathrm{GF}\star} = 6.69(35), 6.60(36)$ and $\gamma^{\star}_g = 0.206(19), 0.199(18)$ for WW and WC, respectively. Globally, results from the WC flow/operator combination appear to be more stable against variations in analysis than the WW combination. As such, I take WC to be the preferred flow/operator combination.

5.5.2 Systematic errors

I consider additional estimates for the systematic error on $\gamma^{\star}_g$ by varying the analysis presented in Secs. 5.4-5.5.1 as follows.

- **Order of the intermediate interpolation:** The highest order polynomial that I can use for the intermediate interpolation ansatz of Eqn. 5.7 before I lose control over the continuum extrapolation due to overfitting is $N = 6$. Therefore, I vary the intermediate interpolation by choosing $N = 6$, which results in a shift of $\delta\gamma^{\star}_g \approx 0.001$. The latter shift is taken as an estimate for the systematic error that is associated with the order of the intermediate interpolation.

- **Flow time range for the continuum extrapolation:** In the bottom panel of Fig. 5.7, I illustrate the dependence of $\gamma^{\star}_g$ on $t_{\mathrm{min}}/a^2 \in [3.0, 4.0]$ at fixed $t_{\mathrm{max}} = 5.0$ (green), 5.5 (tan), and 6.0 (blue) for the WW combination (bottom left) and WC combination (bottom right).

Though the central value for $\gamma_g^\star$ is consistent across all variations in $t_{\rm min}/a^2, t_{\rm max}/a^2$, it stabilizes around $t_{\rm min}/a^2 \gtrsim 3.5$ at $t_{\rm max} \approx 6.0$, lending further justification for the choice of flow time range taken in Sec. 5.4.2.2. I estimate the systematic error that is associated with the continuum flow time range as the difference between the most extreme values for $\gamma_g^\star$ in the bottom panel of Fig. 5.7, yielding $\delta\gamma_g^\star \approx 0.006$.

- **Choosing the W operator over C:** The final variation is simply to choose the WW combination over the WC combination. This yields an estimate for the systematic uncertainty from the choice of flow/operator combination $\delta\gamma_g^\star \approx 0.007$.

As a conservative estimate for the combined statistical/systematic error, I estimate the total error by combining the latter variations with the statistical error linearly. This yields an estimate of

$$\gamma_g^\star = 0.199(32) \tag{5.10}$$

for the leading irrelevant critical exponent. Repeating the latter variations for $g^2_{\rm GF\star}$ yields a systematic error of ≈ 0.12 from the interpolation order, ≈ 0.05 from the continuum flow time range, and ≈ 0.09 from the flow/operator combination. Combining the latter systematic errors for $g^2_{\rm GF\star}$ linearly with its statistical error yields an estimate of

$$g^2_{\rm GF\star} = 6.60(62) \tag{5.11}$$

for the fixed point value of the renormalized coupling $g^2_{\rm GF}$.

5.5.3 Comparison against the literature

There is a small, but non-negligible, literature on calculations of $\gamma_g^\star$ from studies that conclude the twelve flavor system to be conformal on the basis of observing a zero of the RG β-function. The 4-loop β-function in the $\overline{\rm MS}$ scheme produces the estimate $\gamma_g^\star \approx 0.282$ without utilizing resummation techniques [298], while lower orders in the loop expansion tend to yield larger estimates and are

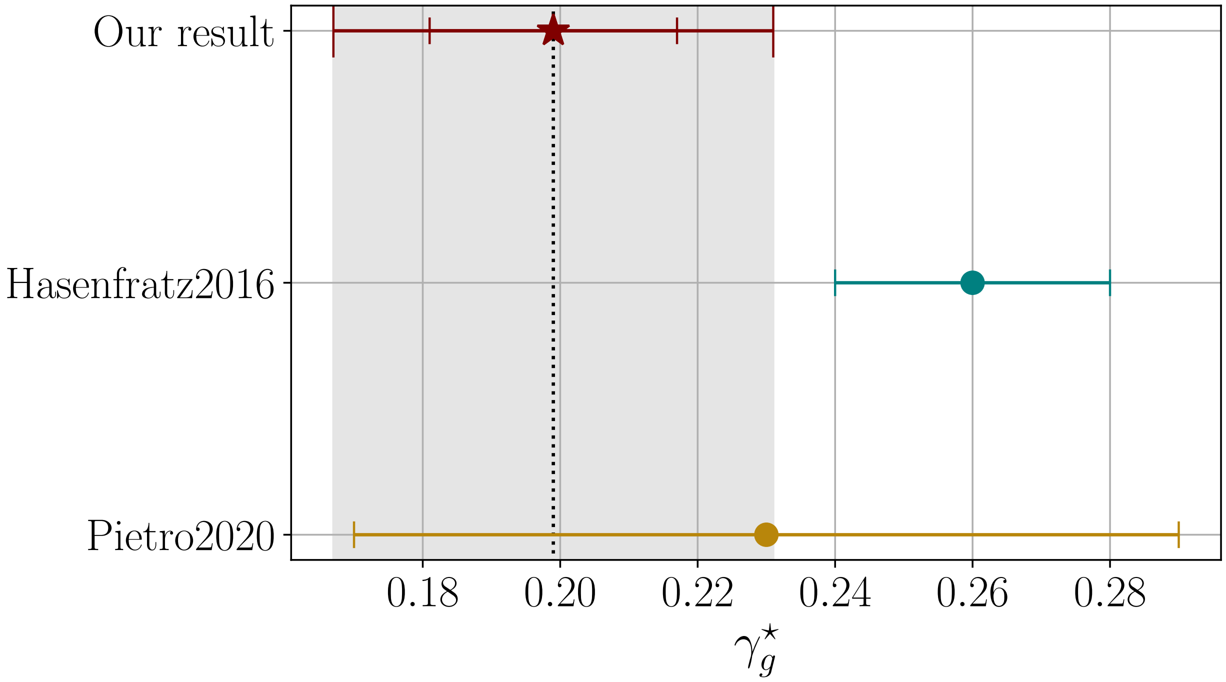

Figure 5.8: From Ref. [293]. Comparison of our value for $\gamma_g^\star$ (maroon errorbar) against Ref. [186] (teal error bar) and Ref. [106] (dark gold error bar). The smaller error bar on our result indicates the error without accounting for systematic effects; the larger error bar indicates our error after accounting for systematic effects. We indicate our total error with a grey band for visualization.

fairly unreliable. After applying Borel resummation to the perturbative 5-loop β-function, Ref. [106] estimates $\gamma_g^\star = 0.23(6)$, where the error is purely systematic. The "scheme-independent" approach of Ref. [321] also estimates $\gamma_g^\star \approx 0.228$ at 5-loop order. The only non-perturbative estimate of $\gamma_g^\star$ that the author of this thesis is aware of comes from Ref. [186], which yields $\gamma_g^\star = 0.26(2)$ using the gradient-flow-based step-scaling method that I briefly touched upon in Chapter 4.3. In Fig. 5.8, I compare the estimate for $\gamma_g^\star$ from Secs. 5.5.1-5.5.2 to the step-scaling estimate $\gamma_g^\star = 0.26(2)$ from Ref. [186] (Hasenfratz2016, blue point) and the Borel-resummed perturbative estimtae $\gamma_g^\star = 0.23(6)$ from Ref. [106] (DiPietro2020, dark gold point). The estimate in this chapter is not consistent with the step-scaling estimate in Ref. [186] within the combined uncertainty of both approaches, but it is consistent with the Borel-resummed perturbative estimate of Ref. [106] (within the combined uncertainty). The "scheme-independent" estimate $\gamma_g^\star \approx 0.228$ of Ref. [321] is also consistent with $\gamma_g^\star = 0.199(32)$; however, it lies on the edge of the combined statistical/systematic error, which I have already stated could be an overestimate. Overall, one can confidently state that the estimate for $\gamma_g^\star$ in this chapter is consistent with the available literature at the 1-2σ leve.

Chapter 6

Interlude: Finite size scaling with radial basis function networks

> *All models are wrong, but some are useful*
>
> George E. P. Box [45]

After having explored the use of renormalization group techniques in systems well outside of the conformal window ($N_{\mathrm{f}} = 0$, Chapter 4) and likely well inside of the conformal window ($N_{\mathrm{f}} = 12$, Chapter 5), the latter part of this thesis is dedicated to trying to understand a system ($N_{\mathrm{f}} = 8$, Chapter 7) that could be on, or at least close to, the edge of the conformal window. This is going to be an extremely difficult task, so before embarking on this journey we shall make a pit stop and explore some techniques that have been developed for tackling the $N_{\mathrm{f}} = 8$ system. This will take us into a world of well-understood two-dimensional spin systems, which may be just what we need before we jump deep into the unknown.

Finite size scaling (FSS, Chapter 2.5) by the method of curve collapse is a simple, yet powerful theoretical tool for probing the properties of a given phase transition, but it must be wielded with care. Assuming that the FSS dataset that one is working with is up to snuff, the primary hurdle to get over when working curve collapse is choosing an appropriate parametric ansatz for the scaling function. This chapter is dedicated to exploring the use of radial basis function networks (RBFNs), a special type of single-layer artificial neural network, are capable of doing the job. To start, Sec. 6.1introduces the method of curve collapse. Sec. 6.2 introduces radial basis function neural networks and how to deployed them in a curve collapse analysis. The RBFN-based curve collapse is tested on the two-dimensional p-state Potts model for $p = 2, 3$ in Sec. 6.3 and the q-state clock model for $q = 4, \infty$ in Sec. 6.4. I end this chapter by speculating other uses of the RBFNs in lattice field theory analyses in Sec. 6.5; as an example, I explore a method for extracting the critical temperature of the ∞-state clock (XY) model from a direct interpolation of the helicity modulus using an RBFN. The content of this chapter is based on Ref. [292].

6.1 Curve collapse

In Chapter 2.5, I discussed the general idea of finite size scaling (FSS) in terms of how the singular part of the free energy $A_{\varphi}^{(\mathrm{s})}\big(K, N_{\mathrm{s}}^{-1}\big)$ transforms under an RG transformation in a finite volume $N_{\mathrm{s}}^{d} \equiv (L/a)^{d}$. As a corollary to that discussion, finite-volume observables $O(K, N_{\mathrm{s}})$ derived from $A_{\varphi}^{(\mathrm{s})}\big(K, N_{\mathrm{s}}^{-1}\big)$ scale within the vicinity of a continuous phase transition as [69, 303]

$$O\big(K, N_{\mathrm{s}}\big) = N_{\mathrm{s}}^{\gamma_O} \mathcal{F}_O\big(\tilde{\xi}(K)/N_{\mathrm{s}}\big) \qquad (K \approx K_{\mathrm{c}}), \tag{6.1}$$

where γ_O is the leading anomalous dimension of O at K_{c}, $\mathcal{F}_O$ is a universal scaling function, and $\tilde{\xi}(K)$ is the infinite volume correlation length in units of the lattice spacing. I have also assumed that the scaling behavior does not differ depending on the direction that we approach K_{c} from; this will be the case for the systems that I study in the rest of this thesis. For a 2nd-order phase transition

$$\tilde{\xi}(K) \sim |k|^{-\nu} \qquad \text{(2nd-order)},$$

where $k \equiv K/K_{\mathrm{c}} - 1$; therefore,

$$O\big(K, N_{\mathrm{s}}\big) = N_{\mathrm{s}}^{\gamma_O} \mathcal{F}_O\big(|k| N_{\mathrm{s}}^{1/\nu}\big) \qquad (K \approx K_{\mathrm{c}},\ \text{2nd-order}). \tag{6.2}$$

Note that the $\mathcal{F}_O$ from Eqn. 6.1 differs from that of Eqn. 6.2 because I have manipulated its arguments; however, I am referring to them as the same function for notational brevity. Meanwhile, for an ∞-order phase transition,

$$\tilde{\xi}(K) \sim \exp\left(\zeta |k|^{-\nu}\right) \qquad (\infty\text{-order});$$

therefore,

$$O\big(K, N_{\mathrm{s}}\big) = N_{\mathrm{s}}^{\gamma_O} \mathcal{F}_O\big(N_{\mathrm{s}} \exp\left(\text{-}\zeta |k|^{-\nu}\right)\big) \qquad (K \approx K_{\mathrm{c}},\ \infty\text{-order}), \tag{6.3}$$

again after a little rearrangement of the argument.

Eqns. 6.2 and 6.3 hint at a method for extracting the critical parameters K_{c}, ν and possibly ζ (depending on the order of the phase transition) from a series of finite volume simulations at various (K, N_{s}) pairs. By simulating the system on multiple volumes N_{s} and couplings K that are in the vicinity of K_{c}, one may estimate the critical parameters by requiring that

$$O\big(K, N_{\mathrm{s}}\big)/N_{\mathrm{s}}^{\gamma_O} = \mathcal{F}_O(x) \tag{6.4}$$

is a unique function of

$$x(K, N_{\mathrm{s}}) \equiv |k| N_{\mathrm{s}}^{1/\nu} \qquad (\text{2nd-order}) \tag{6.5}$$

for a 2nd-order phase transition or

$$x(K, N_{\mathrm{s}}) \equiv N_{\mathrm{s}} \exp\left(\text{-}\zeta |k|^{-\nu}\right) \qquad (\infty\text{-order}) \tag{6.6}$$

for a ∞-order phase transition. When the critical parameters have been identified correctly, the individual curves for $O(K, N_{\mathrm{s}})$ in K at fixed N_{s} are said to *collapse* onto one another when re-expressed in terms of $x(K, N_{\mathrm{s}})$ and rescaled by $N_{\mathrm{s}}^{-\gamma_O}$. The latter method is therefore referred to as *curve collapse.*

As the scaling function $\mathcal{F}_O(x)$ is a priori unknown and (usually) not determined by a finite number of parameters (i.e., it is non-parametric), it is typically estimated from a parametric ansatz whose parameters are determined as part of the curve collapse analysis; for example, a polynomial

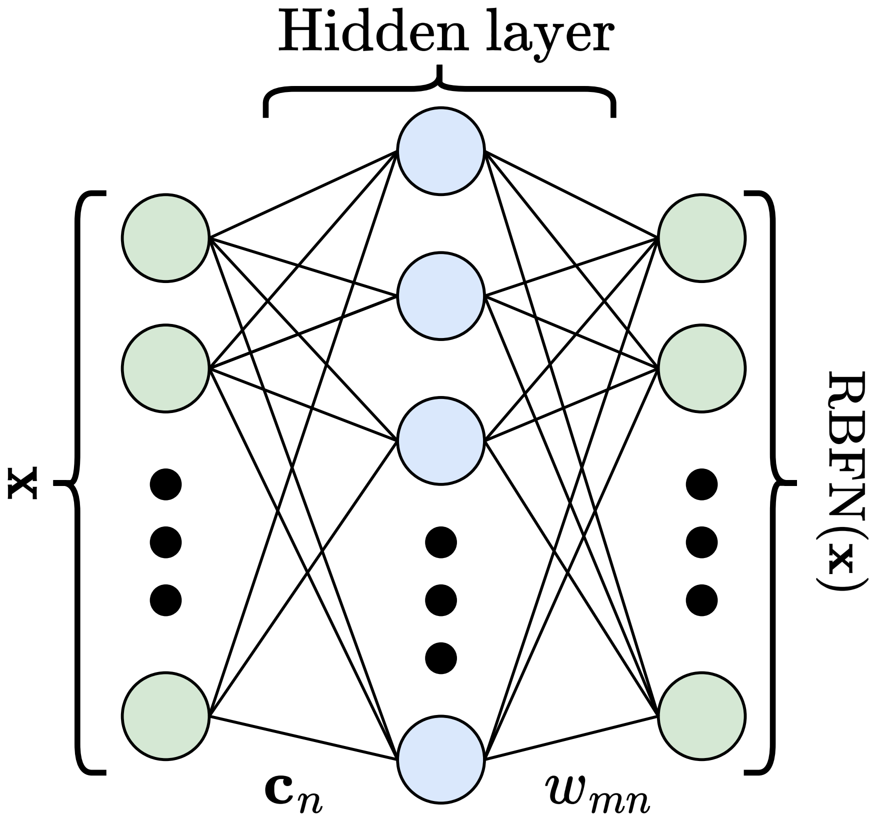

Figure 6.1: From Ref. [292]. Illustration of a radial basis function network (RBFN). The total number of center parameters $\mathbf{c}_n$, counting the components of each $\mathbf{c}_n$, is equal to the number of connections between the input nodes (green circles on left; a.k.a., input *features*) and hidden nodes (blue circles). The number of weights is equal to the number of connections between the hidden nodes and output nodes (green circles on right; a.k.a., output *features*).

or ratio of polynomials. With so much freedom in choosing the ansatz for $\mathcal{F}_O(x)$, it would be nice to have on hand a family of interpolating functions that generally perform well at interpolating over a variety of curves. As universal function approximators, radial basis function networks could be just the right tool for the job. The rest of this chapter is therefore dedicated to exploring the use of radial basis function networks in various finite size scaling problems.

6.2 Radial basis function networks and finite size scaling

A radial basis function network is a special type of single-layer artificial neural network that is specially designed for the purposes of function approximation [145]. In Fig. 6.1, I illustrate the general structure of an RBFN. The inputs $\mathbf{x} \in \mathbb{R}^L$ (left green nodes in Fig. 6.1) of the network are connected to the nodes of the hidden layer (center blue nodes) by the center parameters $\mathbf{c}_n \in \mathbb{R}^L$. The output of each hidden node is then passed through a radial basis function ρ and aggregated linearly into the output nodes (right green nodes) via matrix multiplication with the network weights w_{mn}. The output of the full network is

$$\mathrm{RBFN}_m(\mathbf{x}) = \sum_n w_{mn}\rho\big(\text{-}\beta_n^2||\mathbf{x} - \mathbf{c}_n||^2\big) + b_m, \tag{6.7}$$

where $\mathbf{c}_n$ and w_{mn} are the aforementioned center parameters and network weights, respectively; $\{b_m\}$ are the *network biases*; and $\{\beta_n\}$ are the radial basis function bandwidths. In this chapter, the radial basis functions are exponential

$$\rho(\cdot) = \exp(\cdot) \tag{6.8}$$

and the norm $||\cdot||$ is the standard Euclidean norm; as such, the output of each hidden node has a Gaussian profile. The total number of free parameters $\Theta_{\mathrm{RBFN}} \equiv \{\mathbf{c}_n, \beta_n, w_{mn}, b_m\}$ for an RBFN is the same as for a standard feedforward neural network.

According to the *universal approximation theorem* for RBFNs, an RBFN with exponential activation units in its hidden layer is capable of approximating any continuous function on a compact subset of $\mathbb{R}^M$ with an accuracy that scales with the number of hidden nodes [288]. For problems involving large, complicated datasets, RBFNs have been largely phased out by various kinds of deep neural networks; however, such neural networks are completely unnecessary for the small-scale parameter estimation problems that are needed for most lattice field theory analyses. For such small-scale problems, the simplicity and approximation power of RBFNs could be of great benefit.

6.2.1 Finite size scaling with radial basis function networks

For the purposes of FSS by the method of curve collapse, I estimate the scaling function $\mathcal{F}_O(x)$ defined by Eqns. 6.4-6.6 by parameterizing it with an RBFN that possesses as single input node and a single output node. In other words, I take

$$O\big(K, N_{\rm s}\big)/N_{\rm s}^{\gamma_O} = \mathcal{F}_O(x) \approx \text{RBFN}(x) = \sum_n w_n \exp\left[\text{-}\beta_n^2(x - c_n)^2\right] + b, \tag{6.9}$$

where the input x is given by either Eqn. 6.5 or 6.6 and center parameters c_n are scalar. Because the multi-dimensional hidden layer is collapsed onto a single node in the output layer, I denote the weights in Eqn. 6.9 as w_n. The full set of free parameters for the RBFN is then $\Theta_{\rm RBFN} = \{c_n, \beta_n, w_n, b\}$. The parameters of the RBFN $\Theta_{\rm RBFN}$ are estimated alongside the critical parameters $\Theta_{\rm crit.} \equiv \{K_{\rm c}, \nu, \gamma_O, \zeta\}$ ($\zeta \in \Theta_{\rm crit.}$ only for ∞-order transitions) by fitting Eqn. 6.9 to data for $O\big(K, N_{\rm s}\big)$ via maximum a posteriori (MAP) estimation. For the types of problems that are encountered in lattice gauge theory, MAP estimation reduces to the problem of minimizing an augmented $\chi^2_{\rm aug.}$, which is the sum of the χ^2 of the data $\chi^2_{\rm data}$ and the χ^2 of the prior $\chi^2_{\rm prior}$. Estimating the "mean" (*posterior mode*) of $\Theta \equiv \Theta_{\rm RBFN} \cup \Theta_{\rm crit.}$ will therefore be done in this chapter by minimizing $\chi^2_{\rm aug.}$; uncertainties in the posterior mode of Θ are estimated from a Laplace approximation of the posterior. Details of MAP estimation, the definition of $\chi^2_{\rm aug.}$, and uncertainty estimation about MAP estimates are all described in detail in Appendix D and Chapter 8.2.

6.2.2 Fitting with the basin hopping optimization algorithm

The landscape of $\chi^2_{\rm aug.}(\Theta)$ in Θ can be very rich in structure. Due in part to the symmetries of RBFNs, $\chi^2_{\rm aug.}(\Theta)$ often possess many (possibly degenerate) local optima with their own basins of attraction, along with steep barriers that separate regions of Θ space. MAP estimation assumes that the parameters Θ have been estimated from the *global optimum* of $\chi^2_{\rm aug.}(\Theta)$. If one wishes to use a local optimization algorithm, finding the global optimum requires using an extremely clever (or lucky) procedure for initializing Θ; however, this is difficult to do in practice and by no means efficient.

Algorithm 1: From Ref. [292]. The basin hopping global optimization algorithm, as implemented in Ref. [364]. The `LocalOptimization` step utilizes the trust region reflective local optimization algorithm [49]. We use the `SciPy` library's implementation of both optimization algorithms [361].

```
Input: $\Theta_0, \alpha, T$
$\Theta \leftarrow$ LocalOptimization($\Theta_0$);
$\Theta_{\text{best}} \leftarrow \Theta$;
while $\Theta_{\text{best}}$ not converged do
    $\Theta' \leftarrow$ RandomPerturbation($\Theta, \alpha$);
    $\Theta' \leftarrow$ LocalOptimization($\Theta'$);
    $\Theta \leftarrow$ MetropolisCriterion($\Theta, \Theta', T$) ;
    if new $\Theta_{\text{best}}$ then
        $\Theta_{\text{best}} \leftarrow \Theta$
    end
end
Output: $\Theta_{\text{best}}$
```

In the machine learning literature, the *Adam* variant of stochastic gradient descent optimization algorithms is the industry standard [225], along with its Nesterov-accelerated counterpart [111]. Because it is stochastic, it *can* achieve the task of finding global optima, though it is not *guaranteed* to. It is also specially designed to be efficient for problems involving large datasets. As such, it can be very inefficient for the small-scale problem of fitting a model to data with even less than a few hundred data points. One way out is to utilize a global optimization algorithm; however, many global optimization algorithms tend to be fairly inefficient for one reason or another. After all, there is no free lunch. This is precisely true for optimization algorithms, which perform equally well when averaged over the set of all optimization problems [379]. Unfortunately, this means that one often has to settle with the algorithm that *generally* works best for a particular class of problems.

Typically, I find that utilizing a global optimization algorithm is more efficient for finding even just stable local optima than cleverly engineering a good initial condition for kick-starting a local search. Though I will not go through the details in this thesis, I have exhaustively tested a variety of popular global optimization algorithms, such as dual simulated annealing [357, 383, 384]; various metaheuristic genetic/evolutionary algorithms such as differential evolution [343, 387] and particle swarm [223]; Bayesian optimization [140]; basin hopping [364]; and many others. Of the global optimization algorithms that I have tested, the basin hopping (BH) algorithm combined with the trust region reflecitve (TRF) local optimization algorithm tends to perform the best at finding global optima, or at least stable local optima, in the smallest number of iterations [49, 364].

Conventional BH algorithms combine the benefits of local search strategies via gradient-based optimization techniques, random perturbations in Θ space (hops), and Metropolis acceptance criteria to guide the algorithm's exploration through rugged, potentially funnel-like, $\chi^2_{\text{aug.}}(\Theta)$ landscapes [364]. Each iteration of BH performs the following steps (see Algorithm 1).

(1) `RandomPerturbation`: Pick a random unit direction $\hat{\Theta}$ and update $\Theta \to \Theta'$ as $\Theta' = \Theta + \alpha\hat{\Theta}$ for step size α. I find that choosing α from a uniform distribution can help the algorithm move around Θ space more efficiently.

(2) `LocalOptimization`: Update $\Theta' \to \Theta'$ by running a local optimization algorithm on Θ'. In this thesis, I use the trust region reflective algorithm for the local optimization [49]. The TRF algorithm is a fantastic general-purpose algorithm, and I typically find that it vastly outperforms other popular local optimization algorithms, such as L-BFGS-B [63], Levenberg-Marquardt [242, 259], and nonlinear conjugate gradient methods [341].

(3) `MetropolisCriterion`: Accept the optimized Θ' from Step 2 with probability

$$\text{acc. prob.} = \exp\left[\text{-max}\Big(0, \chi^2_{\text{aug.}}(\Theta') - \chi^2_{\text{aug.}}(\Theta)\Big)\right]^{1/T}, \tag{6.10}$$

where T is a *temperature hyperparameter* that should not be confused with the physical temperature. A good choice for T is the average separation in $\chi^2_{\text{aug.}}$ among local optima [361]; however, I typically find that simply setting $T = 1$ suffices for all intents and purposes. If Θ' is accepted, then $\Theta' \to \Theta$.

(4) **Global optimum test:** If $\chi^2_{\text{aug.}}(\Theta)$ is lower than $\chi^2_{\text{aug.}}(\Theta_{\text{best}})$, set $\Theta_{\text{best}} = \Theta$.

Unfortunately, the BH algorithm for $T \neq 0$ is not guaranteed to converge. Therefore, the termination criterion for BH is often set by the maximum number of iterations that Θ_{best} has not been improved.

Furthermore, I typically chain runs of the BH algorithm with different α to ensure that I have reached as stable of an optimum as a I can. As there is no test of global optimality without knowing what the global optimum explicitly, this is unfortunately the best that one can do. Despite the downsides of the BH algorithm, it nonetheless performs surprisingly well at finding stable optima of $\chi^2_{\text{aug.}}(\Theta)$ in comparison to any other algorithm that I have used. It is no wonder that it is the industry standard for many problems that involve finding the lowest energy state of atomic clusters, crystals, and biological macromolecules [27, 110, 233, 281, 364, 365, 388]. Moreover, in a recent comparative analysis, BH was found to generally outperform many popular evolutionary strategies on various challenging opimization tasks, save for the covariance matrix adaptation evolution strategy (CMA-ES) [160], which slightly outperformed BH [25]. It would be worth looking into CMA-ES in the future as a better alternative to BH.

6.2.3 Empirical Bayes estimation via surrogate-based optimization

One advantage of using MAP estimation over, say, maximum likelihood estimation, is the control that it provides for overfitting through the use of priors. In this chapter, I put a prior on the network weights w_n with a mean of zero and a width λ; such a prior enters $\chi^2_{\text{aug.}}(\Theta)$ as

$$\chi^2_{\text{ridge}}(\Theta) = \frac{1}{\lambda^2} \sum_n w_n^2, \tag{6.11}$$

which I refer to as a *ridge regression prior* in this chapter because it first appeared in the literature on ridge regression [201, 296, 354]. In the machine learning literature, adding terms like Eqn. 6.11 to the loss function of a neural network is referred to as *L2-regularization* or *weight decay* [60, 118, 268]. It is not immediately clear how λ should be set, so as to avoid overfitting. One criterion could simply be to tune λ such that $\chi^2_{\text{aug.}}/\text{d.o.f.} \sim 1$ (see Chapter 8.2); however, this could lead to bias. Another approach could be to set a prior on λ itself, to which λ would be estimated from a doubly-augmented χ^2 via MAP estimation. Such an approach is called *hierarchical Bayesian modelling* [268]. Of course, this only shifts the ambiguity in λ to an ambiguity in its prior. A nice automated procedure for setting λ that is an approximation to hierarchical Bayesian modelling is the *empirical Bayes* procedure [71, 239].

In Appendix D, I briefly defined the marginal likelihood (ML). In short, it is the normalization factor in Bayes' theorem that is proportional to the probability $\Pr\left(\mathcal{D}|\mathcal{M}\right)$ of the data $\mathcal{D}$ given some model $\mathcal{M}$ for the data. The empirical Bayes procedure follows by choosing any unknown priors by extremizing the marginal likelihood. What this means for the goals of this chapter is that we should calculate λ by optimizing the marginal likelihood in λ. In practice, the marginal likelihood can be estimated from a Laplace approximation of the posterior about the MAP estimate Θ^* as

$$-2\log \mathrm{ML} \approx \chi^2_{\mathrm{aug.}}(\Theta^*) + \log \frac{(2\pi)^{\mathrm{d.o.f.}} \det \Sigma_{\overline{X}} \det \Sigma_{\overline{\mathcal{C}}}}{\det \Sigma_{\Theta^*}}, \tag{6.12}$$

where $\Sigma_{\overline{X}}$ is the covariance (of the mean) of the data, $\Sigma_{\overline{\mathcal{C}}}$ is the covariance of the priors, Σ_{Θ^*} is the covariance of the MAP-estimated parameters Θ^*, and "d.o.f." is the number of degrees of freedom of the fit, all of which are defined in Chapter 8.2.

Each value of the approximate marginal likelihood is estimated from a fit at a particular λ. Therefore, extremizing the marginal likelihood (equivalently, minimizing $-\log \mathrm{ML}$) with a gradient-based local optimization algorithm could be prohibitively expensive. A nice way to get around this is via the following *surrogate-based* optimization procedure.

(1) **Calculate ML:** Calculate $\log \mathrm{ML}$ over a grid of λ.

(2) **Interpolate:** Interpolate the estimate for $\log \mathrm{ML}$ in λ from Step 1 with a spline. The monotonic interpolating spline that I've used many times throughout this thesis already will do just fine [342].

(3) **Optimize:** Optimize $-\log \mathrm{ML}$ in λ using a gradient-based local optimization that treats the spline as a *surrogate* for the actual marginal likelihood. The gradient can be calculated easily from the natural derivative of the spline.

The initial grid of $\log \mathrm{ML}$ estimates in λ need not be very fine; ten to twenty or even less often works

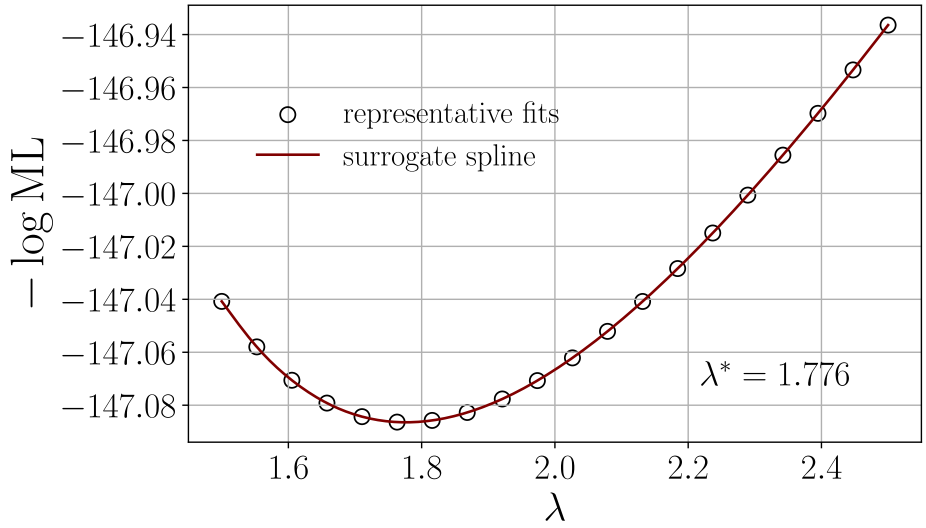

Figure 6.2: From Ref. [292]. Example of an interpolation over the marginal likelihood in Eqn. 6.12 with a cubic spline (red line). Each black circle represents the marginal likelihood calculated from a curve collapse fit of the 2-state Potts model Binder cumulant at a particular value of λ. The minimum of the surrogate spline λ^* is the value for λ suggested by the empirical Bayes procedure.

well. Most important for having the latter procedure work is to make sure that the initial estimates of $\log \mathrm{ML}$ are robust. In practice, this is achieved by calculating $\log \mathrm{ML}$ from one fit at a particular λ at the lower end of the grid, then calculating $\log \mathrm{ML}$ for all other λ by sequentially initializing each fit with the result of the previous fit. As long as the first fit in the sequence is robust, all other fits typically fall in line how they should. Note that, for less complicated fit functions, the surrogate-based empirical Bayes procedure of this section can be implemented in an embarrassingly parallel fashion, which could significantly reduce the amount of time that it takes to go through the entire empirical Bayes procedure. This method could also be improved upon if accuracy is a great concern by embedding it in a bisection algorithm for finding the root of $\mathrm{d}\log \mathrm{ML}/\mathrm{d}\lambda$.

In Fig. 6.2, I illustrate what the a spline interpolation of $-\log \mathrm{ML}$ in λ looks like for one of the example problems that we will encounter in Sec. 6.3. Specifically, the curve in Fig. 6.2 is from a fit of the RBFN-based scaling function for the Binder cumulant of the 2-state Potts (Ising) model. The black circles labelled "representative fits" are from a RBFN-based curve collapse fit at a particular λ. The red curve is the spline-based interpolation over the representative fits. The value

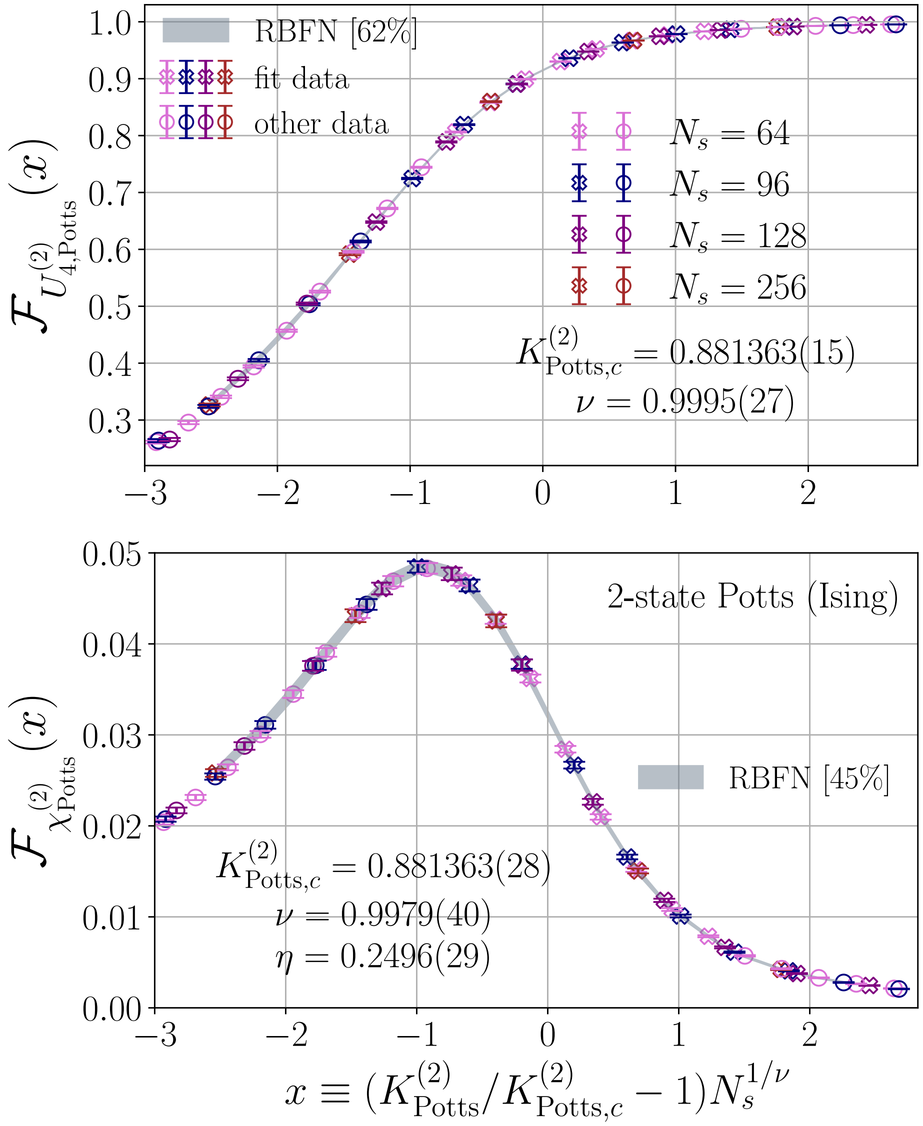

Figure 6.3: From Ref. [292]. BFN-based curve collapse analysis of the 2-state Potts (Ising) model using the Binder cumulant $U^{(2)}_{4,\text{Potts}}$ (top panel) and the magnetic susceptibility $\chi^{(2)}_{\text{Potts}}$ (bottom panel). The curve collapse uses $N_s = 64$ (pink), 96 (blue), 128 (purple), and 256 (red) volumes in the coupling range $0.87 \leq K^{(2)}_{\text{Potts}} \leq 0.90$. Data used in the curve collapse are marked with an open × (fit data); otherwise, they are marked with an open ∘ (other data). The scaling function $\mathcal{F}_O$ predicted by the RBFN is plotted as a grey band. The width of the band corresponds to the predicted error. The RBFN in the top panel has two nodes in its hidden layer and the RBFN in the bottom panel has three.

of the optimum $\lambda^* \approx 1.776$ from the surrogate spline is the prior width that is suggested by the empirical Bayes procedure.

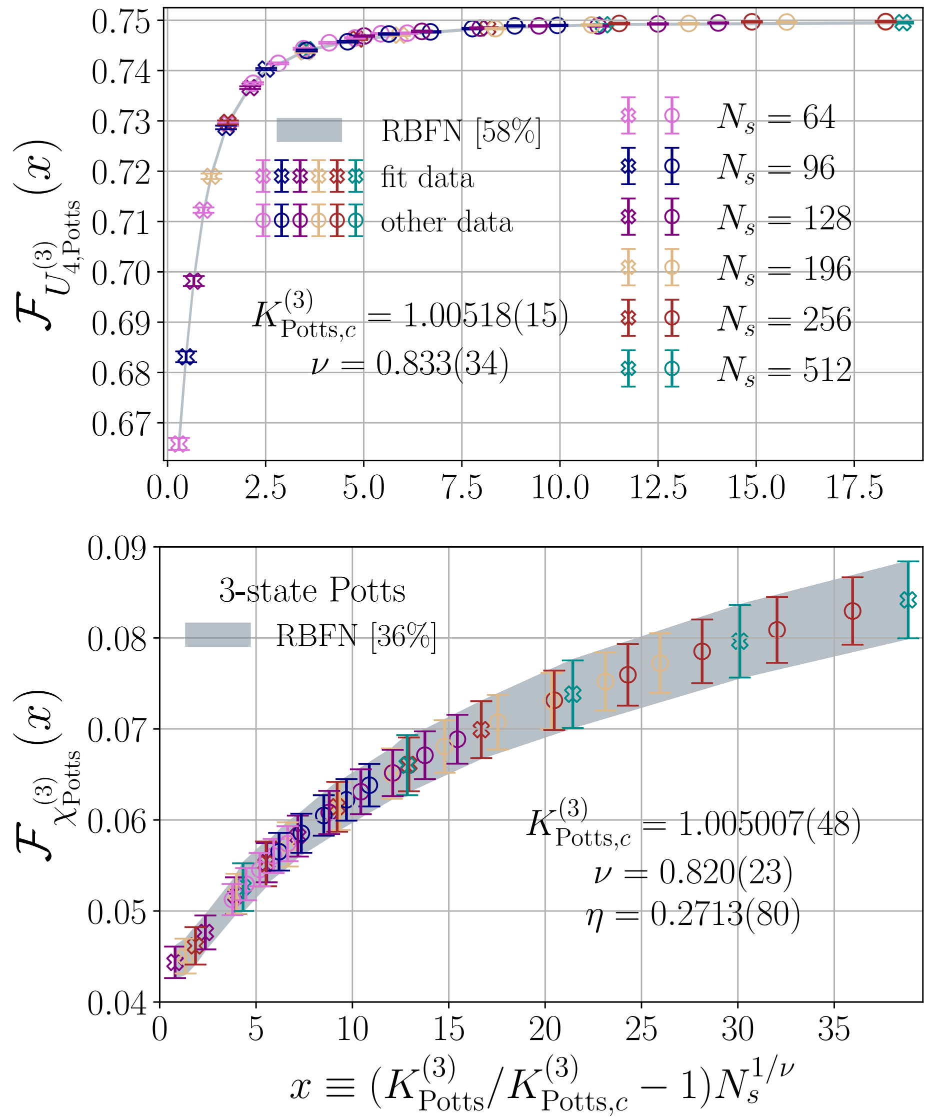

Figure 6.4: From Ref. [292]. RBFN-based curve collapse analysis of the 3-state Potts model using the Binder cumulant $U^{(3)}_{4,\mathrm{Potts}}$ (top panel) and order parameter susceptibility $\chi^{(3)}_{\mathrm{Potts}}$ (bottom panel). The curve collapse uses $N_\mathrm{s} = 64$ (pink), 96 (blue), 128 (purple), 196 (tan), 256 (red), and 512 (cyan) volumes for $U^{(3)}_{4,\mathrm{Potts}}$ and $N_\mathrm{s} = 128, 196, 256$, and 512 volumes for $\chi^{(3)}_{\mathrm{Potts}}$ (same scheme as $U^{(3)}_{4,\mathrm{Potts}}$). The $K^{(3)}_{\mathrm{Potts}}$ values used in both curve collapse analyses are in the range $1.005 \leq K^{(3)}_{\mathrm{Potts}} \leq 1.018$ for $U^{(3)}_{4,\mathrm{Potts}}$ and $1.005 \leq K^{(3)}_{\mathrm{Potts}} \leq 1.026$ for $\chi^{(3)}_{\mathrm{Potts}}$. Data used in the curve collapse are marked with an open × (fit data); otherwise, they are marked with an open ∘ (other data). The scaling function $\mathcal{F}_O$ predicted by the RBFN is plotted as a grey band. The width of the band corresponds to the predicted error. The RBFN in both panels has two nodes in its hidden layer.

	2-state Potts model			3-state Potts model		
Param.	$U_{4,\text{Potts}}^{(q)}$	$\chi_{\text{Potts}}^{(q)}$	Exact	$U_{4,\text{Potts}}^{(q)}$	$\chi_{\text{Potts}}^{(q)}$	Exact
$K_{\text{Potts}}^{(q)}$	0.881363(15)	0.881363(28)	$\log\left(1+\sqrt{2}\right)$	1.00518(15)	1.005007(48)	$\log\left(1+\sqrt{3}\right)$
ν	0.9995(27)	0.9979(40)	1	0.833(34)	0.820(23)	5/6
η	—	0.2496(29)	1/4	—	0.2713(80)	4/15

Table 6.1: From Ref. [292]. Comparison of our RBFN-based estimates of $K_{\text{Potts}}^{(q)}, \nu$ and η critical parameters for $q = 2, 3$ from a curve collapse analysis of $U_{4,\text{Potts}}^{(q)}$ and $\chi_{\text{Potts}}^{(q)}$. Exact critical parameters are from Ref. [382]. Predicted critical parameters from Figs. 6.3-6.4.

6.3 Application: The q-state Potts model for $q = 2, 3$

I first test the RBFN-based method that I introduced in Sec. 6.2 on the two-dimensional isotropic p-state Potts model, which is a generalization of the two-dimensional isotropic Ising model ($q = 2$) to $q \geq 2$ discrete spins. The reduced Hamiltonian for the q-state Potts model is

$$\mathcal{H}_{\text{Potts}}^{(q)} = -K_{\text{Potts}}^{(q)} \sum_{\langle ij \rangle} \delta(s_i, s_j), \tag{6.13}$$

where $s_i \in \{1, ..., q\}$ and the Kronecker delta $\delta(s_i, s_j) = 1$ when $s_i = s_j$ and $\delta(s_i, s_j) = 0$ otherwise. The notation $\langle ij \rangle$ denotes a sum over sites i and nearest-neighbors j. In this section, I consider both $q = 2$ (the Ising model) and $q = 3$. For all $q \geq 2$, the q-state model exhibits a phase transition at

$$K_{\text{Potts,c}}^{(q)} = \log\left(1 + \sqrt{q}\right) \tag{6.14}$$

that is 2nd-order for $q \leq 4$ and first-order for $q > 4$ [113, 114, 382].[1] The order parameter that distinguishes one phase from another is the magnetization

$$M_q\left(K_{\text{Potts}}^{(q)}, N_{\text{s}}\right) \equiv \frac{1}{N_{\text{s}}^2} \sum_i \delta(s_i, 1) - 1/q. \tag{6.15}$$

---

[1] Note that the critical coupling for $q = 2$ differs from the conventional Ising model coupling as $K_{\text{Potts,c}}^{(2)} = 2K_{\text{Ising,c}}$

The critical exponents for $q=2,3$ are known exactly [382], which makes the $q=2,3$ system a great test bed for the RBFN-based curve collapse of this chapter. I simulate both systems using the Wolff cluster algorithm implemented in the beautiful `Julia`-based `SpinMonteCarlo` library [267, 377]. See Appendix J for a brief description of cluster algorithms.

I test the RBFN-based curve collapse of this chapter on the 2- and 3-state Potts model by determining $K^{(q)}_{\text{Potts,c}}$, ν and η[2] from a curve collapse analysis of the Binder cumulant

$$U^{(q)}_{4,\text{Potts}}\big(K^{(q)}_{\text{Potts}}, N_{\text{s}}\big) = \frac{1}{2}\left[3 - \frac{\langle M_q^4\rangle}{\langle M_q^2\rangle^2}\right] \tag{6.16}$$

and the connected magnetic susceptibility

$$\chi^{(q)}_{\text{Potts}}\big(K^{(q)}_{\text{Potts}}, N_{\text{s}}\big) = K^{(q)}_{\text{Potts}} N_{\text{s}}^2 \big\langle (|M_q| - \langle |M_q|\rangle)^2 \big\rangle, \tag{6.17}$$

where $M_q\big(K^{(q)}_{\text{Potts}}, N_{\text{s}}\big)$ is defined in Eqn. 6.15. The Binder cumulant and connected magnetic susceptibility scale as Eqn. 6.4 for $x\big(K^{(q)}_{\text{Potts}}, N_{\text{s}}\big) = \big|k^{(q)}_{\text{Potts}}\big| N_{\text{s}}^{1/\nu}$ (2nd-order scaling) with $\gamma_{U^{(q)}_{4,\text{Potts}}} = 0$ and $\gamma_{\chi^{(q)}_{\text{Potts}}} = 2-\eta$, respectively [69, 303]. In Fig. 6.3, I illustrate the result of the curve collapse for the Binder cumulant (top panel) and connected magnetic susceptibility (bottom panel) for the 2-state Potts (Ising) model. I show the same information for the 3-state Potts model in Fig. 6.4. They grey band in both figures in the prediction for the scaling function $\mathcal{F}_{\mathcal{O}}$ for the Binder cumulant (top panels) and the connected magnetic susceptibility (bottom panels). The multicolored error bars illustrate the data that entering the curve collapse, with each color representing a fixed volume N_{s} (see the respective legends). In Table 6.1, I compare the prediction for the critical parameters from the RBFN-based curve collapse against their exact values from Ref. [382]. From the range of N_{s} and $K^{(q)}_{\text{Potts}}$ entering the curve collapse analysis of both models, the prediction for the critical parameters listed in Table 6.1 agree with their exact counterparts at the 1σ level and the p-values for all fits are in the 36%-62% range.

---

[2] η is the anomalous dimension of the wave function.

Perhaps the most interesting observation of the curve collapse analysis summarized in Figs. 6.3-6.4 is the exceedingly small number of nodes in the hidden layer; each RBFN possesses a mere 2-3 hidden nodes, depending on the observable. Nonetheless, the RBFN manages to accommodate both the S-shaped profile of three out of four of the observables, along with the peaked profile of the connected magnetic susceptibility of the 2-state Potts model. This is to be contrasted with the feedforward-neural-network-based curve collapse of Ref. [385], which deployed neural networks containing two internal layers with twenty nodes per layer. Even using the feedforward neural networks provided by `SwissFit` [290], I find that such a massive neural network is completely unnecessary for the purposes of curve collapse. Even worse, such a model has a negative number of degrees of freedom without the imposition of priors; as such, the χ^2 and p-value are meaningless. The approach advocated in this chapter does not run into such issues.

6.4 Application: The p-state clock model for $p = 4, \infty$

The two-dimensional isotropic p-state clock model is a discrete version of the two-dimensional isotropic XY model. The discrete spin variables are described by angles $\theta_i = 2\pi n_i/p$ that live at each lattice site i for $1 \leq n_i \leq p$. The reduced Hamiltonian for the p-state clock model is

$$\mathcal{H}^{(p)}_{\text{clock}} = -K^{(p)}_{\text{clock}} \sum_{\langle ij \rangle} \cos\left(\theta_i - \theta_j\right). \tag{6.18}$$

As was the case for the q-state Potts model, $p = 2$ is equivalent to the two-dimensional Ising model. In contrast, the $p \to \infty$ limit is equivalent to the famous two-dimensional XY model. It is generally accepted that the p-state clock model exhibits one or more phase transitions for all $p \geq 2$ and that there is some p for which one of the phase transitions switches from being 2nd-order or ∞-order.

In this chapter, I investigate the 4- and ∞-state clock model. Unknown to the author before I started present investigation of this system, the 4-state clock model is in the Ising universality class [117]; that is, it possesses *exactly* the same leading-order critical exponents as the 2-state Potts (Ising) model. As such, it is 2nd-order. Even more, it possesses exactly the same critical temperature

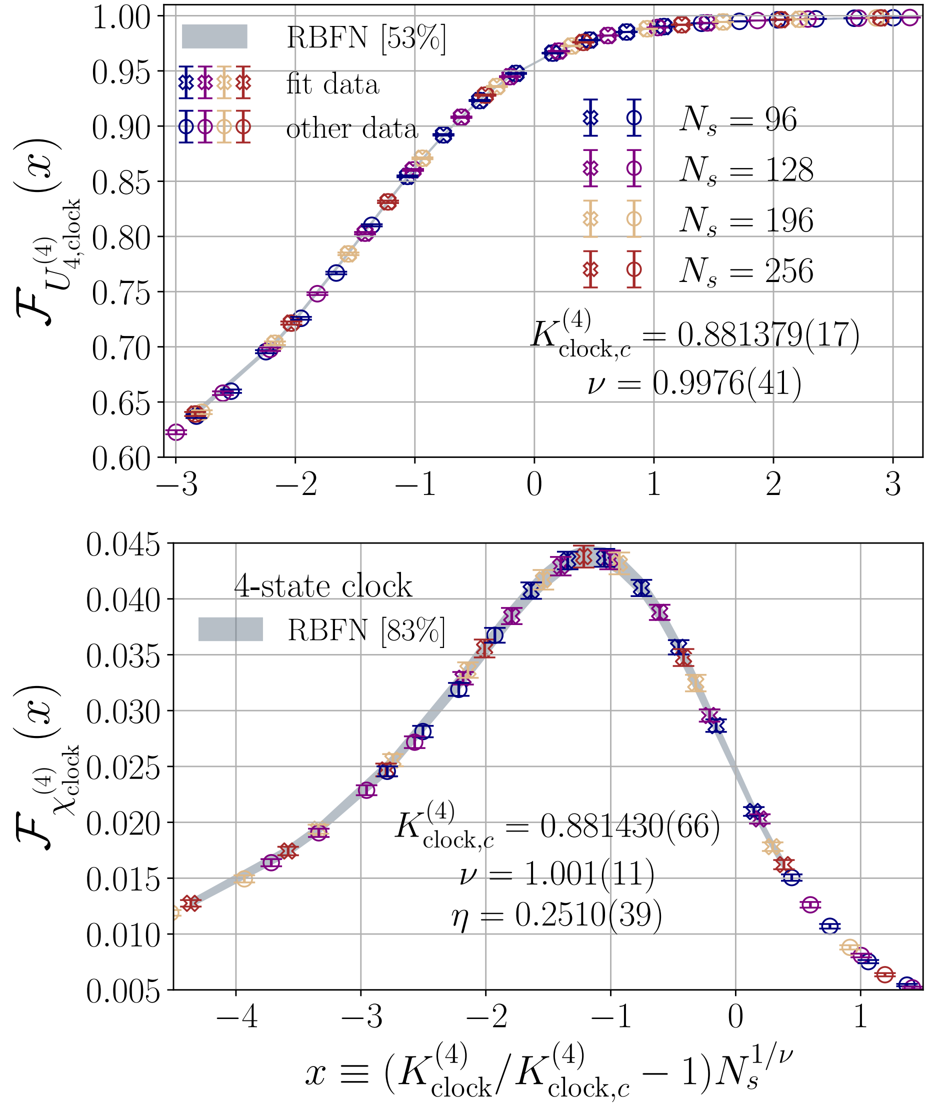

Figure 6.5: From Ref. [292]. RBFN-based curve collapse of the 4-state clock model using the Binder cumulant $U^{(4)}_{4,\text{clock}}$ (top panel) and connected magnetic susceptibility $\chi^{(4)}_{\text{clock}}$ (bottom panel). The curve collapse uses $N_s = 96$ (blue), 128 (purple), 196 (tan), 256 (red). The $K^{(4)}_{\text{clock}}$ values used in the curve collapse of $U^{(4)}_{4,\text{clock}}$ are in the range $0.870 \le K^{(4)}_{\text{clock}} \le 0.893$ and the $K^{(4)}_{\text{clock}}$ values used for $\chi^{(4)}_{\text{clock}}$ are in the range $0.870 \le K^{(4)}_{\text{clock}} \le 0.885$. Data used in the curve collapse is marked with an open × (fit data); otherwise, it is marked with an open ∘ (other data). The scaling function $\mathcal{F}_O$ predicted by the RBFN is plotted as a grey band. The width of the band corresponds to the predicted error. The RBFN in both panels has three nodes in its hidden layer.

[283]

$$K^{(4)}_{\text{clock,c}} = \log\left(1 + \sqrt{2}\right). \tag{6.19}$$

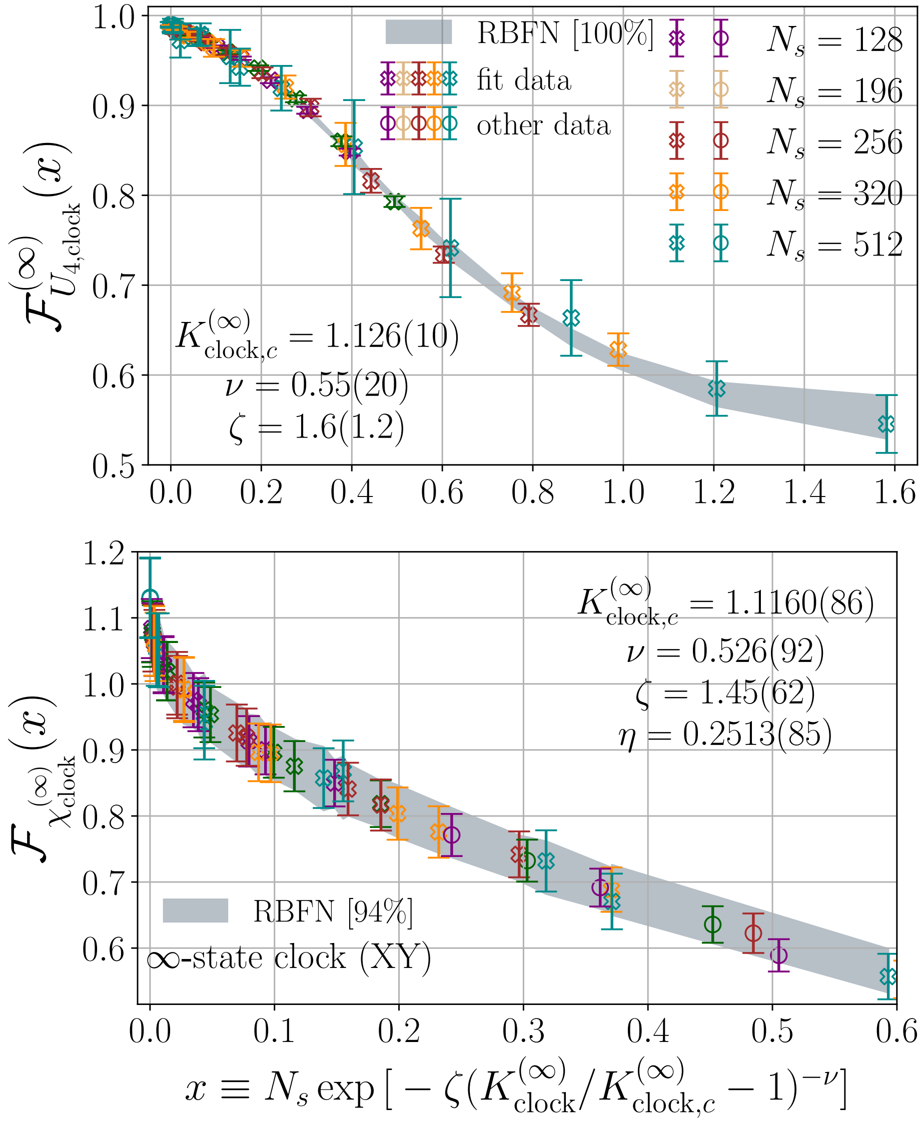

Figure 6.6: From Ref. [292]. RBFN-based curve collapse analysis of the ∞-state clock (XY) model using the Binder cumulant $U^{(\infty)}_{4,\text{clock}}$ (top panel) and connected magnetic susceptibility $\chi^{(\infty)}_{\text{clock}}$ (bottom panel). The curve collapse uses $N_s = 128$ (purple), 196 (tan), 256 (red), 320 (yellow), and 512 (cyan). The $K^{(\infty)}_{\text{clock}}$ values used for both curve collapse analyses between 1.005/1.0 for $U^{(\infty)}_{4,\text{clock}}/\chi^{(\infty)}_{\text{clock}}$ and 1.1, 1.1, 1.102, 1.102, and 1.105 for $N_s = 128, 320, 256, 160$ and 512, respectively for both observables. Data used in the curve collapse is marked with an open $\times$ (fit data); otherwise, it is marked with an open $\circ$ (other data). The scaling function $\mathcal{F}_O$ predicted by the RBFN is plotted as a grey band. The width of the band corresponds to the predicted error. The RBFN in both panels has two nodes in its hidden layer.

The phase transition of the ∞-state clock (XY) model is much more interesting. Because it is a two-dimensional system with *continuous* spins, it cannot exhibit a phase transition in the "conventional"

	4-state clock model			∞-state clock (XY) model		
Param.	$U^{(p)}_{4,\text{clock}}$	$\chi^{(p)}_{\text{clock}}$	Exact	$U^{(p)}_{4,\text{clock}}$	$\chi^{(p)}_{\text{clock}}$	Literature/Exact
$K^{(p)}_{\text{clock}}$	0.881379(17)	0.881430(66)	$\log\left(1+\sqrt{2}\right)$	1.126(10)	1.1160(86)	1.1199...
ζ	—	—	—	1.6(1.2)	1.45(62)	1.5...
ν	0.9976(41)	1.001(11)	1	0.55(20)	0.526(92)	1/2
η	—	0.2510(39)	1/4	—	0.2513(85)	1/4

Table 6.2: From Ref. [292]. Comparison of our RBFN-based estimates of $K^{(p)}_{\text{clock}}, \nu$, η, and ζ critical parameters for $p = 4, \infty$ from a curve collapse analysis of $U^{(p)}_{4,\text{clock}}$ and $\chi^{(p)}_{\text{clock}}$. The exact critical parameters are from Refs. [231, 382]. Values from the literature for $K^{(\infty)}_{\text{clock,c}}$ and ζ are from Refs. [166, 229, 232, 277]. Predicted critical parameters from Figs. 6.5-6.6.

sense (a byproduct of spontaneous symmetry breaking) due to the Mermin-Wagner theorem [260]. The phase transition of the XY model is not such a phase transition. The transition is triggered by a sudden transition from a phase that is dominated by spin-wave dynamics to a phase that is dominated by interacting vortices [222, 231]; in other words, the phase transition is induced by topology and not the spontaneous breakdown of some symmetry. This is the Berezinsky-Kosterlitz-Thouless (BKT) transition [37, 231, 232]. Even more, the transition is ∞-order [231]. The transition temperature is not known exactly; however, a multitude of studies spanning many decades in the literature have pinned it down to

$$1/K^{(\infty)}_{\text{clock,c}} \approx 0.893 \tag{6.20}$$

with a global uncertainty in $1/K^{(\infty)}_{\text{clock,c}}$ that is roughly O(0.001) [166, 229, 277]. This amounts to $K^{(\infty)}_{\text{clock,c}} \approx 1.12$. RG analysis of the ∞-state clock model yields $\nu = 1/2$ and $\zeta \approx 1.5$ [231].

I simulate the 4-state clock model using the cluster algorithm provided by the `SpinMonteCarlo` library [267, 377]. On the other hand, I simulate the ∞-state clock model using the heatbath algorithm provided by the `Quantum EXpressions` (`QEX`) library [284]. The author's reason for using the heatbath algorithm is historical. One the original goals of this project was to use the GF-based renormalized coupling for FSS, as I do in Chapter 7. At the time, I wanted to utilize the tools provided by `QEX` library to accomplish this goal. Conveniently, `QEX` also had an implementation of the heatbath algorithm that allowed me to simultaneously generate heatbath samples and run gradient flow measurements. As little came out of my investigation of the renormalized coupling, I regret not simulating the ∞-state clock model with the cluster algorithm, which is superior to the heatbath algorithm for spin systems near criticality. In any case, I briefly discuss the heatbath algorithm in Appendix J.

As in Sec. 6.3, I test out the RBFN-based curve collapse on the 4- and ∞-state clock model by considering the Binder cumulant and magnetic susceptibility. To this end, I define the magnetization vector

$$\mathbf{M}\big(K^{(p)}_{\text{clock}}, N_{\text{s}}\big) \equiv \frac{1}{N_{\text{s}}^2}\sum_i \big(\cos(\theta_i), \sin(\theta_i)\big) \tag{6.21}$$

in place of the magnetization defined in Eqn. 6.15 for the Potts model. The Binder cumulant is calculated from the magnitude of the magnitization as

$$U^{(p)}_{4,\text{clock}}\big(K^{(p)}_{\text{clock}}, N_{\text{s}}\big) = \frac{1}{2}\left[3 - \frac{\langle|\mathbf{M}|^4\rangle}{\langle|\mathbf{M}|^2\rangle^2}\right]. \tag{6.22}$$

For the magnetic susceptibility of the 4-state clock model, I calculate the connected magnetic susceptibility as

$$\chi^{(4)}_{\text{clock}}\big(K^{(4)}_{\text{clock}}, N_{\text{s}}\big) = K^{(4)}_{\text{clock}} N_{\text{s}}^2 \big\langle(|\mathbf{M}| - \langle|\mathbf{M}|\rangle)^2\big\rangle, \tag{6.23}$$

while for the magnetic susceptibility of the ∞-state clock model, I use the estimator

$$\chi^{(\infty)}_{\text{clock}}\big(K^{(\infty)}_{\text{clock}}, N_{\text{s}}\big) = K^{(\infty)}_{\text{clock}} N_{\text{s}}^2 \Big\langle\big|\mathbf{M}\big(K^{(\infty)}_{\text{clock}}, N_{\text{s}}\big)\big|^2\Big\rangle \tag{6.24}$$

suggested by Refs. [158, 286]. The Binder cumulant and connected magnetic susceptibility scale as Eqn. 6.4 with $x\big(K^{(4)}_{\text{clock}}, N_{\text{s}}\big) = \big|k^{(4)}_{\text{clock}}\big| N_{\text{s}}^{1/\nu}$ for the 4-state model and $x\big(K^{(\infty)}_{\text{clock}}, N_{\text{s}}\big) = N_{\text{s}} \exp\big(\text{-}\zeta\big|k^{(\infty)}_{\text{clock}}\big|^{-\nu}\big)$ for the ∞-state clock model. Moreover, for both models $\gamma_{\chi^{(4)}_{\text{clock}}} = \gamma_{\chi^{(\infty)}_{\text{clock}}} = 2 - \eta$ with $\eta = 1/4$ [232, 382]. In Fig. 6.5, I illustrate the result of the curve collapse for the Binder cumulant (top panel) and connected magnetic susceptibility (bottom panel) for the 4-state clock model. I show the same information for the ∞-state clock (XY) model in Fig. 6.6. In Table 6.2, I compare the prediction for the critical parameters from the RBFN-based curve collapse against what is known from the literature [166, 229, 232, 277].

The RBFN-based curve collapse for the p-state clock model yields predictions for the critical parameters of both models that are consistent with either their exact values or their estimates from the literature over the range of N_s and $K^{(p)}_\mathrm{clock}$ used in each analysis. The p-values for the 4-state clock model curve collapse are 54% and 83% for the Binder cumulant and connected magnetic susceptibility, respectively. For the ∞-state clock model, they are 100% and 94%. Such high p-values indicate either that the model overfits or the errors on the data are overestimated. Given that the size of the error bars appear to be much larger than the fluctuations in the central values of the data, it is much more likely that the errors have been overestimated. As with the q-state Potts model, the RBFN has only 2-3 hidden nodes. The RBFN's ability to accommodate for the differing curvatures of the scaling function while producing estimates for the critical parameters that agree with the literature with such a small number of hidden nodes is quite impressive.

6.5 Other uses of radial basis function networks in lattice field theory

Though I have introduced radial basis function networks in this chapter for the purpose of extracting the critical parameters of a model, there are a variety of problems in lattice field theory and the broader scientific domain that could exploit the expressivity and simplicity of RBFNs. One such class of problems are those which require estimating the non-parametric component of a semi-parametric model. The curve collapse analysis that we have explored in this chapter belongs to the latter category: to estimate the critical parameters of the model, it is necessary to estimate the non-parametric and a priori unknown scaling function $\mathcal{F}_O$. Another familiar example appears in hadron spectroscopy. Spectroscopy calculations often involve estimating the ground state energy $\tilde{E}_0$ and amplitude $\tilde{\mathcal{A}}_0$ (in units of the lattice spacing), along with any excited state energies/amplitudes,

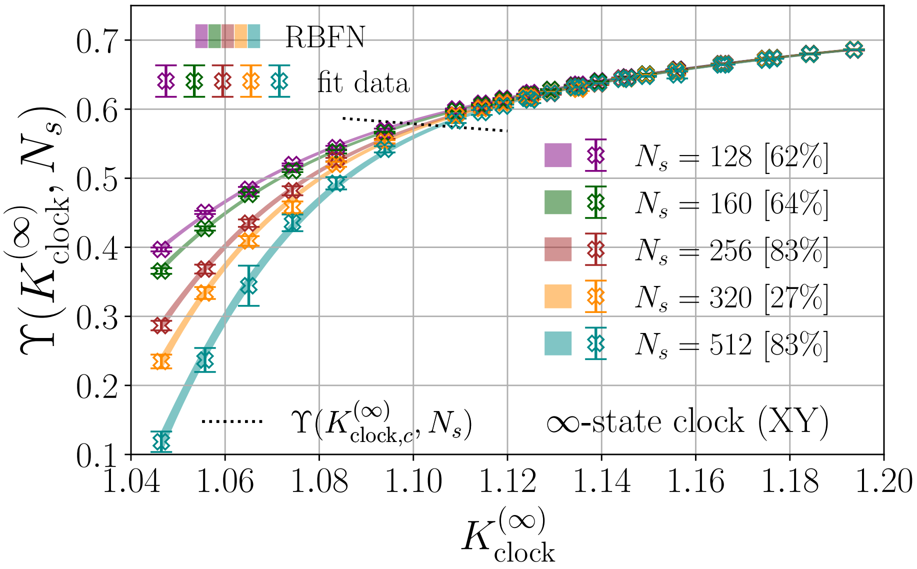

Figure 6.7: From Ref. [292]. RBFN-based interpolation of the helicity modulus $\Upsilon(K^{(\infty)}_{\text{clock}}, N_s)$ for the ∞-state clock (XY) model at fixed N_s. Data included in fit is shown as an errorbar with an open "×" marker. RBFN-based interpolation is shown as a colored band. Interpolation performed on $N_s = 128$ (purple), 160 (dark green), 256 (red), 320 (yellow), and 512 (cyan). The RBFN-based fits are shown as a colored bands, with the width of the band indicating the error. The color of each band indicates the N_s at which the fit was performed. The helicity modulus at $K^{(\infty)}_{\text{clock},c}(N_s)$ given by Eqn. 6.29 is indicated by a dotted black line. The RBFN has 2 nodes in its hidden layer.

from a fit of two-point function data to the ansatz

$$G(\tilde{x}_4) = \tilde{\mathcal{A}}_0 \exp\left(\text{-}\tilde{E}_0\tilde{x}_4\right) + \sum_{i=1}^{\infty} \tilde{\mathcal{A}}_i \exp\left(\text{-}\tilde{E}_i\tilde{x}_4\right), \tag{6.25}$$

where $\tilde{x}_4$ is the temporal extent of the lattice in units of the lattice spacing. The excited state sum is non-parametric in the sense that requires knowledge of an infinite number of $\tilde{\mathcal{A}}_i, \tilde{E}_i$. Estimating $\tilde{E}_0$ and $\tilde{\mathcal{A}}_0$ requires truncating the excited state sum, and the stability of the estimate is assessed by the observation of a stable plateau of the ground state parameters in the number of states contribution to the sum. The need to choose a cutoff in the number of excited states can introduce an additional source of systematic uncertainty. Parameterizing the excited state sum with a radial basis function network could improve the estimate of at least the ground state energy/amplitude by removing this ambiguity; however, one of course still needs to choose how many parameters the RBFN should

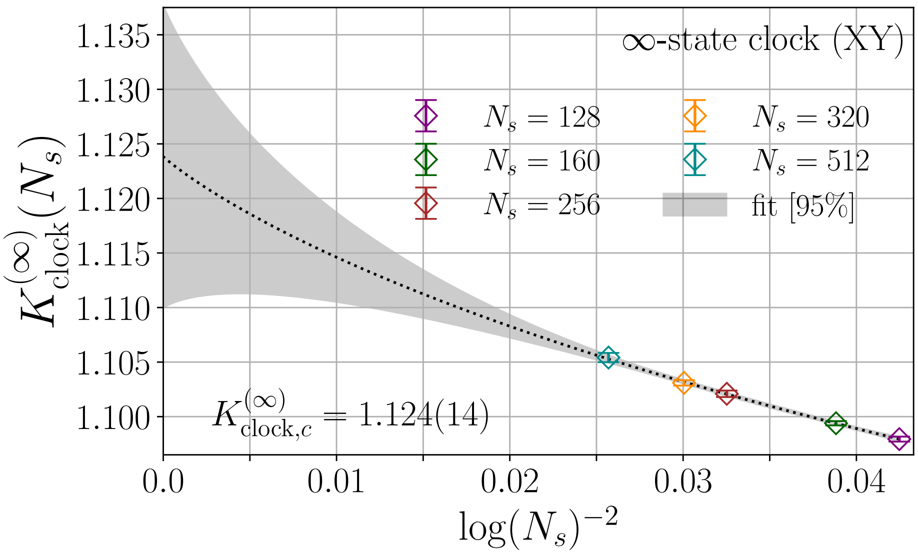

Figure 6.8: From Ref. [292]. Extrapolation of the pseudocritical temperature $K^{(\infty)}_{\mathrm{clock},c}(N_s)$ calculated from the intersection of our RBFN-based interpolation (colored bands in Fig. 6.7) with the universal jump condition (dotted line in Fig. 6.7) to $1/N_s \to 0$ using Eqn. 6.30. The pseudocritical temperatures are indicated by multi-colored errorbars with open diamond markers "$\diamond$" and utilize the same color scheme as Fig. 6.7 for different N_s (see caption). Result of fit to Eqn. 6.30 is shown as a grey band and the central value of the fit prediction is shown as a dotted black line.

posses.

A more trivial class of problems are those that require directly interpolating some observable. We have already encountered these kinds of problems many times in this thesis; see, for example, Chapters 4.5.3.1 and 5.4.2.1. Let's now end this chapter with an example calculation that utilizes an RBFN as a tool for direct interpolation. From the RBFN-based interpolation, I will extract the critical temperature of the ∞-state clock model.

6.5.1 Example: The critical temperature of the ∞-state clock model

My goal in this section is to estimate the critical temperature $K^{(\infty)}_{\mathrm{clock,c}}$ of the BKT transition from the helicity modulus, defined as

$$\Upsilon\left(K^{(\infty)}_{\mathrm{clock}}, N_{\mathrm{s}}\right) = \frac{1}{2}\sum_{\mu=1}^{2}\left\langle e_\mu - N_{\mathrm{s}}^2 K^{(p)}_{\mathrm{clock}} s_\mu^2\right\rangle, \tag{6.26}$$

with

$$e_\mu = \frac{1}{N_{\mathrm{s}}^2} \sum_{\langle ij \rangle_\mu} \cos(\theta_i - \theta_j), \tag{6.27}$$

$$s_\mu = \frac{1}{N_{\mathrm{s}}^2} \sum_{\langle ij \rangle_\mu} \sin(\theta_i - \theta_j), \tag{6.28}$$

where $\langle ij \rangle_\mu$ denotes a sum of lattice sites i along the μ-direction and their nearest-neighbors j [229, 277, 358, 359]. Next, I interpolate $\Upsilon\big(K^{(p)}_{\mathrm{clock}}, N_{\mathrm{s}}\big)$ in $K^{(\infty)}_{\mathrm{clock}}$ at fixed N_{s} with an RBFN possessing two hidden nodes. In Fig. 6.7, I show the result of the interpolation at $N_{\mathrm{s}} = 128, 160, 256, 320,$ and 512 as multicolored bands. The p-values for the fits are in the $27\% - 83\%$ range. At $K^{(\infty)}_{\mathrm{clock,c}}$, there is a universal jump condition

$$\Upsilon\big(K^{(\infty)}_{\mathrm{clock},c}, N_{\mathrm{s}}\big) = 2f_r/\pi K^{(\infty)}_{\mathrm{clock,c}}(N_{\mathrm{s}}), \tag{6.29}$$

where $f_r = 1 - 16\pi \exp(\text{-}4\pi)$ [277]. The jump condition is shown as a dotted black line in Fig. 6.7. I calculate the pseudocritical temperature $K^{(\infty)}_{\mathrm{clock,c}}(N_{\mathrm{s}})$ from the intersection of $\Upsilon\big(K^{(p)}_{\mathrm{clock,c}}, N_{\mathrm{s}}\big)$ intersects the right hand side of Eqn. 6.29. At leading order, the pseudocritical temperature $K^{(\infty)}_{\mathrm{clock,c}}(N_{\mathrm{s}})$ depends upon N_{s} as

$$K^{(\infty)}_{\mathrm{clock,c}}(N_{\mathrm{s}}) = K^{(\infty)}_{\mathrm{clock,c}} + \zeta^{-1/\nu} \log\big(\kappa N_{\mathrm{s}}\big)^{-1/\nu}. \tag{6.30}$$

I then extract $K^{(\infty)}_{\mathrm{clock,c}}$ from Eqn. 6.30 using the pseudocritical coupling $K^{(\infty)}_{\mathrm{clock,c}}(N_{\mathrm{s}})$ extracted from the universal jump condition. The data is not precise enough to determine both $K^{(\infty)}_{\mathrm{clock,c}}$ and ν, so I fix $\nu = 1/2$. I illustrate the extrapolation in Fig. 6.8. The estimate $K^{(\infty)}_{\mathrm{clock,c}} = 1.124(14)$ is consistent with the result that I obtained from the RBFN-based curve collapse and the literature (see Table 6.2). It is clear from Fig. 6.8 that the error in $K^{(\infty)}_{\mathrm{clock,c}}(N_{\mathrm{s}})$ fans out as $1/N_{\mathrm{s}} \to 0$ because the data entering the extrapolation are situated far away from the infinite volume limit. The large distance over which the extrapolation must be performed is a consequence of the logarithmic scaling of certain observables with N_{s} for ∞-order transitions. Unfortunately, one must simulate large volumes.

Chapter 7

Finite size scaling and β-function of the massless eight flavor system

Up to this point, we have explored a variety of renormalization group methods and physical systems. My goal in this chapter is to apply the methods that we have developed and subsequently deployed to study the infrared properties of the massless SU(3) gauge-fermion system with $N_{\mathrm{f}} = 8$ fermions in the fundamental representation of SU(3). The $N_{\mathrm{f}} = 8$ system is the least well-understood and, consequentially, the most controversial/challenging system that I explore in this thesis. As such, the content of this chapter reflects an ongoing effort to understand its infrared behavior. In Sec. 7.1, I review the literature on $N_{\mathrm{f}} = 8$. I summarize the simulations that have been performed for the content of this chapter in Sec. 7.2. The rest of this chapter is dedicated to understanding the zero-temperature phase structure of the $N_{\mathrm{f}} = 8$ using the radial basis function network-based finite size scaling method described/deployed in Chapter 6 (Sec. 7.3) and the renormalization group β-function using the methods described/deployed in Chapters 4 and 5 (Sec. 7.4).

7.1 Overview of the eight flavor system

Much of the literature on $N_{\mathrm{f}} = 8$ system suggests that it could be close to the edge of the conformal window. Early investigations of this system were partially motivated by the presence of an IRFP in the perturbative two-loop β-function (see Chapter 1), which predicts an IRFP for $8.05 \lesssim N_{\mathrm{f}} \lesssim 16.5$ [346]. However, the two-loop fixed point at $N_{\mathrm{f}} \approx 8.05$ is observed at such an absurdly large renormalized coupling that it cannot be described by perturbation theory. The most recent non-perturbative determinations of the β-function from lattice simulations of the

$N_\mathrm{f} = 8$ system have not observed an IRFP over the range of renormalized couplings that have been investigated thusfar [18, 19, 124, 185, 187]. As the value of the fixed point coupling $g^2_{X\star}$ in any RG scheme X increases as $N_\mathrm{f} \to N_\mathrm{f}^\mathrm{c}$ from the conformal phase, it is possible that such studies simply could not reach large enough g^2_X to see signs of an IRFP. Interestingly enough, the few studies that have managed to reach renormalized couplings in the range $8.0 \lesssim g^2_c \lesssim 16.0$ using gradient-flow-based step-scaling schemes (see Chapter 4.3) do observe signs of an inflection point in the continuum β-function [185, 187]; however, the presence of a bulk phase transition (see Chapter 5.2) in such studies prevented them from reaching further into the infrared regime. Note additionally that an observation of the β-function turning around could indicate either that the β-function possesses an IRFP or that there is some range of renormalized couplings over which the renormalized coupling is *slowly walking*; without a conclusive observation of either behavior, it is impossible to tell from the turnaround of the β-function alone. It is also possible that the β-function turning around, or at least exhibiting an inflection point, is due to some scheme-dependent quirk, unlike the existence of an infrared fixed point.

As investigations of the non-perturbative RG β-function have thus far been largely inconclusive, much of what is known about the $N_\mathrm{f} = 8$ system from the literature has been obtained by other means. Finite-temperature investigations tend to suggest that the system is chirally broken; however, such investigations have largely been unable to conclusively establish the existence of chiral symmetry breaking due to the presence of bulk first-order phase transitions that prevent a reliable extrapolation to the chiral limit [15, 104, 262]. The Monte Carlo RG calculation of Ref. [168] indicates that the $N_\mathrm{f} = 8$ system is QCD-like, while the Dirac eigenspectrum of Ref. [79] suggests that the mass anomalous dimension is O(1), which could indicate either the presence of an IRFP or walking. Of the various observables/methods that have been deployed to understand the $N_\mathrm{f} = 8$ system, investigations of the hadron spectrum have arguably revealed the most useful/interesting results.

Large-scale studies of the $N_\mathrm{f} = 8$ hadron spectrum tend to conclude that it is very close to the confined/chirally-broken side of the conformal sill [7, 8, 10, 13, 15, 17, 53, 127]. The Lattice Strong Dynamics (LSD) and Lattice Kobayashi–Maskawa Institute (LatKMI) collaborations are the big dogs on this front. Both have observed that the lightest scalar meson appears to be nearly degenerate with the lightest pseudoscalar meson, possibly even all of the way down to the chiral limit. This is to be contrasted with QCD, where the mass of the broad σ resonance (≈ 500 MeV) is certainly not degenerate with the mass of the neutral pion (≈ 135 MeV). Moreover, many of the qualitative features of the hadron spectrum appear to be described well by dilaton effective chiral perturbation theory (dχPT) [14, 20, 149–151, 208], such as possible approximate conformal hyperscaling away from the chiral limit [21, 332]. See also Refs. [101, 390] for an emerging alternative to dχPT.

While a sizeable portion of the high energy physics community has been arguing over low-energy behavior of the $N_\mathrm{f} = 8$ system, the condensed matter community has been working on a separate development that could be relevant to the $N_\mathrm{f} = 8$ system and BSM model building as a whole. Roughly over the last two decades, a new mechanism for generating fermion masses without spontaneous symmetry breaking has been discovered and subsequently understood in terms of anomaly cancellation. Such *symmetric mass generation* (SMG) is possible *if* the system of fermions is *strictly* free of quantum anomalies [367]. It is becoming increasingly clear that SMG is intimately related to the physics of chiral edge modes, which underlies the domain wall discretization of the Dirac operator that I briefly described in Chapter 3.2.2; see, for example, Refs. [366, 386]. In fact, SMG has even been proposed as a machanism for evading the Nielsen–Ninomiya theorem (see Chapter 3.2) altogether [355]. As far as the $N_\mathrm{f} = 8$ system is concerned, it is now understood that 4-dimensional systems with $N_\mathrm{f} = 8$ continuum Dirac fermions, or two species of Khäler-Dirac fermion,[1] are free of anomalies; hence, they are capable of exhibiting SMG [62, 76]. Such anomaly cancellation can also be understood by identifying 4-dimensional Kähler-Dirac fermions with the boundary of a 5-dimensional bosonic symmetry-protected topological phase [156]. This opens up the fascinating possibility that the $N_\mathrm{f} = 8$ system could possess an SMG phase that is triggered by strong dynamics.

Evidence for the existence of a strongly-coupled SMG phase in the $N_\mathrm{f} = 8$ system has been presented in Ref. [170], though it had been known for some time already that lattice simulations of

---

[1] Lattice-discretized Khäler-Dirac fermions are equivalent to free staggered fermions [35].

	L/a (weak coupling)						
β_b	16	20	24	30	32	36	40
8.80	252	238	211	250	115	129	151
8.85	263	219	196	246	185	142	150
8.90	211	212	180	201	121	281	151
9.00	286	203	230	242	248	445	158
9.10	182	169	197	213	213	177	–
9.20	306	187	215	275	160	185	151
9.30	249	176	182	205	182	164	–
9.40	171	178	250	207	183	175	150
9.60	247	141	281	210	206	132	151
9.90	–	–	241	236	229	168	151
10.4	–	–	252	172	197	127	–
11.0	–	–	181	202	232	176	–

Table 7.1: The number of thermalized configurations analyzed at each bare coupling β_b and volume L/a. The configurations are separated by 10 MDTUs.

the $N_{\mathrm{f}} = 8$ system possessed a very strange strong coupling phase [80]. Staggered simulations refer to the latter phase as an "$\not{S}_4$ phase" because it is identified in staggered simulations from a spontaneous breakdown of single-site shift symmetry. The system in the $\not{S}_4$ phase exhibits confinement, but *not* chiral symmetry breaking. In the confined/chirally-broken phase of a QCD-like system (that is, below the conformal window), it is impossible to simulate at zero bare fermion mass am_f because the Dirac operator develops a zero mode; however, in the $\not{S}_4$ phase of the $N_{\mathrm{f}} = 8$ system, nothing prevents simulating at $am_f = 0$ because chiral symmetry is preserved. Some other mechanism must be responsible for generating confinement with $am_f = 0$ in this phase, and a plausible candidate could be SMG. The $\not{S}_4$ phase was originally identified as a bulk phase (see Chapter 5.2), as the transition into the $\not{S}_4$ phase appeared to be first-order and had been observed in both the $N_{\mathrm{f}} = 8$ and $N_{\mathrm{f}} = 12$ systems. However, with the advent of Pauli-Villars improvement, which I have already discussed in Chapter 5.2, it was soon realized that the transition into the $\not{S}_4$ phase in the $N_{\mathrm{f}} = 8$

system *could* be continuous, at least from the perspective of a gradient-flow-based finite size scaling (FSS) analysis of the transition [170]. Such extraordinary claims require extraordinary evidence, so I will scrutinize the results of Ref. [170] with an updated dataset using the RBFN-based curve collapse of Chapter 6 (Sec. 7.3). This will be followed up to a non-perturbative determination of the β-function using the method described in Chapter 4 and subsequently deployed in the $N_{\mathrm{f}} = 0$ and $N_{\mathrm{f}} = 12$ systems in Chapters 4 and 5, respectively (Sec. 7.4).

7.2 Simulation details

The simulations in this chapter utilize an adjoint-plaquette gauge action (see Appendix F), a massless ($am_f = 0$) nHYP-smeared staggered fermion action, and an nHYP-smeared Pauli-Villars (PV) action with $N_{\mathrm{PV}} = 64$ staggered PV fields (8 PV fields per fermion field) with mass $am_{\mathrm{PV}} = 0.75$. Both the PV and fermion files have completely anti-periodic boundary conditions in all four directions. The same action was deployed in the finite size scaling study of Ref. [170]. All simulations are performed either using a modified version of the `MILC` library (`KS_nHYP_FS`)[2] or `qex_staghmc` (see Sec. 8.1 for details). The same is true for all gauge flow measurements. Details of how gauge flow measurements are performed are discussed in Sec. 8.1 and Chapter 8.1.2.

The datasets in this chapter are split into two categories: *weak coupling* and *strong coupling*; see Tables 7.1 and 7.2 for the total number of thermalized configurations at each $(\beta_b, L/a)$ pair making up both, respectively. The weak coupling ensembles (Table 7.1) are used to determine the continuum β-function from the weak coupling phase (Sec. 7.4); though the ensembles for this dataset have been generated from a mix of `KS_nHYP_FS` and `qex_staghmc`, all gauge flow measurements utilize `qex_staghmc`. Many of the strong coupling ensembles with $16 \leq L/a \leq 24$ were utilized in Ref. [170], though a number of $L/a = 24$ volumes are either new or have been updated and the $L/a = 32$ volumes are completely new. The statistical errors in the strong coupling dataset are likely underestimated due to not properly accounting for the autocorrelation time. As such, the statistical errors have been increased by a factor of two on all strong coupling ensembles. A more

---

[2] `KS_nHYP_FS` can be found at https://github.com/daschaich/KS_nHYP_FA

	L/a (strong coupling)						
β_b	16	20	24	30	32	36	40
8.470	168	–	–	–	–	–	–
8.500	129	–	–	–	–	–	–
8.540	–	111	–	–	–	–	–
8.525	257	–	–	–	–	–	–
8.550	221	–	–	–	–	–	–
8.560	–	145	–	–	–	–	–
8.575	171	–	61	–	–	–	–
8.580	–	156	–	–	125	–	–
8.590	238	139	–	–	100	–	–
8.600	488	131	90	–	80	–	–
8.610	457	293	164	–	–	–	–
8.620	–	–	–	–	145	–	–
8.625	717	388	107	–	–	–	–
8.630	–	–	172	–	–	–	–
8.640	–	190	142	–	121	–	–
8.650	410	321	117	–	–	–	–
8.660	–	–	114	–	114	–	–
8.670	–	–	–	–	107	–	–
8.680	–	–	–	–	60	–	–
8.700	132	277	85	–	–	–	–
8.750	291	216	117	–	38	–	–

Table 7.2: The number of thermalized configurations analyzed at each bare coupling β_b and volume L/a. The configurations are separated by 10 MDTUs.

careful scrutiny of the statistical errors is underway.

In Fig. 7.1, I illustrate the dependence of

$$g_c^2(L, g_0^2) \equiv g_{\mathrm{GF}}^2(t; L, g_0^2)\Big|_{t=(cL)^2/8}$$

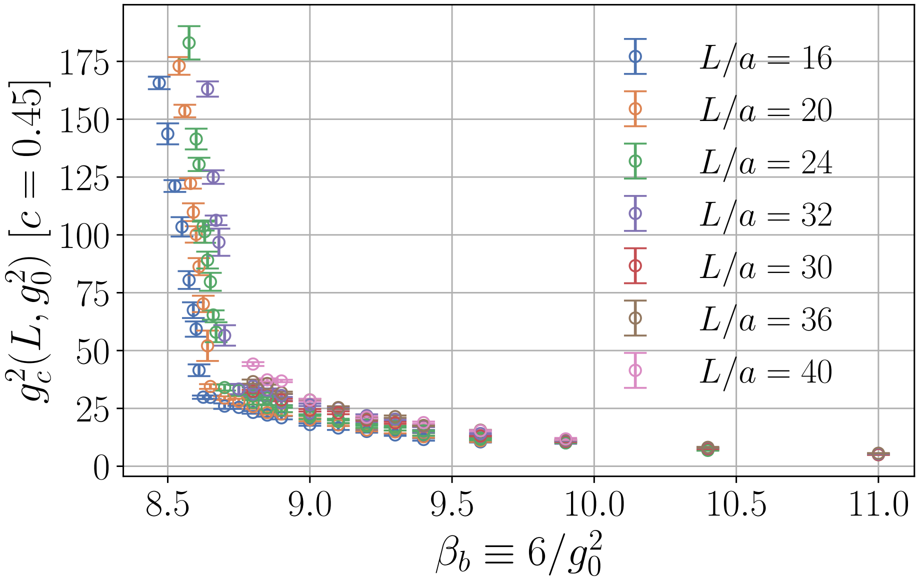

Figure 7.1: The gradient flow coupling $g^2_{\mathrm{GF}}(t; L, g_0^2)$ at $8t/a^2 = (cL/a)^2$ ($c = 0.45$) for each $\beta_b \equiv 6/g_0^2$ at fixed $16 \leq L/a \leq 40$. The coupling on each volume is indicated by multicolored error bars. See legend for color coding.

at $c = 0.45$ on β_b at fixed $16 \leq L/a \leq 40$ in both the weak and strong coupling phases. The gradient flow coupling $g^2_{\mathrm{GF}}(t; L, g_0^2)$ is defined in Eqn. 4.24 with $\delta(t, L)$ set by Eqn. 4.16. I already defined $g_c^2(L, g_0^2)$ in Chapter 4.3, where it was introduced in the context of gradient-flow-based determinations of the RG β-function that utilize step-scaling. The coupling $g_c^2(L, g_0^2)$ in Fig. 7.1 is determined from Wilson flow ($\mathcal{S}^{\mathrm{f}} = \mathrm{W}$) and the operator ($\mathcal{S}^{\mathrm{e}}$) is the following combination of $\mathcal{S}^{\mathrm{e}} = \mathrm{W}, \mathrm{C}$:

$$g_c^2(L, g_0^2) \equiv \frac{3}{2} g^2_{c,\mathrm{W}}(L, g_0^2) - \frac{1}{2} g^2_{c,\mathrm{C}}(L, g_0^2) \qquad \text{(this chapter)}, \tag{7.1}$$

where $g^2_{c,\mathrm{W}}(L, g_0^2), g^2_{c,\mathrm{C}}(L, g_0^2)$ are g_c^2 from the $\mathcal{S}^{\mathrm{f}}\mathcal{S}^{\mathrm{e}} = \mathrm{WW}, \mathrm{WC}$ combinations, respectively. The combination of Eqn. 7.1 is observed empirically to reduce cutoff effects and was utilized in Ref. [170]. The same combination with $c = 0.45$ will be deployed in Sec. 7.3 for the purposes of finite size scaling. The dependence of $g_c^2(L, g_0^2)$ on β_b is much more mild in the weak coupling phase than it is in the strong coupling phase. Once the simulations cross the phase boundary from weak to strong coupling, $g_c^2(L, g_0^2)$ sharply increases in magnitude. As discussed at length in Ref. [170], the onset of the strong coupling phase is also characterized by a sudden increase in fluctuations of the topological charge (see Chapter 4.4.1). As I have already mentioned in Sec. 7.1, the strong coupling phase is confining, which explains both the strong dependence of $g_c^2\big(L, g_0^2\big)$ on β_b and the observation of fluctuations in the topological charge. Nonetheless, all simulations in the strong coupling phase have $am_f = 0$, which is made possible by the apparent preservation of chiral symmetry in the strong coupling phase. Large-scale spectroscopy calculations in the strong coupling phase are currently being performed by the LSD collaboration using `qex_staghmc`. These should reveal more information about the chiral properties of the strong coupling phase. All simulations in the weak coupling phase have $am_f = 0$ as well.

7.3 Finite size scaling in the eight flavor system

The coupling $g_c^2\big(L, g_0^2\big)$ is a dimensionless scaling variable; as such, it is expected to scale as Eqn. 6.4 within the vicinity of a phase transition ($\beta_b \approx \beta_{b,\mathrm{c}}$) with $\gamma_{g_c^2} = 0$ and $N_\mathrm{s} \equiv L/a$. In other words,

$$g_c^2\big(L, g_0^2\big) \approx \mathcal{F}_{g_c^2}(x) \qquad (\beta_b \approx \beta_{b,\mathrm{c}}), \tag{7.2}$$

where

$$x(\beta_b, L) \equiv |\beta_b/\beta_{b,\mathrm{c}} - 1|(L/a)^{1/\nu} \qquad \text{(2nd-order)} \tag{7.3}$$

or

$$ax(\beta_b, L) \equiv L\exp\big(\text{-}\zeta|\beta_b/\beta_{b,\mathrm{c}} - 1|^{-\nu}\big) \qquad (\infty\text{-order}). \tag{7.4}$$

Note that the phase transition, should it exist, is not a finite-temperature phase transition. It is a zero-temperature phase transition, also referred to as a *quantum phase transition* (QPT). Without PV fields, FSS predicts $\nu \approx 1/d$ ($d = 4$), which is consistent with the discontinuity fixed point prediction resulting from a first-order phase transition [120, 278]. Introducing PV fields has the effect of smoothing out the transition; scaling with Eqn. 7.3 yields a prediction for ν that is not consistent with a first-order transition and the data is even consistent with ∞-order scaling (Eqn.

7.4) [170].

It is important to note that the non-observation of $\nu = 1/d$ does not necessarily imply that the phase transition is not first-order. If the correlation length $\tilde{\xi}(\beta_b)$ is much greater than the volumes L/a deployed in an FSS analysis of the phase transition, then FSS *can* predict a ν that is not consistent with first-order scaling, even if the phase transition is first-order. Such *pseudo*-critical scaling can only be detected by either simulating on volumes $L/a \gg \tilde{\xi}(\beta_b)$ or establishing that the correlation length diverges as $\beta_b \to \beta_{b,\mathrm{c}}$. First-order transitions with large correlation lengths are referred to as *weakly first-order* because the infinite volume discontinuity is smoothed out significantly for $L/a \ll \tilde{\xi}(K)$. For an example of such a potentially dangerous situation, one need not look further than the q-state Potts model already explored in Chapter 6.3. The finite-temperature phase transition of the q-state Potts model is first-order for $q \geq 5$ [113, 114]. The correlation length at the first-order phase transition of the 5-state Potts model is an incredible $\tilde{\xi}\big(K^{(5)}_{\mathrm{Potts,c}}\big) \approx 2512$ and decreases to $\tilde{\xi}\big(K^{(10)}_{\mathrm{Potts,c}}\big) \approx 11$ by $q = 10$ [58, 207]. If $L/a \ll \tilde{\xi}\big(K^{(5)}_{\mathrm{Potts}}\big)$, a FSS analysis of the 5-state Potts model will predict ν that reflects the *pseudo*-critical behavior of the model, which can lead to *incorrectly* concluding that it is continuous if caution is not taken.

7.3.1 Finite size scaling with radial basis function networks

I explore the zero-temperature phase transition of the $N_{\mathrm{f}} = 8$ system by approximating the scaling function $\mathcal{F}_{g_c^2}(x)$ in Eqn. 7.2 with a radial basis function network (RBFN). Finite size scaling with an RBFN has already been explored extensively in Chapter 6 for the 2nd-order transition of the $q = 2, 3$ state Potts model and the $p = 4$ state clock model, along with the ∞-order (BKT) transition of the ∞-state clock (XY) model. The RBFN-based FSS analysis of this chapter is carried over verbatim, save for the imposition of priors on the RBFN centers $(0(100))$, bandwidths $(0(100))$, and bias $(0(100))$ for the purposes of stabilizing the curve collapse. The width λ of the prior on the weights of the RBFN is determined by the surrogate-based empirical Bayes analysis described in Chapter 6.2.3. I perform the FSS with a variety of scaling ansatz. I test for 2nd-order scaling using Eqn. 7.3 with ν free, while 1st-order scaling is tested similarly using Eqn. 7.3 with $\nu = 1/4$ explicitly.

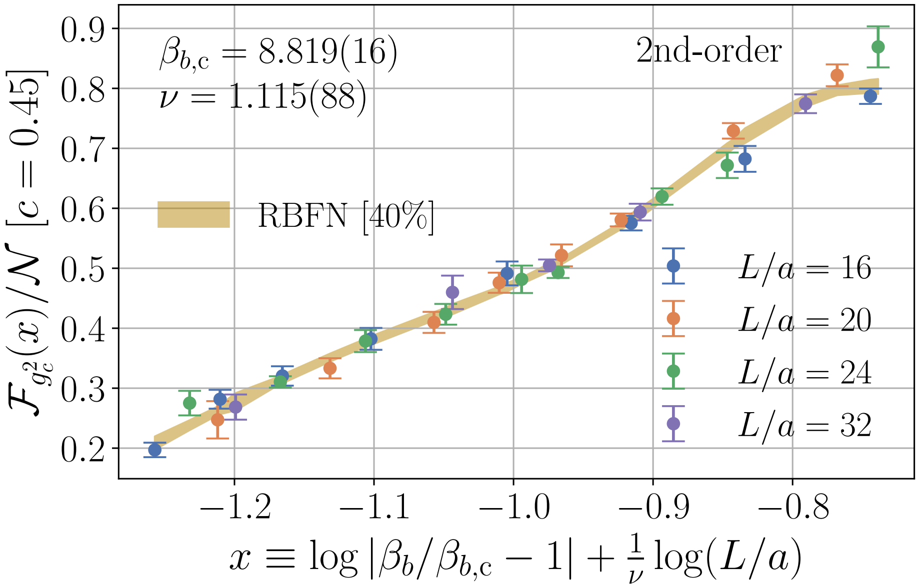

Figure 7.2: 2nd-order curve collapse (Eqn. 7.3) for g_c^2 at $c = 0.45$. Scaling function normalized by $\mathcal{N} = 128\pi^2/(3N^2 - 3)$ for visualization purposes. Prediction from RBFN indicated by a gold band with the width of the band indicating the error. Data entering the curve collapse indicated bu multicolored error bars for $L/a = 16$ (blue), 20 (orange), 24 (green), 32 (purple). Percent on RBFN label indicates the p-value of the fit.

The data is not precise enough to be able to determine ζ and ν simultaneously for ∞-order scaling without introducing a dangerous amount of bias into the prior for ζ. Hence, I compare BKT-like scaling using Eqn. 7.4 with $\nu = 1/2$ against "walking" scaling using Eqn. 7.4 with $\nu = 1$. The prior for $\beta_{b,\mathrm{c}}$ for all four fit ansatz is 8.7(7), based on the results of Ref. [170]; such a wide prior is chosen to stabilize the fit while reducing bias as much as possible. The prior for ν from 2nd-order scaling is 1(1); the same is true for the prior on ζ for the two ∞-order ansatz with $\nu = 1/2$ and $\nu = 1$. Priors on the critical parameters $\beta_{b,\mathrm{c}}, \nu$ and ζ are enforced logarithmically, so as to explicitly enforce positivity. The RBFN that fits the data best with the least number of parameters possesses two nodes in its internal layer. Curiously, the width λ^* that maximizes the marginal likelihood (Eqn. 6.12) is close to $\lambda^* \approx 0.529$ for all four scaling ansatz. In Table 7.3, I summarize the result of the RBFN-based FSS at λ^* for each of the aforementioned scaling ansatz.

According to Table 7.3, 1st-order scaling is strongly disfavored compared to both 2nd-order

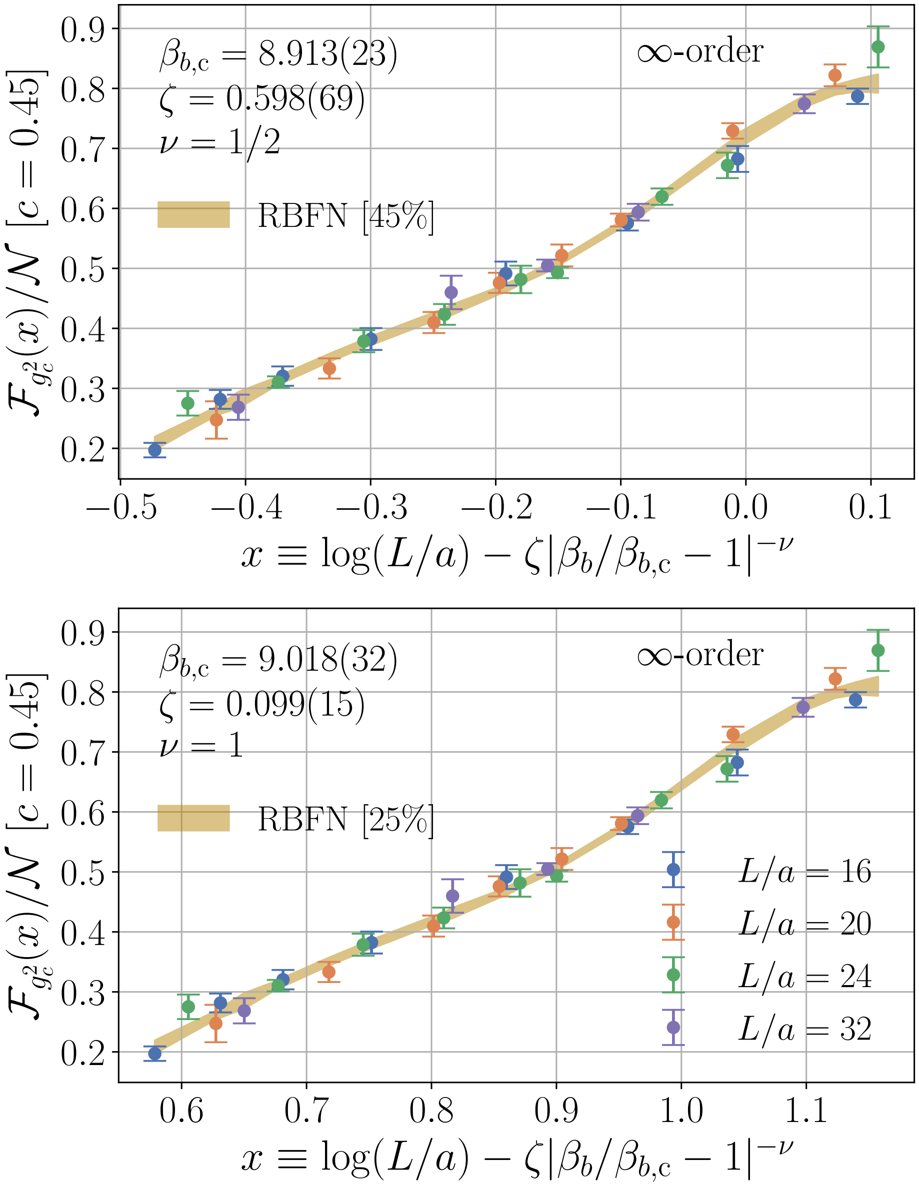

Figure 7.3: ∞-order curve collapse (Eqn. 7.4) for g_c^2 at $c = 0.45$ with fixed $\nu = 1/2$ (top panel) and $\nu = 1$ (bottom panel). Scaling function normalized by $\mathcal{N} = 128\pi^2/(3N^2 - 3)$ for visualization purposes. Prediction from RBFN indicated by a gold band with the width of the band indicating the error. Data entering the curve collapse indicated bu multicolored error bars for $L/a = 16$ (blue), 20 (orange), 24 (green), 32 (purple). Percent on RBFN label indicates the p-value of the fit.

scaling and ∞-order scaling. The augmented χ^2 ($\chi^2_{\mathrm{aug.}}$) and the χ^2 of the data χ^2_{data} per their respective degrees of freedom are ≈ 20× larger for 1st-order scaling than all three continuous scaling ansatz. The marginal likelihood for the 1st-order scaling ansatz is also significantly lower. As I have stated already, it is nevertheless impossible to completely rule out 1st-order scaling based on very poor fit quality alone; more information from $\tilde{\xi}(\beta_b)$ is needed.

In Fig. 7.2, I illustrate the result of the RBFN-based FSS with 2nd-order scaling. The scaling function $\mathcal{F}_{g_c^2}$ is plotted as a dark gold band and the width of the band indicates the statistical error of $\mathcal{F}_{g_c^2}$. The $(\beta_b, L/a)$ pairs entering the curve collapse are shown as multi-colored error bars, with the color indicating the volume. The prediction for $\beta_{b,\mathrm{c}}$ is just above the strongest β_b entering the β-function analysis of Sec. 7.4. The prediction for $\nu = 1.115(88)$ is $\approx 10\%$ away from unity in central value and statistically consistent with unity at the $\approx 1.35\sigma$ level. Note additionally that the critical parameters $\beta_{b,\mathrm{c}}, \nu = 8.819(16), 1.115(88)$ are not consistent with those of Ref. [170] and the $\chi^2_{\mathrm{aug.}}$/d.o.f. is much more reasonable. The reason for this discrepancy is likely attributed to the improved analysis procedure of this chapter: I do not have to rely on matching to a single bare gauge coupling, the RBFN is a better interpolator than the smoothing spline used in Ref. [170], I am using a global optimization algorithm, I am controlling for overfitting with empirical Bayes, and the error estimation is more reliable. Moreover, removing the $L/a = 32$ data yields a consistent prediction for the critical parameters; therefore, it is unlikely that the discrepancy can be attributed to an improvement in the underlying dataset.

Similarly, the result of the RBFN-based ∞-order curve collapse with $\nu = 1/2$ and $\nu = 1$ are shown in the top and bottom panels of Fig. 7.3, respectively. The prediction for $\beta_{b,\mathrm{c}} = 8.913(23), 9.018(32)$ from $\nu = 1/2, 1$, respectively, is weaker (smaller g_0^2) than it is for 2nd-order scaling and increases in magnitude with ν. Moreover, the prediction for ζ is negatively correlated with ν; $\zeta = 0.598(69)$ from $\nu = 1/2$ and $\zeta = 0.099(15)$ from $\nu = 1$. As was the case for 2nd-order scaling, the prediction for β_b from either ∞-order ansatz is both higher than and not consistent with the value for β_b reported in Ref. [170].

From $\chi^2_{\mathrm{aug.}}$, χ^2_{data}, and $\log\mathrm{ML}$ (listed in Table 7.3) alone, it is not possible to determine if any one continuous scaling ansatz is unambiguously preferred over the others. The ∞-order ansatz with $\nu = 1/2$ is *slightly* preferred by $\chi^2_{\mathrm{aug.}}$ and $\log\mathrm{ML}$, while ∞-order scaling with $\nu = 1$ is *slightly* preferred by $\chi^2_{\mathrm{aug.}}$. The lead that ∞-order scaling with $\nu = 1/2$ has over 2nd-order scaling as far as

	Fit result					
Fit	$\beta_{b,\mathrm{c}}$	ν	ζ	$\chi^2_{\mathrm{aug.}}/\mathrm{d.o.f.}$	$\chi^2_{\mathrm{data}}/\mathrm{d.o.f.}^*$	$\log \mathrm{ML}$
1st-order	8.65101(20)	1/4	–	20.77	28.12	-267.70
2nd-order	8.819(16)	1.115(88)	–	1.05	1.44	30.20
∞-order	8.913(23)	1/2	0.598(69)	1.01	1.38	30.74
∞-order	9.018(32)	1	0.099(15)	1.16	1.34	27.04

Table 7.3: Juxtaposition of 1st-, 2nd-, and ∞-order scaling ansatz. See Chapter 8.2 and Appendix D for definition of $\chi^2_{\mathrm{aug.}}$, χ^2_{data}, and marginal likelihood (ML). d.o.f. $\equiv$ "# data $-$ # parameters $+$ # priors" and d.o.f.* $\equiv$ "# data $-$ # parameters". "# data $= 30$" fixed for all fits, "# parameters $=$ # priors $= 9$" for 2nd-/∞-order and 8 for 1st-order.

the marginal likelihood is concerned is so small that even stating that it is "preferred" over 2nd-order scaling generous at best. Taken together, the best that one can state is that ∞-order scaling is *slightly* preferred over 2nd-order scaling; however, the possibility of 2nd-order being preferred over ∞-order cannot be confidently ruled out with the dataset and analysis of this chapter. The most definitive statement that can be made is that continuous scaling is *strongly* preferred over 1st-order scaling, which indicates that the phase transition is likely *not* 1st-order *if* subsequent measurements of the correlation length support that it is *not* finite at $\beta_{b,\mathrm{c}}$. Such investigations are underway.

7.4 Calculation of the continuum β-function

In Chapters 4 and 5, I calculated the continuum gradient flow β-function for the $N_{\mathrm{f}} = 0$ and $N_{\mathrm{f}} = 12$ systems, respectively. In both cases, I was able to calculate some quantity from the β-function that more or less verified the legitimacy of the infinite volume RG scheme that I have been using throughout this thesis, up to irregularities in the literature and systematic effects in the analysis. In this section, I will apply the same methods to extract the continuum β-function of the $N_{\mathrm{f}} = 8$ system. As such, statistical uncertainties are estimated and kept track of using the automatic error propagation tools provided by the `gvar` library [241]. Fits are performed using either the `SwissFit` library (see Sec. 8.2) or the `lsqfit` library [240, 290]. I have also tested using the

Γ-method to account for accounting for autocorrelation in correlated uncertainties [378]. This is accomplished by first calculating the ordinary correlation matrix, then calculating the covariance matrix from the correlation matrix by rescaling it with the statistical error estimated from the Γ-method, as suggested by Ref. [215]. However, doing so introduces numerical noise into the analysis that is difficult to control for, so I continue the practice of accounting for autocorrelation by binning.

In the CBFM analysis of Refs. [178, 179], multiple flows were used to cover the entire renormalized trajectory from the UVFP to the IRFP of the models investigated in those works. The flows deployed in Refs. [178, 179] are parameterized by the plaquette c_{p} and rectangle c_{r} coefficient of the Lüscher-Weiss action briefly discussed in Chapter F [248, 369, 370], which are constrained by

$$c_{\mathrm{p}} + 8c_{\mathrm{r}} = 1 \tag{7.5}$$

For example, the Wilson action is obtained from $c_{\mathrm{r}} = 0$, while the tree-level improved Lüscher-Weiss action is obtained from $c_{\mathrm{r}} = -1/12$. In this chapter, I utilize different flows for the same purpose. I refer to each flow as $\mathcal{S}^{\mathrm{f}} = \mathrm{P}XY$, where X is the numerator and Y is the denominator of the c_{p} coefficient; e.g., “P11=W” (*Wilson flow*) has $c_{\mathrm{p}} = 1$, while “P13” has $c_{\mathrm{p}} = 1/3$. The continuum β-function in this chapter utilizes four flows: P11≡W ($c_{\mathrm{p}} = 1$), P23 ($c_{\mathrm{p}} = 2/3$), P13 ($c_{\mathrm{p}} = 1/3$), and P16 ($c_{\mathrm{p}} = 1/6$). As in Chapters 4 and 5, I refer to the β-function determined from a specific flow/operator combination as “$\mathcal{S}^{\mathrm{f}}\mathcal{S}^{\mathrm{e}}$”; e.g., P13 flow with the clover operator is “P13C”. For each flow/operator combination, I calculate $g^2_{\mathrm{GF}}(t; L, g_0^2)$ using Eqn. 4.24 with $\delta(t, L)$ set by Eqn. 4.16. The corresponding finite-volume β-function $\beta_{\mathrm{GF}}(t; L, g_0^2)$ is calculated from $g^2_{\mathrm{GF}}(t; L, g_0^2)$ using the 5-point stencil of Eqn. 4.37 with $\delta t/a^2 = 0.05$.

7.4.1 Infinite volume extrapolation

I extrapolate both $g^2_{\mathrm{GF}}(t; L, g_0^2)$ and $\beta_{\mathrm{GF}}(t; L, g_0^2)$ to the $a/L \to 0$ limit at fixed β_b and t/a^2 using Eqn. 4.38, as in Chapters 4 and 5. The uncertainty in $k_i(t; g_0^2)$ coefficients of Eqn. 4.38 includes both the statistical uncertainty in $g^2_{\mathrm{GF}}(t; L, g_0^2)$ and $\beta_{\mathrm{GF}}(t; L, g_0^2)$, respectively, along

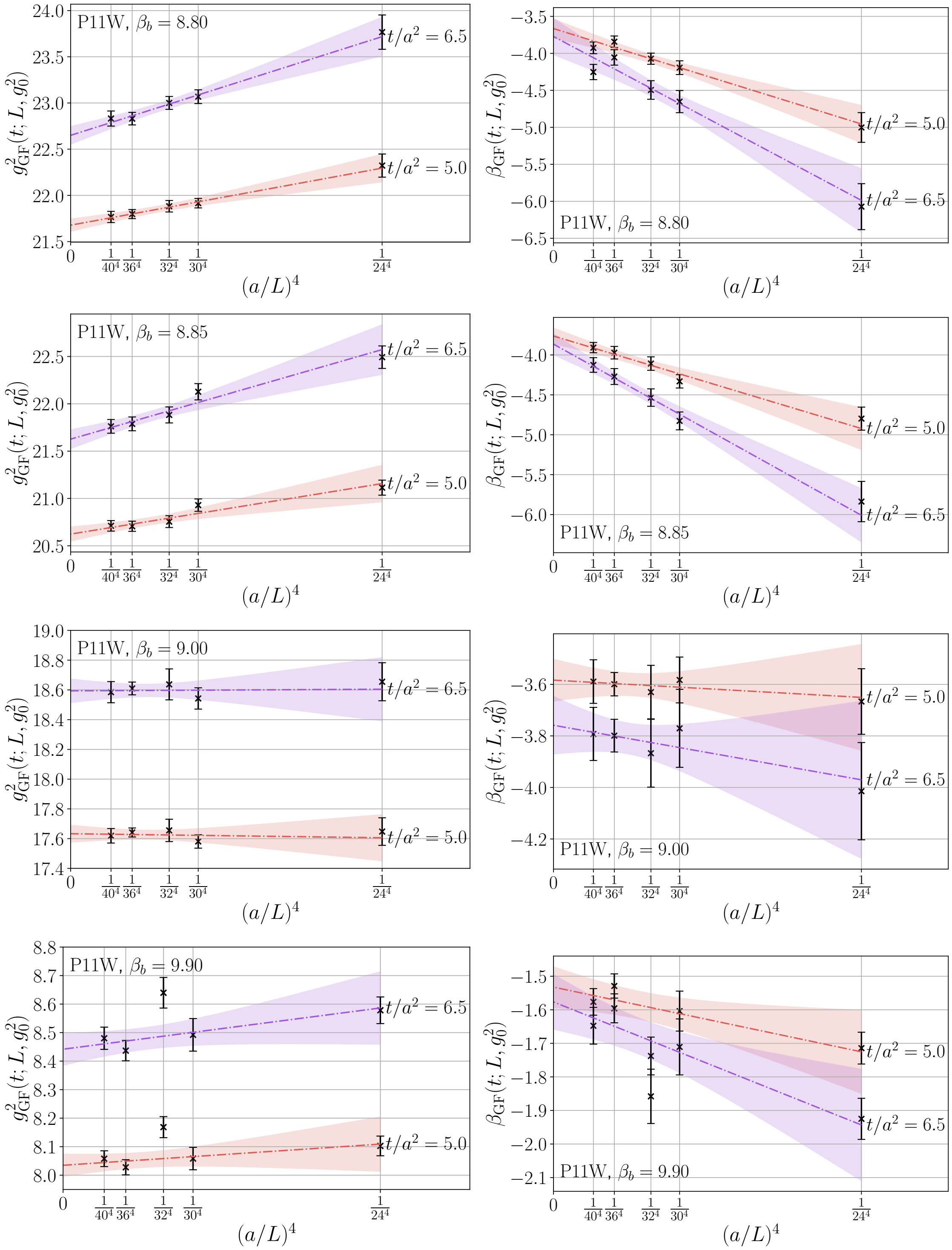

Figure 7.4: Example of infinite volume extrapolation of $g^2_{\rm GF}\big(t;L,g_0^2\big)$ (left panels) and $\beta_{\rm GF}\big(t;L,g_0^2\big)$ (right panels) for P11W (WW in previous chapters) flow/operator combination at $\beta_b = 8.85$ (top panels), 9.00 (middle panels), and 9.90 (bottom panels). Extrapolation shown at only $t/a^2 = 5.0$ (red) and 6.5 (purple) for visualization purposes.

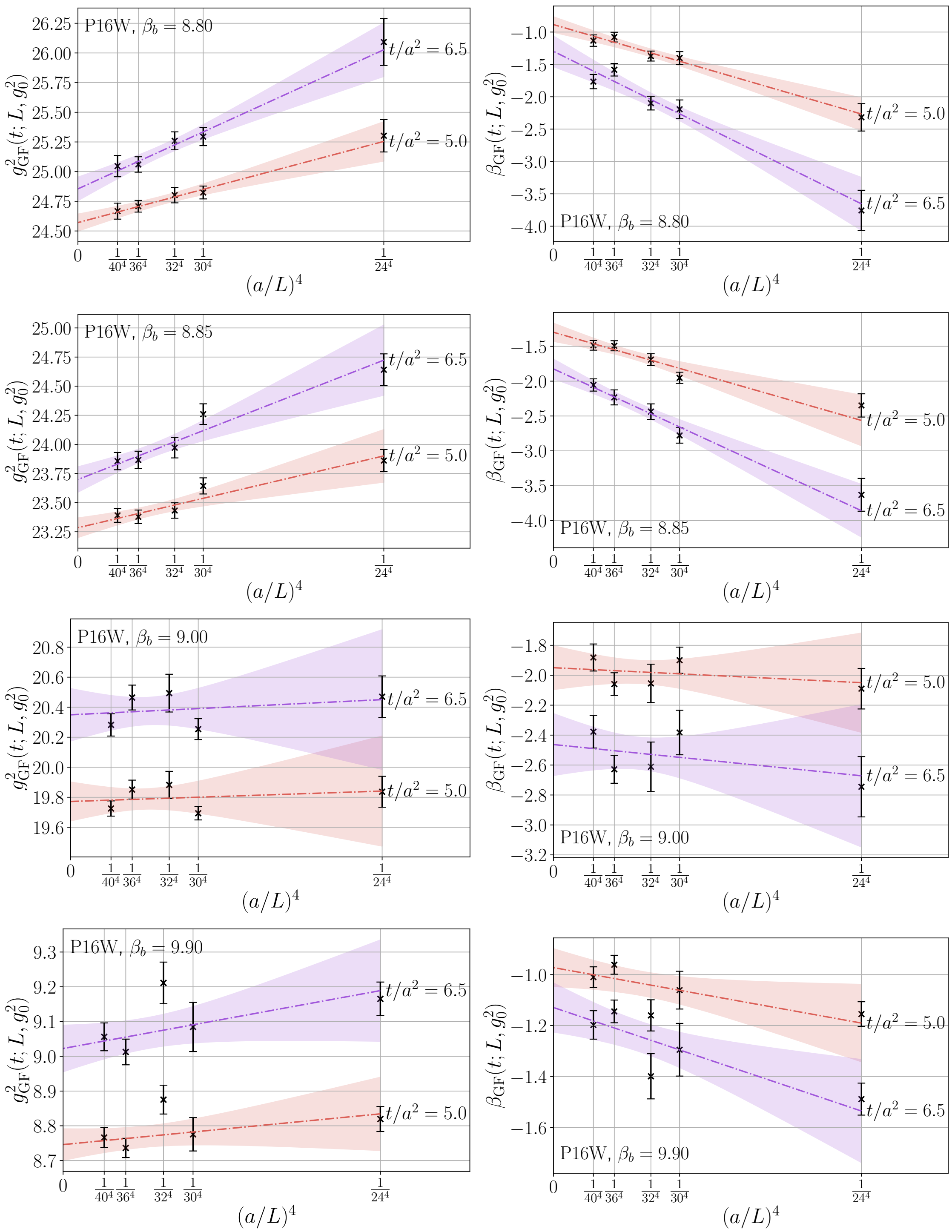

Figure 7.5: Example of infinite volume extrapolation of $g^2_{\mathrm{GF}}\big(t; L, g_0^2\big)$ (left panels) and $\beta_{\mathrm{GF}}\big(t; L, g_0^2\big)$ (right panels) for P16W flow/operator combination at $\beta_b = 8.85$ (top panels), 9.00 (middle panels), and 9.90 (bottom panels). Extrapolation shown at only $t/a^2 = 5.0$ (red) and 6.5 (purple) for visualization purposes.

with the systematic uncertainty that is associated with extrapolating over any subsets of volumes $L/a \in \{24, 30, 32, 36, 40\}$ using the Bayesian model averaging procedure outlined/deployed in the CBFM analysis of the $N_{\rm f} = 12$ system in Chapter 5. All model variations contain at least three volumes, so that the linear extrapolation in a^4/L^4 have at least one degree of freedom left.

I illustrate the infinite volume extrapolation with $\mathcal{S}^{\rm f}\mathcal{S}^{\rm e} = $ P11W (WW) in Fig. 7.4 and $\mathcal{S}^{\rm f}\mathcal{S}^{\rm e} = $ P16W in Fig. 7.5 for $g^2_{\rm GF}\left(t; L, g_0^2\right)$ (left panels) and $\beta_{\rm GF}\left(t; L, g_0^2\right)$ (right panels) at $\beta_b = 8.80$ (top panels), 8.85 (middle-top), 9.00 (middle-bottom), and 9.90 (bottom). For the purposes of illustration, I only show the extrapolation at $t/a^2 = 5.0, 6.5$, which are the lower/upper bound on the flow time entering the continuum extrapolation in Sec. 7.4.2. As β_b decreases, finite volume effects become more pronounced, perhaps due to being within the vicinity of the phase transition. Nonetheless, finite volume effects at the strongest couplings ($8.8 \leq \beta_b \leq 8.9$) appear to be described quite well by a linear dependence on a^4/L^4. See, for example, the extrapolation at $\beta_b = 8.80$ and 8.85 in Figs. 7.4 and 7.5. As β_b increases, finite-volume effects tend to be comparatively mild; however, statistical fluctuations in the ensembles at different L/a tend muddy the appearance of any particular trend in a^4/L^4, perhaps because the extrapolation tends to be fairly flat. As such, the combined statistical/systematic error predicted by Bayesian model averaging can, and often does, cover the spread in the central values of the data entering the infinite volume extrapolation for $9.00 \leq \beta_b \leq 11.0$. See, for example, the extrapolations at $\beta_b = 9.00, 9.90$ in Figs. 7.4 and 7.5.

7.4.2 Continuum extrapolation

I extrapolate $\beta_{\rm GF}\left(t; g_0^2\right)$ to the $a^2/t \to 0$ limit at fixed $g^2_{\rm GF}$ using the ansatz of Eqn. 4.39. As in Chapters 4 and 5, doing so requires first interpolating $\beta_{\rm GF}\left(t; g_0^2\right)$ in $g^2_{\rm GF}\left(t; g_0^2\right)$ at fixed t/a^2.

7.4.2.1 Intermediate interpolation

The fast running β-function of the $N_{\rm f} = 0$ system in Chapter 4 required the use of an interpolating function that is able to accommodate different asymptotic behaviors of the β-function in the weak/strong coupling regime (Eqn. 4.42). The intermediate interpolation for the slow running

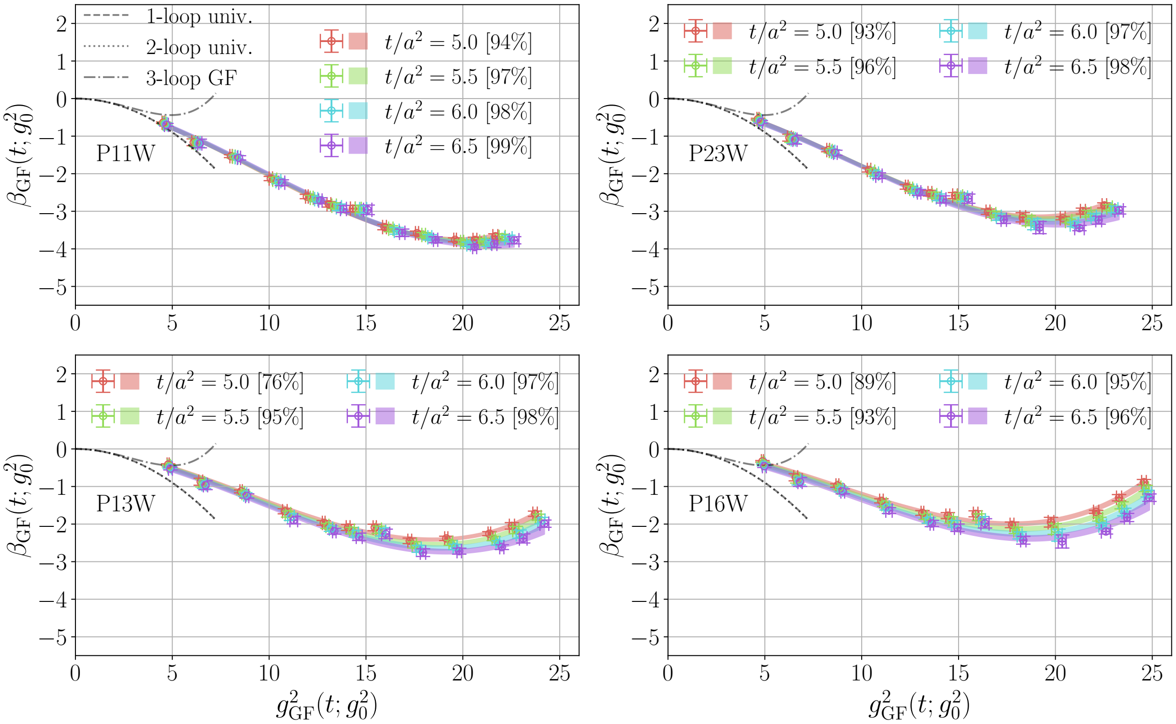

Figure 7.6: Example of interpolation of $\beta_{\mathrm{GF}}\left(t; g_0^2\right)$ in $g^2_{\mathrm{GF}}\left(t; g_0^2\right)$ for P11W (WW, top left), P23W (top right), P13W (bottom left), and P16W (bottom right). Result of interpolation indicated by multicolored bands at fixed $t/a^2 = 5.0$ (red), 5.5 (green), 6.0 (cyan), and 6.5 (purple). Data entering interpolation indicated by multicolored error bars with corresponding colors for each t/a^2. Interpolation juxtaposed against 1- (dashed), 2- (dotted), and 3-loop (dashed-dotted) continuum β-function from perturbation theory [161].

β-function of the $N_{\mathrm{f}} = 12$ system in Chapter 5 is described well by a perturbative ansatz (Eqn. 5.7). The β-function of the $N_{\mathrm{f}} = 8$ system does not run as fast as the $N_{\mathrm{f}} = 0$ system, nor does it exhibit signs of asymptotic linearity in g^2_{GF} for $g^2_{\mathrm{GF}}/4\pi \gg 1$. However, the running of the β-function in the $N_{\mathrm{f}} = 8$ system is fast enough so as to prevent it from being described well by a perturbative ansatz over the entire range of renormalized couplings explored in this chapter. As such, the interpolating function $\mathcal{I}_N$ is chosen to be a generic polynomial

$$\mathcal{I}_N\left(g^2_{\mathrm{GF}}\right) = \sum_{n=0}^{N-1} p_n g^{2n}_{\mathrm{GF}}. \tag{7.6}$$

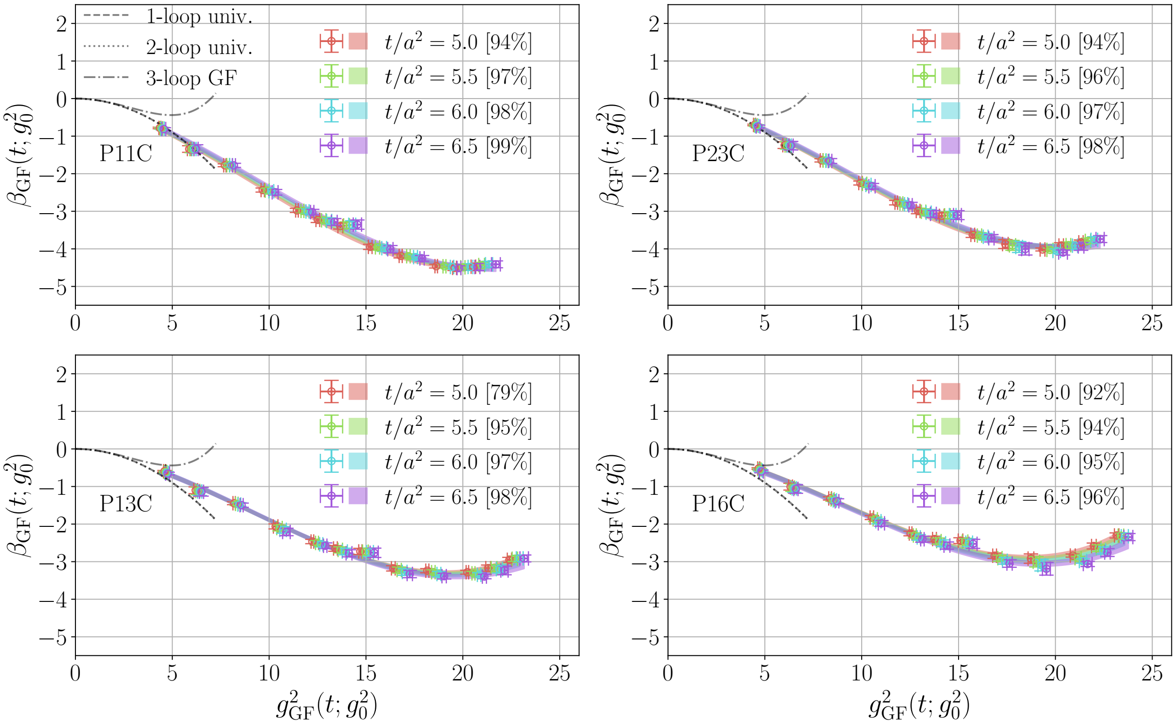

Figure 7.7: Example of interpolation of $\beta_{\rm GF}(t; g_0^2)$ in $g^2_{\rm GF}(t; g_0^2)$ for P11C (WC, top left), P23C (top right), P13C (bottom left), and P16C (bottom right). Result of interpolation indicated by multicolored bands at fixed $t/a^2 = 5.0$ (red), 5.5 (green), 6.0 (cyan), and 6.5 (purple). Data entering interpolation indicated by multicolored error bars with corresponding colors for each t/a^2. Interpolation juxtaposed against 1- (dashed), 2- (dotted), and 3-loop (dashed-dotted) continuum β-function from perturbation theory [161].

As in Chapter 5, the statistical error in $g^2_{\rm GF}(t; g_0^2)$ is accounted for by treating $g^2_{\rm GF}(t; g_0^2)$ as a Gaussian prior (see Appendix D). Moreover, a prior of 0.0(0.1) is imposed on the p_n coefficients to stabilize the fit. The lowest order N that fits the data well is $N = 4$; i.e., a cubic polynomial. In Fig. 7.6, I illustrate the intermediate interpolation for P11W (top left panel), P23W (top right panel), P13W (bottom left panel), and P16W (bottom right panel) at $t/a^2 = 5.0, 6.0, 5.5$ and 6.0 (different colors). The same information for PXYC is illustrated in Fig. 7.7. In all cases, the p-values are typically quite high. This was also the case in the $N_{\rm f} = 12$ system and likely indicates that the combined statistical/systematic error is overestimated. At the weakest couplings ($g^2_{\rm GF}/4\pi \lesssim 1$), discretization effects appear to be quite mild. As $g^2_{\rm GF}$ enters the strong coupling $g^2_{\rm GF}/4\pi \gtrsim 1$ regime, discretization effects appear to increase; see, for example, the bottom right panel of Fig. 7.6 (P16W).

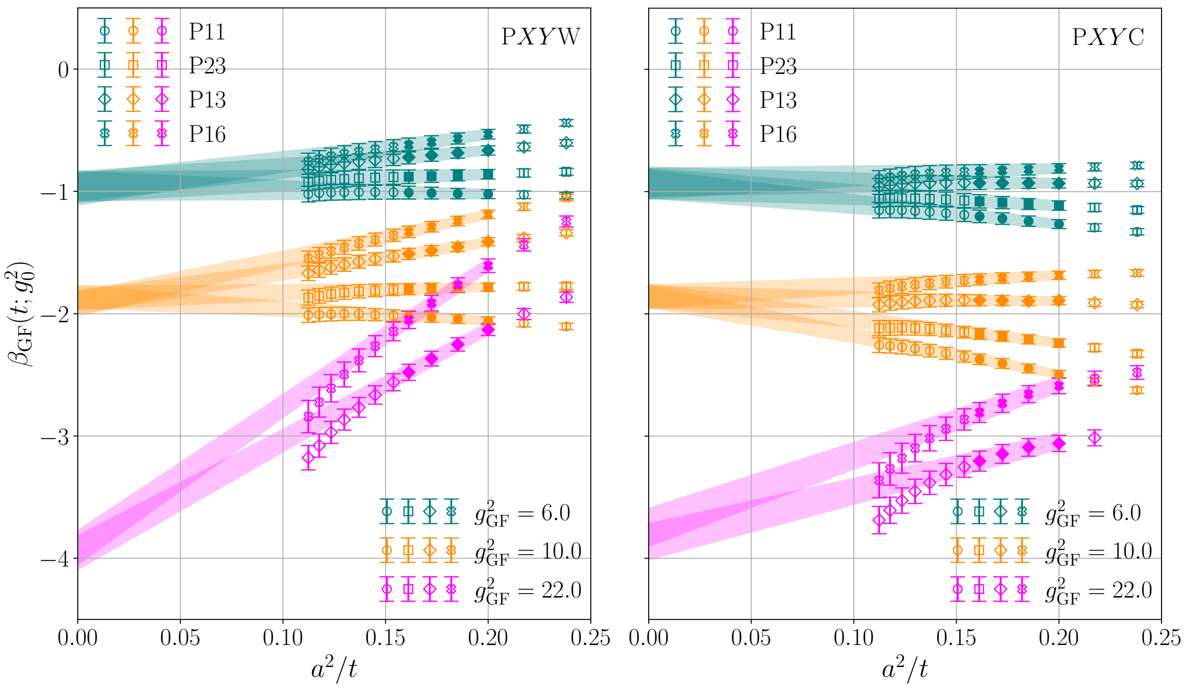

Figure 7.8: Sample of continuum extrapolation for W (left panel) and C (right panel) operator at fixed $g^2_{\mathrm{GF}} = 6.0$ (teal), 10.0 (orange), and 22.0 (magenta). Different flows indicated by different symbols: P11 (circle), P23 (box), P13 (diamond), and P16 (×). Data entering continuum extrapolation indicated by filled symbols and data not entering continuum extrapolation are open.

7.4.2.2 Continuum extrapolation

With the intermediate interpolation on hand, the continuum limit is taken by extrapolating $\beta_{\mathrm{GF}}(t; g_0^2)$ to the $a^2/t \to 0$ limit over a set of fixed g^2_{GF} by fitting $\beta_{\mathrm{GF}}(t; g_0^2)$ linearly in a^2/t at each g^2_{GF}. The values for $[t_{\mathrm{min.}}, t_{\mathrm{max.}}]/a^2$ over which the continuum extrapolation is performed is chosen based upon the overlap of the continuum prediction for $\beta_{\mathrm{GF}}(g^2_{\mathrm{GF}})$ from different flow/operator combinations; the choice of $[t_{\mathrm{min.}}, t_{\mathrm{max.}}]/a^2 = [5.0, 6.5]$ yields reasonable agreement over the range of accessible renormalized couplings $5 \leq g^2_{\mathrm{GF}} \lesssim 24$. The same issues with statistical correlations that had to be dealt with in Chapters 4 and 5 are also present in the continuum extrapolation of this chapter. As before, they are dealt with by estimating the central value of the continuum β-function $\beta_{\mathrm{GF}}(g^2_{\mathrm{GF}})$ from an "uncorrelated" fit and the error in $\beta_{\mathrm{GF}}(g^2_{\mathrm{GF}})$ from the half difference of the central value of fits to $\beta_{\mathrm{GF}}(t; g_0^2) \pm 1\sigma$.

In Fig. 7.8, I illustrate the continuum extrapolation for PXYW (left panel) and PXYW (right panel) at $g_{\mathrm{GF}}^2 = 6.0$ (blue), 10.0 (orange), 22.0 (pink). The raw data for each flow is indicated by its own marker: P11 ($\circ$), P23 ($\square$), P13 ($\diamond$), P16 ($\times$). The continuum prediction for $\beta_{\mathrm{GF}}\left(g_{\mathrm{GF}}^2\right)$ from the W operator is consistent for all flows between $5.0 \leq g_{\mathrm{GF}}^2 \lesssim 15.0$ and $5.0 \leq g_{\mathrm{GF}}^2 \lesssim 12.5$ from the C operator. The P11 flow peels off first, followed by P23; the P13 and P16 flows remain consistent over the entire range of g_{GF}^2. The point in g_{GF}^2 at which flows with larger c_{p} peel off is accompanied by nonlinear effects of $\beta_{\mathrm{GF}}\left(t; g_0^2\right)$ in a^2/t. At first, the observation of flows with different c_{p} peeling off earlier in g_{GF}^2 appears to be a cause for concern; however, it may be the case that different flows tend to confidently cover different sections of the renormalized trajectory as far as the CBFM is concerned. The same phenomenon has been observed in the CBFM analysis of the $N_{\mathrm{f}} = 10$ system in Ref. [179] and the CBFM analysis of the massless SU(4) system with four fermions in the fundamental representation and four fermions in the two-index antisymmetric representation of Ref. [178].

7.4.3 The continuum β-function

In Fig. 7.9, I show the result for the continuum β-function from PXYW (top panel) and PXYC (bottom panel). The continuum β-function is juxtaposed against the 1-, 2-, and 3-loop perturbative β-function and the continuum β-function from the CBFM applied to simulations of domain wall fermions (DWF, briefly discussed in Chapter 3.2.2) [161, 192]. For $5.0 \leq g_{\mathrm{GF}}^2 \lesssim 7.5$, the prediction for $\beta_{\mathrm{GF}}\left(g_{\mathrm{GF}}^2\right)$ from each flow/operator combination overlaps with the DWF prediction. The continuum β-function extends past the DWF prediction up to $g_{\mathrm{GF}}^2 \approx 24$. As I have already mentioned, the continuum prediction from the P11 and P23 flows peels off earlier in g_{GF}^2, while the P13 and P16 flows are consistent over the entire range of g_{GF}^2.

Intriguingly, the P13/P16 flows predict that the continuum β-function could curve upwards. If the upward curvature is not either a systematic effect or an artefact of the RG scheme, then the upward curvature could indicate either the presence of an IRFP or that the β-function will eventually exhibit slow walking. If the results of the FSS analysis in Sec. 7.3 are to be believed, then it is a bit surprising that the continuum β-function of this section is not already showing stronger

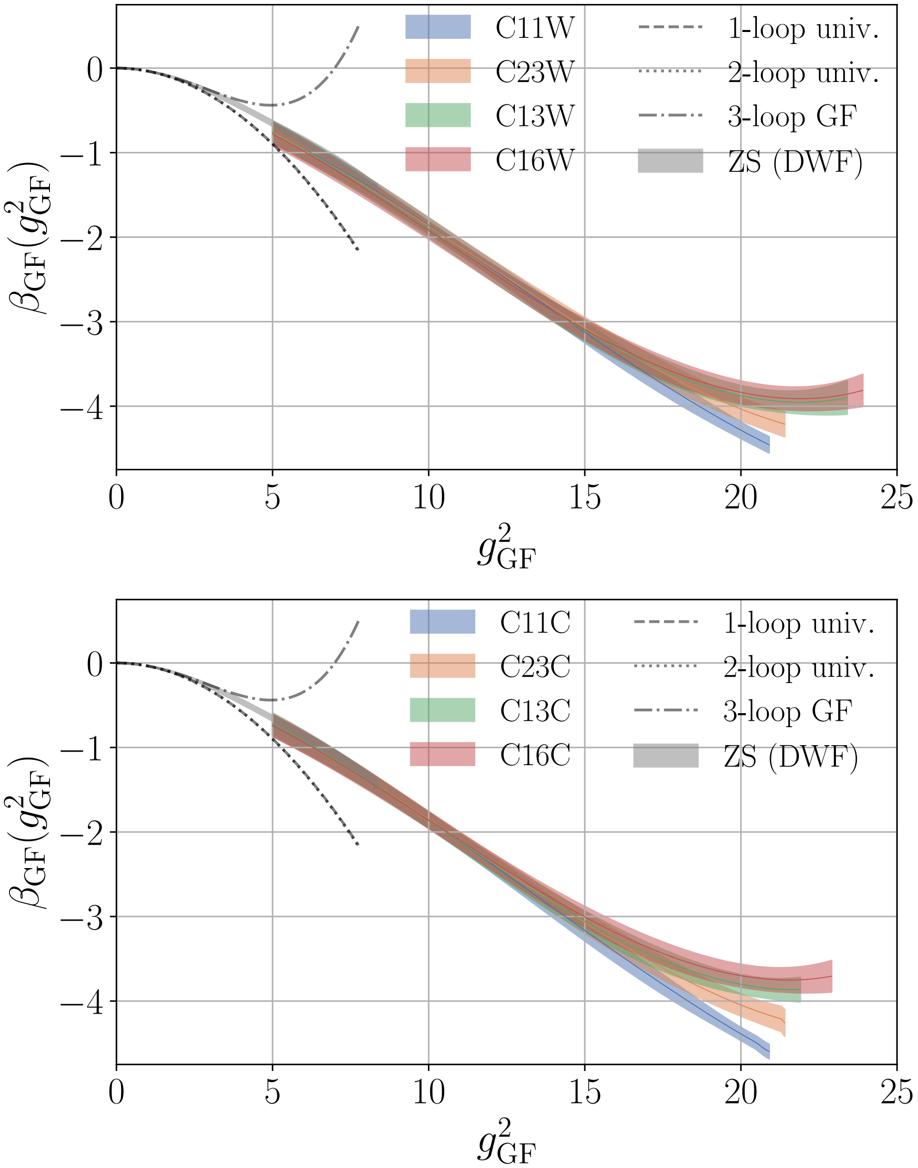

Figure 7.9: Continuum $N_{\mathrm{f}} = 8$ β-function from W (top panel) and C (bottom panel) operators for each flow: C11 (blue), C23 (orange), C13 (green), and C16 (red). Width of band for continuum β-function indicates the error. Juxtaposed against continuum prediction from domain wall simulations using the same RG scheme [192] and 1- (dashed), 2- (dotted), and 3-loop (dashed-dotted) continuum β-function from perturbation theory [161].

signs of an IRFP or slow walking. With the current action, it is unlikely that flows with smaller c_{p} will significantly extend the reach of the continuum β-function in g^2_{GF}, as they had for the $N_{\mathrm{f}} = 10$ system in Ref. [179]. As is evident from Ref. [188], adding more PV fields is also not likely to extend the reach in g^2_{GF}, as the current action already saturates the PV improvement. Nonetheless, if the continuum β-function of this chapter is to be taken seriously, or, even better, reproduced by another group, then it is likely that simulations of the $N_{\mathrm{f}} = 8$ system that have been performed up to the date of writing this thesis have only probed the "weakly coupling" region of the $N_{\mathrm{f}} = 8$ system. The best that one can say now is that the $N_{\mathrm{f}} = 8$ system likely has many surprises in store for the future.

Chapter 8

Other developments

The topics covered in this chapter can generally be considered as “bonus content”, as they are not core to the primary objective to this thesis, which has been covered in Chapters 4-7. I discuss the important software development that made the results of Chapters 5-7 possible; namely, the `Quantum EXpressions` (`QEX`-based) staggered Hamiltonian Monte Carlo code `qex_staghmc` software suite (Sec. 8.1) and the `SwissFit` library (Sec. 8.2) [284, 290], both of which I wrote and am currently responsible for maintaining. In Sec. 8.3, I discuss the extraction of running operator anomalous dimensions in the $N_{\mathrm{f}} = 10$ system using the RG scheme/methods deployed in Chapters 4-7. The latter project will likely never make it past preliminary results; hence, what has been achieved is summarized in this chapter.

8.1 The `Quantum EXpressions`-based `qex_staghmc` suite

Chapters 5 and 7 utilize Pauli-Villars (PV) improved Hamiltonian Monte Carlo simulations with $N_{\mathrm{st.}} = 2$ and 3 species of staggered fermion, respectively. The first simulations of both systems with PV improvement utilized a modified version of the `MILCv7` library [170, 188].[1] For the small volume ($8 \leq L/a \leq 24$) simulations that were targeted by such studies, the modified `MILC` library was efficient enough for all intents and purposes. However, simulating larger volumes ($L/a = 36, 40$) over a reasonable amount of both human and computer time on modern machines required a more modern code that is capable of fully utilizing the innovations in hardware/algorithms that have been made

---

[1] The modified `MILC` library can be found at `https://github.com/daschaich/KS_nHYP_FA`

over the last decade. A natural target in the modern era of lattice calculations was to implement PV improvement in the beautiful, highly-optimized `GRID` library [47], which was used in Chapter 4 to simulate the pure Yang-Mills system. Even better, `GRID` was designed around domain wall fermions, which implement Pauli-Villars fields already for a different purpose. However, I eventually came to learn that modifying the `GRID` library to fit the needs of the calculations performed in Chapters 5 and 7 would be fairly difficult to achieve over a reasonable time frame. Instead, I decided to implement PV improvement in the `Quantum EXpressions` (`QEX`) library, which already had many of the foundational tools that I needed for the specific action that I intended to simulate [284]. Even more, `QEX` was already in use by the Lattice Strong Dynamics (LSD) collaboration and is actively supported by its developers at Argonne National Laboratory.

`QEX`[2] is a high-level framework for performing lattice field theory computations using the beautiful `Nim` programming language [284, 299]. `QEX` supports a combination of task-based parallelism, thread-based parallelism, and single instruction/multiple data (SIMD-based) parallelism. It is also capable of utilizing graphics processing units (GPUs) when deployed with `QUDA` as a backend [82].[3] `Nim` is a high-level, multi-paradigm systems programming language that supports a variety of meta-programming features that make it an extremely powerful/versatile tool for high performance computing (HPC) [299]. `qex_staghmc`[4] builds upon the extensive tools of `QEX` to provide a staggered HMC suite that is both modular and readily/easily deployed with or without PV improvement. `qex_staghmc` also ships with a production-ready code for integrating gradient flow equations and measuring gradient flow observables with any Lüscher-Weiss gauge flow [248, 369, 370].

---

[2] `QEX` is available at `https://github.com/jcosborn/qex`.

[3] `QUDA` is available at `https://github.com/lattice/quda`

[4] `qex_staghmc` is available at `https://github.com/ctpeterson/qex_staghmc`.

8.1.1 Hamiltonian Monte Carlo

After integrating out the *staggered* fermion fields using the properties of Berezin integrals, the partition function of Eqn. 3.29 can be written for certain flavor numbers as [227]

$$\mathcal{Z} = \int \left[\mathrm{d}\mathcal{U}_\mu\right] \exp\left(\text{-}\mathcal{S}_{\mathrm{G}}\left[\mathcal{U}_\mu\right]\right) \prod_{s=0}^{N_{\mathrm{st.}}-1} \det\left(\mathfrak{D}_{\mathrm{S}}^{(s)}\mathfrak{D}_{\mathrm{S}}^{(s)\dagger}\right), \tag{8.1}$$

where $\mathcal{S}_{\mathrm{G}}$ is some lattice discretization of the pure Yang-Mills action (Eqn. 3.2) and $\mathfrak{D}_{\mathrm{S}}^{(s)}$ is the staggered Dirac operator of staggered species s defined by Eqn. 3.17. The operator $\mathfrak{D}_{\mathrm{S}}^{(s)}\mathfrak{D}_{\mathrm{S}}^{(s)\dagger}$ is positive-definite. Working with $\mathfrak{D}_{\mathrm{S}}^{(s)}\mathfrak{D}_{\mathrm{S}}^{(s)\dagger}$ as opposed to $\mathfrak{D}_{\mathrm{S}}^{(s)}$ directly will not be a problem, as number of fermion species represented by Eqn. 8.1 can be halved by using staggered "half" fields, which I will discuss in Sec. 8.1.1.2. The coupling of the fermions to the gauge fields is contained within $\mathfrak{D}_{\mathrm{S}}^{(s)}$, as I have discussed in Sec. 3.2.3 already. Note, also, that $\mathfrak{D}_{\mathrm{S}}^{(s)}$ typically includes smearing; see Appendix G for details. The determinant in Eqn. 8.1 can be represented by pseudofermion fields $\phi(n) \in \mathbb{C}$ as

$$\det\left(\mathfrak{D}_{\mathrm{S}}^{(s)}\mathfrak{D}_{\mathrm{S}}^{(s)\dagger}\right) \propto \int \left[\mathrm{d}\phi^\dagger \mathrm{d}\phi\right] \prod_{n,s} \exp\Big(-\frac{1}{2}\phi^{(s)}(n)^\dagger \left[\mathfrak{D}_{\mathrm{S}}^{(s)}\mathfrak{D}_{\mathrm{S}}^{(s)\dagger}\right]^{-1} \phi^{(s)}(n)\Big), \tag{8.2}$$

which follows from the properties of Gaussian integrals. Combining Eqn. 8.1 with Eqn. 8.2, the lattice-discretized staggered action can be represented in terms of purely bosonic fields as

$$\mathcal{Z} = \int \left[\mathrm{d}\mathcal{U}_\mu\right]\left[\mathrm{d}\phi^\dagger \mathrm{d}\phi\right] \exp\left(\text{-}\mathcal{S}_{\mathrm{lat.}}\left[\mathcal{U}_\mu, \phi, \phi^\dagger\right]\right), \tag{8.3}$$

where

$$\mathcal{S}_{\mathrm{lat.}}\left[\mathcal{U}_\mu, \phi, \phi^\dagger\right] = \mathcal{S}_{\mathrm{G}}\left[\mathcal{U}_\mu\right] + \frac{1}{2}\sum_{n,s} \phi^{(s)}(n)^\dagger \left[\mathfrak{D}_{\mathrm{S}}^{(s)}\mathfrak{D}_{\mathrm{S}}^{(s)\dagger}\right]^{-1} \phi^{(s)}(n) \tag{8.4}$$

The state-of-the-art algorithm for sampling from the joint distribution

$$\Pr\left(\mathcal{U}_\mu, \phi, \phi^\dagger\right) \equiv \mathcal{Z}^{-1} \exp\left(\text{-}\mathcal{S}_{\mathrm{lat.}}\left[\mathcal{U}_\mu, \phi, \phi^\dagger\right]\right) \tag{8.5}$$

since the late 1980s is the *Hamiltonian (hybrid) Monte Carlo* (HMC) algorithm [112].

HMC is a Markov chain Monte Carlo algorithm for sampling from an arbitrary continuous multi-dimensional probability distribution Pr. Each iteration of the HMC algorithm updates the variables of the distribution Pr via discretized molecular dynamics evolution followed by a Metropolis accept/reject test [142, 213, 227, 268]. Suppressing the species index, the field

$$\varphi(n) \equiv \mathfrak{D}_{\mathrm{S}}^{-1}\phi(n) \tag{8.6}$$

is distributed as

$$\mathrm{Pr}\left(\varphi,\varphi^{\dagger}\right) \propto \prod_{n} \exp\left(-\frac{1}{2}\varphi(n)^{\dagger}\varphi(n)\right);$$

in other words, it is Gaussian-distributed. Therefore, sampling from the pseudofermion fields is as simple as drawing samples from $\mathrm{Pr}\left(\varphi,\varphi^{\dagger}\right)$ using an efficient algorithm for sampling from multi-dimensional Gaussian distributions, such as the Box-Muller algorithm [46], and multiplying each sample $\varphi(n)$ by $\mathfrak{D}_{\mathrm{S}}$ to obtain a sample for $\phi(n)$. Once the sample for $\phi(n)$ has been obtained according to the latter procedure, updating the gauge field $\mathcal{U}_\mu$ follows by first generating momenta $\mathfrak{P}_\mu \in \mathfrak{su}(N)$ from the distribution,

$$\log \mathrm{Pr}\big(\mathfrak{P}_\mu\big) \propto \frac{1}{2}\mathrm{Tr}_{\mathrm{c}}\mathfrak{P}_\mu^2, \tag{8.7}$$

then evolving $\mathcal{U}_\mu \to \mathcal{U}_\mu'$ via discrete Hamiltonian evolution (molecular dynamics). The proposed gauge configuration $\mathcal{U}_\mu'$ is then accepted with probability

$$\alpha = \exp\left(-\max\Big[0, \mathcal{H}\big(\mathcal{U}_\mu'\big) - \mathcal{H}\big(\mathcal{U}_\mu\big)\Big]\right), \tag{8.8}$$

where

$$\mathcal{H}\big(\mathcal{U}_\mu\big) = \frac{1}{2}\mathrm{Tr}_{\mathrm{c}}\mathfrak{P}_\mu^2 + \mathcal{S}_{\mathrm{lat.}}\big[\mathcal{U}_\mu, \phi, \phi^{\dagger}\big] \tag{8.9}$$

is the “Hamiltonian” that governs the molecular dynamics evolution. Because the evolution of $\mathcal{U}_\mu$ is discretized, α is not necessarily equal to unity; instead, it is largely governed by a combination of the molecular dynamics trajectory length τ, the number of molecular dynamics time steps N_{MD}, and the order of the molecular dynamics integration algorithm. The optimal acceptance rate α also depends upon the molecular dynamics integration algorithm, which I discuss next.

The HMC algorithm combined with the sampling algorithm for ϕ yields $(\mathcal{U}, \phi, \phi^{\dagger})$ that are distributed according to $\Pr\left(\mathcal{U}_{\mu}, \phi, \phi^{\dagger}\right)$ if the molecular dynamics (MD) algorithm preserves the volume of phase space and the dynamics is time-reversible; in other words, it is *symplectic* [227]. Symplectic MD integration algorithms were explored extensively in Ref. [282]. The latter integrators are often referred to in the lattice gauge theory literature as *Omelyan integrators* as an homage to the lead author of Ref. [282]; note, however, that there are many other symplectic integrators that do not fall under the umbrella of Omelyan integrators. To the best of the author's knowledge, higher-order Omelyan integrators were first explored in the context of lattice gauge theory in Ref. [353]. The lowest-order Omelyan integrator is the famous *leap frog* algorithm, which had been used for decades before higher-order integrators were systematically introduced. All `QEX`-based simulations of this thesis utilize a 2nd-order Omelyan integrator. Further details regarding Omelyan integrators, particularly 2nd-order Omelyan integrators, can be found in Appendix H.

8.1.1.1 Even/odd decomposition

The staggered Dirac operator (Eqn. 3.17) connects only even sites to odd sites (and vice versa). As such, it admits a decomposition into even/odd subspaces as [98, 227]

$$\mathfrak{D}_{\mathrm{S}} = \begin{bmatrix} D_{\mathrm{ee}} & D_{\mathrm{eo}} \\ D_{\mathrm{oe}} & D_{\mathrm{oo}} \end{bmatrix}, \tag{8.10}$$

where

$$D_{\mathrm{ee}} = D_{\mathrm{oo}} = m \text{ and } D \equiv D_{\mathrm{oe}} = -{D_{\mathrm{eo}}}^{\dagger}. \tag{8.11}$$

Furthermore, the expression for $\mathfrak{D}_{\mathrm{S}}$ in Eqn. 8.10 admits a *Schur decomposition* of the form [98, 227]

$$\begin{bmatrix} D_{\mathrm{ee}} & D_{\mathrm{eo}} \\ D_{\mathrm{oe}} & D_{\mathrm{oo}} \end{bmatrix} = \begin{bmatrix} m & D \\ -D^{\dagger} & m \end{bmatrix} = \begin{bmatrix} 1 & D/m \\ 0 & 1 \end{bmatrix} \begin{bmatrix} m + DD^{\dagger}/m & 0 \\ 0 & m \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -D^{\dagger}/m & 1 \end{bmatrix}. \tag{8.12}$$

A corollary to the decomposition of Eqn. 8.12 is

$$\mathfrak{D}_{\mathrm{S}}^{\dagger}\mathfrak{D}_{\mathrm{S}} = \begin{bmatrix} \widehat{D} & 0 \\ 0 & \overline{D} \end{bmatrix}, \tag{8.13}$$

where

$$\widehat{D} \equiv m^2 + DD^{\dagger} \ \ \text{and} \ \ \overline{D} \equiv m^2 + D^{\dagger}D. \tag{8.14}$$

Another corollary of Eqn. 8.12 is that, for any two pseudofermion fields related by $\mathfrak{D}\varphi = \phi$, the following relations hold [227]

$$\widehat{D}\varphi_{\mathrm{e}} = m\phi_{\mathrm{e}} - D\varphi_{\mathrm{o}} \ \ \text{and} \ \ m\varphi_{\mathrm{o}} = \phi_{\mathrm{o}} + D^{\dagger}\varphi_{\mathrm{e}}, \tag{8.15}$$

where

$$\varphi = (\varphi_{\mathrm{e}}, \varphi_{\mathrm{o}}) \text{ and } \phi = (\phi_{\mathrm{e}}, \phi_{\mathrm{o}})$$

is the decomposition of φ and ϕ into their respective even/odd subspaces. The even/odd decomposition, which I stress again is a consequence of having a nearest-neighbor Dirac operator, is extremely useful, as we shall learn in Secs. 8.1.1.2-8.1.1.3. Note also that the $D_{\mathrm{ee}}/D_{\mathrm{oo}}$ components of $\mathfrak{D}_{\mathrm{S}}$ are not always equal to the mass. For example, the Dirac operator for Wilson fermions $\mathfrak{D}_{\mathrm{W}}$ (Eqn. 3.25) admits the decomposition of Eqn. 8.10; however, the $D_{\mathrm{ee}}/D_{\mathrm{oo}}$ components are slightly more complicated due to the explicit presence of the d'Alembertian.

8.1.1.2 Hamiltonian Monte Carlo with staggered half fields

Eqns. 8.13 and 8.15 make staggered fermions, or, really, any fermion formulation with a nearest-neighbor Dirac operator (e.g., Wilson fermions), incredibly nice to work with as far as HMC is concerned. For one, Eqn. 8.12 combined with Eqn. 8.13 implies that setting $\phi_\mathrm{o} = 0$ and integrating over ϕ_e in Eqn. 8.3 yields

$$\int \left[\mathrm{d}\phi_\mathrm{e}^\dagger \mathrm{d}\phi_\mathrm{e}\right] \exp\Big(-\frac{1}{2}\phi^\dagger \Big[\mathfrak{D}_\mathrm{S}\mathfrak{D}_\mathrm{S}^\dagger\Big]^{-1}\phi\Big) \propto \det \widehat{D} = \det \mathfrak{D}_\mathrm{S}. \tag{8.16}$$

In other words, the action of Eqn. 8.4 restricted to just the even sites is equivalent to the action of a single staggered field at the level of expectation values. This does not change if $\mathfrak{D}_\mathrm{S}$ is smeared. The pseudofermion field ϕ restricted to just its even components is referred to as a *half field*. `qex_staghmc` works *exclusively* with half fields.

So that we can better understand the differences between simulating with fermion fields and Pauli-Villars fields, let us consider the force $\mathfrak{F}_\mu^{(\mathrm{ferm.})}(n) \in \mathfrak{su}(N)$ that is derived from the half field action of Eqn. 8.16. From the definition of the link derivative $\partial_{n,\mu}$ in Sec. 4.1, the force takes the form (up to factors of 2)

$$2m\mathfrak{F}_\mu^{(\mathrm{ferm.})}(n) \propto -T^a \Re\mathrm{Tr}_\mathrm{c}\Big[\Big(\partial^a_{n,\mu}\mathfrak{D}_\mathrm{S}\Big)\overline{\varphi}\ \overline{\varphi}^\dagger\Big] \qquad (m \neq 0), \tag{8.17}$$

where

$$\overline{\varphi} \equiv \mathfrak{D}_\mathrm{S}^{-1}\phi \tag{8.18}$$

for $\phi = (\phi_\mathrm{e}, 0)$ and T^a are the conventional $\mathfrak{su}(N)$ generators defined in Sec. 4.1. The components of $\overline{\varphi}$ are obtained from the even/odd reconstruction of Eqn. 8.15. For massless fields, no reconstruction is needed and the force takes the equivalent form

$$2\mathfrak{F}_\mu^{(\mathrm{ferm.})}(n) \propto -T^a \Re\mathrm{Tr}_\mathrm{c}\Big[\Big(\partial^a_{n,\mu}\mathfrak{D}_\mathrm{S}\Big)\widehat{\varphi}\ \widehat{\varphi}^\dagger\Big] \qquad (m = 0), \tag{8.19}$$

where

$$\widehat{\varphi} \equiv \begin{bmatrix} \overline{\varphi}_\mathrm{e} \\ -D^\dagger\overline{\varphi}_\mathrm{e} \end{bmatrix} \tag{8.20}$$

with $\overline{\varphi}_\mathrm{e}$ the even component of $\overline{\varphi} = (\overline{\varphi}_\mathrm{e}, \overline{\varphi}_\mathrm{o})$ defined in Eqn. 8.18. In `qex_staghmc`, the fermion force for any massive field is calculated using Eqn. 8.17; for massless fields, it is calculated with Eqn. 8.19. In both cases, it is necessary to invert $\widehat{D}$ to obtain $\overline{\varphi}$. Because $\widehat{D}$ is positive definite, this is done using the conjugate gradient algorithm [199]. More information about conjugate gradient can be found in Appendix I. The reader may be concerned that I am inverting the Dirac operator for two half fields when I could be inverting for one full field ($N_{\mathrm{f}} \geq 8$); however, half fields are typically half as expensive as full fields. Therefore, the computational cost of simulating two half fields is typically as expensive as simulating one full field. Note, also, that the molecular dynamics evolution of $\mathcal{U}_\mu$ keeps the original φ field (Eqn. 8.6) fixed.

8.1.1.3 Including Pauli-Villars fields

In some very real sense, Pauil-Villars fields are the opposite of pseudofermion fields. As far as HMC is concerned, for any operation applied to the pseudofermions ϕ, the inverse operation is applied to the PV fields $\upphi$. The action for the PV half fields is

$$\mathcal{S}_{\mathrm{PV}}\left[\upphi, \upphi^\dagger\right] = \frac{1}{2}\upphi^\dagger \mathfrak{D}_{\mathrm{S}}^\dagger \mathfrak{D}_{\mathrm{S}} \upphi, \tag{8.21}$$

where $\upphi = (\upphi_{\mathrm{e}}, 0)$. As was the case for the pseudofermion fields,

$$\int \left[\mathrm{d}\upphi_{\mathrm{e}} \mathrm{d}\upphi_{\mathrm{e}}^\dagger\right] \exp\left(\text{-}\mathcal{S}_{\mathrm{PV}}\left[\upphi, \upphi^\dagger\right]\right) \propto \det \widehat{D}^{-1} = \det \mathfrak{D}_{\mathrm{S}}^{-1}. \tag{8.22}$$

The $\upvarphi(n) \equiv \mathfrak{D}_{\mathrm{S}}\upphi(n)$ fields are distributed as

$$\Pr\left(\upvarphi, \upvarphi^\dagger\right) \propto \prod_n \exp\left(-\frac{1}{2}\upvarphi(n)^\dagger \upvarphi(n)\right); \tag{8.23}$$

therefore, sampling the PV fields $\upphi$ is as simple as drawing $\upvarphi$ from a multi-dimensional Gaussian distribution and inverting $\mathfrak{D}_{\mathrm{S}}$ to obtain a sample of $\upphi$.

		2 nodes		4 nodes		8 nodes	
$(L/a)^4$	β	hr.	cr.-hr.	hr.	cr.-hr.	hr.	cr.-hr.
32^4	8.8	1.67	106.9	0.94	120.3	0.48	122.9
32^4	9.4	0.39	25.0	0.20	25.6	0.11	28.2
40^4	9.4	1.21	96.8	0.60	96.0	0.32	102.4

Table 8.1: Timing (in hours and core-hours) for a single unit molecular dynamics trajectory on two volumes $(L/a)^4 = 32^4$ $(\beta = 8.8, 9.4)$ and 40^4 $(\beta = 9.4)$ on 2, 4 and 8 Fermilab Cascade Lake LQ1 nodes.

For updating the gauge fields, the force $\mathfrak{F}^{(\mathrm{PV})}_{\mu}(n)$ from the Pauli-Villars fields is of the form

$$2\mathfrak{F}^{(\mathrm{PV})}_{\mu}(n) \propto T^a \Re \mathrm{Tr}_{\mathrm{c}}\left[\left(\partial^a_{n,\mu}\mathfrak{D}_{\mathrm{S}}\right)\widehat{\phi}\ \widehat{\phi}^{\dagger}\right], \tag{8.24}$$

where

$$\widehat{\phi} \equiv \begin{bmatrix} \phi_{\mathrm{e}} \\ -D^{\dagger}\phi_{\mathrm{e}} \end{bmatrix}. \tag{8.25}$$

That is all there is to it. The addition of PV fields incurs a negligible additional cost to the HMC simulation, as the Dirac operator only needs to be inverted when obtaining ϕ from the distribution of $\mathfrak{D}_{\mathrm{S}}\phi$; this pales in comparison to the cost of inverting $\mathfrak{D}_{\mathrm{S}}$ O(10-100) times to calculate the fermion force along the MD trajectory. In fact, PV fields have been observed empirically to improve the condition number of the Dirac operator quite dramatically; as such, the cost of including PV fields could be offset dramatically by the effect that they have on reducing the number of conjugate gradient iterations needed to invert $\mathfrak{D}_{\mathrm{S}}$ on the fermion fields [188]. For readers familiar with Hasenbusch preconditioning, this is perhaps not too surprising [163, 165]. In fact, the first implementation of PV improvement in `KS_nHYP_FS` simply modified the Hasenbusch preconditioning that was present in `MILC` at the time.

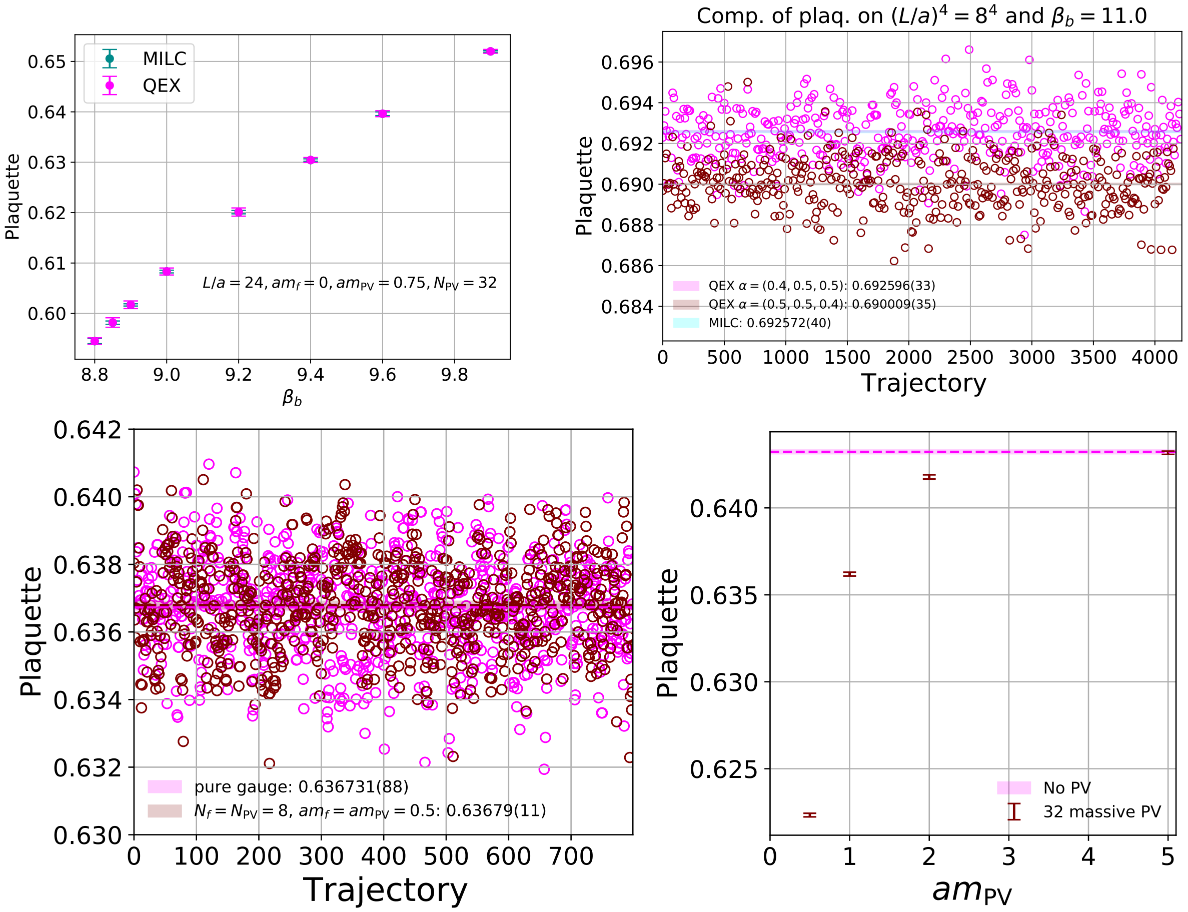

Figure 8.1: (Top left panel) Comparison of plaquette from `qex_staghmc` (teal) against `KS_nHYP_FS` (`QEX`, magenta) for $(am_f, am_{\rm PV}) = (0.0, 0.75)$ with $(L/a, N_{\rm PV}) = (24, 32)$ and $8.8 \leq \beta_b \leq 9.9$. (Top right panel) comparison of `qex_staghmc` with smearing parameters $\alpha = (0.4, 0.5, 0.5)$ (magenta) against `qex_staghmc` with smearing parameters $\alpha = (0.5, 0.5, 0.4)$ (maroon) and `KS_nHYP_FS` (`MILC`, cyan) for $(L/a, am_f, am_{\rm PV}, N_{\rm PV}) = (8, 0.0, 0.75, 32)$ and $\beta_b = 11.0$. (Bottom left panel) Comparison of pure gauge against $(L/a, N_{\rm f}, N_{\rm PV}, am_f, am_{\rm PV}) = (8, 8, 8, 0.5, 0.5)$, both using `qex_staghmc`. (Bottom right panel) Comparison of $(N_{\rm PV}, am_f) = (0, 0.0)$ against $(N_{\rm PV}, am_f) = (32, 0.0)$ with $am_{\rm PV} \in \{0.5, 1.0, 2.0, 5.0\}$ and $(L/a, N_f) = (8, 8)$, both using `qex_staghmc`.

8.1.1.4 Simple tests

Fig. 8.1 illustrates simple tests verifying some expected behaviors of the PV-improved HMC simulation performed by `qex_staghmc` using the plaquette (Eqn. 3.4). These tests are *not* exhaustive by any means and the plaquette can only reveal so much; `qex_staghmc` has undergone much more testing than what is shown in Fig. 8.1. Perhaps the most important test is to check that `qex_staghmc`

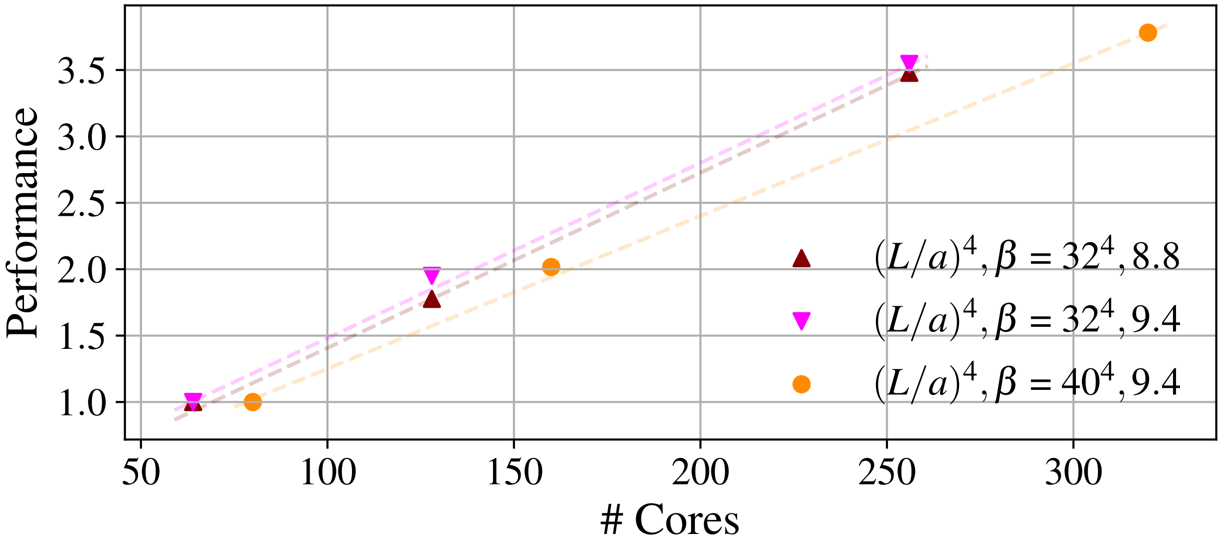

Figure 8.2: Strong scaling plot `qex_staghmc`'s performance (normalized to the first data point) as a function of the number of computing cores on Fermilab's Cascade Lake cluster (`LQ1`).

reproduces PV-improved `KS_nHYP_FS` simulations. The top left panel of Fig. 8.1 compares the plaquette from a simulation with the action of Chapter 7 on $L/a = 24$ for $8.8 \leq \beta_b \leq 9.9$ produced by `qex_staghmc` ("QEX", magenta) and `KS_nHYP_FS` ("MILC", dark cyan). The central value for the plaquette from both codes is consistent within their combined statistical precision. Additionally, the order in which the nHYP smearing parameters $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ are specified is reversed in `qex_staghmc` relative to `KS_nHYP_FS` (see Appendix G). The top right panel of Fig. 8.1 verifies that the reversed ordering of α in `qex_staghmc` ("`QEX`", magenta) reproduces the plaquette from the regular ordering of α in `KS_nHYP_FS` ("`MILC`", cyan) on $(L/a, \beta_b) = (8, 11.0)$.

According to Eqn. 5.2, the effective action induced by *heavy* PV fields should exhibit the following easily-verifiable properties.

- If $N_\text{f} = N_\text{PV}$ and $am_f = am_\text{PV}$, then expectation values of observables drawn from the joint distribution $\Pr\left(\mathcal{U}_\mu, \phi, \phi^\dagger, \phi, \phi^\dagger\right)$ should agree with observables generated from the pure Yang-Mills distribution $\Pr\left(\mathcal{U}_\mu\right)$. In other words, the ensembled-averaged effective action induced by the pseudofermion fields should cancel out the ensemble-averaged effective action induced by the PV fields. What's left is a pure Yang-Mills system.

- If $am_{\text{PV}} \gg 1$, then observables drawn from the joint distribution $\Pr\left(\mathcal{U}_\mu, \phi, \phi^\dagger, \upphi, \upphi^\dagger\right)$ should agree with observables generated from the joint distribution $\Pr\left(\mathcal{U}_\mu, \phi, \phi^\dagger\right)$ up to finite-am_{PV} effects. In other words, the PV fields *decouple* in the $am_{\text{PV}} \to \infty$ limit.

The ensemble-averaged cancellation of the pseudofermion/PV effective actions is illustrated in the bottom left panel of Fig. 8.1 for $N_{\text{f}} = N_{\text{PV}} = 8$ and $am_f = am_{\text{PV}} = 0.5$ $(L/a = 8)$. The plaquette, 0.636731(88), from 800 samples of $\Pr\left(\mathcal{U}_\mu\right)$ is consistent with the plaquette, 0.63679(11), from the same number of samples from $\Pr\left(\mathcal{U}_\mu, \phi, \phi^\dagger, \upphi, \upphi^\dagger\right)$, verifying the latter expectation. The decoupling of the PV fields in the $am_{\text{PV}} \to \infty$ limit is illustrated in the bottom right panel of Fig. 8.1, which compares the expectation value of a $(L/a, \beta_b) = (8, 9.0)$ simulation with $(N_{\text{f}}, am_f) = (8, 0.0)$ and no PV fields ("No PV", magenta line) against a series of $(L/a, \beta_b) = (8, 9.0)$ with $(N_{\text{f}}, am_{\text{f}}) = (8, 0.0)$ simulations and $N_{\text{PV}} = 32$ with $am_{\text{PV}} = 0.5, 1.0, 2.0,$ and 5.0. Indeed, the plaquette for simulations with increasing am_{PV} are observed to converge upon the simulation with no PV fields, as expected from decoupling.

Table 8.1 summarizes the result of timing tests on 2, 4 and 8 nodes of Fermilab's Cascade Lake cluster (48 cores/node) from a single HMC trajectory on $(L/a, \beta_b) = (32, 8.8), (32, 9.4)$, and $(40, 9.4)$ using the action of Chapter 7. In the context of high-performance computing, parallel software is said to exhibit *strong scaling* if its performance is linear in the number of cores at fixed "problem size" (in our case, fixed $L/a, \beta_b$). Fig. 8.2 shows the result of a strong scaling test of `qex_staghmc` using the information from Table 8.1. The "performance" in Fig. 8.2 is the time taken to complete a single MD trajectory, normalized by the performance of the lowest core count at each fixed $L/a, \beta_b$; moreover, the simulations in Fig. 8.2 only utilizes task- and SIMD-based parallelism (i.e., no multithreading). The vector length for SIMD-based parallelism is set to 8 by default. All three test cases exhibit a reasonable degree of linearity of the performance in the number of cores; hence, `qex_staghmc` exhibits favorable properties under strong scaling, at least without multithreading, which deserves its own scrutiny.

8.1.1.5 Performing Hamiltonian Monte Carlo simulations with `qex_staghmc`

I designed `qex_staghmc` around the solid foundation of `QEX` with ease-of-use and portability in mind, even if doing so incurred minor penalties to performance. I wanted `qex_staghmc` to be both flexible and performant, so as to address the needs of researchers working on small-medium (and some large) scale lattice calculations that tackle problems in field theory and beyond Standard Model physics. This is opposed to the large-scale collaborative efforts targeting precision QCD observables, which require every bit of optimization that they can get, even if at the cost of making software difficult to understand/modify by non-professionals. The massive amount of effort that the developers of `QEX` have poured into it makes this possible, and their example `QEX`-based HMC codes made the task of designing `qex_staghmc` much less painful than it could have been.

Compiling `qex_staghmc` is simple. After building/installing `QEX` according to the information provided in the main `Git` branch of `QEX` (see Footnote 2), compilation proceeds by running

```
make staghmc_spv :FUELCompat=1
```

under the build directory that is created as part of build/installation process. The `:FUELCompat=1` flag ensures backwards compatibility with the `FUEL` (`QHMC`) code.[5] Specifying `:FUELCompat=1` is recommended, as all `qex_staghmc` features that have been tested from binaries that have been compiled with the `FUEL` compatibility flag. If one wishes to change the number of colors N, simply add the `:nc` flag to the `make` command; e.g., `:nc=2` for SU(2). As much of `QEX` is N-agnostic, the HMC code of `qex_staghmc` is as well (including the nHYP smearing).

Running the `qex_staghmc` HMC binary requires only an `XML` file, which is read in by `qex_staghmc` to determine important properties of the HMC simulations, such as the number of MD integration steps for each field (gauge, fermion, and PV); the MD trajectory length; the Omelyan integrator for each field (gauge, fermion, and PV); the gauge action (Wilson, Lüscher-Weiss, or adjoint-plaquette); the number of fermions and their masses; the number of PV fields and their masses; the lattice geometry (L/a for any direction in any number of dimensions); the

---

[5] `QHMC` is available at https://github.com/jcosborn/qhmc.

boundary conditions for the fermion/PV fields (periodic or anti-periodic in any direction); the nHYP searing parameters; and much more, such as the observables to be measured (and the measurement frequency), along with any tests that should be performed (such as reversibility tests). A sample `XML` file can be found under `src/stagg_pv_hmc/input_hmc.xml` in the `Git` development branch of `qex_staghmc` (see Footnote 4). To run the `qex_staghmc` HMC binary, simply execute

```
<parallel_executor> <path_to_binary> -start_config=<start_config>
-end_config=<end_config> -config_space=<config_space> -save_freq=<save_freq>
-xml=<path_to_xml> -rank_geom=<rank_geom> -filename=<filename_convention>
-path=<path_to_read_write>
```

where `<parallel_executor>`=“`mpiexec -n <n_tasks> -bind-to user:0,1`”, “`srun -mpi=pmi2`”, etc.; `<path_to_binary>` is the global path to the `staghmc_spv` binary; `<start_config>` is the starting configuration number and `<end_config>` is the ending configuration number; `<config_space>` is the number of full HMC iterations between configurations; `<save_freq>` is the frequency (in configuration number) at which configurations are to be saved; `<path_to_xml>` is the path to the aforementioned configuration file; `<rank_geom>` is the geometry of the MPI ranks; `<filename_convention>` is the default tag for naming configuration files; and `<path_to_read_write>` is the directory that `qex_staghmc` is directed to read/write from. The files that are read/written from disk are the gauge configurations (`<filename_convention>_<configuration_number>.lat`), parallel random number generator files (`<filename_convention>_<configuration_number>.rng`), and serial random number generator files (`<filename_convention>_<configuration_number>.global_rng`). With all three files, it is possible to easily checkpoint the HMC and regenerate each configuration for measurements, up to ordering differences in parallel reductions.

8.1.2 Gradient flow

`qex_staghmc` *lightly* extends the gradient flow code already present in `QEX` to allow for integrating flow equations with arbitrary Lüscher-Weiss gauge actions (see Chapter 7.4 or Appendix F) and ease-of-deployment in production running. Numerical integration of the gradient flow equation (Eqn. 4.2) is performed using the 5th-order Runge-Kutta (RK5) scheme suggested in Ref. [252]. In brief, the gradient flow equation of Eqn. 7.4 is of the form

$$\frac{\mathrm{d}\mathcal{U}_t}{\mathrm{d}t} = Z(\mathcal{U}_t)\mathcal{U}_t, \tag{8.26}$$

where $\mathcal{U}_t \in \mathrm{SU}(N)$ is a generic gauge variable. Define the integration step $\epsilon = \delta t/a^2$. Then, following Ref. [252], the RK5 integration proceeds recursively as

$$\begin{aligned} W_0 &= \mathcal{U}_t, \\ W_1 &= \exp\left[\frac{1}{4}Z_0\right]W_0, \\ W_2 &= \exp\left[\frac{8}{9}Z_1 - \frac{17}{36}Z_0\right]W_1, \\ \mathcal{U}_{t+\epsilon} &= \exp\left[\frac{3}{4}Z_2 - \frac{8}{9}Z_1 + \frac{17}{36}Z_0\right]W_2, \end{aligned} \tag{8.27}$$

where

$$Z_i = \epsilon Z(W_i) \qquad (i = 0, 1, 2). \tag{8.28}$$

The RK5 scheme proposed by Eqns. 8.1.2-8.28 generally works well for $0.05 \lesssim \epsilon \lesssim 0.1$ at early flow times. It is possible to achieve a considerable speed up by switching to a larger ϵ at late flow times, where the transient effects are suppressed. In the future, I wish to test out the geometric integration scheme proposed in Ref. [34]. I also intend to implement "Zeuthen flow" [305], which amounts choosing $Z(\mathcal{U}_\mu)$ in Eqn. 8.26 as

$$Z(\mathcal{U}_\mu) = -g_0^2\left(1 + \frac{a^2}{12}\partial_\mu^*\partial_\mu\right)\partial_{x,\mu}\mathcal{S}_{\mathrm{LW}}\left[\mathcal{U}_\mu\right] \tag{8.29}$$

where ∂_μ and ∂_μ^* have are defined in Eqn. 3.27 and Eqn. 3.28, respectively, and $\mathcal{S}_{\mathrm{LW}}$ is the tree-level-improved Lüscher-Weiss gauge action (see Appendix F). Zeuthen flow was used in Chapter 4 with great success.

8.1.2.1 Running gradient flow with `qex_staghmc`

Compiling `qex_staghmc`'s gradient follows as in Sec. 8.1.1.5. After installing/building `QEX`, simply execute

```
make gauge_flow :FUELCompat=1
```

in the build directory created by the build/installation process. Running the gauge flow also requires specifying an `XML` file, which tells `qex_staghmc` which c_{p} to use; the value of ϵ to be used over any arbitrary range of flow times; the lattice geometry; and various specifications for measuring gradient flow observables, such as $E(t)$ ($\mathcal{S}^{\mathrm{e}} = \mathrm{W}, \mathrm{C}$), the Polyakov loop (Eqn. 4.32), and the topological charge (Eqn. 4.35). A sample `XML` file can be found under `src/flow/input_gf.xml` in the `Git` development branch of `qex_staghmc` (see Footnote 4). Executing the gauge flow is as simple as running

```
<parallel_executor> <path_to_binary> -start_config=<start_config>
-end_config=<end_config> -xml=<path_to_xml> -filename=<filename_convention>
-rank_geom=<rank_geom>
```

where `<parallel_executor>`="`mpiexec -n <n_tasks> -bind-to user:0,1`", "`srun -mpi=pmi2`", etc.; `<path_to_binary` is the global path to the `gauge_flow` binary; `<start_config>` is the starting configuration number and `<end_config>` is the ending configuration number; `<path_to_xml>` is the global path to the aforementioned `XML` file; `<rank_geom>` is the geometry of the MPI ranks; and `<filename_convention>` is the name convention for the gauge configurations to be flowed (e.g., `<filename_convention>_<configuration_number>.lat`).

8.2 The `SwissFit` library

For the neural-network-based fits of Chapters 6-7, I needed both an efficient, but not overly-optimized, implementation of feedfoward and radial basis function neural networks that I could easily modify. I also needed the freedom to choose the optimization algorithm for MAP estimation (see

Appendix D) freely while retaining the ability to utilize the powerful tools of the `gvar` library for propagating correlated uncertainties of MAP estimates Θ^* for model parameters into both derived quantities and the underlying dataset [241]. Preferably, the neural networks would also support `gvar` data types. `SwissFit` was developed to meet these needs; however, over time it morphed into the early workings of a fully-fledged scientific library written in `Python`. In this section, I will describe the numerical details of features that `SwissFit` currently supports, though some minor details are subject to change in future versions of `SwissFit` [290].[6]

8.2.1 Maximum a posteriori estimation by nonlinear least squares

In Appendix D.2, I introduce maximum a posteriori estimation (MAP) from the perspective of Bayesian statistics and information theory. In short, the most probable *constrained* estimate for the parameters Θ^* of a model $\mathcal{M}$ that describes a dataset $\boldsymbol{X} \equiv (x_0, ..., x_{M\text{-}1})$ with mean $\langle X \rangle \equiv \overline{X}$ and covariance of the mean $\Sigma_{\overline{X}}$ is given by the maximum a posteriori estimate (MAP)

$$\Theta^* \equiv \arg\min_{\Theta} \chi^2_{\text{aug.}}(\Theta),$$

where

$$\chi^2_{\text{aug.}}(\Theta) \equiv \chi^2_{\text{data}}(\Theta) + \chi^2_{\text{prior}}(\Theta)$$

is the *augmented* χ^2 with

$$\chi^2_{\text{data}}(\Theta) \equiv \Big(\boldsymbol{\mathcal{M}}(\Theta) - \overline{\mathbf{X}}\Big)^{\mathrm{T}} \Sigma^{-1}_{\overline{X}} \Big(\boldsymbol{\mathcal{M}}(\Theta) - \overline{\mathbf{X}}\Big)$$

defined by Eqn. D.26 and

$$\chi^2_{\text{prior}}(\Theta) \equiv \Big(\boldsymbol{\mathcal{C}}(\Theta) - \overline{\boldsymbol{\mathcal{C}}}\Big)^{\mathrm{T}} \Sigma^{-1}_{\overline{\mathcal{C}}} \Big(\boldsymbol{\mathcal{C}}(\Theta) - \overline{\boldsymbol{\mathcal{C}}}\Big)$$

---

[6] `SwissFit` is available at `https://github.com/ctpeterson/SwissFit`. Future versions of `SwissFit` will be available for installation via `pip` through `PyPI` [304].

define by Eqn. D.27. See Appendix D.2 for a discussion of individual terms in χ^2_{data} and χ^2_{prior}. For the sake of notational simplicity, define

$$\Sigma \equiv \Sigma_{\overline{X}} \oplus \Sigma_{\overline{\mathcal{C}}} \tag{8.30}$$

and

$$\boldsymbol{\delta}(\Theta) \equiv \boldsymbol{\delta}_{\text{data}}(\Theta) \oplus \boldsymbol{\delta}_{\text{prior}}(\Theta), \tag{8.31}$$

where

$$\boldsymbol{\delta}_{\text{data}}(\Theta) \equiv \boldsymbol{\mathcal{M}}(\Theta) - \overline{\mathbf{X}} \tag{8.32}$$

and

$$\boldsymbol{\delta}_{\text{prior}}(\Theta) \equiv \boldsymbol{\mathcal{C}}(\Theta) - \overline{\boldsymbol{\mathcal{C}}}. \tag{8.33}$$

In terms of the notation of Eqns. 8.30-8.33, the augmented χ^2 reads

$$\chi^2_{\text{aug.}}(\Theta) = \boldsymbol{\delta}(\Theta)^{\text{T}} \Sigma^{-1} \boldsymbol{\delta}(\Theta)$$

and the leading-order statistical uncertainty in Θ^* from Eqn. D.24 is

$$\left[\Sigma_{\Theta^*}\right]_{ij} \equiv \frac{\partial \boldsymbol{\delta}(\Theta)^{\text{T}}}{\partial \Theta_i} \Sigma^{-1} \frac{\partial \boldsymbol{\delta}(\Theta)}{\partial \Theta_j}\bigg|_{\Theta=\Theta^*}$$

if Σ is independent of Θ. Note that Σ could depend on Θ in the case of hierarchical Bayesian modelling; as `SwissFit` does not currently fully support hierarchical Bayesian modelling, such cases are not considered in this section.

Before optimizing $\chi^2_{\text{aug.}}(\Theta)$, it is both numerically advantageous and computationally cost effective to calculate the inverse square root $\Sigma^{-1/2}$ of the covariance Σ. Because Σ is positive-definite, $\Sigma^{-1/2}$ is well-defined in terms of the singular value decomposition (SVD) of Σ

$$\Sigma = U^{\text{T}} S U \tag{8.34}$$

as

$$\Sigma^{-1/2} = S^{-1/2}U, \tag{8.35}$$

where S is a diagonal matrix of singular values and U is an orthogonal matrix. The problem of optimizing $\chi^2_{\text{aug.}}(\Theta)$ is then recast completely in terms of

$$\tilde{\boldsymbol{\delta}}(\Theta) \equiv \Sigma^{-1/2}\boldsymbol{\delta}_{\text{prior}}(\Theta), \tag{8.36}$$

whereby

$$\chi^2_{\text{aug.}}(\Theta) = \tilde{\boldsymbol{\delta}}(\Theta)^{\mathrm{T}}\tilde{\boldsymbol{\delta}}(\Theta) \tag{8.37}$$

and

$$\left[\Sigma^{-1}_{\Theta^*}\right]_{ij} = \frac{\partial\tilde{\boldsymbol{\delta}}(\Theta)^{\mathrm{T}}}{\partial\Theta_i}\frac{\partial\tilde{\boldsymbol{\delta}}(\Theta)}{\partial\Theta_j}\bigg|_{\Theta=\Theta^*}. \tag{8.38}$$

`SwissFit` utilizes the SVD algorithm implemented in `NumPy` to calculate $\Sigma^{-1/2}$ [162]. `NumPy` itself utilizes the `_gesvd` routine provided by `LAPACK`, which implements a generalized (quotient) singular value decomposition algorithm [5]. On the other hand, the parameter covariance Σ_{Θ^*} of Eqn. 8.38 is calculated from the Moore-Penrose pseudoinverse provided by `SciPy` [361], which is little more than a repackaged SVD decomposition [289].

`SwissFit` supports the trust region reflective [49], dogbox [362], and Levenberg-Marquardt [242, 259] local optimization algorithms and the basin hopping global optimization algorithm [364], all of which are provided by the `SciPy` library [361]. Gradients are calculated using the automatic differentiation tools provided by the `gvar` library [241]; hence, all that the user needs to provide `SwissFit` is the model function in the form of a `Python` function that accepts `gvar` data types. Most standard `NumPy` functions are compatible with `gvar` data types. `gvar` itself contains a variety of wrapper functions for most standard functions. `gvar` also provides infrastructure for creating custom `gvar`-compatible functions, so long as the derivative of the desired function is directly accessible (analytically or numerically) [241].

8.2.2 Quality of fit and model selection criteria

From the MAP estimate Θ^*, `SwissFit` calculates a number of standard measures for the quality of fit. Directly accessible from the fit is the "Bayesian" $\chi^2_{\text{aug.}}/\text{d.o.f.}$ with

$$\text{d.o.f.} \equiv |X| - |\Theta| + |\mathcal{C}| \tag{8.39}$$

and the "frequentist" $\chi^2_{\text{data}}/\text{d.o.f.}^*$ with

$$\text{d.o.f.}^* \equiv |X| - |\Theta|, \tag{8.40}$$

where $|X|$ is the number of data points, $|\Theta|$ is the number of model parameters, and $|\mathcal{C}|$ is the number of priors (constraints). The standard criterion for a "good" fit is

$$\chi^2_{\text{aug.}}/\text{d.o.f.} \sim 1 \text{ and } \chi^2_{\text{data}}/\text{d.o.f.}^* \sim 1. \tag{8.41}$$

A χ^2 per the respective degrees of freedom that is much less than unity suggests either that the model overfits the data or the uncertainty from Σ is much greater than the actual fluctuations in the data (i.e., the uncertainties are overestimated). On the other hand, if the χ^2 per the respective degrees of freedom is much greater than unity, then the model either overfits or the uncertainty from Σ is much smaller than the actual fluctuations in the data (i.e., the uncertainties are underestimated). The Bayesian p-value

$$p\text{-value} = \frac{1}{\Gamma(\text{d.o.f.}/2)} \int_{\chi^2_{\text{aug.}}/2}^{\infty} \mathrm{d}t \; t^{\frac{\text{d.o.f.}}{2}-1} e^{-t} \qquad \text{(Bayesian)} \tag{8.42}$$

is the probability that $\chi^2_{\text{aug.}}$ could have by larger by chance and is easily calculated from `SciPy`'s implementation of the regularized upper incomplete gamma function [361]. A "good" fit is conventionally considered to have a p-value that is $\gtrsim 10\%$, assuming that the statistical error is not overestimated. The corresponding “frequentist” p-value is

$$p\text{-value} = \frac{1}{\Gamma(\text{d.o.f.}^*/2)} \int_{\chi^2_{\text{data}}/2}^{\infty} \mathrm{d}t\; t^{\frac{\text{d.o.f.}^*}{2}-1} e^{-t} \qquad \text{(frequentist)}. \tag{8.43}$$

The interpretation of the frequentist p-value is slightly more complicated in the presence of priors. Without priors, it of course equivalent to the Bayesian p-value. `SwissFit` calculates both p-values.

The marginal likelihood (Eqn. D.16) and Akaike information criterion (Eqn. D.33) are useful for the purposes of model selection. Given a finite collection of candidate models $M \equiv \{\mathcal{M}^{(1)}, \mathcal{M}^{(2)}, ..., \mathcal{M}^{(N_{\text{M}})}\}$, the model $\mathcal{M}^{(\eta)} \in M$ with the highest marginal likelihood $\Pr\left(X \middle| \mathcal{M}^{(\eta)}\right)$ is more likely to have given rise to the data X than any other model $\mathcal{M}^{(\eta')} \in M$. The Laplace-approximated marginal likelihood (Eqn. 6.12) about the MAP estimate Θ^* is

$$-2\log\Pr\left(X \middle| \mathcal{M}^{(\eta)}\right) \approx \chi^2_{\text{aug.}}(\Theta^*) + \log\frac{(2\pi)^{\text{d.o.f.}} \det\Sigma_{\overline{X}} \det\Sigma_{\overline{C}}}{\det\Sigma_{\Theta^*}}.$$

It should be noted that the Laplace-approximated marginal likelihood is a point estimate. As such, it could be a very poor estimate of the actual marginal likelihood, especially if the distribution possesses many (possibly degenerate) peaks. In any case, the full marginal likelihood can be used for the purposes of model selection if it is known that the *correct model* (i.e., the model that X arises from) exists in M. As I have discussed in Chapter 6.2.3, the marginal likelihood is also useful for inferring priors from data via the *empirical Bayes* procedure. Empirical Bayes is useful both for estimating priors in the face of uncertainty and controlling for overfitting. If the correct model is not in M and one simply wishes to select the model that best describes the data with the least number of fit parameters, then the marginal likelihood is not appropriate. Instead, one should use a quantity that weights fit quality against information loss, such as the Akaike information criterion

$$\text{AIC} = \chi^2_{\text{aug.}}\left(\Theta^*\right) + 2|\Theta|. \tag{8.44}$$

As is briefly discussed in Appendix D.2.4 and derived explicitly in Ref. [210], the AIC arises naturally from the *posterior probability* $\Pr(X|\mathcal{M})$, which is related to the marginal likelihood by Bayes' theorem (Eqn. D.17). The model with the lowest AIC fits the data best with the least number of parameters (information loss); in other words, it is preferred by Occam's razor. Both the AIC and the Laplace-approximated marginal likelihood are estimated by `SwissFit`.

8.2.3 Consistent uncertainty propagation

Ensuring that all uncertainties are accounted for in algebraic operatations of Gaussian random variables with Θ^* using `gvar` is a surprisingly difficult task. Luckily, the `lsqfit` library has implemented an ingenious solution that `SwissFit` utilizes to achieve the same goal [240]. The idea is to utilize the estimate [44]

$$\nabla_{\boldsymbol{X}}\boldsymbol{\Theta}^* = \Sigma_{\Theta^*}\left[\nabla_{\boldsymbol{\Theta}}\tilde{\boldsymbol{\delta}}(\Theta)\right]^{\mathrm{T}}\Sigma^{-1/2}\Big|_{\Theta=\Theta^*} \tag{8.45}$$

for the derivative $\nabla_{\boldsymbol{X}}\boldsymbol{\Theta}^*$ to both propagate the uncertainty of Θ^* into derived quantities and automatically account for correlations that any Gaussian random variable has with Θ^* through X; see Appendix D.1.3 for a brief discussion of Gaussian error propagation. For example, `gvar` calculates the uncertainty in the scaling function of Chapters 6-7 using Eqn. 8.45. Similarly, Eqn. 8.45 also allows `gvar` to account for the correlation between γ_O and $O(K, N_{\mathrm{s}})$ in estimating the uncertainty of $O(K, N_{\mathrm{s}})/N_{\mathrm{s}}{}^{\gamma_O}$ in the curve collapse of magnetic susceptibilities in Chapter 6.

8.2.4 Incorporating neural networks

`SwissFit` provides utilities from implementing `gvar`-compatible radial basis function networks (RBFN, Chapters 6-7) and feedforward neural networks (FNN) in fit functions through the `RadialBasisNeuralNetwork` and `FeedforwardNeuralNetwork` class, respectively. Both classes inherit methods from the `NeuralNetwork` class. The topology and activation function for each layer of either neural network is specified as a `Python` dictionary when instantiating a child of the `NeuralNetwork` class. Exponential, multi-quadratic, inverse-quadratic, and multi-inverse-quadratic activations are supported for RBFN layers, along with both L1 and L2 norms for the argument of the RBFs of each layer. The `FeedforwardNeuralNetwork` class supports linear, rectified linear unit (ReLU), Gaussian error linear unit (GELU), exponential linear unit (ELU), hyperbolic tanh, and sigmoidal activations, along with user-provided custom activations [84, 197]. Examples are provided under the `examples` folder in the development branch of `SwissFit` (see Footnote 6).

8.3 Non-perturbative running anomalous dimensions in the ten flavor system

In Chapter 2.4, I discussed running operator anomalous dimensions in the context of Wilsonian RG. In Chapter 1, I briefly touched upon the relevance that the fixed point value of running operator anomalous dimensions has for beyond Standard Model physics phenomenology. In this section, I am going to apply the gradient-flow-based methods of Chapters 4, 5, and 7 to calculate the running pseudoscalar, tensor, and proton anomalous dimension of the $N_{\mathrm{f}} = 10$ system over $2.0 \lesssim g^2_{\mathrm{GF}} \lesssim 8.5$ from the running of the wave function renormalization (Eqn. 2.24) of all three operators. If the $N_{\mathrm{f}} = 10$ system indeed lies within the conformal window, then the recent CBFM-based prediction for the fixed point GF coupling $g^2_{\mathrm{GF}\star} \approx 15.0$ of Ref. [179] unfortunately indicates that the range of couplings covered in this section is too small to be able to estimate any of the fixed point operator anomalous dimensions for the $N_{\mathrm{f}} = 10$ system. Similar calculations for the $N_{\mathrm{f}} = 2$ system are currently underway by collaborations that I am not currently involved in; see, for example, the work of Ref. [177]. See also Ref. [74] for an early development of similar ideas.

8.3.1 Including fermions in gradient flow

In Chapters 4, 5, and 7, only the gradient flow of the gauge field was needed. Correlation functions for fermionic observables are required to extract the running anomalous dimension of the pseudoscalar, tensor and proton; hence, the gradient flow equation must be extended to include the fermion fields. Including fermions in the gradient flow equations turns out to be fairly simple and is laid out clearly in Ref. [249]. In short, the evolution equation for the gauge field is left untouched

(Eqn. 4.2), while the lattice fermion field ψ is evolved according to

$$\frac{\mathrm{d}\psi(n,t)}{\mathrm{d}t} = \sum_{\mu} \Box_{\mu}\psi(n,t), \tag{8.46}$$

where the box operator $\Box_{\mu}$ has been defined in Eqn. 3.23 and is coupled to the flowed gauge field $\mathcal{U}_{\mu}(n,t)$ as in Chapter 3.2.3. The gradient flow is them performed by evolving Eqns. 4.2 and 8.46 simultaneously. In this section, the fermion fields are flowed alongside the gauge fields using the "adjoint flow" first discussed in Ref. [249] and implemented in the `QLUA` library [300].

8.3.2 Operator anomalous dimensions from gradient-flowed correlation functions

Note that, unlike the GF equation for the gauge fields, the GF equation for the fermion fields is *linear*. Because it is linear, the wave function picks up an additional contribution from the wave function renormalization of the fermion fields Z_{ψ} that needs to be cancelled off in the definition of $R_{\mathcal{O}}$ in Eqn. 2.28. Additionally, it is numerically advantageous to flow only one of the operators $\mathcal{O}$ in any given two-point function. As such, define the *partially-flowed two-point function* in infinite volume as

$$G_{\mathcal{O}}\big(t,\hat{x}_4;g_0^2\big) = \int \mathrm{d}^3\hat{\mathbf{x}}\ \langle \mathcal{O}\big(\hat{\mathbf{x}},\hat{x}_4,t\big)\mathcal{O}\big(\hat{\mathbf{x}},0\big)\rangle, \tag{8.47}$$

where the operator $\mathcal{O}\big(\hat{\mathbf{x}},\hat{x}_4,t\big)$ is composed of "GF-blocked" fermion fields (Chapter 4.2), $\mathcal{O}\big(\hat{\mathbf{x}},0\big)$ is composed of unblocked fermion fields, and I've changed notation slightly to $\big(\hat{\mathbf{x}},\hat{x}_4\big) = \hat{x} \equiv n$ to remain consistent with the noation of Ref. [177]. The integral over the spatial components $\mathbf{x}$ of $\hat{x}$ projects the two-point function onto zero spatial momentum. Eqn. 2.23 in Chapter 2.4 for the partially-flowed two-point function then takes the form

$$G_{\mathcal{O}}\big(t_0,\hat{x}_4;g_0^2\big) = Z_{\psi}^{n_{\mathcal{O}}}(b)Z_{\mathcal{O}}(b)G_{\mathcal{O}}\big(t,\hat{x}_4;g_0^2\big), \tag{8.48}$$

where $b^2 \equiv t/t_0$ and $n_{\mathcal{O}}$ is the number of fermion fields in $\mathcal{O}$. To calculate the running of $Z_{\mathcal{O}}$ from $\gamma_{\mathcal{O}}$, the contribution of $Z_{\psi}^{n_{\mathcal{O}}}$ to Eqn. 8.48 needs to be cancelled off. Because the vector current is exactly conserved in the $N_\mathrm{f} = 10$ system, it only picks up a contribution from Z_ψ; hence, the definition of $R_\mathcal{O}$ in Eqn. 2.28 can be modified to cancel off the contribution of $Z_\psi^{n_\mathcal{O}}$ as

$$R_\mathcal{O}\big(t_0, \hat{x}_4; g_0^2\big) \equiv \frac{G_\mathcal{O}\big(t, \hat{x}_4; g_0^2\big)}{G_\mathcal{V}\big(t, \hat{x}_4; g_0^2\big)^{n_\mathcal{O}/n_\mathcal{V}}}, \tag{8.49}$$

where $G_\mathcal{V}\big(t, \hat{x}_4; g_0^2\big)$ is the two-point function of the vector current. Though unflowed two-point functions were included in the definition of Eqn. 2.28, they do not contribute to $\gamma_\mathcal{O}$; hence, they are not included in the definition of $R_\mathcal{O}$ in Eqn. 8.49. From $R_\mathcal{O}$, the calculation of $\gamma_\mathcal{O}$ follows as (see Eqn. 2.29 in Chapter 2.4):

$$-2t\frac{\mathrm{d}}{\mathrm{d}t}\log R_\mathcal{O}\big(t_0, \hat{x}_4; g_0^2\big) = \gamma_\mathcal{O}\big(t, \hat{x}_4; g_0^2\big). \tag{8.50}$$

For $\hat{x}_4^2 \gg 8t/a^2$, the logarithmic derivative should be indepndent of $\hat{x}_4$, as should $\gamma_\mathcal{O}$. Note that the factor of two in Eqn. 8.50 arises from a different source as in Eqn. 2.28. With the finite volume version of the running anomalous dimension $\gamma_\mathcal{O}\big(t; L, g_0^2\big)$ determined by Eqns. 8.48-8.50, calculating the continuum prediction for $\gamma_\mathcal{O}\big(g_\mathrm{GF}^2\big)$ follows exactly the same steps as the continuous β-function method that I have deployed exhaustively in Chapters 4, 5, and 7.

8.3.3 Simulation details

The results of this section utilize a tree-level improved Lüscher-Weiss gauge action (Appendix F) for the gauge sector [248, 251] and a stout smeared Möbius domain wall (DWF, Chapter 3.2.2) fermion action for the fermion sector [54]. The stout-smeared pseudofermion action utilizes six levels of stout smearing with $\rho = 0.1$; see Appendix G.2 for an overview of stout smearing [266]. Gauge configurations are generated using the Hamiltonian Monte Carlo algorithm provided by the `GRID` library [47, 112]. Gauge/fermion flow measurements are performed using the `QLUA` library [300]. The flow is Wilson flow. The full dataset is composed of six bare gauge couplings $\beta_b \equiv 6/g_0^2 = 4.05, 4.10, 4.20, 4.60, 5.00, 6.00$ on two volumes $(L/a)^3 \times (T/a) = 24^3 \times 64$ and $32^3 \times 64$. The length of the fifth dimension is $N_5 = 16$

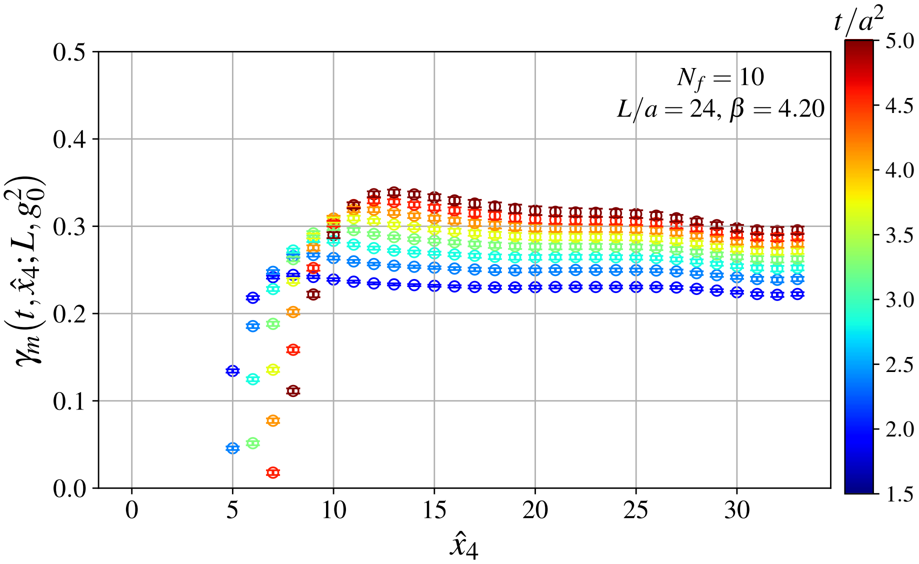

Figure 8.3: Pseudoscalar mass anomalous dimension $\gamma_m(t, \hat{x}_4; L, g_0^2)$ at fixed $(L/a, N_\mathrm{f}, \beta_b) = (24, 10, 4.2)$ and $1.5 \leq t/a^2 \leq 5.0$ (blue to red) against $\hat{x}_4$.

(Chapter 3.2.2), which yields a residual mass $am_{\mathrm{res.}} \lesssim 10^{-6}$.[7]

8.3.4 Calculation of operator anomalous dimensions

I start by calculating the finite volume renormalized coupling $g^2_{\mathrm{GF}}(t; L, g_0^2)$ using Eqn. 4.19 with $\delta(t, L)$ chosen to include tree-level corrections for cutoff effects using Eqn. 4.27. The Yang-Mills energy density $E(t; L)$ is estimated from Wilson flow the tree-level-correct clover operator ($\mathcal{S}^\mathrm{f}\mathcal{S}^\mathrm{e} = \mathrm{WC}$). I then calculate the finite volume anomalous dimension $\gamma_{\mathcal{O}}(t, \hat{x}_4; L, g_0^2)$ for the pseudoscalar, tensor, and proton operator using the finite-volume version of Eqns. 8.47-8.50. The derivative in Eqn. 8.50 is calculated using the 5-point stencil of Eqn. 4.37 and has been crosschecked against improved finite-difference schemes for logarithmic derivatives of logarithmic functions.

Fig. 8.3 illustrates the variation of the finite-volume pseudoscalar anomalous dimension with $\hat{x}_4$ at fixed $1.5 \leq t/a^2 \leq 5.0$ (different colors), $L/a = 24$, and $\beta_b = 4.2$. The pseudoscalar anomalous dimension is denoted as "$\gamma_m(t, \hat{x}_4; L, g_0^2)$" because the preservation of chiral symmetry in the system forces the scalar and pseudscalar to be degenerate. As such, the pseudoscalar anomalous dimension

---

[7] The size of the residual mass indicates the degree of chiral symmetry breaking.

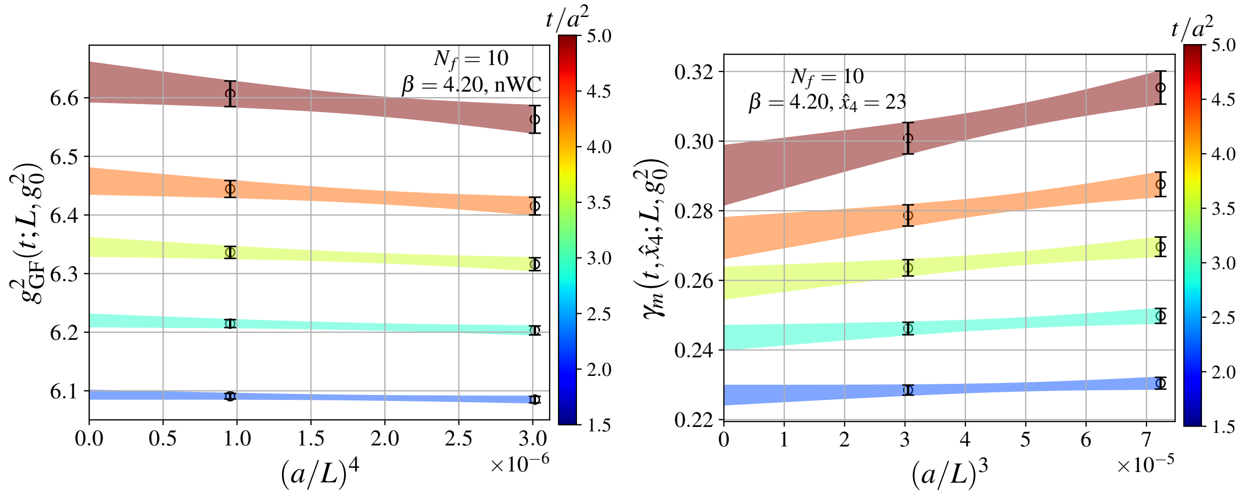

Figure 8.4: Sample of infinite volume extrapolation of $g^2_{\rm GF}(t;L,g_0^2)$ (left panel) and mass pseudoscalar anomalous dimension $\gamma_m(t,\hat{x}_4;L,g_0^2)$ at fixed $(L/a,N_{\rm f},\beta_b)=(24,10,4.2)$ and $1.5 \leq t/a^2 \leq 5.0$ (blue to red). Extrapolation indicated by multicolored bands with the width of the band indicating the error and data entering interpolation indicated by corresponding multicolored error bar.

is the mass anomalous dimension. For each t/a^2, there is a plateau over which $\gamma_{\mathcal{O}}(t,\hat{x}_4;L,g_0^2)$ is independent of $\hat{x}_4$. As t/a^2 increases, the region in $\hat{x}_4$ over which the anomalous dimension is flat decreases because the short-distance overlap between the flowed and unflowed operators in the partially-flowed two point function increases with t/a^2. Such overlap distorts the correlation function over distances $\hat{x}_4 \lesssim 8t/a^2$.

8.3.4.1 Infinite volume extrapolation

The infinite volume extrapolation is performed in the same manner as the infinite volume extrapolation step of the continuous β-function method in Chapters 4, 5, and 7, only with a minor modifications to the finite volume scaling of $\gamma_{\mathcal{O}}(t,\hat{x}_4;L,g_0^2)$ as compared to $g^2_{\rm GF}(t;L,g_0^2)$. I extrapolate $g^2_{\rm GF}(t;L,g_0^2)$ to $a/L \to 0$ limit at fixed $\beta_b, t/a^2$ by fitting the $L/a = 24, 32$ volumes to the ansatz of Eqn. 4.38. The operator anomalous dimension is extrapolated to the infinite volume limit via the same procedure; however, the finite volume ansatz is taken to be

$$\mathrm{FV}_{\mathcal{O}}(t;L,g_0^2) = k_1(t;g_0^2) + k_2(t;g_0^2)(a/L)^{\delta_{\mathcal{O}}/2} \qquad \text{(fixed } t/a^2 \text{ and } \beta_b) \qquad (8.51)$$

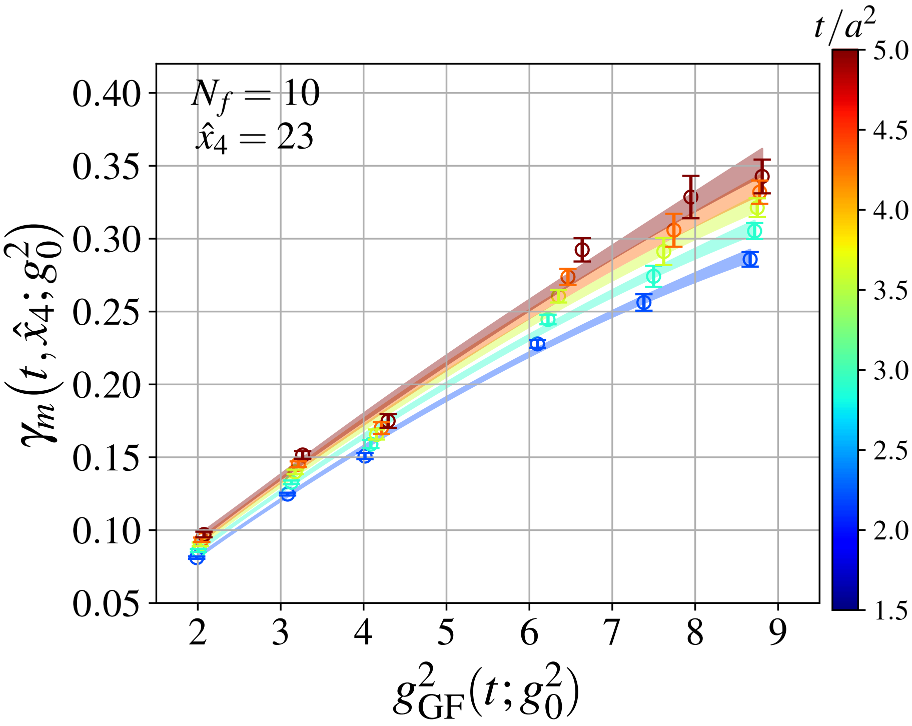

Figure 8.5: Sample of quadratic interpolation of pseudoscalar anomalous dimension $\gamma_m\big(t,\hat{x}_4;g_0^2\big)$ in gradient flow coupling $g^2_{\mathrm{GF}}\big(t,\hat{x}_4;g_0^2\big)$ at fixed $(N_{\mathrm{f}},\hat{x}_4) = (10,23)$ and $1.5 \leq t/a^2 \leq 5.0$. Interpolation indicated by multicolored bands with the width of the band indicating the error and data entering interpolation indicated by corresponding multicolored error bar.

based on the mass dimension of $\mathcal{O}$. For mesonic operators, $\delta_{\mathcal{O}} \approx 3$ (e.g., the pseudoscalar), while for baryonic operators $\delta_{\mathcal{O}} \approx 9/2$ (e.g., the proton). I illustrate the infinite volume extrapolation for $g^2_{\mathrm{GF}}\big(t;L,g_0^2\big)$ in the left panel of Fig. 8.7 and $\gamma_m\big(t,\hat{x}_4;L,g_0^2\big)$ (pseudoscalar) in the right panel of Fig. 8.7 for $1.5 \leq t/a^2 \leq 5.0$ (different colors) and $\beta_b = 4.2$. Because there are only two volumes to do the extrapolation over, the resulting fit unfortunately has zero degrees of freedom. Though the finite-volume effects appear to be mild, little can be said about the control that I have over the infinite volume limit without at least one more volume.

8.3.4.2 Continuum extrapolation

As with the β-function, taking the continuum $a^2/t \to 0$ limit at fixed g^2_{GF} requires first interpolating $\gamma_{\mathcal{O}}\big(t,\hat{x}_4;g_0^2\big)$ in $g^2_{\mathrm{GF}}\big(t;L,g_0^2\big)$ at fixed t/a^2 for each $\mathcal{O}$. The interpolating function is

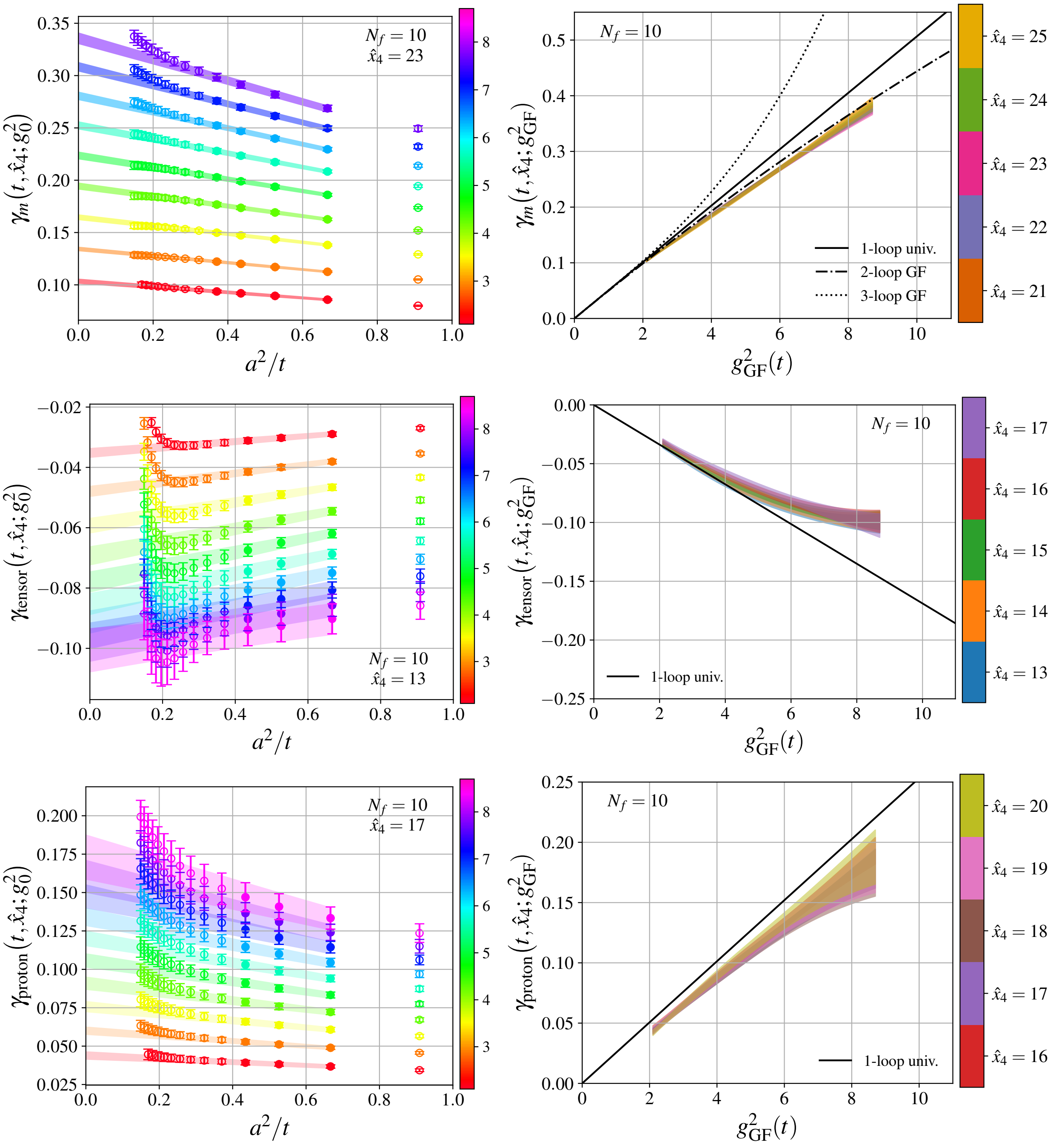

Figure 8.6: Continuum extrapolation (left panels) and final continuum result (right panels) for pseudoscalar (top panels), tensor (middle panels), and proton (bottom panels). Left panels show multiple fixed $2.0 \lesssim g_{\mathrm{GF}}^2 \lesssim 9.0$ (red to purple). Right panels show the final result from multiple $\hat{x}_4$ (see color bar on right). Continuum prediction for pseudoscalar juxtaposed against 1-, 2-, and 3-loop perturbative mass anomalous dimension from Ref. [23]. Tensor and proton are only compared against their corresponding 1-loop perturbative results.

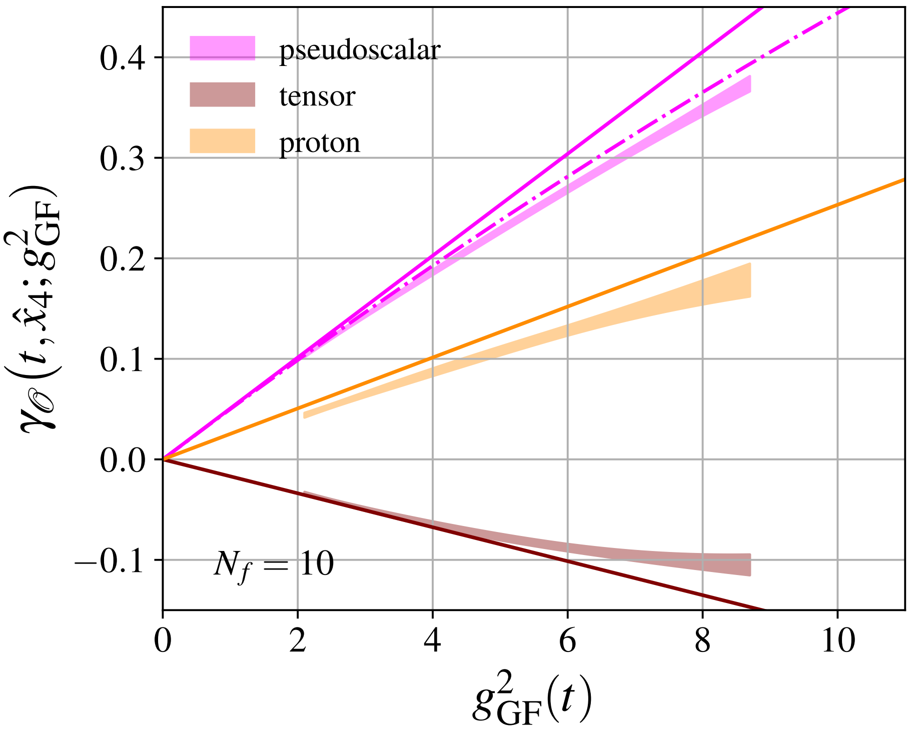

Figure 8.7: Continuum prediction for the pseudoscalar (pink), tensor (maroon), and proton (dark yellow) operator anomalous dimension. Pseudoscalar juxtaposed against 1- and 2-loop mass anomalous dimension from Ref. [23]. Tensor and proton are only compared against their corresponding 1-loop perturbative results.

chosen to be the simple polynomial ansatz of Eqn. 7.6 with $N = 3$. In other words, the interpolating function is quadratic in g^2_{GF}. No priors are imposed on the fit parameters. The interpolation is illustrated in Fig. 8.5 for $\gamma_m(t, \hat{x}_4; g_0^2)$ at $\hat{x}_4$ over $1.5 \leq t/a^2 \leq 5.0$ (different colors). The interpolation fits through most of the data well, save for $\beta_b = 4.2$ and 4.6, which fluctuate away from the interpolating band. It is reasonable to suspect that such fluctuations are be attributed to a lack of control over the infinite volume extrapolation.

The continuum extrapolation is performed by fitting $\gamma_{\mathcal{O}}(t, \hat{x}_4; g_0^2)$ to the linear a^2/t ansatz of Eqn. 4.39 at each fixed g^2_{GF}. The continuum extrapolation is illustrated in the left panels of Fig. 8.6, while the continuum prediction for $\gamma_{\mathcal{O}}(g^2_{\mathrm{GF}})$ from various $\hat{x}_4$ is illustrated in the right panels of 8.6. The latter information is shown for the pseudoscalar in the top panels, tensor in the middle panels, and proton in the bottom panels. The pseudoscalar (mass) anomalous dimension is juxtaposed against the 1-loop universal and 2-/3-loop gradient flow prediction from Ref. [23] in the top right panel of Fig. 8.6, while the tensor (middle right) and proton (bottom right) are juxtaposed only against their corresponding 1-loop perturbative anomalous dimension. Finite volume effects are observed to heavily distort the trend of $\gamma_{\mathcal{O}}\big(t, \hat{x}_4; g_0^2\big)$ in a^2/t for $t/a^2 \gtrsim 1.5$. Moreover, nonlinear cutoff effects in a^2/t appear to be suppressed for $t/a^2 \gtrsim 3.0$, though they are not as harsh as they can be for the β-function. Hence, there is a small region of linearity between $1.5 \lesssim t/a^2 \lesssim 3.0$ over which the continuum extrapolation is performed and the continuum prediction from various $\hat{x}_4$ are consistent. Note that different operators can have different regions in $\hat{x}_4$ over which the anomalous dimension of the corresonding operator is observed to plateau; e.g., the pseudoscalar anomalous dimension reasonably plateus about $21 \leq \hat{x}_4 \leq 25$, while the proton plateaus about $16 \leq \hat{x}_4 \leq 20$, though with more variation in the continuum prediction from different $\hat{x}_4$.

8.3.4.3 The continuum operator anomalous dimensions

In Fig. 8.7, I show the continuum prediction for the pseudoscalar (pink), tensor (maroon), and proton (dark yellow). The pseudoscalar is juxtaposed against the 1-loop universal and 2-loop gradient flow prediction from the perturbative calculation of Ref. [23], while the tensor and proton are juxtaposed against only their 1-loop values. To little surprise, the continuum prediction for the pseudoscalar is the most well resolved and exhibits the least significant systematic effects from the choice in $\hat{x}_4$. In the other hand, the tensor and proton anomalous dimensions exhibit a considerable degree of uncertainty in $\hat{x}_4$. Both correlation functions tend exhibit a considerable degree of statistical noise; hence, the latter observation is also not particularly surprising. In any case, all three operator anomalous dimensions indicate that they are approaching their perturbative counterparts in the $g_{\mathrm{GF}}^2 \to 0$ limit, with the proton most significant deviating from the 1-loop universal curve. It is reasonable to suspect that the latter deviation is due to finite volume effects and that the combined statistical/systematic error of the proton is much higher than the statistical error that is shown in Figs. 8.6-8.7. On the other hand, the pseudoscalar (mass) anomalous dimension both converges nicely onto the 1- and 2-loop curves as $g_{\mathrm{GF}}^2 \to 0$ and closely follows the 2-loop curve up to, and possibly beyond, $g^2_{\mathrm{GF}} \approx 10$. Ref. [177] has also calculated the mass and tensor anomalous dimension using a Pauli-Villars improved Wilson fermion action. The results of this section appear to agree reasonably well with Ref. [177], though the mass anomalous dimension in Ref. [177] appears to converge onto the 1-loop curve much earlier in g^2_{GF} and the upward curvature in the tensor is less dramatic about $6 \lesssim g^2_{\mathrm{GF}} \lesssim 8.5$. Of course, as I have already mentioned, it is not possible to tell if the infinite volume limit in this section is controlled without more information from larger volumes.

Chapter 9

Recapitulation and an eye toward the future

We have covered quite a bit of ground in this thesis. Unfortunately, due to constraints from time, space, and relevance, I was not able to cover my contributions to muon g_μ-2 [32, 94], nor have I covered the long list of failed projects, such as my investigation of the thermodynamic properties of the $4+6$ mass-split system and its relevance to stochastic gravitational wave production [376]; Markov Chain Monte Carlo sampling of field theories with affine invariant Monte Carlo; including effects from Pauli-Villars fields and fermion loops into corrections for discretization effects in the gradient flow coupling; gradient flow renormalization applied to the BKT transition of the XY model; the application finite size scaling with radial basis function networks to first-order phase transitions simulated with multicanonical Monte Carlo algorithms; the application of feedfoward neural networks to finite size scaling; and much more. In any case, I hope that you enjoyed the journey as much as I have enjoyed both working on the content of this thesis and writing about it. Before we sign off, let me summarize the lessons have been learned from this thesis and where they might take us in the future.

I have calculated the continuum/infinite-volume renormalization group β-function of the massless SU(3) gauge-fermion system for $N_{\mathrm{f}} = 0$, 8, and 12 fermions in the fundamental representation of SU(3) in Chapters 4, 7, and 5, respectively. All β-function calculations in this thesis utilize gradient flow and deploy the continuous β-function method (CBFM). I have also developed a method for finite size scaling by the method of curve collapse that utilizes a radial basis function network to approximate the scaling function. The latter method has been tested on the finite-temperature phase transition of various classical spin systems in Chapter 6. In Chapter 7, I applied the RBFN-based curve collapse to the zero-temperature phase transition of the $N_{\mathrm{f}} = 8$ system.

9.1 Non-perturbative β-functions

The state-of-the-art determinations of the β-function in this thesis have achieved unprecedented reach into the infrared regime of each system investigated. In the $N_{\mathrm{f}} = 0$ system, this is attributed to the CBFM itself, which allows the β-function to be calculated in the confined regime; such is not true for determinations of the β-function based on step scaling. The considerable reach into the infrared regime of the $N_{\mathrm{f}} = 8$ and 12 systems is attributed to the use of Pauli-Villars improvement, which dampen the ultraviolet fluctuations that typically trigger first-order phase transitions in simulations of many-flavor systems.

The $N_{\mathrm{f}} = 0$ β-function exhibits interesting features in the strong coupling regime. For one, it is linear in the renormalized coupling g^2_{GF} for sufficiently large $g^2_{\mathrm{GF}}/4\pi \gg 1$; i.e., $\beta_{\mathrm{GF}}\left(g^2_{\mathrm{GF}}\right) \sim c_0 + c_1 g^2_{\mathrm{GF}}$ at strong coupling. The slope c_1 of the β-function in the strong coupling regime determines the dependence of the expected Yang-Mills energy density $t^2\langle E(t)\rangle$ on the flow time as $t^2\langle E(t)\rangle \propto t^{-c_1}$. The expectation for c_1 from the literature is $c_1 = -1$, which is merely empirical; however, I calculate a slope for the β-function of $c_1 = -1.320(10)$, which is not consistent with the latter expectation. The linearity of the β-function in the strong coupling regime could be due to instantons or even a universal feature of asymptotic RG β-functions (based on Wilsonian RG) in confined systems. Note, additionally, that tree-level corrections are used to determine the $N_{\mathrm{f}} = 0$ β-function. At weak coupling, they work as expected. They are also surprisingly effective in the strong coupling regime, where they are not expected work as well, if at all.

As the $N_{\mathrm{f}} = 0$ β-function of Chapter 4 has been determined between $1.2 \lesssim g^2_{\mathrm{GF}} \lesssim 27.0$, I am able to extract from it the Λ-parameter in the $\overline{\mathrm{MS}}$ using an exact relation between the $\overline{\mathrm{MS}}$ Λ-parameter and the Λ-parameter of the infinite volume GF scheme of this thesis. Unfortunately, doing so requires matching the non-perturbative β-function at the weakest accessible couplings to the perturbative β-function, which is known up to 3-loops. The matching procedure is performed by appending onto the 3-loop β-function an additional $\mathrm{O}\big(g_{\mathrm{GF}}^{10}\big)$ term with a coefficient that is determined by the matching procedure. An alternative non-perturbative matching procedure is proposed but not deployed. The β-function is then determined from $g_{\mathrm{GF}}^2 = 0$ to $g_{\mathrm{GF}}^2(t_0) = 0.3\mathcal{N}$ where the scale setting parameter t_0 is defined. From the full β-function, I determine the $\overline{\mathrm{MS}}$ Λ-parameter of the $N_{\mathrm{f}} = 0$ to be $\sqrt{8t_0}\Lambda_{\overline{\mathrm{MS}}} = 0.622(10)$, which is in fantastic agreement with the most high-precision recent gradient-flow-based determinations, but in tension with other methods appearing in the results compiled by the Flavor Lattice Averaging Group (FLAG) report. The reason for this tension is unknown at the moment; however, it is reasonable to speculate that it could be due to a poor determination of the ratio of t_0 to the Sommer scale r_0, which can be quite difficult to calculate accurately. In the near future, the method deployed in Chapter 4 for calculating the Λ-parameter could be utilized for high-precision determinations of the strong coupling constant α_{s}.

The $N_{\mathrm{f}} = 12$ β-function reported in Chapter 5 exhibits a clear infrared fixed point (IRFP) at $g_{\mathrm{GF}\star}^2 = 6.60(62)$. Additional systematic effects from the finite volume extrapolation are accounted for with Bayesian model averaging. The leading irrelevant critical exponent at the IRFP $\gamma_g^\star = 0.199(32)$ is calculated from the slope of the continuum β-function at $g_{\mathrm{GF}\star}^2$ and is in reasonable agreement with estimates from perturbation theory at the 1σ level and the only other non-perturbative estimate at the 2σ level. Though this calculation is the second non-perturbative β-function to have exhibited a fixed point, it extends far beyond what has been calculated previously and is far more controlled due to Pauli-Villars improvement.

Determination of the $N_{\mathrm{f}} = 8$ β-function in Chapter 7 was by far the most challenging on a technical level. As with $N_{\mathrm{f}} = 12$, Bayesian model averaging is deployed to account for additional systematic effects from the infinite volume extrapolation automatically. Multiple flows are deployed and a reasonable degree of consistency between different flows is observed. Even better, the β-function in the weak coupling agrees with determinations based on simulations with domain wall fermions using the same RG scheme. The β-function initially decreases, then begins to show signs of turning around around $15 \lesssim g_{\mathrm{GF}}^2 \lesssim 22$. There are a number of reasons that the β-function could begin to turn around. The least exciting reason is that it is an artefact of the RG scheme; though it is not possible to rule this out completely, it is unlikely. If the $N_{\mathrm{f}} = 8$ system sits either below or at the edge of the conformal window, the such a turn around could be indicative of slow walking. On the other hand, if the $N_{\mathrm{f}} = 8$ system is in the conformal phase, then the turnaround is likely a sign of the β-function beginning to flow into an IRFP. In any case, such a turnaround could indicate that large-scale simulations of the $N_{\mathrm{f}} = 8$ system have yet to probe is behavior outside of the "weakly coupling" region.

When combined with the $N_{\mathrm{f}} = 8$ results of Ref. [179] obtained from the same RG scheme and the $N_{\mathrm{f}} = 16$ β-function that is accessible from perturbation theory, a consistent picture of the conformal window of SU(3) gauge-fermion systems begins to take shape. For one, the value of $g^2_{\mathrm{GF}\star}$ increases with decreasing N_{f} toward the location of the conformal sill at $N_{\mathrm{f}}^{\mathrm{c}}$. Should the results of this thesis and that of Ref. [179] be replicated by other lattice gauge theory collaborations, this puts a probable location for $N_{\mathrm{f}}^{\mathrm{c}}$ at $N_{\mathrm{f}}^{\mathrm{c}} \lesssim 9$ from Ref. [179] and possibly $N_{\mathrm{f}}^{\mathrm{c}} \gtrsim 8$ from the calculation of Chapter 7, though the lower bound imposed by the $N_{\mathrm{f}} = 8$ calculation of Chapter 7 is certainly up for debate. I have taken steps toward aiding in replication efforts by making all gradient flow data for the $N_{\mathrm{f}} = 12$ system publicly available at Ref. [291]. Data for the $N_{\mathrm{f}} = 8$ system will be made available upon publication.

Further developments of the CBFM that offer refined control over systematic errors and estimates of statistical errors in the infinite volume and continuum extrapolations are needed. Though the analysis of the β-function after the infinite volume extrapolation in Chapters 5 and 7 benefits immensely from implementing Bayesian model averaging (BMA) in the infinite volume extrapolation, it is not immediately clear how to keep track of correlations between BMA quantities and the quantities that they are derived from. This is needed to consistently keep track of correlations throughout the interpolation step and the continuum extrapolation step. Otherwise, one has to settle with taking the diagonal approximation of the covariance matrix in the continuum extrapolation and estimate the statistical error from the 1σ bands of the data (assuming 100% correlation).

Additionally, treating the data of the continuum extrapolation as discrete is a fiction that introduces a number of systematic effects into the continuum extrapolation that would better be rid of. For one, individual data points are highly-correlated with one another and it is typically impossible to perform reliable correlated fits. By taking the diagonal approximation, the χ^2 and p-value lose their meaning. As I have already mentioned in Chapter 4, the latter problem could be partially alleviated by estimating the χ^2 and p-value in the diagonal approximation as suggested by Ref. [57]; however, it remains unclear to how properly estimate the statistical error without reverting to ad-hoc methods, such as assuming 100% correlation and fitting above/below the 1σ band. Additionally, it is not possible to implement conventional BMA for the continuum extrapolation, where it could be very useful, because there is no immediately obvious meaning to subset selection for continuous data. Note that continuous Gaussian variables are well-defined as Gaussian processes. Hence, it may be possible to instead leverage methods deployed for analyzing Gaussian processes in the continuum extrapolation; however, where to start with that is not yet clear.

9.2 Finite size scaling with radial basis function networks

In Chapter 6, I develop a method for finite size scaling curve collapse based on radial basis function networks (RBFN), which are a type of single-layer artificial neural network that is specially-designed for function approximation. Curve collapse fits with the RBFN utilize the basin hopping global optimization algorithm paired with the trust region reflective local optimization algorithm to get around the issue of having to either engineer initial conditions or utilize local optimization algorithms that are not guaranteed to converge to reasonable optima. The parameters of the RBFN are constrained using the empirical Bayes method, which proceeds by optimizing the marginal likelihood. Optimization of the marginal likelihood is efficiently achieved using a surrogated-based optimization algorithm that I develop in Chapter 6. The empirical Bayes procedure takes care of issues introduced by overfitting automatically and is extensible to higher-dimensional problems.

The RBFN-based curve collapse is tested on curve collapse analyses of the finite-temperature phase transition of the two-dimensional q-state Potts model and p-state clock model for $q = 2, 3$ and $p = 4, \infty$. The critical parameters estimated from the Binder cumulant and magnetic susceptibility of each model agree with either their exact values in the case of $p = 2, 3$ and $q = 4$ or their estimates from the literature for the case of $q = \infty$. Intriguingly, the size of the RBFN need not be unreasonably large, with 2-3 internal nodes performing just fine. Such a network is incredibly small compared to the networks deployed in larger and obviously much more difficult problems; however, there seems to be a pervasive mode of thought that that network should be much larger than is actually required by the complexity of the problem at hand. I wrap Chapter 6 up by speculating on other uses of RBFNs for analyses of lattice field theory data; as an example, I extract the critical temperature of the ∞-state clock (XY) model from a direct interpolation of helicity modulus using an RBFN.

The RBFN-based curve collapse method of Chapter 6 is applied to the $N_{\mathrm{f}} = 8$ system in Chapter 7 to scrutinize claims of a possible continuous phase transition into a symmetric mass generation phase made in Ref. [170]. The finite size scaling data in Chapter 7 extends the dataset of Ref. [170] by including extra $L/a = 24$ volumes and a new set of $L/a = 32$ volumes. In agreement with Ref. [170], 1st-order scaling is strongly disfavored by the curve collapse; however, it is not possible to distinguish a preference for ∞-order scaling ($\nu = 1/2, 1$) over 2nd-order scaling using standard statistical tests. Moreover, the predictions for the critical parameters in Chapter 7 are in tension with those of Ref. [170]; however, the curve collapse analysis of Chapter 7 is more reliable on many levels due to both an improved dataset and an improved method. It is absolutely crucial to stress that, despite first-order scaling being strongly disfavored by the curve collapse, it cannot be ruled out without measurements of the correlation length over the same region of parameter space. This will hopefully be possible with future investigations of the spectroscopy of the SMG-like phase in the $N_{\mathrm{f}} = 8$ system.

There much room for improvement in the RBFN-based FSS method. For one, it would be highly desireable to have on hand a better global optimization algorithm than basin hopping for the maximum a posteriori estimation. Though basin hopping vastly outperforms every other global optimization algorithm that I tried, it is still not guaranteed to converge and coaxing the algorithm into the global optimum can be a pill for the analyst. Improvements could also come in the way of implementing basin hopping steps that are better tailored to neural networks. The empirical Bayes procedure also leaves much room for improvement. For one, it is easily improved in one dimension by embedding the surrogate-based optimization procedure into a bisection method for finding the optimum of the marginal likelihood. Though it is also possible to extend the surrogate based method to more than one dimension, as would be desired for implementing the empirical Bayes procedure for more than just the network weights. It may be more efficient to simply utilize the full machinery of Bayesian optimization, which is able to implement more intelligent moves for surrogate-based sampling of the objective function in question.

References

[1] G. ’t Hooft, “When was asymptotic freedom discovered? or the rehabilitation of quantum field theory”, Nucl. Phys. B Proc. Suppl. **74**, edited by S. Narison, 413 (1999) (Cited on p. 5).

[2] Y. Abe and M. Fukuma, “Gradient flow and the renormalization group”, PTEP **2018**, 083B02 (2018) (Cited on p. 55).

[3] M. Albanese et al. (APE), “Glueball Masses and String Tension in Lattice QCD”, Phys. Lett. B **192**, 163 (1987) (Cited on pp. 252, 253).

[4] C. Alexandrou, A. Athenodorou, K. Cichy, A. Dromard, E. Garcia-Ramos, K. Jansen, U. Wenger, and F. Zimmermann, “Comparison of topological charge definitions in Lattice QCD”, Eur. Phys. J. C **80**, 424 (2020) (Cited on p. 66).

[5] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *LAPACK users’ guide*, Third (Society for Industrial and Applied Mathematics, Philadelphia, PA, 1999) (Cited on p. 183).

[6] Y. Aoki et al. (Flavour Lattice Averaging Group (FLAG)), “FLAG Review 2021”, Eur. Phys. J. C **82**, 869 (2022) (Cited on pp. 30, 52, 91, 94).

[7] Y. Aoki et al. (LatKMI), “Light composite scalar in eight-flavor QCD on the lattice”, Phys. Rev. D **89**, 111502 (2014) (Cited on p. 143).

[8] Y. Aoki et al. (LatKMI), “Light flavor-singlet scalars and walking signals in $N_f = 8$ QCD on the lattice”, Phys. Rev. D **96**, 014508 (2017) (Cited on p. 143).

[9] Y. Aoki, T. Aoyama, M. Kurachi, T. Maskawa, K.-i. Nagai, H. Ohki, E. Rinaldi, A. Shibata, K. Yamawaki, and T. Yamazaki (LatKMI), “Light composite scalar in twelve-flavor QCD on the lattice”, Phys. Rev. Lett. **111**, 162001 (2013) (Cited on p. 99).

[10] Y. Aoki, T. Aoyama, M. Kurachi, T. Maskawa, K.-i. Nagai, H. Ohki, A. Shibata, K. Yamawaki, and T. Yamazaki (LatKMI), “Walking signals in $N_f = 8$ QCD on the lattice”, Phys. Rev. D **87**, 094511 (2013) (Cited on p. 143).

[11] Y. e. a. Aoki (LatKMI), “Lattice study of conformality in twelve-flavor QCD”, Phys. Rev. **D86**, 054506 (2012) (Cited on p. 99).

[12] T. Aoyama et al., “The anomalous magnetic moment of the muon in the Standard Model”, Phys. Rept. **887**, 1 (2020) (Cited on p. 51).

[13] T. Appelquist et al. (LSD), “Lattice simulations with eight flavors of domain wall fermions in SU(3) gauge theory”, Phys. Rev. D **90**, 114502 (2014) (Cited on p. 143).

[14] T. Appelquist et al. (LSD), “Linear Sigma EFT for Nearly Conformal Gauge Theories”, Phys. Rev. D **98**, 114510 (2018) (Cited on pp. 16, 144).

[15] T. Appelquist et al. (Lattice Strong Dynamics), “Nonperturbative investigations of SU(3) gauge theory with eight dynamical flavors”, Phys. Rev. D **99**, 014509 (2019) (Cited on p. 143).

[16] T. Appelquist, G. Fleming, M. Lin, E. Neil, and D. Schaich, “Lattice Simulations and Infrared Conformality”, Phys.Rev. **D84**, 054501 (2011) (Cited on p. 99).

[17] T. Appelquist et al., “Strongly interacting dynamics and the search for new physics at the LHC”, Phys. Rev. D **93**, 114514 (2016) (Cited on p. 143).

[18] T. Appelquist, G. T. Fleming, and E. T. Neil, “Lattice study of the conformal window in QCD-like theories”, Phys.Rev.Lett. **100**, 171607 (2008) (Cited on pp. 99, 143).

[19] T. Appelquist, G. T. Fleming, and E. T. Neil, “Lattice Study of Conformal Behavior in SU(3) Yang-Mills Theories”, Phys. Rev. D **79**, 076010 (2009) (Cited on pp. 99, 143).

[20] T. Appelquist, J. Ingoldby, and M. Piai, “Dilaton potential and lattice data”, Phys. Rev. D **101**, 075025 (2020) (Cited on pp. 16, 144).

[21] T. Appelquist, J. Ingoldby, and M. Piai, “Dilaton Effective Field Theory”, Universe **9**, 10 (2023) (Cited on p. 144).

[22] T. Appelquist, A. Ratnaweera, J. Terning, and L. C. R. Wijewardhana, “The Phase structure of an SU(N) gauge theory with N(f) flavors”, Phys. Rev. D **58**, 105017 (1998) (Cited on p. 99).

[23] J. Artz, R. V. Harlander, F. Lange, T. Neumann, and M. Prausa, “Results and techniques for higher order calculations within the gradient-flow formalism”, JHEP **06**, [Erratum: JHEP 10, 032 (2019)], 121 (2019) (Cited on pp. 193–195).

[24] M. F. Atiyah and I. M. Singer, “The Index of elliptic operators. 5.”, Annals Math. **93**, 139 (1971) (Cited on p. 66).

[25] M. Baioletti, V. Santucci, and M. Tomassini, “A performance analysis of basin hopping compared to established metaheuristics for global optimization”, Journal of Global Optimization, 1 (2024) (Cited on p. 126).

[26] T. Balaban, “Renormalization Group Approach to Lattice Gauge Fields Theories. 1. Generation of Effective Actions in a Small Fields Approximation and a Coupling Constant Renormalization in Four-dimensions”, Commun. Math. Phys. **109**, 249 (1987) (Cited on p. 39).

[27] A. Banerjee, D. Jasrasaria, S. P. Niblett, and D. J. Wales, “Crystal structure prediction for benzene using basin-hopping global optimization”, The Journal of Physical Chemistry A **125**, 3776 (2021) (Cited on p. 126).

[28] T. Banks and A. Zaks, “On the Phase Structure of Vector-Like Gauge Theories with Massless Fermions”, Nucl. Phys. B **196**, 189 (1982) (Cited on p. 7).

[29] O. Bar and M. Golterman, “Chiral perturbation theory for gradient flow observables”, Phys. Rev. D **89**, [Erratum: Phys.Rev.D 89, 099905 (2014)], 034505 (2014) (Cited on p. 61).

[30] A. Bashir, A. Raya, and J. Rodriguez-Quintero, “QCD: Restoration of Chiral Symmetry and Deconfinement for Large N_f”, Phys. Rev. D **88**, 054003 (2013) (Cited on p. 99).

[31] A. Bazavov et al. (MILC), “Gradient flow and scale setting on MILC HISQ ensembles”, Phys. Rev. D **93**, 094510 (2016) (Cited on pp. 52, 91).

[32] A. Bazavov et al. (Fermilab Lattice, HPQCD,, MILC), “Light-quark connected intermediate-window contributions to the muon g-2 hadronic vacuum polarization from lattice QCD”, Phys. Rev. D **107**, 114514 (2023) (Cited on pp. 52, 197).

[33] A. Bazavov and B. A. Berg, “Program package for multicanonical simulations of u(1) lattice gauge theory”, Computer Physics Communications **180**, 2339 (2009) (Cited on p. 35).

[34] A. Bazavov and T. Chuna, “Efficient integration of gradient flow in lattice gauge theory and properties of low-storage commutator-free Lie group methods”, (2021) (Cited on p. 179).

[35] P. Becher and H. Joos, “The Dirac-Kahler Equation and Fermions on the Lattice”, Z. Phys. C **15**, 343 (1982) (Cited on pp. 44, 144).

[36] A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Y. S. Tyupkin, “Pseudoparticle Solutions of the Yang-Mills Equations”, Phys. Lett. B **59**, edited by J. C. Taylor, 85 (1975) (Cited on p. 64).

[37] V. L. Berezinsky, “Destruction of long range order in one-dimensional and two-dimensional systems having a continuous symmetry group. I. Classical systems”, Sov. Phys. JETP **32**, 493 (1971) (Cited on p. 136).

[38] B. A. Berg, “Introduction to multicanonical monte carlo simulations”, (1999) (Cited on p. 35).

[39] B. A. Berg, “Multicanonical simulations step by step”, Computer Physics Communications **153**, 397 (2003) (Cited on p. 35).

[40] M. Berry, “Making light of mathematics”, American Mathematical Society Josiah Willard Gibbs Lectures, 2002 (Cited on p. 24).

[41] D. Bleecker, *GAUGE THEORY AND VARIATIONAL PRINCIPLES* (1981) (Cited on p. 65).

[42] S. Borsanyi et al., “Leading hadronic contribution to the muon magnetic moment from lattice QCD”, Nature **593**, 51 (2021) (Cited on p. 52).

[43] S. Borsányi, S. Dürr, Z. Fodor, C. Hoelbling, S. D. Katz, S. Krieg, T. Kurth, L. Lellouch, T. Lippert, and C. McNeile (BMW), “High-precision scale setting in lattice QCD”, JHEP **09**, 010 (2012) (Cited on pp. 52, 91).

[44] C. M. Bouchard, G. P. Lepage, C. Monahan, H. Na, and J. Shigemitsu, “$B_s \to K\ell\nu$ form factors from lattice QCD”, Phys. Rev. D **90**, 054506 (2014) (Cited on p. 186).

[45] G. E. P. Box and N. R. Draper, *Empirical model-building and response surfaces.* (John Wiley & Sons, 1987) (Cited on p. 118).

[46] G. E. P. Box and M. E. Muller, “A Note on the Generation of Random Normal Deviates”, The Annals of Mathematical Statistics **29**, 610 (1958) (Cited on p. 168).

[47] P. Boyle, G. Cossu, A. Portelli, and A. Yamaguchi, *Grid*, 2015 (Cited on pp. 62, 166, 189).

[48] N. Brambilla, X. Garcia i Tormo, J. Soto, and A. Vairo, “Precision determination of $r_0\Lambda_{\overline{\rm MS}}$ from the QCD static energy”, Phys. Rev. Lett. **105**, [Erratum: Phys.Rev.Lett. 108, 269903 (2012)], 212001 (2010) (Cited on pp. 94, 95).

[49] M. A. Branch, T. F. Coleman, and Y. Li, “A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems”, SIAM Journal on Scientific Computing **21**, 1 (1999) (Cited on pp. 69, 124, 125, 183).

[50] J. Braun, C. S. Fischer, and H. Gies, “Beyond Miransky Scaling”, Phys. Rev. D **84**, 034045 (2011) (Cited on pp. 13, 99).

[51] J. Braun and H. Gies, “Chiral phase boundary of QCD at finite temperature”, JHEP **06**, 024 (2006) (Cited on p. 99).

[52] R. P. Brent, *Algorithms for minimization without derivatives* (Courier Corporation, 2013) (Cited on p. 114).

[53] R. C. Brower et al. (Lattice Strong Dynamics), "Light Scalar Meson and Decay Constant in SU(3) Gauge Theory with Eight Dynamical Flavors", (2023) (Cited on p. 143).

[54] R. C. Brower, H. Neff, and K. Orginos, "The Möbius domain wall fermion algorithm", Comput. Phys. Commun. **220**, 1 (2017) (Cited on pp. 47, 189).

[55] F. R. Brown, F. P. Butler, H. Chen, N. H. Christ, Z.-h. Dong, W. Schaffer, L. I. Unger, and A. Vaccarino, "On the existence of a phase transition for QCD with three light quarks", Phys. Rev. Lett. **65**, 2491 (1990) (Cited on p. 64).

[56] M. Bruno, M. Dalla Brida, P. Fritzsch, T. Korzec, A. Ramos, S. Schaefer, H. Simma, S. Sint, and R. Sommer (ALPHA), "QCD Coupling from a Nonperturbative Determination of the Three-Flavor Λ Parameter", Phys. Rev. Lett. **119**, 102001 (2017) (Cited on p. 60).

[57] M. Bruno and R. Sommer, "On fits to correlated and auto-correlated data", Comput. Phys. Commun. **285**, 108643 (2023) (Cited on pp. 82, 201).

[58] E. Buddenoir and S. Wallon, "The correlation length of the potts model at the first-order transition point", Journal of Physics A: Mathematical and General **26**, 3045 (1993) (Cited on p. 150).

[59] C. P. Burgess, "Introduction to Effective Field Theory", Ann. Rev. Nucl. Part. Sci. **57**, 329 (2007) (Cited on p. 1).

[60] A. Burkov, *The hundred-page machine learning book*, Vol. 1 (Andriy Burkov Quebec City, QC, Canada, 2019) (Cited on p. 126).

[61] A. Bussone, M. Della Morte, V. Drach, and C. Pica, "Tuning the hybrid Monte Carlo algorithm using molecular dynamics forces' variances", Comput. Phys. Commun. **234**, 179 (2019) (Cited on p. 258).

[62] N. Butt, S. Catterall, A. Pradhan, and G. C. Toga, "Anomalies and symmetric mass generation for Kähler-Dirac fermions", Phys. Rev. D **104**, 094504 (2021) (Cited on pp. 16, 144).

[63] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization", SIAM Journal on Scientific Computing **16**, 1190 (1995) (Cited on p. 125).

[64] G. Cacciapaglia, A. Deandrea, and K. Sridhar, "Review of fundamental composite dynamics", Eur. Phys. J. ST **231**, 1221 (2022) (Cited on p. 15).

[65] C. G. Callan Jr., "Broken scale invariance in scalar field theory", Phys. Rev. D **2**, 1541 (1970) (Cited on pp. 3, 33).

[66] I. Campos, P. Fritzsch, C. Pena, D. Preti, A. Ramos, and A. Vladikas (ALPHA), "Non-perturbative quark mass renormalisation and running in $N_f = 3$ QCD", Eur. Phys. J. C **78**, 387 (2018) (Cited on p. 60).

[67] S. Capitani, M. Lüscher, R. Sommer, and H. Wittig, "Non-perturbative quark mass renormalization in quenched lattice QCD", Nucl. Phys. B **544**, [Erratum: Nucl.Phys.B 582, 762–762 (2000)], 669 (1999) (Cited on pp. 94, 95).

[68] F. Capponi, L. Del Debbio, S. Ehret, R. Pellegrini, A. Portelli, and A. Rago, "Renormalisation of the scalar energy-momentum tensor with the Wilson flow", PoS **LATTICE2016**, 341 (2016) (Cited on p. 55).

[69] J. L. Cardy, ed., *FINITE SIZE SCALING* (1988) (Cited on pp. 36, 119, 132, 229).

[70] J. L. Cardy, *Scaling and renormalization in statistical physics* (1996) (Cited on pp. 26, 36, 38, 229).

[71] B. P. Carlin and T. A. Louis, “Empirical bayes: past, present and future”, Journal of the American Statistical Association **95**, 1286 (2000) (Cited on p. 127).

[72] A. Carosso, “Novel Approaches to Renormalization Group Transformations in the Continuum and on the Lattice”, PhD thesis (Colorado U., 2020) (Cited on pp. 53, 55, 56).

[73] A. Carosso, “Stochastic Renormalization Group and Gradient Flow”, JHEP **01**, 172 (2020) (Cited on pp. 31, 53, 55, 56).

[74] A. Carosso, A. Hasenfratz, and E. T. Neil, “Nonperturbative Renormalization of Operators in Near-Conformal Systems Using Gradient Flows”, Phys. Rev. Lett. **121**, 201601 (2018) (Cited on pp. 31, 33, 34, 53, 55, 187).

[75] W. E. Caswell, “Asymptotic Behavior of Nonabelian Gauge Theories to Two Loop Order”, Phys. Rev. Lett. **33**, 244 (1974) (Cited on pp. 5, 7).

[76] S. Catterall, “’t Hooft anomalies for staggered fermions”, Phys. Rev. D **107**, 014501 (2023) (Cited on pp. 16, 144).

[77] M. Chaichian and M. Frasca, “Condition for confinement in non-Abelian gauge theories”, Phys. Lett. B **781**, 33 (2018) (Cited on pp. 12, 88).

[78] A. Cheng, A. Hasenfratz, Y. Liu, G. Petropoulos, and D. Schaich, “Finite size scaling of conformal theories in the presence of a near-marginal operator”, Phys. Rev. D **90**, 014509 (2014) (Cited on p. 99).

[79] A. Cheng, A. Hasenfratz, G. Petropoulos, and D. Schaich, “Scale-dependent mass anomalous dimension from Dirac eigenmodes”, JHEP **07**, 061 (2013) (Cited on pp. 99, 143).

[80] A. Cheng, A. Hasenfratz, and D. Schaich, “Novel phase in SU(3) lattice gauge theory with 12 light fermions”, Phys. Rev. D **85**, 094509 (2012) (Cited on pp. 100, 145).

[81] H. S. Chung and D. Nogradi, “$f\varrho/m\varrho$ and $f\pi/m\varrho$ ratios and the conformal window”, Phys. Rev. D **107**, 074039 (2023) (Cited on p. 99).

[82] M. A. Clark, R. Babich, K. Barros, R. C. Brower, and C. Rebbi (QUDA), “Solving Lattice QCD systems of equations using mixed precision solvers on GPUs”, Comput. Phys. Commun. **181**, 1517 (2010) (Cited on p. 166).

[83] M. A. Clark, A. D. Kennedy, and P. J. Silva, “Tuning HMC using Poisson brackets”, PoS **LATTICE2008**, edited by C. Aubin, S. Cohen, C. Dawson, J. Dudek, R. Edwards, B. Joo, H.-W. Lin, K. Orginos, D. Richards, and H. Thacker, 041 (2008) (Cited on p. 258).

[84] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, “Fast and accurate deep network learning by exponential linear units (elus)”, arXiv preprint arXiv:1511.07289 (2015) (Cited on p. 187).

[85] A. G. Cohen and H. Georgi, “Walking Beyond the Rainbow”, Nucl. Phys. B **314**, 7 (1989) (Cited on p. 15).

[86] R. Contino, “The Higgs as a Composite Nambu-Goldstone Boson”, in Theoretical Advanced Study Institute in Elementary Particle Physics: Physics of the Large and the Small (2011), pp. 235–306 (Cited on p. 16).

[87] M. Creutz, “Global Monte Carlo algorithms for many-fermion systems”, Phys. Rev. D **38**, 1228 (1988) (Cited on p. 257).

[88] M. Creutz, *Quarks, Gluons and Lattices* (Oxford University Press, 1983) (Cited on pp. 39, 230).

[89] D. d'Enterria et al., "The strong coupling constant: State of the art and the decade ahead", (2022) (Cited on p. 89).

[90] M. Dalla Brida, "Past, present, and future of precision determinations of the QCD parameters from lattice QCD", Eur. Phys. J. A **57**, 66 (2021) (Cited on p. 89).

[91] M. Dalla Brida, P. Fritzsch, T. Korzec, A. Ramos, S. Sint, and R. Sommer (ALPHA), "A non-perturbative exploration of the high energy regime in $N_\mathrm{f} = 3$ QCD", Eur. Phys. J. C **78**, 372 (2018) (Cited on pp. 60, 95, 96).

[92] M. Dalla Brida and A. Ramos, "The gradient flow coupling at high-energy and the scale of SU(3) Yang–Mills theory", Eur. Phys. J. C **79**, 720 (2019) (Cited on pp. 18, 60, 67, 79, 90, 91, 94–96).

[93] C. T. H. Davies et al. (Fermilab Lattice, LATTICE-HPQCD, MILC), "Hadronic-vacuum-polarization contribution to the muon's anomalous magnetic moment from four-flavor lattice QCD", Phys. Rev. D **101**, 034512 (2020) (Cited on p. 52).

[94] C. T. H. Davies et al. (Fermilab Lattice, MILC, HPQCD), "Windows on the hadronic vacuum polarization contribution to the muon anomalous magnetic moment", Phys. Rev. D **106**, 074509 (2022) (Cited on p. 197).

[95] Z. Davoudi et al., "Report of the Snowmass 2021 Topical Group on Lattice Gauge Theory", in Snowmass 2021 (2022) (Cited on p. 89).

[96] K. Decker, A. Hasenfratz, and P. Hasenfratz, "Singular Renormalization Group Transformations and First Order Phase Transitions: Monte Carlo Renormalization Group Results", Nucl. Phys. B **295**, 21 (1988) (Cited on pp. 20, 37).

[97] T. DeGrand, "Finite-size scaling tests for spectra in SU(3) lattice gauge theory coupled to 12 fundamental flavor fermions", Phys. Rev. D **84**, 116901 (2011) (Cited on p. 99).

[98] T. DeGrand and C. E. Detar, *Lattice methods for quantum chromodynamics* (2006) (Cited on pp. 39, 44–46, 169, 170, 262).

[99] T. A. DeGrand, A. Hasenfratz, P. Hasenfratz, and F. Niedermayer, "The Classically perfect fixed point action for SU(3) gauge theory", Nucl. Phys. B **454**, 587 (1995) (Cited on pp. 25, 251).

[100] L. Del Debbio and R. Zwicky, "Hyperscaling relations in mass-deformed conformal gauge theories", Phys. Rev. D **82**, 014502 (2010) (Cited on p. 8).

[101] L. Del Debbio and R. Zwicky, "Dilaton and massive hadrons in a conformal phase", JHEP **08**, 007 (2022) (Cited on p. 144).

[102] A. Deuzeman, M. P. Lombardo, and E. Pallante, "Evidence for a conformal phase in SU(N) gauge theories", Phys. Rev. D **82**, 074503 (2010) (Cited on p. 99).

[103] A. Deuzeman, M. P. Lombardo, T. Nunes Da Silva, and E. Pallante, "The bulk transition of QCD with twelve flavors and the role of improvement", Phys. Lett. B **720**, 358 (2013) (Cited on p. 100).

[104] A. Deuzeman, M. P. Lombardo, and E. Pallante, "The Physics of eight flavours", Phys. Lett. B **670**, 41 (2008) (Cited on p. 143).

[105] C. DeWitt-Morette, M. Dillard-Bleick, and Y. Choquet-Bruhat, *Analysis, manifolds and physics* (North-Holland, 1978) (Cited on p. 39).

[106] L. Di Pietro and M. Serone, "Looking through the QCD Conformal Window with Perturbation Theory", JHEP **07**, 049 (2020) (Cited on pp. 19, 99, 116).

[107] J. Dimock, "The Renormalization Group According to Balaban - II. Large fields", J. Math. Phys. **54**, 092301 (2013) (Cited on p. 39).

[108] J. Dimock, "The Renormalization Group According to Balaban, I. Small fields", Rev. Math. Phys. **25**, 1330010 (2013) (Cited on p. 39).

[109] J. Dimock, "The Renormalization Group According to Balaban III. Convergence", Annales Henri Poincare **15**, 2133 (2014) (Cited on p. 39).

[110] J. P. Doye, R. H. Leary, M. Locatelli, and F. Schoen, "Global optimization of morse clusters by potential energy transformations", INFORMS Journal on Computing **16**, 329 (2004) (Cited on p. 126).

[111] T. Dozat, "Incorporating nesterov momentum into adam", (2016) (Cited on p. 124).

[112] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, "Hybrid Monte Carlo", Phys. Lett. B **195**, 216 (1987) (Cited on pp. 62, 168, 189).

[113] H. Duminil-Copin, M. Gagnebin, M. Harel, I. Manolescu, and V. Tassion, "Discontinuity of the phase transition for the planar random-cluster and potts models with $q > 4$", arXiv preprint arXiv:1611.09877 (2016) (Cited on pp. 131, 150).

[114] H. Duminil-Copin, V. Sidoravicius, and V. Tassion, "Continuity of the phase transition for planar random-cluster and potts models with $1 \leq q \leq 4$", Communications in Mathematical Physics **349**, 47 (2017) (Cited on pp. 131, 150).

[115] S. Durr et al. (BMW), "Ab-Initio Determination of Light Hadron Masses", Science **322**, 1224 (2008) (Cited on pp. 51, 52).

[116] S. Elitzur, "Impossibility of Spontaneously Breaking Local Symmetries", Phys. Rev. D **12**, 3978 (1975) (Cited on p. 14).

[117] S. Elitzur, R. B. Pearson, and J. Shigemitsu, "Phase structure of discrete abelian spin and gauge systems", Phys. Rev. D **19**, 3698 (1979) (Cited on p. 133).

[118] T. Evgeniou, M. Pontil, and T. Poggio, "Regularization networks and support vector machines", Advances in computational mathematics **13**, 1 (2000) (Cited on p. 126).

[119] E. Farhi and L. Susskind, "Technicolor", Phys. Rept. **74**, 277 (1981) (Cited on p. 14).

[120] M. E. Fisher and A. N. Berker, "Scaling for first-order phase transitions in thermodynamic and finite systems", Phys. Rev. B **26**, 2507 (1982) (Cited on pp. 20, 37, 149).

[121] R. Foadi, M. T. Frandsen, T. A. Ryttov, and F. Sannino, "Minimal Walking Technicolor: Set Up for Collider Physics", Phys. Rev. D **76**, 055005 (2007) (Cited on p. 15).

[122] Z. Fodor and C. Hoelbling, "Light Hadron Masses from Lattice QCD", Rev. Mod. Phys. **84**, 449 (2012) (Cited on p. 51).

[123] Z. Fodor, K. Holland, J. Kuti, S. Mondal, D. Nogradi, and C. H. Wong, "The lattice gradient flow at tree-level and its improvement", JHEP **09**, 018 (2014) (Cited on p. 59).

[124] Z. Fodor, K. Holland, J. Kuti, S. Mondal, D. Nogradi, and C. H. Wong, "The running coupling of 8 flavors and 3 colors", JHEP **06**, 019 (2015) (Cited on pp. 60, 143).

[125] Z. Fodor, K. Holland, J. Kuti, S. Mondal, D. Nogradi, and C. H. Wong, "The running coupling of the minimal sextet composite Higgs model", JHEP **09**, 039 (2015) (Cited on p. 60).

[126] Z. Fodor, K. Holland, J. Kuti, S. Mondal, D. Nogradi, and C. H. Wong, "Fate of the conformal fixed point with twelve massless fermions and SU(3) gauge group", Phys. Rev. **D94**, 091501 (2016) (Cited on pp. 99, 100).

[127] Z. Fodor, K. Holland, J. Kuti, D. Nogradi, and C. Schroeder, "Nearly conformal gauge theories in finite volume", Phys. Lett. B **681**, 353 (2009) (Cited on pp. 99, 143).

[128] Z. Fodor, K. Holland, J. Kuti, D. Nogradi, and C. Schroeder, "Twelve massless flavors and three colors below the conformal window", Phys. Lett. **B703**, 348 (2011) (Cited on pp. 99, 100).

[129] Z. Fodor, K. Holland, J. Kuti, D. Nogradi, C. Schroeder, and C. H. Wong, "Conformal finite size scaling of twelve fermion flavors", PoS **LATTICE2012**, edited by D. Leinweber, W. Kamleh, S. Mahbub, H. Matevosyan, A. Thomas, A. G. Williams, R. Young, and J. Zanotti, 279 (2012) (Cited on pp. 99, 100).

[130] Z. Fodor, K. Holland, J. Kuti, D. Nogradi, and C. H. Wong, "The Yang-Mills gradient flow in finite volume", JHEP **11**, 007 (2012) (Cited on pp. 57, 60).

[131] Z. Fodor, K. Holland, J. Kuti, D. Nogradi, and C. H. Wong, "A new method for the beta function in the chiral symmetry broken phase", EPJ Web Conf. **175**, edited by M. Della Morte, P. Fritzsch, E. Gámiz Sánchez, and C. Pena Ruano, 08027 (2018) (Cited on pp. 60–62).

[132] Z. Fodor, K. Holland, J. Kuti, D. Nogradi, and C. H. Wong, "Extended investigation of the twelve-flavor β-function", Phys. Lett. B **779**, 230 (2018) (Cited on pp. 60, 99).

[133] Z. Fodor, K. Holland, J. Kuti, D. Nogradi, and C. H. Wong, "Fate of a recent conformal fixed point and β-function in the SU(3) BSM gauge theory with ten massless flavors", PoS **LATTICE2018**, 199 (2018) (Cited on p. 60).

[134] Z. Fodor, K. Holland, J. Kuti, D. Nogradi, and C. H. Wong, "The twelve-flavor β-function and dilaton tests of the sextet scalar", EPJ Web Conf. **175**, edited by M. Della Morte, P. Fritzsch, E. Gámiz Sánchez, and C. Pena Ruano, 08015 (2018) (Cited on pp. 60, 99).

[135] Z. Fodor, K. Holland, J. Kuti, D. Nogradi, and C. H. Wong, "Case studies of near-conformal β-functions", PoS **LATTICE2019**, 121 (2019) (Cited on p. 100).

[136] P. de Forcrand, M. Pepe, and U.-J. Wiese, "Walking near a Conformal Fixed Point: the 2-d O(3) Model at theta near pi as a Test Case", Phys. Rev. D **86**, 075006 (2012) (Cited on p. 12).

[137] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, "emcee: The MCMC Hammer", **125**, 306 (2013) (Cited on p. 234).

[138] Y. Frishman and J. Sonnenschein, *Non-Perturbative Field Theory*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, 2023) (Cited on pp. 7, 65).

[139] P. Fritzsch and A. Ramos, "The gradient flow coupling in the Schrödinger functional", JHEP **10**, 008 (2013) (Cited on p. 59).

[140] R. Garnett, *Bayesian Optimization* (Cambridge University Press, 2023) (Cited on p. 124).

[141] J. Gasser and H. Leutwyler, "Chiral Perturbation Theory to One Loop", Annals Phys. **158**, 142 (1984) (Cited on p. 10).

[142] C. Gattringer and C. B. Lang, *Quantum chromodynamics on the lattice*, Vol. 788 (Springer, Berlin, 2010) (Cited on pp. 10, 39, 40, 44, 47, 58, 168, 235).

[143] M. Gell-Mann and F. E. Low, “Quantum electrodynamics at small distances”, Phys. Rev. **95**, 1300 (1954) (Cited on p. 3).

[144] M. Gell-Mann, R. J. Oakes, and B. Renner, “Behavior of current divergences under SU(3) x SU(3)”, Phys. Rev. **175**, 2195 (1968) (Cited on p. 10).

[145] J. Ghosh and A. Nag, “An overview of radial basis function networks”, in *Radial basis function networks 2: new advances in design*, edited by R. J. Howlett and L. C. Jain (Physica-Verlag HD, Heidelberg, 2001), pp. 1–36 (Cited on p. 122).

[146] P. H. Ginsparg and K. G. Wilson, “A Remnant of Chiral Symmetry on the Lattice”, Phys. Rev. D **25**, 2649 (1982) (Cited on p. 47).

[147] M. Gockeler, R. Horsley, A. C. Irving, D. Pleiter, P. E. L. Rakow, G. Schierholz, and H. Stuben, “A Determination of the Lambda parameter from full lattice QCD”, Phys. Rev. D **73**, 014513 (2006) (Cited on pp. 94, 95).

[148] N. Goldenfeld, *Lectures on phase transitions and the renormalization group* (1992) (Cited on pp. 37, 229).

[149] M. Golterman, E. T. Neil, and Y. Shamir, “Application of dilaton chiral perturbation theory to $N_f = 8$, SU(3) spectral data”, Phys. Rev. D **102**, 034515 (2020) (Cited on pp. 16, 144).

[150] M. Golterman and Y. Shamir, “Large-mass regime of the dilaton-pion low-energy effective theory”, Phys. Rev. D **98**, 056025 (2018) (Cited on pp. 16, 144).

[151] M. Golterman and Y. Shamir, “Explorations beyond dilaton chiral perturbation theory in the eight-flavor SU(3) gauge theory”, Phys. Rev. D **102**, 114507 (2020) (Cited on pp. 16, 144).

[152] S. Grable and P. Romatschke, “Elements of Confinement for QCD with Twelve Massless Quarks”, (2023) (Cited on pp. 99, 100).

[153] J. Greensite, *An introduction to the confinement problem*, Vol. 821 (2011) (Cited on pp. 14, 64).

[154] D. J. Gross and F. Wilczek, “Asymptotically Free Gauge Theories - I”, Phys. Rev. D **8**, 3633 (1973) (Cited on p. 5).

[155] D. J. Gross and F. Wilczek, “Ultraviolet Behavior of Nonabelian Gauge Theories”, Phys. Rev. Lett. **30**, edited by J. C. Taylor, 1343 (1973) (Cited on p. 5).

[156] Y. Guo and Y.-Z. You, “Symmetric mass generation of kähler-dirac fermions from the perspective of symmetry-protected topological phases”, Phys. Rev. B **108**, 115139 (2023) (Cited on pp. 16, 144).

[157] R. Gupta, “Monte carlo renormalization group: a review”, in *Lattice gauge theory: a challenge in large-scale computing*, edited by B. Bunk, K. H. Mütter, and K. Schilling (Springer US, Boston, MA, 1986), pp. 37–65 (Cited on p. 31).

[158] R. Gupta and C. F. Baillie, “Critical behavior of the two-dimensional xy model”, Phys. Rev. B **45**, 2883 (1992) (Cited on p. 137).

[159] M. Hamilton J. D., *Mathematical Gauge Theory: With Applications to the Standard Model of Particle Physics*, Universitext (Springer International Publishing, Cham, 2017) (Cited on p. 39).

[160] N. Hansen and A. Ostermeier, “Completely derandomized self-adaptation in evolution strategies”, Evolutionary Computation **9**, 159 (2001) (Cited on p. 126).

[161] R. V. Harlander and T. Neumann, "The perturbative QCD gradient flow to three loops", JHEP **06**, 161 (2016) (Cited on pp. 17, 18, 79, 80, 85, 91, 92, 96, 108, 110, 111, 159, 160, 162, 163).

[162] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy", Nature **585**, 357 (2020) (Cited on p. 183).

[163] M. Hasenbusch and K. Jansen, "Speeding up lattice QCD simulations with clover improved Wilson fermions", Nucl. Phys. B **659**, 299 (2003) (Cited on p. 173).

[164] M. Hasenbusch and S. Necco, "SU(3) lattice gauge theory with a mixed fundamental and adjoint plaquette action: Lattice artifacts", JHEP **08**, 005 (2004) (Cited on p. 249).

[165] M. Hasenbusch, "Speeding up the hybrid Monte Carlo algorithm for dynamical fermions", Phys. Lett. B **519**, 177 (2001) (Cited on p. 173).

[166] M. Hasenbusch, "The Two dimensional XY model at the transition temperature: A High precision Monte Carlo study", J. Phys. A **38**, 5869 (2005) (Cited on pp. 136, 138).

[167] A. Hasenfratz, C. Rebbi, and O. Witzel, "Nonperturbative determination of β functions for SU(3) gauge theories with 10 and 12 fundamental flavors using domain wall fermions", Phys. Lett. B **798**, 134937 (2019) (Cited on p. 99).

[168] A. Hasenfratz, "Conformal or Walking? Monte Carlo renormalization group studies of SU(3) gauge models with fundamental fermions", Phys. Rev. D **82**, 014506 (2010) (Cited on pp. 99, 143).

[169] A. Hasenfratz, "Infrared fixed point of the 12-fermion SU(3) gauge model based on 2-lattice MCRG matching", Phys. Rev. Lett. **108**, 061601 (2012) (Cited on p. 99).

[170] A. Hasenfratz, "Emergent strongly coupled ultraviolet fixed point in four dimensions with eight Kähler-Dirac fermions", Phys. Rev. D **106**, 014513 (2022) (Cited on pp. 16, 21, 144, 146, 148, 150, 151, 153, 165, 202).

[171] A. Hasenfratz and T. A. DeGrand, "Heavy dynamical fermions in lattice QCD", Phys. Rev. D **49**, 466 (1994) (Cited on p. 100).

[172] A. Hasenfratz and T. A. DeGrand, "The Role of heavy fermions", Nucl. Phys. B Proc. Suppl. **34**, 317 (1994) (Cited on p. 100).

[173] A. Hasenfratz and P. Hasenfratz, "The Connection Between the Lambda Parameters of Lattice and Continuum QCD", Phys. Lett. B **93**, edited by J. Julve and M. Ramón-Medrano, 165 (1980) (Cited on pp. 90, 94).

[174] A. Hasenfratz and P. Hasenfratz, "The Equivalence of the SU(N) Yang-Mills theory with a purely fermionic model", Phys. Lett. B **297**, 166 (1992) (Cited on p. 100).

[175] A. Hasenfratz, R. Hoffmann, and S. Schaefer, "Hypercubic smeared links for dynamical fermions", JHEP **05**, 029 (2007) (Cited on pp. 252, 253).

[176] A. Hasenfratz and F. Knechtli, "Flavor symmetry and the static potential with hypercubic blocking", Phys. Rev. D **64**, 034504 (2001) (Cited on p. 252).

[177] A. Hasenfratz, C. J. Monahan, M. D. Rizik, A. Shindler, and O. Witzel, "A novel nonperturbative renormalization scheme for local operators", PoS **LATTICE2021**, 155 (2022) (Cited on pp. 18, 34, 61, 62, 187, 188, 196).

[178] A. Hasenfratz, E. T. Neil, Y. Shamir, B. Svetitsky, and O. Witzel, "Infrared fixed point and anomalous dimensions in a composite Higgs model", Phys. Rev. D **107**, 114504 (2023) (Cited on pp. 61, 62, 101, 155, 162).

[179] A. Hasenfratz, E. T. Neil, Y. Shamir, B. Svetitsky, and O. Witzel, "Infrared fixed point of the SU(3) gauge theory with Nf=10 flavors", Phys. Rev. D **108**, L071503 (2023) (Cited on pp. 61, 62, 101, 155, 162, 163, 187, 200).

[180] A. Hasenfratz and C. T. Peterson, "Infrared fixed point in the massless twelve-flavor SU(3) gauge-fermion system", (2024) (Cited on pp. 61, 62).

[181] A. Hasenfratz, C. T. Peterson, J. van Sickle, and O. Witzel, "Λ Parameter of the su(3) yang-mills theory from the continuous β function", Phys. Rev. D **108**, 014502 (2023) (Cited on pp. 17, 53, 61–64, 70, 72–74, 77, 78, 80, 81, 85–87, 91, 92, 95).

[182] A. Hasenfratz, C. Rebbi, and O. Witzel, "Gradient flow step-scaling function for SU(3) with twelve flavors", Phys. Rev. D **100**, 114508 (2019) (Cited on p. 99).

[183] A. Hasenfratz, C. Rebbi, and O. Witzel, "Gradient flow step-scaling function for SU(3) with ten fundamental flavors", Phys. Rev. D **101**, 114508 (2020) (Cited on p. 60).

[184] A. Hasenfratz, C. Rebbi, and O. Witzel, "Gradient flow step-scaling function for SU(3) with Nf=6 or 4 fundamental flavors", Phys. Rev. D **106**, 114509 (2022) (Cited on p. 60).

[185] A. Hasenfratz, C. Rebbi, and O. Witzel, "Gradient flow step-scaling function for SU(3) with Nf=8 fundamental flavors", Phys. Rev. D **107**, 114508 (2023) (Cited on pp. 60, 143).

[186] A. Hasenfratz and D. Schaich, "Nonperturbative β function of twelve-flavor SU(3) gauge theory", JHEP **02**, 132 (2018) (Cited on pp. 19, 60, 99, 110–112, 116).

[187] A. Hasenfratz, D. Schaich, and A. Veernala, "Nonperturbative β function of eight-flavor SU(3) gauge theory", JHEP **06**, 143 (2015) (Cited on pp. 60, 143).

[188] A. Hasenfratz, Y. Shamir, and B. Svetitsky, "Taming lattice artifacts with Pauli-Villars fields", Phys. Rev. D **104**, 074509 (2021) (Cited on pp. 100, 101, 163, 165, 173).

[189] A. Hasenfratz and O. Witzel, "Continuous β function for the SU(3) gauge systems with two and twelve fundamental flavors", PoS **LATTICE2019**, 094 (2019) (Cited on pp. 17, 61, 62).

[190] A. Hasenfratz and O. Witzel, "Continuous renormalization group β function from lattice simulations", Phys. Rev. D **101**, 034514 (2020) (Cited on pp. 17, 61, 62).

[191] A. Hasenfratz and O. Witzel, "Dislocations under gradient flow and their effect on the renormalized coupling", Phys. Rev. D **103**, 034505 (2021) (Cited on p. 67).

[192] A. Hasenfratz and O. Witzel, "In preparation", (Cited on pp. 162, 163).

[193] P. Hasenfratz and F. Niedermayer, "Perfect lattice action for asymptotically free theories", Nucl. Phys. B **414**, 785 (1994) (Cited on pp. 50, 251).

[194] P. Hasenfratz, "MONTE CARLO RENORMALIZATION GROUP METHODS AND RESULTS IN QCD", in International School of Subnuclear Physics: Quarks, Leptons and their Constituents (1984) (Cited on pp. 25, 50).

[195] P. Hasenfratz, "Lattice QCD without tuning, mixing and current renormalization", Nucl. Phys. B **525**, 401 (1998) (Cited on p. 47).

[196] P. Hasenfratz, V. Laliena, and F. Niedermayer, "The Index theorem in QCD with a finite cutoff", Phys. Lett. B **427**, 125 (1998) (Cited on p. 47).

[197] D. Hendrycks and K. Gimpel, "Gaussian error linear units (gelus)", (2016) (Cited on p. 187).

[198] F. Herzog, B. Ruijl, T. Ueda, J. A. M. Vermaseren, and A. Vogt, "The five-loop beta function of Yang-Mills theory with fermions", JHEP **02**, 090 (2017) (Cited on p. 89).

[199] M. R. Hestenes, E. Stiefel, et al., *Methods of conjugate gradients for solving linear systems*, Vol. 49, 1 (NBS Washington, DC, 1952) (Cited on p. 172).

[200] C. T. Hill and E. H. Simmons, "Strong Dynamics and Electroweak Symmetry Breaking", Phys. Rept. **381**, [Erratum: Phys.Rept. 390, 553–554 (2004)], 235 (2003) (Cited on p. 14).

[201] A. E. Hoerl and R. W. Kennard, "Ridge regression: biased estimation for nonorthogonal problems", Technometrics **42**, 80 (2000) (Cited on p. 126).

[202] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky, "Bayesian model averaging: a tutorial", Statistical Science **14**, 382 (1999) (Cited on p. 240).

[203] B. Holdom, "Large N flavor beta-functions: a recap", Phys. Lett. B **694**, 74 (2011) (Cited on p. 6).

[204] K. Holland, A. Ipp, D. I. Müller, and U. Wenger, "Machine learning a fixed point action for SU(3) gauge theory with a gauge equivariant convolutional neural network", (2024) (Cited on pp. 25, 251, 253).

[205] T. J. Hollowood, *Renormalization group and fixed points: in quantum field theory*, Vol. 48 (Springer, 2013) (Cited on p. 51).

[206] N. Husung, P. Marquard, and R. Sommer, "Asymptotic behavior of cutoff effects in Yang–Mills theory and in Wilson's lattice QCD", Eur. Phys. J. C **80**, 200 (2020) (Cited on p. 246).

[207] S. Iino, S. Morita, N. Kawashima, and A. W. Sandvik, "Detecting signals of weakly first-order phase transitions in two-dimensional potts models", Journal of the Physical Society of Japan **88**, 034006 (2019) (Cited on p. 150).

[208] J. Ingoldby (Lattice Strong Dynamics), "Hidden Conformal Symmetry from Eight Flavors", in 40th International Symposium on Lattice Field Theory (2023) (Cited on pp. 16, 144).

[209] K.-I. Ishikawa, I. Kanamori, Y. Murakami, A. Nakamura, M. Okawa, and R. Ueno, "Non-perturbative determination of the Λ-parameter in the pure SU(3) gauge theory from the twisted gradient flow coupling", JHEP **12**, 067 (2017) (Cited on pp. 94, 95).

[210] W. I. Jay and E. T. Neil, "Bayesian model averaging for analysis of lattice field theory results", Phys. Rev. D **103**, 114502 (2021) (Cited on pp. 83, 106, 185, 240).

[211] E. T. Jaynes, "Information Theory and Statistical Mechanics", Physical Review **106**, 620 (1957) (Cited on p. 237).

[212] E. T. Jaynes, "Information Theory and Statistical Mechanics. II", Physical Review **108**, 171 (1957) (Cited on p. 237).

[213] A. Joseph, "Markov Chain Monte Carlo Methods in Quantum Field Theories: A Modern Primer", in , SpringerBriefs in Physics (2019) (Cited on p. 168).

[214] F. Joswig, S. Kuberski, J. T. Kuhlmann, and J. Neuendorf, "pyerrors: A python framework for error analysis of Monte Carlo data", Comput. Phys. Commun. **288**, 108750 (2023) (Cited on p. 103).

[215] F. Joswig, S. Kuberski, J. T. Kuhlmann, and J. Neuendorf, "pyerrors: A python framework for error analysis of Monte Carlo data", Computer Physics Communications **288**, 108750 (2023) (Cited on pp. 103, 155, 235).

[216] L. P. Kadanoff, “Scaling laws for ising models near T_c”, Physics Physique Fizika **2**, 263 (1966) (Cited on pp. 4, 36).

[217] D. B. Kaplan, “Chiral fermions on the lattice”, Nucl. Phys. B Proc. Suppl. **30**, 597 (1993) (Cited on p. 46).

[218] D. B. Kaplan, “Flavor at SSC energies: A New mechanism for dynamically generated fermion masses”, Nucl. Phys. B **365**, 259 (1991) (Cited on p. 15).

[219] D. B. Kaplan, “A Method for simulating chiral fermions on the lattice”, Phys. Lett. B **288**, 342 (1992) (Cited on p. 46).

[220] D. B. Kaplan, H. Georgi, and S. Dimopoulos, “Composite Higgs Scalars”, Phys. Lett. B **136**, 187 (1984) (Cited on p. 14).

[221] D. B. Kaplan, J.-W. Lee, D. T. Son, and M. A. Stephanov, “Conformality Lost”, Phys. Rev. D **80**, 125005 (2009) (Cited on pp. 12, 13).

[222] Kardar, Mehran, *Statistical physics of fields* (Cambridge University Press, 2007) (Cited on pp. 25, 26, 136).

[223] J. Kennedy and R. Eberhart, “Particle swarm optimization”, in Proceedings of icnn’95 - international conference on neural networks, Vol. 4 (1995), 1942–1948 vol.4 (Cited on p. 124).

[224] K. Kikuchi and T. Onogi, “Generalized Gradient Flow Equation and Its Application to Super Yang-Mills Theory”, JHEP **11**, 094 (2014) (Cited on p. 243).

[225] D. P. Kingma and J. Ba, “Adam: a method for stochastic optimization (2014)”, arXiv preprint arXiv:1412.6980 **15** (2017) (Cited on p. 124).

[226] M. Kitazawa, T. Iritani, M. Asakawa, T. Hatsuda, and H. Suzuki, “Equation of State for SU(3) Gauge Theory via the Energy-Momentum Tensor under Gradient Flow”, Phys. Rev. D **94**, 114512 (2016) (Cited on pp. 94, 95).

[227] F. Knechtli, M. Günther, and M. Peardon, *Lattice Quantum Chromodynamics: Practical Essentials*, SpringerBriefs in Physics (Springer, 2017) (Cited on pp. 39, 42, 90, 167–170, 230, 253).

[228] J. B. Kogut and L. Susskind, “Hamiltonian Formulation of Wilson’s Lattice Gauge Theories”, Phys. Rev. D **11**, 395 (1975) (Cited on p. 44).

[229] Y. Komura and Y. Okabe, “Large-scale monte carlo simulation of two-dimensional classical xy model using multiple gpus”, Journal of the Physical Society of Japan **81**, 113001 (2012) (Cited on pp. 136, 138, 141).

[230] P. Kopietz, L. Bartosch, and F. Schütz, *Introduction to the functional renormalization group*, Vol. 798 (2010) (Cited on pp. 25–27, 37, 229).

[231] J. M. Kosterlitz, “The critical properties of the two-dimensional xy model”, Journal of Physics C: Solid State Physics **7**, 1046 (1974) (Cited on p. 136).

[232] J. M. Kosterlitz and D. J. Thouless, “Ordering, metastability and phase transitions in two-dimensional systems”, J. Phys. C **6**, 1181 (1973) (Cited on pp. 136–138).

[233] M. Kucharík, I. L. Hofacker, P. F. Stadler, and J. Qin, “Basin Hopping Graph: a computational framework to characterize RNA folding landscapes”, Bioinformatics **30**, 2009 (2014) (Cited on p. 126).

[234] S. Kullback and R. A. Leibler, “On information and sufficiency”, The Annals of Mathematical Statistics **22**, 79 (1951) (Cited on p. 237).

[235] J. Kuti, Z. Fodor, K. Holland, and C. H. Wong, “From ten-flavor tests of the β-function to α_s at the Z-pole”, PoS **LATTICE2021**, 321 (2022) (Cited on pp. 61, 62).

[236] K. Lane, “Two Lectures on Technicolor”, (2002) (Cited on pp. 14, 15).

[237] J.-W. Lee, “Conformal window from conformal expansion”, Phys. Rev. D **103**, 076006 (2021) (Cited on p. 99).

[238] J. K. L. Lee, E. Dobson, A. Jüttner, B. Kitching-Morley, V. Nourry, A. Portelli, H. B. Rocha, K. Skenderis, and L. Del Debbio, “Renormalisation of the 3D SU(N) scalar energy-momentum tensor using the Wilson flow”, PoS **LATTICE2021**, 313 (2022) (Cited on p. 55).

[239] G. P. Lepage, B. Clark, C. T. H. Davies, K. Hornbostel, P. B. Mackenzie, C. Morningstar, and H. Trottier (HPQCD), “Constrained curve fitting”, Nucl. Phys. B Proc. Suppl. **106**, edited by M. Muller-Preussker, W. Bietenholz, K. Jansen, F. Jegerlehner, I. Montvay, G. Schierholz, R. Sommer, and U. Wolff, 12 (2002) (Cited on pp. 127, 239).

[240] P. Lepage, *Lsqfit*, 2014 (Cited on pp. 69, 104, 154, 186).

[241] P. Lepage, *Gvar*, 2015 (Cited on pp. 69, 86, 93, 104, 113, 114, 154, 181, 183, 235).

[242] K. LEVENBERG, “A method for the solution of certain non-linear problems in least squares”, Quarterly of Applied Mathematics **2**, 164 (1944) (Cited on pp. 125, 183).

[243] Z. Li and D. Poland, “Searching for gauge theories with the conformal bootstrap”, JHEP **03**, 172 (2021) (Cited on p. 99).

[244] C. .-. D. Lin, K. Ogawa, H. Ohki, and E. Shintani, “Lattice study of infrared behaviour in SU(3) gauge theory with twelve massless flavours”, JHEP **08**, 096 (2012) (Cited on pp. 60, 99).

[245] C. .-. D. Lin, K. Ogawa, and A. Ramos, “The Yang-Mills gradient flow and SU(3) gauge theory with 12 massless fundamental fermions in a colour-twisted box”, JHEP **12**, 103 (2015) (Cited on pp. 60, 99, 100).

[246] M. P. Lombardo, K. Miura, T. J. N. da Silva, and E. Pallante, “On the particle spectrum and the conformal window”, JHEP **12**, 183 (2014) (Cited on p. 99).

[247] M. Luscher, “Some Analytic Results Concerning the Mass Spectrum of Yang-Mills Gauge Theories on a Torus”, Nucl. Phys. B **219**, 233 (1983) (Cited on p. 57).

[248] M. Luscher and P. Weisz, “Computation of the Action for On-Shell Improved Lattice Gauge Theories at Weak Coupling”, Phys. Lett. B **158**, 250 (1985) (Cited on pp. 59, 62, 155, 166, 189).

[249] M. Luscher, “Chiral symmetry and the Yang–Mills gradient flow”, JHEP **04**, 123 (2013) (Cited on pp. 187, 188).

[250] M. Luscher and P. Weisz, “Perturbative analysis of the gradient flow in non-abelian gauge theories”, JHEP **02**, 051 (2011) (Cited on pp. 55, 56).

[251] M. Lüscher and P. Weisz, “On-Shell Improved Lattice Gauge Theories”, Commun. Math. Phys. **97**, [Erratum: Commun. Math. Phys.98,433(1985)], 59 (1985) (Cited on pp. 62, 189, 246–248).

[252] M. Lüscher, “Properties and uses of the Wilson flow in lattice QCD”, JHEP **1008**, 071 (2010) (Cited on pp. 17, 18, 53, 54, 56, 66, 86, 87, 91, 95, 179).

[253] M. Lüscher, “Trivializing maps, the Wilson flow and the HMC algorithm”, Commun.Math.Phys. **293**, 899 (2010) (Cited on pp. 53, 54).

[254] M. Lüscher, "Step scaling and the Yang-Mills gradient flow", JHEP **06**, 105 (2014) (Cited on pp. 53, 54, 60).

[255] S.-k. Ma, "Introduction to the renormalization group", Rev. Mod. Phys. **45**, 589 (1973) (Cited on pp. 25, 27).

[256] H. Makino, O. Morikawa, and H. Suzuki, "Gradient flow and the Wilsonian renormalization group flow", PTEP **2018**, 053B02 (2018) (Cited on pp. 31, 55, 56).

[257] H. Makino and H. Suzuki, "Lattice energy–momentum tensor from the Yang–Mills gradient flow—inclusion of fermion fields", PTEP **2014**, [Erratum: PTEP 2015, 079202 (2015)], 063B02 (2014) (Cited on p. 55).

[258] H. Makino and H. Suzuki, "Renormalizability of the gradient flow in the 2D $O(N)$ non-linear sigma model", PTEP **2015**, 033B08 (2015) (Cited on pp. 32, 243).

[259] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters", Journal of the Society for Industrial and Applied Mathematics **11**, 431 (1963) (Cited on pp. 125, 183).

[260] N. D. Mermin and H. Wagner, "Absence of ferromagnetism or antiferromagnetism in one-dimensional or two-dimensional isotropic Heisenberg models", Phys. Rev. Lett. **17**, 1133 (1966) (Cited on p. 136).

[261] V. A. Miransky and K. Yamawaki, "Conformal phase transition in gauge theories", Phys. Rev. D **55**, [Erratum: Phys.Rev.D 56, 3768 (1997)], 5051 (1997) (Cited on pp. 12, 13).

[262] K. Miura and M. P. Lombardo, "Lattice Monte-Carlo study of pre-conformal dynamics in strongly flavoured QCD in the light of the chiral phase transition at finite temperature", Nucl. Phys. B **871**, 52 (2013) (Cited on p. 143).

[263] Y. Miyakawa, H. Sonoda, and H. Suzuki, "Manifestly gauge invariant exact renormalization group for quantum electrodynamics", PTEP **2022**, 023B02 (2022) (Cited on p. 55).

[264] Y. Miyakawa, H. Sonoda, and H. Suzuki, "Chiral anomaly as a composite operator in the gradient flow exact renormalization group formalism", PTEP **2023**, 063B03 (2023) (Cited on p. 55).

[265] I. Montvay and G. Munster, *Quantum fields on a lattice*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, 1997) (Cited on pp. 39, 40, 45, 88).

[266] C. Morningstar and M. J. Peardon, "Analytic smearing of SU(3) link variables in lattice QCD", Phys. Rev. D **69**, 054501 (2004) (Cited on pp. 189, 253).

[267] Y. Motoyama, M. Piibeleht, and S. Karpinski, *Spinmontecarlo*, `https://github.com/yomichi/SpinMonteCarlo.jl`, 2019 (Cited on pp. 132, 136).

[268] K. P. Murphy, *Probabilistic machine learning: advanced topics* (MIT Press, 2023) (Cited on pp. 126, 168, 236).

[269] A. Nada and A. Ramos, "An analysis of systematic effects in finite size scaling studies using the gradient flow", Eur. Phys. J. C **81**, 1 (2021) (Cited on pp. 60, 94, 95).

[270] M. Nakahara, *Geometry, topology and physics* (2003) (Cited on pp. 39, 65).

[271] Y. Nakamura and G. Schierholz, "The strong CP problem solved by itself due to long-distance vacuum effects", Nucl. Phys. B **986**, 116063 (2023) (Cited on pp. 86–88).

[272] R. Narayanan and H. Neuberger, "Infinite N phase transitions in continuum Wilson loop operators", JHEP **0603**, 064 (2006) (Cited on p. 53).

[273] E. T. Neil and J. W. Sitison, “Model averaging approaches to data subset selection”, Phys. Rev. E **108**, 045308 (2023) (Cited on pp. 83, 106).

[274] E. T. Neil and J. W. Sitison, “Improved information criteria for Bayesian model averaging in lattice field theory”, Phys. Rev. D **109**, 014510 (2024) (Cited on pp. 83, 106).

[275] H. Neuberger, “More about exactly massless quarks on the lattice”, Phys. Lett. B **427**, 353 (1998) (Cited on p. 47).

[276] H. Neuberger, “Vector - like gauge theories with almost massless fermions on the lattice”, Phys. Rev. D **57**, 5417 (1998) (Cited on p. 47).

[277] P. H. Nguyen and M. Boninsegni, “Superfluid transition and specific heat of the 2d x-y model: monte carlo simulation”, Applied Sciences **11**, `10.3390/app11114931` (2021) (Cited on pp. 136, 138, 141).

[278] B. Nienhuis and M. Nauenberg, “First-order phase transitions in renormalization-group theory”, Phys. Rev. Lett. **35**, 477 (1975) (Cited on pp. 149, 229).

[279] V. A. Novikov, M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, “Exact Gell-Mann-Low Function of Supersymmetric Yang-Mills Theories from Instanton Calculus”, Nucl. Phys. B **229**, 381 (1983) (Cited on pp. 12, 88).

[280] T. Nunes da Silva and E. Pallante, “The strong coupling regime of twelve flavors QCD”, PoS **LATTICE2012**, edited by D. Leinweber, W. Kamleh, S. Mahbub, H. Matevosyan, A. Thomas, A. G. Williams, R. Young, and J. Zanotti, 052 (2012) (Cited on p. 100).

[281] B. Olson, I. Hashmi, K. Molloy, and A. Shehu, “Basin hopping as a general and versatile optimization framework for the characterization of biological macromolecules. advances in artificial intelligence, 2012, 1–19”, Advances in Artificial Intelligence, `10.1155/2012/674832` (2012) (Cited on p. 126).

[282] I. Omelyan, I. Mryglod, and R. Folk, “Symplectic analytically integrable decomposition algorithms: classification, derivation, and application to molecular dynamics, quantum and celestial mechanics simulations”, Computer Physics Communications **151**, 272 (2003) (Cited on pp. 169, 256).

[283] G. Ortiz, E. Cobanera, and Z. Nussinov, “Dualities and the phase diagram of the p-clock model”, Nucl. Phys. B **854**, 780 (2012) (Cited on p. 134).

[284] J. Osborn and X.-Y. Jin, “Introduction to the Quantum EXpressions (QEX) framework”, PoS **LATTICE2016**, 271 (2017) (Cited on pp. 23, 136, 165, 166, 245, 252, 253).

[285] K. Osterwalder and R. Schrader, “AXIOMS FOR EUCLIDEAN GREEN’S FUNCTIONS”, Commun. Math. Phys. **31**, 83 (1973) (Cited on p. 49).

[286] S. Ota, S. B. Ota, and M. Fahnle, “Microcanonical monte carlo simulations for the two-dimensional xy model”, Journal of Physics: Condensed Matter **4**, 5411 (1992) (Cited on p. 137).

[287] G. Panico and A. Wulzer, *The Composite Nambu-Goldstone Higgs*, Vol. 913 (Springer, 2016) (Cited on pp. 14, 16).

[288] J. Park and I. W. Sandberg, “Universal approximation using radial-basis-function networks”, Neural Computation **3**, 246 (1991) (Cited on p. 122).

[289] R. Penrose, “On best approximate solutions of linear matrix equations”, Mathematical Proceedings of the Cambridge Philosophical Society **52**, 17 (1956) (Cited on p. 183).

[290] C. Peterson, *SwissFit* (Cited on pp. 104, 133, 154, 165, 181).

[291] C. Peterson and A. Hasenfratz, *Twelve flavor SU(3) gradient flow data for the continuous beta-function* (Zenodo, 2024) (Cited on p. 200).

[292] C. T. Peterson and A. Hasenfratz, "Constrained curve fitting for semi-parametric models with radial basis function networks", (2024) (Cited on pp. 119, 121, 124, 128–131, 134–136, 139, 140).

[293] C. T. Peterson and A. Hasenfratz, "Infrared fixed point in the massless twelve-flavor SU(3) gauge-fermion system", (2024) (Cited on pp. 17, 98, 101, 104, 105, 108–110, 113, 116).

[294] C. T. Peterson, A. Hasenfratz, J. van Sickle, and O. Witzel, "Determination of the continuous β function of SU(3) Yang-Mills theory", in 38th International Symposium on Lattice Field Theory (2021) (Cited on pp. 53, 60–62).

[295] P. Petreczky and H. .-. Schadler, "Renormalization of the Polyakov loop with gradient flow", Phys. Rev. D **92**, 094517 (2015) (Cited on p. 65).

[296] D. L. Phillips, "A technique for the numerical solution of certain integral equations of the first kind", J. ACM **9**, 84 (1962) (Cited on p. 126).

[297] M. Piai, "Lectures on walking technicolor, holography and gauge/gravity dualities", Adv. High Energy Phys. **2010**, 464302 (2010) (Cited on p. 15).

[298] C. Pica and F. Sannino, "UV and IR Zeros of Gauge Theories at The Four Loop Order and Beyond", Phys. Rev. D **83**, 035013 (2011) (Cited on pp. 99, 115).

[299] D. Picheta, *Nim in action*, 1st (Manning Publications Co., USA, 2017) (Cited on p. 166).

[300] A. Pochinsky et al., *Qlua*, 2008 (Cited on pp. 63, 188, 189).

[301] H. D. Politzer, "Reliable Perturbative Results for Strong Interactions?", Phys. Rev. Lett. **30**, edited by J. C. Taylor, 1346 (1973) (Cited on p. 5).

[302] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes 3rd edition: the art of scientific computing* (Cambridge university press, 2007) (Cited on pp. 68, 93).

[303] V. Privman, *Finite size scaling and numerical simulation of statistical systems* (World Scientific, 1990) (Cited on pp. 119, 132).

[304] *Python package index - pypi*, `https://pypi.org/` (Cited on p. 181).

[305] A. Ramos and S. Sint, "Symanzik improvement of the gradient flow in lattice gauge theories", Eur. Phys. J. C **76**, 15 (2016) (Cited on p. 179).

[306] A. Ramos, "The gradient flow running coupling with twisted boundary conditions", JHEP **11**, 101 (2014) (Cited on p. 60).

[307] A. Ramos, "The Yang-Mills gradient flow and renormalization", PoS **LATTICE2014**, 017 (2015) (Cited on p. 60).

[308] A. Ramos, "Automatic differentiation for error analysis", PoS **TOOLS2020**, 045 (2021) (Cited on p. 103).

[309] M. e. a. Ratsimbazafy, *Arraymancer*, 2021 (Cited on p. 245).

[310] C. Ratti and R. Bellwied, *The Deconfinement Transition of QCD: Theory Meets Experiment*, Vol. 981, Lecture Notes in Physics (2021) (Cited on p. 64).

[311] M. Reuter and F. Saueressig, *Quantum Gravity and the Functional Renormalization Group: The Road towards Asymptotic Safety* (Cambridge University Press, 2019) (Cited on p. 6).

[312] P. Romatschke, "An alternative to perturbative renormalization in 3+1 dimensional field theories", (2024) (Cited on pp. 99, 100).

[313] H. J. Rothe, *Lattice Gauge Theories : An Introduction (Fourth Edition)*, Vol. 43 (World Scientific Publishing Company, 2012) (Cited on pp. 88, 230).

[314] S. Rychkov, *EPFL Lectures on Conformal Field Theory in D>= 3 Dimensions*, SpringerBriefs in Physics (2016) (Cited on p. 7).

[315] T. A. Ryttov and F. Sannino, "Supersymmetry inspired QCD beta function", Phys. Rev. D **78**, 065001 (2008) (Cited on pp. 12, 88).

[316] T. A. Ryttov and R. Shrock, "Higher-loop corrections to the infrared evolution of a gauge theory with fermions", Phys. Rev. **D83**, 056011 (2011) (Cited on p. 99).

[317] T. A. Ryttov and R. Shrock, "An Analysis of Scheme Transformations in the Vicinity of an Infrared Fixed Point", Phys. Rev. D **86**, 085005 (2012) (Cited on p. 97).

[318] T. A. Ryttov and R. Shrock, "Scheme Transformations in the Vicinity of an Infrared Fixed Point", Phys. Rev. D **86**, 065032 (2012) (Cited on p. 97).

[319] T. A. Ryttov and R. Shrock, "Infrared Zero of β and Value of γ_m for an SU(3) Gauge Theory at the Five-Loop Level", Phys. Rev. **D94**, 105015 (2016) (Cited on p. 99).

[320] T. A. Ryttov and R. Shrock, "Scheme-Independent Series Expansions at an Infrared Zero of the Beta Function in Asymptotically Free Gauge Theories", Phys. Rev. D **94**, 125005 (2016) (Cited on p. 99).

[321] T. A. Ryttov and R. Shrock, "Higher-order scheme-independent series expansions of $\gamma_{\bar{\psi}\psi,IR}$ and β'_{IR} in conformal field theories", Phys. Rev. **D95**, 105004 (2017) (Cited on pp. 19, 99, 116).

[322] T. A. Ryttov and R. Shrock, "Effect of Scheme Transformations on a Beta Function with Vanishing One-Loop Term", Phys. Rev. D **102**, 056016 (2020) (Cited on pp. 96, 97).

[323] F. Sannino, *Dynamical stabilization of the fermi scale, Towards a composite universe* (Springer, 2013) (Cited on pp. 11, 13–15).

[324] F. Sannino, "Jumping Dynamics", Mod. Phys. Lett. A **28**, 1350127 (2013) (Cited on p. 11).

[325] S. Schaefer, R. Sommer, and F. Virotta (ALPHA), "Critical slowing down and error analysis in lattice QCD simulations", Nucl. Phys. B **845**, 93 (2011) (Cited on p. 103).

[326] T. Schäfer and E. V. Shuryak, "Instantons in QCD", Rev. Mod. Phys. **70**, 323 (1998) (Cited on p. 65).

[327] D. Schaich, A. Cheng, A. Hasenfratz, and G. Petropoulos, "Bulk and finite-temperature transitions in SU(3) gauge theories with many light fermions", PoS **LATTICE2012**, edited by D. Leinweber, W. Kamleh, S. Mahbub, H. Matevosyan, A. Thomas, A. G. Williams, R. Young, and J. Zanotti, 028 (2012) (Cited on p. 100).

[328] G. Schierholz, "Dynamical solution of the strong CP problem within QCD?", EPJ Web Conf. **274**, 01009 (2022) (Cited on pp. 86–88).

[329] E. Seiler, *Gauge Theories as a Problem of Constructive Quantum Field Theory and Statistical Mechanics*, Vol. 159 (1982) (Cited on p. 39).

[330] J. C. Sexton and D. H. Weingarten, “Hamiltonian evolution for the hybrid Monte Carlo algorithm”, Nucl. Phys. B **380**, 665 (1992) (Cited on p. 255).

[331] Y. Shamir, “Chiral fermions from lattice boundaries”, Nucl. Phys. B **406**, 90 (1993) (Cited on p. 46).

[332] Y. Shamir and M. Golterman, “Dilaton chiral perturbation theory and application”, PoS **LATTICE2021**, 372 (2022) (Cited on p. 144).

[333] M. A. Shifman and A. I. Vainshtein, “Solution of the Anomaly Puzzle in SUSY Gauge Theories and the Wilson Operator Expansion”, Nucl. Phys. B **277**, 456 (1986) (Cited on pp. 12, 88).

[334] R. Shrock, “Generalized Scheme Transformations for the Elimination of Higher-Loop Terms in the Beta Function of a Gauge Theory”, Phys. Rev. D **90**, 045011 (2014) (Cited on p. 97).

[335] J. Smit, *Introduction to Quantum Fields on a Lattice*, Vol. 15, Cambridge Lecture Notes in Physics (Cambridge University Press, 2023) (Cited on pp. 90, 230, 231).

[336] R. Sommer, “Scale setting in lattice QCD”, PoS **LATTICE2013**, 015 (2014) (Cited on pp. 52, 91).

[337] H. Sonoda and H. Suzuki, “Derivation of a gradient flow from the exact renormalization group”, PTEP **2019**, 033B05 (2019) (Cited on p. 55).

[338] H. Sonoda and H. Suzuki, “Gradient flow exact renormalization group”, PTEP **2021**, 023B05 (2021) (Cited on p. 55).

[339] H. Sonoda and H. Suzuki, “One-particle irreducible Wilson action in the gradient flow exact renormalization group formalism”, PTEP **2022**, 053B01 (2022) (Cited on p. 55).

[340] F. Springer, D. Schaich, and E. Rinaldi (Lattice Strong Dynamics (LSD)), “First-order bulk transitions in large-N lattice Yang–Mills theories using the density of states”, (2023) (Cited on p. 100).

[341] P. S. Stanimirović, B. Ivanov, H. Ma, and D. Mosić, “A survey of gradient methods for solving nonlinear optimization”, Electronic Research Archive **28**, 1573 (2020) (Cited on p. 125).

[342] M. Steffen, “A simple method for monotonic interpolation in one dimension.”, **239**, 443 (1990) (Cited on pp. 86, 93, 113, 114, 127).

[343] R. Storn and K. Price, “Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces”, Comput. Phys. Commun. **11**, 341 (1997) (Cited on p. 124).

[344] R. F. Streater and A. S. Wightman, *PCT, spin and statistics, and all that* (1989) (Cited on p. 49).

[345] H. Suzuki, “Energy–momentum tensor from the Yang–Mills gradient flow”, PTEP **2013**, [Erratum: PTEP 2015, 079201 (2015)], 083B03 (2013) (Cited on p. 55).

[346] B. Svetitsky, “Looking behind the Standard Model with lattice gauge theory”, EPJ Web Conf. **175**, edited by M. Della Morte, P. Fritzsch, E. Gámiz Sánchez, and C. Pena Ruano, 01017 (2018) (Cited on p. 142).

[347] R. H. Swendsen, “Monte Carlo Renormalization Group”, Phys. Rev. Lett. **42**, 859 (1979) (Cited on p. 31).

[348] R. H. Swendsen and J.-S. Wang, “Nonuniversal critical dynamics in monte carlo simulations”, Phys. Rev. Lett. **58**, 86 (1987) (Cited on p. 264).

[349] K. Symanzik, “Small distance behavior in field theory and power counting”, Commun. Math. Phys. **18**, 227 (1970) (Cited on pp. 3, 33).

[350] K. Symanzik, “Small distance behavior analysis and Wilson expansion”, Commun. Math. Phys. **23**, 49 (1971) (Cited on p. 33).

[351] K. Symanzik, “Continuum Limit and Improved Action in Lattice Theories. 1. Principles and φ^4 Theory”, Nucl. Phys. B **226**, 187 (1983) (Cited on p. 246).

[352] K. Symanzik, “Continuum Limit and Improved Action in Lattice Theories. 2. O(N) Nonlinear Sigma Model in Perturbation Theory”, Nucl. Phys. B **226**, 205 (1983) (Cited on p. 246).

[353] T. Takaishi and P. de Forcrand, “Testing and tuning new symplectic integrators for hybrid Monte Carlo algorithm in lattice QCD”, Phys. Rev. E **73**, 036706 (2006) (Cited on pp. 169, 255–258).

[354] A. N. Tikhonov, “On the solution of ill-posed problems and the method of regularization”, in Doklady akademii nauk, Vol. 151, 3 (Dokl. Akad. Nauk SSSR, 1963), pp. 501–504 (Cited on p. 126).

[355] D. Tong, “Comments on symmetric mass generation in 2d and 4d”, JHEP **07**, 001 (2022) (Cited on p. 144).

[356] G. Toulouse and P. Pfeuty, *Introduction to the renormalization group and to critical phenomena*, A Wiley-interscience publication (Wiley, 1977) (Cited on pp. 25, 26).

[357] C. Tsallis and D. A. Stariolo, “Generalized simulated annealing”, Physica A: Statistical Mechanics and its Applications **233**, 395 (1996) (Cited on p. 124).

[358] L. M. Tuan, T. T. Long, D. X. Nui, P. T. Minh, N. D. Trung Kien, and D. X. Viet, “Binder ratio in the two-dimensional q-state clock model”, Phys. Rev. E **106**, 034138 (2022) (Cited on p. 141).

[359] J. E. Van Himbergen and S. Chakravarty, “Helicity modulus and specific heat of classical XY model in two dimensions”, Phys. Rev. B **23**, 359 (1981) (Cited on p. 141).

[360] D. Vats, J. M. Flegal, and G. L. Jones, “Multivariate output analysis for Markov chain Monte Carlo”, Biometrika **106**, 321 (2019) (Cited on p. 234).

[361] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”, Nature Methods **17**, 261 (2020) (Cited on pp. 69, 124, 125, 183, 184).

[362] C. Voglis and I. Lagaris, “A rectangular trust region dogleg approach for unconstrained and bound constrained nonlinear optimization”, in Wseas international conference on applied mathematics, Vol. 7 (2004) (Cited on p. 183).

[363] P. M. Vranas, “Domain wall fermions and applications”, Nucl. Phys. B Proc. Suppl. **94**, edited by T. Bhattacharya, R. Gupta, and A. Patel, 177 (2001) (Cited on p. 46).

[364] D. J. Wales and J. P. K. Doye, “Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms”, The Journal of Physical Chemistry A **101**, 5111 (1997) (Cited on pp. 124–126, 183).

[365] D. J. Wales and H. A. Scheraga, "Global optimization of clusters, crystals, and biomolecules", Science **285**, 1368 (1999) (Cited on p. 126).

[366] J. Wang and X.-G. Wen, "Nonperturbative regularization of $(1+1)$-dimensional anomaly-free chiral fermions and bosons: on the equivalence of anomaly matching conditions and boundary gapping rules", Phys. Rev. B **107**, 014311 (2023) (Cited on pp. 16, 144).

[367] J. Wang and Y.-Z. You, "Symmetric mass generation", Symmetry **14**, `10.3390/sym14071475` (2022) (Cited on pp. 16, 144).

[368] S. Weinberg, "Pion scattering lengths", Phys. Rev. Lett. **17**, 616 (1966) (Cited on p. 10).

[369] P. Weisz, "Continuum Limit Improved Lattice Action for Pure Yang-Mills Theory. 1.", Nucl. Phys. B **212**, 1 (1983) (Cited on pp. 59, 155, 166, 246).

[370] P. Weisz and R. Wohlert, "Continuum Limit Improved Lattice Action for Pure Yang-Mills Theory. 2.", Nucl. Phys. B **236**, [Erratum: Nucl.Phys.B 247, 544 (1984)], 397 (1984) (Cited on pp. 59, 155, 166).

[371] B. Widom, "Equation of State in the Neighborhood of the Critical Point", The Journal of Chemical Physics **43**, 3898 (1965) (Cited on p. 36).

[372] K. G. Wilson, "The renormalization group and critical phenomena", Rev. Mod. Phys. **55**, 583 (1983) (Cited on p. 4).

[373] K. G. Wilson, "Renormalization group and critical phenomena. 1. Renormalization group and the Kadanoff scaling picture", Phys. Rev. B **4**, 3174 (1971) (Cited on p. 4).

[374] K. G. Wilson, "Renormalization group and critical phenomena. 2. Phase space cell analysis of critical behavior", Phys. Rev. B **4**, 3184 (1971) (Cited on p. 4).

[375] K. G. Wilson, "Confinement of Quarks", Phys. Rev. D **10**, edited by J. C. Taylor, 2445 (1974) (Cited on pp. 40, 41, 46).

[376] O. Witzel, A. Hasenfratz, and C. T. Peterson (Lattice Strong Dynamics), "Composite Higgs scenario in mass-split models", PoS **ICHEP2020**, 675 (2021) (Cited on p. 197).

[377] U. Wolff, "Collective Monte Carlo Updating for Spin Systems", Phys. Rev. Lett. **62**, 361 (1989) (Cited on pp. 132, 136, 264).

[378] U. Wolff (ALPHA), "Monte Carlo errors with less errors", Comput. Phys. Commun. **156**, [Erratum: Comput.Phys.Commun. 176, 383 (2007)], 143 (2004) (Cited on pp. 103, 155, 234).

[379] D. Wolpert and W. Macready, "No free lunch theorems for optimization", IEEE Transactions on Evolutionary Computation **1**, 67 (1997) (Cited on p. 124).

[380] C. H. Wong, S. Borsanyi, Z. Fodor, K. Holland, and J. Kuti, "Toward a novel determination of the strong QCD coupling at the Z-pole", PoS **LATTICE2022**, 043 (2023) (Cited on pp. 18, 61, 62, 94, 95).

[381] R. L. Workman et al. (Particle Data Group), "Review of Particle Physics", PTEP **2022**, 083C01 (2022) (Cited on pp. 30, 52, 90).

[382] F. Y. Wu, "The Potts model", Rev. Mod. Phys. **54**, [Erratum: Rev.Mod.Phys. 55, 315–315 (1983)], 235 (1982) (Cited on pp. 131, 132, 136, 137).

[383] Y. Xiang, D. Sun, W. Fan, and X. Gong, "Generalized simulated annealing algorithm and its application to the thomson model", Physics Letters A **233**, 216 (1997) (Cited on p. 124).

[384] Y. Xiang and X. G. Gong, “Efficiency of generalized simulated annealing”, Phys. Rev. E **62**, 4473 (2000) (Cited on p. 124).

[385] R. Yoneda and K. Harada, “Neural network approach to scaling analysis of critical phenomena”, Phys. Rev. E **107**, 044128 (2023) (Cited on p. 133).

[386] M. Zeng, Z. Zhu, J. Wang, and Y.-Z. You, “Symmetric mass generation in the 1+1 dimensional chiral fermion 3-4-5-0 model”, Phys. Rev. Lett. **128**, 185301 (2022) (Cited on pp. 16, 144).

[387] J. Zhang and A. C. Sanderson, “Jade: adaptive differential evolution with optional external archive”, IEEE Transactions on Evolutionary Computation **13**, 945 (2009) (Cited on p. 124).

[388] C. Zhou, C. Ieritano, and W. S. Hopkins, “Augmenting basin-hopping with techniques from unsupervised machine learning: applications in spectroscopy and ion mobility”, Frontiers in Chemistry **7**, `10.3389/fchem.2019.00519` (2019) (Cited on p. 126).

[389] J. Zinn-Justin, *Quantum field theory and critical phenomena*, Vol. 77, International Series of Monographs on Physics (Oxford University Press, 2021) (Cited on p. 55).

[390] R. Zwicky, “QCD with an infrared fixed point: The pion sector”, Phys. Rev. D **109**, 034009 (2024) (Cited on p. 144).

Appendix A

Characterizing a fixed point by its local topology under an RG flow

Defining $\delta \boldsymbol{K} \equiv \boldsymbol{K} - \boldsymbol{K}^\star$ and taking $\delta K_i / K_i^\star \ll 1$, we can linearize $\mathbf{R}_b$ about $\boldsymbol{K}^\star$ as

$$\mathbf{R}_b(\delta \boldsymbol{K}) = \boldsymbol{\mathfrak{R}}_b \delta \boldsymbol{K} + \mathrm{O}\big(\delta \boldsymbol{K}^2\big), \tag{A.1}$$

where

$$\mathfrak{R}_b^{ij} = \left.\frac{\partial \mathrm{R}_b(\boldsymbol{K})_i}{\partial K_j}\right|_{\boldsymbol{K}=\boldsymbol{K}^\star}. \tag{A.2}$$

is a linear operator. The left *normalized* eigenvectors $\mathbf{e}_{\mathrm{L}}^{(\alpha)}$ of $\boldsymbol{\mathfrak{R}}_b$, defined by

$$\mathbf{e}_{\mathrm{L}}^{(\alpha)} \boldsymbol{\mathfrak{R}}_b = \mathbf{e}_{\mathrm{L}}^{(\alpha)} \lambda_b^{(\alpha)}, \tag{A.3}$$

characterize the local topology of the flow within the vicinity of $\boldsymbol{K}^\star$, as they are the principal axes of $\boldsymbol{\mathfrak{R}}_b$. The component δK_α of $\delta \boldsymbol{K}$ along any one of the principal axes is referred to as a *scaling variable*. Due to the semi-group composition law of Eqn. 1,

$$\lambda_{b'}^{(\alpha)} \lambda_b^{(\alpha)} = \lambda_{b'b}^{(\alpha)}, \tag{A.4}$$

which implies

$$\lambda_b^{(\alpha)} = b^{y_\alpha}, \tag{A.5}$$

for some b-independent variable y_α, which is imprecisely called a *renormalization group eigenvalue*. Therefore, the scaling variables δK_α transform under an RG transformation as

$$\delta K'_\alpha = b^{y_\alpha} \delta K_\alpha + \mathrm{O}\big(K_\alpha^2\big). \tag{A.6}$$

Whether or not the component δK_α of $\delta \boldsymbol{K}$ along the principal axis $\mathbf{e}_{(\alpha)}$ flows into or out of the fixed point $\boldsymbol{K}^\star$ is completely determined by the sign of y_α. If $y_\alpha > 0$, then it is repelled; in this case, we say that δK_α is *relevenat*. On the other hand, if $y_\alpha < 0$, it is attracted to $\delta K_\alpha = 0$, and we refer to δK_α as *irrelevant*. If $y_\alpha = 0$, we refer to δK_α as *marginal* and higher-order terms in Eqn. A.6 are required to determine whether or not δK_α is *marginally relevant* or *marginally irrelevant*.

Appendix B

Global topology of an RG flow

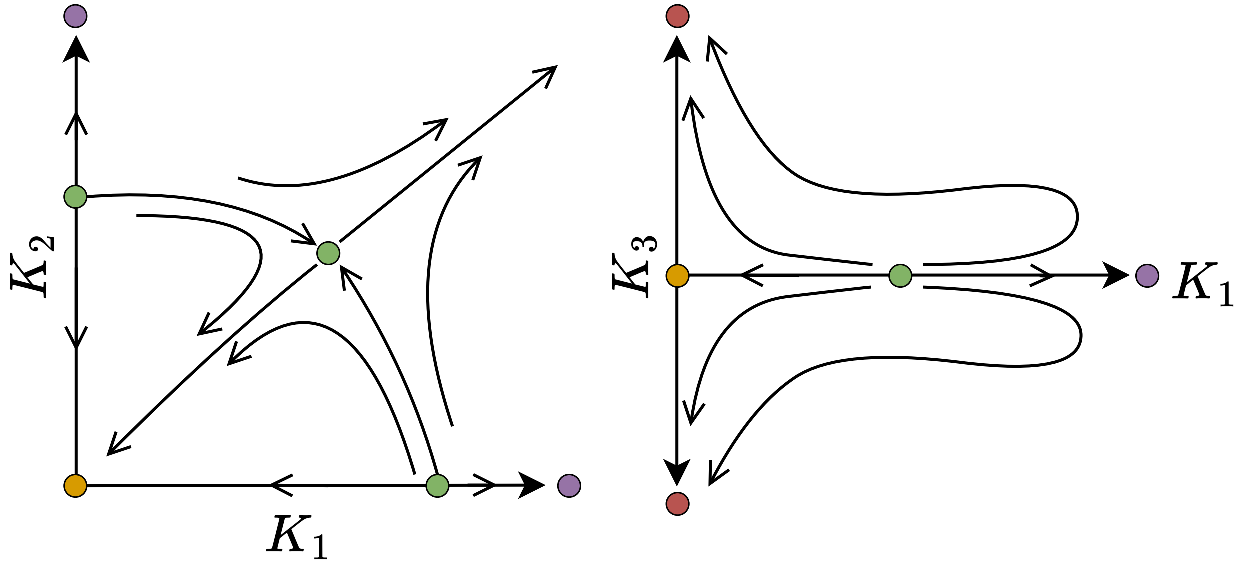

Figure B.1: Renormalization group flow diagram of the 2-dimensional Ising model in the (K_1, K_2) subspace (left panel) and (K_1, K_3) subspace (right panel).

Let us now explore the global topology of an RG flow by examining the Ising model. In terms of the notation that we have set up in Chapter 2, the fields of the Ising model take values in $\varphi_{\mathfrak{f}}(n) = \varphi(n) \in \{-1, +1\}$ and interact via the action

$$\mathcal{S}[\varphi] = K_1 \sum_{\langle nm \rangle_{\mathrm{n}}} \varphi(n)\varphi(m) + K_2 \sum_{\langle nm \rangle_{\mathrm{nn}}} \varphi(n)\varphi(m) + K_3 \sum_{n} \varphi(n) + ..., \tag{B.1}$$

where $\langle nm \rangle_{\mathrm{n}}$ denotes a sum over lattice sites n and nearest-neigbors m, while $\langle nm \rangle_{\mathrm{nn}}$ denotes a sum over lattice sites n and next-nearest-neighbors m. The interested reader is advised to consult Refs. [70, 148, 230] for a treatment of constructing RG transformations for the Ising model. In Fig. B.1, I show the flow diagram of the Ising model in the (K_1, K_2) subspace with $K_3 = 0$ and the (K_1, K_3) subspace with $K_2 = 0$. The green dots in both diagrams represent critical fixed points. The critical fixed point in the $K_2 = K_3 = 0$ subspace represents the famous ferromagnetic phase transition of the Ising model, and the value of K_1 at the fixed point is proportional to the inverse of the Curie temperature. In the (K_1, K_2) subspace, the critical fixed point at the center of the diagram on the right panel of Fig. B.1 possess a critical surface that extends to the critical fixed points on the K_1- and K_2-axes. The orange dots represent trivial fixed points, which occur at infinite temperature, where the system is non-interacting. The purple fixed point in the (K_1, K_3) subspace diagram represents what is known as a *discontinuity fixed point*; these generally describe first-order phase transitions and lie on a phase boundary [69, 148, 278]. Finally, the red circles indicate sinks, which have no relevant directions.

Appendix C

Tuning the bare gauge coupling to a critical surface

I stated in Sec. 3.4 that the existance of a continuum limit is intimately tied to the existence of a critical fixed point; we call such fixed point an *ultraviolet fixed point*. As the correlation length in units of the lattice spacing depends on the bare gauge coupling g_0^2, one should be able to tune g_0^2 to a critical surface where the correlation length diverges. Assuming that this relationship is invertible, one expects the bare gauge coupling to depend on the correlation length and therefore on the lattice spacing. Close to a critical surface, any physical quantity $Q(a)$ with a well-defined continuum limit scales with the lattice spacing as [227],

$$-a^2\frac{\mathrm{d}}{\mathrm{d}a^2}Q(a) = \left(a^2\frac{\partial}{\partial a^2} - \beta\big(g_0^2\big)\frac{\partial}{\partial g_0^2}\right)Q(a) = \mathrm{O}\big((a\mu)^p\ln(a\mu)^k\big) \tag{C.1}$$

for some $p, k \in \mathbb{Z}$. The $\mathrm{O}\big((a\mu)^p\ln(a\mu)^k\big)$ terms are called *scaling violations*. Eqn. C.1 is a lattice version of the Callan-Symanzik equation and

$$-a^2\frac{\mathrm{d}}{\mathrm{d}a^2}g_0^2(a) \equiv \beta\big(g_0^2\big) \tag{C.2}$$

is the β-function of the bare coupling; it characterizes the dependence of the bare gauge coupling $g_0^2(a)$ on the lattice spacing a. One can calculate the β-function $\beta\big(g_0^2\big)$ in the weak bare gauge coupling limit using perturbation theory [88, 227, 313, 335]. The β-function at two-loop order is

$$\beta\big(g_0^2\big) \sim -\frac{g_0^4}{(4\pi)^2}\left[b_0 + b_1\frac{g_0^2}{(4\pi)^2}\right] \qquad \big(g_0^2/4\pi \to 0\big), \tag{C.3}$$

where

$$3b_0 = 11N - 2N_{\mathrm{f}},$$
$$b_1 = \frac{2N(17N - 5N_{\mathrm{f}})}{3} - \frac{\big(N^2 - 1\big)N_{\mathrm{f}}}{N},$$

which were already defined in in Chapter 1. In terms of b_0, b_1, the solution to Eqn. C.2 is

$$-2\ln(a\Lambda_{\mathrm{Lat.}}) = \frac{1}{b_0 g_0^2} + \frac{b_1}{b_0^2}\ln\big(b_0 g_0^2\big) + \mathrm{O}\big(g_0^2\big), \tag{C.4}$$

where $\Lambda_{\mathrm{Lat.}}$ is an dimensionful integration constant known as the *lattice* Λ*-parameter*. Eqn. C.4 tells us that we should take the continuum limit $a\Lambda_{\mathrm{Lat.}} \to 0$ by tuning $g_0^2(a)$ to zero when $N > 2N_f/11$. Note that the above solution is not unique; the form that I have chosen is referred to as the "$\ln b_0$" convention [335].

Appendix D

Important ideas from statistics

Most modern lattice field theory calculations, including those in this thesis, are done by running "computer experiments" using Monte Carlo algorithms. As such, making/testing predictions from lattice data requires using statistical methods. The goal of this appendix is to *briefly* describe some of the foundational ideas/methods from statistics that are applied throughout this thesis.

D.1 Estimating observables from Markov chains

Consider a collection of random variables $X = \{x_0, x_1, ...\}$ and samples

$$\boldsymbol{X}_i \in S(X) \equiv \left\{ \boldsymbol{X}_j = \left(x_0^{(j)}, x_1^{(j)}, ... \right) \middle| 1 \leq j \leq M-1 \right\} \tag{D.1}$$

of X that have been generated from some Markov chain Monte Carlo (MCMC) algorithm. From $S(X)$, we wish to estimate the mean $\overline{X}$ and covariance of the mean $\Sigma_{\overline{X}}$ (uncertainty) of X. As the samples $\boldsymbol{X}_i$ are members of a Markov chain, the optimal estimate for the mean is simply

$$\overline{\boldsymbol{X}} = \frac{1}{M} \sum_{i=0}^{M-1} \boldsymbol{X}_i. \tag{D.2}$$

Because the samples in $S(X)$ are not necessarily statistically independent, the standard sample covariance, defined as

$$M\Sigma_{\overline{X}} \equiv \frac{1}{M-1} \sum_i \left(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_i \right) \otimes \left(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_i \right) \tag{D.3}$$

where "⊗" is the dyadic (tensor) product, *underestimates* the statistical uncertainty in $\overline{X}$. Accounting for autocorrelation is crucial for properly estimating statistical errors from Markov chain Monte Carlo data.

D.1.1 Autocorrelation time

The autocorrelation time $\tau(x_k)$ of some observable $x_k \in X$ is defined as the average separation between statistically independent samples $x_k^{(j)} \to x_k^{(j+\tau(x_k))}$ of the Markov chain. The effect of autocorrelation is to modify the variance

$$M\sigma_k^2 = \frac{1}{M-1}\sum_i \left(x_k^{(i)} - \overline{x}_k\right)^2 \tag{D.4}$$

as

$$\sigma_k^2 \to \tau(x_k)\sigma_k^2, \tag{D.5}$$

thereby reducing M as

$$M_{\text{eff.}} \to M/\tau(x_k). \tag{D.6}$$

$M_{\text{eff.}}$ is referred to as the *effective sample size.* There are a number of methods for estimating $\tau(x_k)$; I will outline two that I utilize regularly.

D.1.1.1 Batch means

One way to estimate $\tau(x_k)$ is from the effective sample (ESS) size itself. For a sufficiently long Markov chain, the effective sample size can be estimated reliably from the variance

$$\tilde{\sigma}_k^2 = \frac{1}{N_{\text{b}}-1}\sum_{j=0}^{N_{\text{b}}-1} B_j\left(\overline{x}_k^{(j)} - \overline{x}_k\right)^2 \tag{D.7}$$

of batch means

$$\overline{x}_k^{(j)} \equiv \frac{1}{B_j}\sum_{i=0}^{B_j-1} x_k^{(i)} \tag{D.8}$$

over batches of size B_j satisfying

$$M = \sum_{j=0}^{N_\mathrm{b}} B_j. \tag{D.9}$$

From $\tilde{\sigma}_k^2$, the effective sample size is

$$\mathrm{ESS}_k = \frac{\sigma_k^2}{\tilde{\sigma}_k^2}, \tag{D.10}$$

from which the autocorrelation time is estimated as

$$\tau(x_k) \approx \frac{M}{\mathrm{ESS}_k} \tag{D.11}$$

The asymptotically optimal batch size is $B_j = M^{1/3}$; for a multivariate definition of the effective sample size, see Ref. [360]. The ESS method for estimating $\tau(x_k)$ is particularly useful when more robust methods are not available or as a crosscheck of other methods.

D.1.1.2 Integrated autocorrelation time

The integrated autocorrelation time

$$\tau_{\mathrm{int.}}(x_k) = 1 + 2\sum_{\tau=0}^{M-1} \frac{\Gamma_k(\tau)}{\sigma_k^2} \tag{D.12}$$

defined in terms of the autocorrelation function

$$\Gamma_k(\tau) \equiv \frac{1}{M-1}\sum_j \left(x_k^{(j)} - \overline{x}_k\right)\left(x_k^{(j+\tau)} - \overline{x}_k\right) \tag{D.13}$$

is an estimator for the autocorrelation time. Note that $\Gamma_k(0) = \sigma_k^2$. The noise in the estimate for $\tau_{\mathrm{int.}}(x_k)$ increases with incresing τ. Therefore, the upper bound on the sum in Eqn. D.12 is typically cut off at some value W that is calculated from an automatic windowing procedure [378]. The statistical error in $\tau_{\mathrm{int}}(x_k)$ can also be directly estimated [378]. A particularly efficient alternative method for calculating $\tau_{\mathrm{int.}}(x_k)$ from $\Gamma_k(\tau)$ is by fast Fourier transform; see the `autocorr` module of Ref. [137] for details.

D.1.2 Accounting for autocorrelation

From an estimate of the autocorrelation time $\tau(x_k)$ for x_k of interest, the covariance can be estimated by rescaling the covariance matrix as

$$\left[\Sigma_{\overline{X}}\right]_{kl} \to \tau(x_k)^{1/2}\left[\Sigma_{\overline{X}}\right]_{kl}\tau(x_l)^{1/2}. \tag{D.14}$$

Note, however, that Eqn. D.14 is not guaranteed to be positive semi-definite and can introduce a considerable degree of noise into $\Sigma_{\overline{X}}$. Alternatively, one can simply use the estimate for $\tau(x_k)$ to estimate bin size from which $\Sigma_{\overline{X}}$ is estimated by binning [142]. Past $\tau(x_k)$, the statistical error should stabilize up to finite sample size effects. If no plateau in the statistical error is observed about $\tau(x_k)$, then the size M of the dataset is too small to reliably estimate the statistical error. Unfortunately, binning cannot account for autocorrelation in the full covariance matrix, as each observable has its own autocorrelation time. Hence, if binning is used to account for autocorrelation for multiple observables, the bin size should be chosen to be at least as large as the largest autocorrelation time.

D.1.3 Propagation of uncertainties

From the uncertainty of primary observables $x_k \in X$, the uncertainty of an observable $f(X)$ derived from X is easily estimated from the derivative $[\partial f(\overline{X})/\partial \overline{x}_k]$ and covariance $\Sigma_{\overline{X}}$ using standard Gaussian error propagation

$$\sigma_f^2 = \sum_{kl} \frac{\partial f(\overline{X})}{\partial \overline{x}_k}\left[\Sigma_{\overline{X}}\right]_{kl}\frac{\partial f(\overline{X})}{\partial \overline{x}_l}. \tag{D.15}$$

The derivative $[\partial f(\overline{X})/\partial \overline{x}_k]$ can easily be calculated using automatic differentiation. This opens the door to automatic error propagation by automatic differentiation, which is implemented in the beautiful `gvar` and `pyerrors` libraries [215, 241].

D.2 Inference with Bayesian statistics

Throughout this thesis, probability distributions Pr are interpreted as representing the degree to which a hypothesis taking the form of a model $\mathcal{M}$ with model parameters Θ is to be believed. In other words, *probabilities* represent *beliefs*. A hypothesis (model) $\mathcal{M}$ is tested by confronting it with data X. Prior beliefs regarding $\mathcal{M}$ and its parameters Θ before being confronted with data are represented by the joint distribution $\Pr(\mathcal{M}, \Theta)$ (*prior*). *Expectations* regarding the data X given the model $\mathcal{M}$ and its parameters Θ are represented by the conditional joint distribution $\Pr(X|\mathcal{M}, \Theta)$ (*likelihood*). Given the prior $\Pr(\mathcal{M}, \Theta)$ and likelihood $\Pr(X|\mathcal{M}, \Theta)$, the *marginal likelihood*

$$\Pr(X|\mathcal{M}) \equiv \int [\mathrm{d}\Theta] \Pr(X|\mathcal{M}, \Theta) \Pr(\mathcal{M}, \Theta) \tag{D.16}$$

represents expectations regarding X given $\mathcal{M}$ taking all Θ into consideration. The integration measure $[\mathrm{d}\Theta]$ is a multidimensional integration measure over each $\Theta_i \in \Theta$. The marginal likelihood is also referred to as the *evidence*, and it plays an important role in both model selection when confronted with competing models and prior selection when an appropriate choice for the prior is unknown. Expectations regarding Θ given X and $\mathcal{M}$ are *inferred* from the prior, likelihood, and marginal likelihood according to *Bayes' theorem*

$$\Pr(\Theta|X, \mathcal{M}) = \frac{\Pr(X|\mathcal{M}, \Theta) \Pr(\mathcal{M}, \Theta)}{\Pr(X|\mathcal{M})}, \tag{D.17}$$

which follows from elementary properties of conditional distributions [268]. The distribution $\Pr(\Theta|X, \mathcal{M})$ is referred to as the *posterior*.

D.2.1 Principle of maximum entropy

Let X be a collection of random variables. Given the first M *central* moments of X

$$\mu_X \equiv \left\{ \mu_X^{(1)} \equiv \overline{X}, \mu_X^{(2)} \equiv \Sigma_{\overline{X}}, ..., \mu_X^{(M)} \right\}, \tag{D.18}$$

where $\overline{X} = \langle X \rangle$ is the mean of X and Σ_X is the covariance of X, the distribution $\Pr\big(X|\mu_X\big)$ that maximizes the differential entropy functional

$$\mathcal{S}_{\mathrm{DE}}[\Pr] \equiv -\int [\mathrm{d}X] \Pr\big(X|\mu_X\big) \log \Pr\big(X|\mu_X\big) \tag{D.19}$$

is the best representation of the current state of knowledge about X from the perspective of information theory [211, 212]. It is interesting to note that the differential entropy is the diagonal component of the *cross entropy* entering the *Kullback-Liebler* divergence [234].

In lattice field theory calculations based on Markov chain Monte Carlo (see Sec. D.1), it is often the case that one wishes to estimate the parameters Θ of a model $\mathcal{M}$ from an estimate of the mean $\overline{X}$ and covariance of the mean $\Sigma_{\overline{X}}$ that is derived from a finite number of samples for X. Defining $\mathcal{M}_i(\Theta)$ to be an estimate for the estimate of $\overline{x}_i \in \overline{X}$ from the model $\mathcal{M}$, the principle of maximum entropy suggests that we should model the *likelihood* in Eqn. D.17 as

$$\Pr(X|\mathcal{M},\Theta) = (2\pi)^{-|X|/2} \det\big(\Sigma_{\overline{X}}\big)^{-1/2} \exp\left[-\frac{1}{2}\Big(\boldsymbol{\mathcal{M}}(\Theta) - \overline{\mathbf{X}}\Big)^{\mathrm{T}} \Sigma_{\overline{X}}^{-1} \Big(\boldsymbol{\mathcal{M}}(\Theta) - \overline{\mathbf{X}}\Big)\right], \tag{D.20}$$

where $|X|$ is the size (cardinality) of X, $\overline{\mathbf{X}} \equiv (\overline{x}_0, \overline{x}_1, ...)$ is the collection of $\overline{x}_i \in \overline{X}$ represented as a vector that is ordered according to the entries of $\Sigma_{\overline{X}}$, and the same is true of the respective model estimates $\boldsymbol{\mathcal{M}}(\Theta) = \big(\mathcal{M}_0(\Theta), \mathcal{M}_1(\Theta), ...\big)$. Suppose also that any prior beliefs (e.g., constraints, domain-specific knowledge, etc.) $\mathcal{C}(\Theta)$ about the parameters Θ of $\mathcal{M}$ can be represented by their mean $\overline{C}$ (expectation) and covariance of the mean $\Sigma_{\overline{C}}$ (uncertainty). Then the principle of maximum entropy suggests that the prior $\Pr(\mathcal{M},\Theta)$ should be chosen as

$$\Pr(\mathcal{M},\Theta) \equiv (2\pi)^{-|\mathcal{C}|/2} \det\big(\Sigma_{\overline{\mathcal{C}}}\big)^{-1/2} \exp\left[-\frac{1}{2}\Big(\boldsymbol{\mathcal{C}}(\Theta) - \overline{\boldsymbol{\mathcal{C}}}\Big)^{\mathrm{T}} \Sigma_{\overline{\mathcal{C}}}^{-1} \Big(\boldsymbol{\mathcal{C}}(\Theta) - \overline{\boldsymbol{\mathcal{C}}}\Big)\right], \tag{D.21}$$

where $|\mathcal{C}|$ is the cardinality of $\mathcal{C}$, $\overline{\boldsymbol{\mathcal{C}}} \equiv (\overline{\mathcal{C}}_0, \overline{\mathcal{C}}_1, ...)$ is the collection of $\overline{\mathcal{C}}_i \in \overline{\mathcal{C}}$ represented as a vector that is ordered according to the entries of $\Sigma_{\overline{\mathcal{C}}}$, and $\boldsymbol{\mathcal{C}}(\Theta) = \big(\mathcal{C}_0(\Theta), \mathcal{C}_1(\Theta), ...\big)$ is the respective estimate for the constraints $\overline{\mathcal{C}}$ at a particular value of Θ.

D.2.2 Maximum a posteriori estimation

The posterior derived from Eqns. D.20-D.21 yields the expected value

$$\langle\Theta\rangle \equiv \overline{\Theta} = \int [\mathrm{d}\Theta]\ \Theta\,\mathrm{Pr}(\Theta|X,\mathcal{M})$$

for Θ. However, it is quite often the case that the expected value $\overline{\Theta}$ is difficult to estimate, as it requires numerically evaluating an integral over all Θ. What is typically done instead is to calculate the *posterior mode* Θ^* of Θ as

$$\Theta^* \equiv \arg\max_{\Theta}\big[\,\mathrm{Pr}(\Theta|X,\mathcal{M})\big]. \tag{D.22}$$

If most of the posterior mass of $\mathrm{Pr}(\Theta|X,\mathcal{M})$ is concentrated about the posterior mode Θ^*, then Θ^* may be a good estimate for $\overline{\Theta}$. About the posterior mode Θ^*,

$$\log\mathrm{Pr}(\Theta|X,\mathcal{M}) = \log\mathrm{Pr}\left(\Theta^*\middle|X,\mathcal{M}\right) - \frac{1}{2}\left(\boldsymbol{\Theta}-\boldsymbol{\Theta}^*\right)^{\mathrm{T}}\Sigma^{-1}_{\Theta^*}\left(\boldsymbol{\Theta}-\boldsymbol{\Theta}^*\right) + \mathrm{O}\big((\Theta-\Theta^*)^3\big), \tag{D.23}$$

where

$$\left[\Sigma^{-1}_{\Theta^*}\right]_{ij} \equiv \left.\frac{\partial^2\log\mathrm{Pr}(\Theta|X,\mathcal{M})}{\partial\Theta_i\partial\Theta_j}\right|_{\Theta=\Theta^*} \tag{D.24}$$

is an estimate of the *posterior covariance matrix* and $\Theta_i, \Theta_j \in \Theta$. The approximation of Eqn. D.23 is referred to as a *Laplace approximation*. If the $\mathrm{O}\big((\Theta-\Theta^*)^3\big)$ terms in Eqn. D.23 are negligible, then $\mathrm{Pr}(\Theta|X,\mathcal{M})$ about Θ^* is approximately Gaussian and Σ_{Θ^*} is a good estimate for the covariance. Note that the estimate for the marginal likelihood $\mathrm{Pr}(X|\mathcal{M})$ in Eqn. 6.12 of Chapter 6 is obtained by directly plugging Eqn. D.23 into the definition of the marginal likelihood (Eqn. D.16) with $\mathrm{O}\big((\Theta-\Theta^*)^3\big)$ terms dropped.

Unfortunately, real world posterior distributions are not always the nicely-peaked, unimodal distributions that practitioners of statistical methods dream of in their sleep. For example, the model

$\mathcal{M}$ could possess symmetries that show up in $\Pr(\Theta|X, \mathcal{M})$ as degenerate extrema[1] or $\Pr(\Theta|X, \mathcal{M})$ itself could possess a variety steep valleys, flat directions, sharp "cliffs" separating disparate regions of Θ space, etc. for a variety of reasons. In such cases, all that one can state is that the *maximum a posteriori* (MAP) estimate Θ^* is a *probable* estimate for Θ^* with an uncertainty Σ_{Θ^*} that is estimated from the local curvature of $\log \Pr(\Theta|X, \mathcal{M})$ about Θ^*. Still, many statistical practitioners, including the author, will loosely refer Θ^* as the central value of Θ estimated from *fitting* $\mathcal{M}$ to X.

D.2.3 The augmented χ^2

Obtaining the MAP estimate Θ^* of Eqn. D.22 is typically achieved by minimizing an augmented χ^2 [239], defined as

$$-2\log \Pr(\Theta|X, \mathcal{M}) \propto \chi^2_{\text{aug.}}(\Theta) \equiv \chi^2_{\text{data}}(\Theta) + \chi^2_{\text{prior}}(\Theta), \tag{D.25}$$

where

$$-2\log \Pr(X|\mathcal{M}, \Theta) \propto \Big(\boldsymbol{\mathcal{M}}(\Theta) - \overline{\mathbf{X}}\Big)^{\mathrm{T}} \Sigma_{\overline{X}}^{-1} \Big(\boldsymbol{\mathcal{M}}(\Theta) - \overline{\mathbf{X}}\Big) \equiv \chi^2_{\text{data}}(\Theta) \tag{D.26}$$

and

$$-2\log \propto \Pr(\mathcal{M}, \Theta) \propto \Big(\boldsymbol{\mathcal{C}}(\Theta) - \overline{\boldsymbol{\mathcal{C}}}\Big)^{\mathrm{T}} \Sigma_{\overline{\mathcal{C}}}^{-1} \Big(\boldsymbol{\mathcal{C}}(\Theta) - \overline{\boldsymbol{\mathcal{C}}}\Big) \equiv \chi^2_{\text{prior}}(\Theta). \tag{D.27}$$

The factors of $\log 2\pi$, $\log \Pr(X|\mathcal{M})$, $\log\det \Sigma_{\overline{X}}$, and $\log\det \Sigma_{\overline{\mathcal{C}}}$ typically do not enter the χ^2 definitions of Eqns. D.25-D.27 because they do not depend upon Θ. Note, however, that a modified definition of $\chi^2_{\text{aug.}}$ is needed if one wishes to implement hierarchical Bayesian modelling. The problem of estimating Θ^* is then recast as a problem of finding the Θ that minimizes $\chi^2_{\text{aug.}}(\Theta)$; in other words,

$$\Theta^* = \arg\min_{\Theta} \chi^2_{\text{aug.}}(\Theta). \tag{D.28}$$

The term "augmented χ^2" is a little bit silly, as it refers to idea of augmenting onto the frequentist χ^2 (i.e., χ^2_{data}) an additional term that is meant to constrain Θ (as in the case of ridge regression,

---

[1] Sometimes, such symmetries can broken explicitly with priors.

see Chapter 6). We see now from our short discussion in this appendix that the definition of $\chi^2_{\text{aug.}}$ is founded in combining ideas of Bayesian statistics with those of information theory. Chapter 8.2 discusses the optimization of $\chi^2_{\text{aug.}}$ to obtain Θ^*, along with measures of *goodness of fit* and *model selection criteria* in the context of the open source `SwissFit` software written by the author.

D.2.4 Bayesian model averaging

The problem of inferring a set of parameters Θ that are common to a collection of competing models $\mathcal{M}^{(\eta)}$ and/or data subsets $X^{(\eta)} \subseteq X$ is achieved by extending Bayes' theorem as [202]

$$\Pr(\Theta|X) = \sum_{\eta} \Pr\left(\Theta\middle|\mathcal{M}^{(\eta)}, X^{(\eta)}\right) \Pr\left(\mathcal{M}^{(\eta)}\middle|X^{(\eta)}\right), \tag{D.29}$$

where the *posterior probability* $\Pr\left(\mathcal{M}^{(\eta)}|X^{(\eta)}\right)$ is related to the marginal likelihood $\Pr\left(X^{(\eta)}\middle|\mathcal{M}^{(\eta)}\right)$ according to Bayes' theorem as

$$\Pr\left(\mathcal{M}^{(\eta)}\middle|X^{(\eta)}\right) = \frac{\Pr\left(X^{(\eta)}\middle|\mathcal{M}^{(\eta)}\right)\Pr\left(\mathcal{M}^{(\eta)}\right)}{\sum_{\sigma}\Pr\left(X^{(\sigma)}\middle|\mathcal{M}^{(\sigma)}\right)\Pr\left(\mathcal{M}^{(\sigma)}\right)}. \tag{D.30}$$

The leading-order, bias-corrected estimate for the posterior probability $\Pr\left(\mathcal{M}^{(\eta)}\middle|X^{(\eta)}\right)$ about the MAP estimate Θ^* has been derived in Ref. [210], yielding

$$-2\log\Pr\left(\mathcal{M}^{(\eta)}\middle|X^{(\eta)}\right) = -2\log\Pr\left(\mathcal{M}^{(\eta)}\right) + \chi^2_{\text{aug.}}\left(\Theta^*\right) + 2|\Theta| + 2d_\eta + ..., \tag{D.31}$$

where

$$d_\eta \equiv |X| - \left|X^{(\eta)}\right|. \tag{D.32}$$

The last three terms in Eqn. D.31 make up the Akaike information criterion (AIC)

$$\text{AIC} \equiv \chi^2_{\text{aug.}}\left(\Theta^*\right) + 2|\Theta| + 2d_\eta \tag{D.33}$$

Assuming that all models $\mathcal{M}^{(\eta)}$ are equally probable before being confronted with data $X^{(\eta)}$, the model-averaged posterior mode $\overline{\Theta}$ is

$$\overline{\Theta} = \sum_{\eta} \Theta^*_\eta \Pr\left(\mathcal{M}^{(\eta)} \middle| X^{(\eta)}\right), \tag{D.34}$$

while the model-averaged posterior covariance Σ_Θ is

$$\Sigma_\Theta = \sum_{\eta} \Sigma_{\Theta^*_\eta} \Pr\left(\mathcal{M}^{(\eta)} \middle| X^{(\eta)}\right) + \sum_{\eta} \boldsymbol{\Theta}^*_\eta \otimes \boldsymbol{\Theta}^*_\eta \Pr\left(\mathcal{M}^{(\eta)} \middle| X^{(\eta)}\right) - \overline{\boldsymbol{\Theta}} \otimes \overline{\boldsymbol{\Theta}}, \tag{D.35}$$

with $\Pr\left(\mathcal{M}^{(\eta)} \middle| X^{(\eta)}\right)$ given by Eqn. D.31; Θ^*_η the MAP estimate for Θ^* from model/subset η; $\boldsymbol{\Theta}^*_\eta$ the corresponding vector of map estimates ordered according to the covariance $\Sigma_{\Theta^*_\eta}$ of $\boldsymbol{\Theta}^*_\eta$; and $\overline{\boldsymbol{\Theta}}$ the vector of model-averaged MAP estimates (Eqn. D.34). The first term in Eqn. D.35 is merely the posterior covariance averaged against $\Pr\left(\mathcal{M}^{(\eta)} \middle| X^{(\eta)}\right)$, while the latter two terms contribute an additional uncertainty to Σ_Θ from variations in the models/subsets. In other words, if the posterior probability $\Pr\left(\mathcal{M}^{(\eta)} \middle| X^{(\eta)}\right)$ is the same for all models/subsets, then the latter two terms in Eqn. D.35 do not contribute to the total uncertainty.

Appendix E

Gradient flow equation for the XY model

E.1 Gradient flow equation with an explicit constraint

In this section, I am going to derive a gradient flow equation for the XY model that builds in some set of constraints. I briefly introduce the XY model as the "∞-state clock model" in Chapter 6.4. First, define the rescaled reduced Hamiltonian $\tilde{\mathcal{H}}^{(\infty)}_{\text{clock}}[n]$ as

$$\tilde{\mathcal{H}}^{(\infty)}_{\text{clock}}[n] \equiv \mathcal{H}^{(\infty)}_{\text{clock}}[n]/K^{(\infty)}_{\text{clock}} = -\sum_{x,\mu,j} n^j(x)n^j(x+\hat{\mu}) \tag{E.1}$$

where $x \in \Lambda \subset \mathbb{Z} \times \mathbb{Z}$, $\mu, j \in \{0,1\}$, and, according to Eqn. 6.18 in Chapter 6.4,

$$\mathbf{n}(x)^2 = 1 \ \forall x \in \Lambda. \tag{E.2}$$

To write down a gradient flow equation that respects the constraint of Eqn. E.2, I use the Lagrange multiplier method. Computing the gradient of the rescaled Hamiltonian with respect to the flowed spin $n^i_t(y)$ (not yet taking into account the constraint), I get

$$\frac{\delta}{\delta n^i_t(y)} \sum_{x,\mu,j} n^j_t(x)n^j_t(x+\hat{\mu}) = \sum_{\mu} \left[n^i(y-\hat{\mu}) + (y+\hat{\mu})\right] \equiv f^i_t(y), \tag{E.3}$$

where I've defined $f^i_t(y)$ as the value of the (negative) gradient for convenience. The gradient flow equation with a Lagrange multiplier taking into account the constraint will have a constraint force of the form

$$\frac{\lambda(t)}{2}\frac{\delta}{\delta n_t^i(y)}\sum_x\left(\sum_j n_t^j(x)^2-1\right)=\lambda(t)n_t^i(y),$$

where the Lagrange multiplier was defined with a factor of 1/2 out front for convenience. We therefore have the following equations

$$\frac{\mathrm{d}n_t^i(y)}{\mathrm{d}t}=f_t^i(y)+\lambda(t)n_t^i(y), \tag{E.4}$$

and

$$\sum_j n_t^j(y)\frac{\mathrm{d}n_t^j(y)}{\mathrm{d}t}=0, \tag{E.5}$$

where Eqn. E.5 is obtained by differentiating Eqn. E.2 with respect to the gradient flow time t. Plugging the right hand side of Eqn. E.4 into Eqn. E.5, we obtain $\lambda(t)$ as

$$\lambda(t)=-\sum_j n_t^j(y)f_t^j(y),$$

so that the full gradient flow equation is

$$\frac{\mathrm{d}n_t^i(y)}{\mathrm{d}t}=f_t^i(y)-n_t^i(y)\sum_j n_t^j(y)f_t^j(y)=\sum_j\left(\delta^{ij}-n_t^i(y)n_t^j(y)\right)f_t^j(y), \tag{E.6}$$

which is of the same form as the O(N) model gradient flow equations obtained in Refs. [224, 258] by different methods. Explicitly, we have for the XY model

$$\frac{\mathrm{d}\mathbf{n}_t(y)}{\mathrm{d}t}=\begin{bmatrix} n_t^1(y)^2 & -n_t^0(y)n_t^1(y) \\ -n_t^0(y)n_t^1(y) & n_t^0(y)^2\end{bmatrix}\begin{bmatrix} f_t^0(y) \\ f_t^1(y)\end{bmatrix}\equiv\mathbf{M}_t(y)\mathbf{f}_t(y). \tag{E.7}$$

E.2 Gradient flow equation with an implicit constraint

Define

$$z_t(y) \equiv n_t^0(y) + i n_t^1(y) \text{ and } f_t(y) = f_t^0(y) + i f_t^1(y). \tag{E.8}$$

Also, just to set up notation, $\bar{z} = z^* = z^\dagger$. Then from the explicit form of the gradient flow equation given by E.7, we have

$$\frac{\mathrm{d}z_t(y)}{\mathrm{d}t} = -\frac{1}{2} z_t(y)^2 \bar{f_t}(y) + \frac{1}{2} f_t(y) = \frac{1}{2}\Big(\bar{z}_t(y) f_t(y) - z_t(y)\bar{f}_t(y)\Big) z_t(y) = i\mathrm{Im}\Big(\bar{z}_t(y) f_t(y)\Big) z_t(y).$$

This is nice and compact. Let's write this result down in its final form.

$$\frac{\mathrm{d}z_t(y)}{\mathrm{d}t} = i\mathrm{Im}\Big(\bar{z}_t(y) f_t(y)\Big) z_t(y). \tag{E.9}$$

Now let's take the GF equation expressed as Eqn. 4.2 and see what we get out of that. First, I have to write down a quick rule for taking derivatives with respect to U(1) group elements. This rule is

$$\partial_y z_t(y) \equiv i \frac{\mathrm{d}}{\mathrm{d}s} e^{is} z_t(y)\Big|_{s=0} \delta_{x,y} = -z_t(y)\delta_{x,y}. \tag{E.10}$$

Eqn. E.10 is just a specialization of the natural derivative on Lie groups from the left-invariant vector field, which I discuss in Chapter 4.1. The derivative satisfies

$$\partial_y \bar{z}_t(y) = \bar{z}_t(y). \tag{E.11}$$

Now, the U(1) analogue of Eqn. 4.2 reads

$$\frac{\mathrm{d}z_t(y)}{\mathrm{d}t} = -\partial_y \tilde{\mathcal{H}}^{(\infty)}_{\text{clock}}\big[z_t, \bar{z}_t\big] z_t(y), \tag{E.12}$$

where the XY Hamiltonian $\tilde{\mathcal{H}}^{(\infty)}_{\text{clock}}\big[z_t, \bar{z}_t\big]$ is derived from Eqn. E.1 using the reparameterization in Sec. E.1 as

$$\tilde{\mathcal{H}}^{(\infty)}_{\text{clock}}\big[z_t, \bar{z}_t\big] = -\Re\sum_{x,\mu} z_t(x)\bar{z}_t(x+\hat{\mu}). \tag{E.13}$$

Working the derivative out with our rule from Eqn. E.10, I get

$$\partial_y \tilde{\mathcal{H}}^{(\infty)}_{\text{clock}}\big[z_t, \bar{z}_t\big] = -i\text{Im}\sum_{\mu} \bar{z}_t(y)\big(z_t(y+\hat{\mu}) + z_t(y-\hat{\mu})\big) = -i\text{Im}\Big(\bar{z}_t(y)f_t(y)\Big),$$

so that Eqn. 4.2 turns into

$$\frac{\mathrm{d}z_t(y)}{\mathrm{d}t} = i\text{Im}\Big(\bar{z}_t(y)f_t(y)\Big)z_t(y). \tag{E.14}$$

Therefore, we see that these two equations are exactly the same. I have implemented the gradient flow equation of Eqn. E.14 in the `Quantum EXpressions` (`QEX`)-based `qex_staghmc` library with the help of the beautiful `Arraymancer` library in `Nim` [284, 309].

Appendix F

Improved actions

Perturbative on-shell (classical) improvement attempts to remove cutoff effects explicitly. Improvement of the action and the observables derived from the action tend to be intertwined. As such, if one wishes to implement classical improvment, it is often the case that both the action and target observables must be considered. In the RG picture of this thesis, namely that of Fig. 3.1 in Chapter 3.4, classical improvement can be thought of as moving actions that are in the vicinity of the critical surface closer to the renormalized trajectory. Far away from the critical surface, quantum effects can kick in and one must consider quantum effects in their improvement programme; see, for example, the Symanzik effective field theory approach of Ref. [206]. In this appendix, I am going to describe only the gauge action improvement used in this thesis.

F.1 Lüscher-Weiss (Symanzik) actions

Classical improvement is outlined beautifully in the seminal work of Ref. [251] by Martin Lüscher and Peter Weisz, though many of the details of classical improvement had been worked out already by Peter Weisz and Kurt Symanzik [351, 352, 369]. To start, consider that the classical gauge field $\mathfrak{A}_\mu(x)$ can be approximated arbitrarily well by a link variable $\mathcal{U}_\mu(n)$ as

$$ia\mathfrak{A}_\mu(an) \sim \log \mathcal{U}_\mu(n) \qquad (a \to 0),$$

as in Eqn. 3.5 of Chapter 3.1. As such, any gauge-invariant observable $\mathcal{O}$ that is also invariant under the lattice equivalent of rotations and reflections can be expanded asymptotically in the lattice spacing a as

$$\mathcal{O} \sim \sum_{\text{even } k>0} a^k \mathcal{O}_k \qquad (a \to 0), \tag{F.1}$$

where, schematically,

$$\begin{aligned} \mathcal{O}_2 &= 0 \qquad (1), \\ \mathcal{O}_4 &= r \sum_{\mu\nu} \mathrm{Tr_c}\left[\mathfrak{F}_{\mu\nu}\mathfrak{F}_{\mu\nu}\right] \qquad (2), \\ \mathcal{O}_6 &= r_1 \sum_{\mu\nu} \mathrm{Tr_c}\left[\mathfrak{D}_\mu\mathfrak{F}_{\mu\nu}\mathfrak{D}_\mu\mathfrak{F}_{\mu\nu}\right] + r_2 \sum_{\mu\nu\rho} \mathrm{Tr_c}\left[\mathfrak{D}_\mu\mathfrak{F}_{\nu\rho}\mathfrak{D}_\mu\mathfrak{F}_{\nu\rho}\right] + r_3 \sum_{\mu\nu\rho} \mathrm{Tr_c}\left[\mathfrak{D}_\mu\mathfrak{F}_{\mu\rho}\mathfrak{D}_\nu\mathfrak{F}_{\nu\rho}\right] \\ &\quad + \text{“contractions over } \mathrm{Tr_c}\left[\mathfrak{F}_{\mu\nu}\mathfrak{D}_\rho\mathfrak{F}_{\sigma\delta}\right]\text{''} \qquad (3), \end{aligned} \tag{F.2}$$

with

$$\mathfrak{D}_\rho\mathfrak{F}_{\mu\nu} = \partial_\rho\mathfrak{F}_{\mu\nu} + \left[\mathfrak{A}_\rho, \mathfrak{F}_{\mu\nu}\right]. \tag{F.3}$$

Now take

$$\mathcal{S}_{\mathrm{G}}\left[\mathcal{U}_\mu\right] = \beta_b \sum_{l=0}^{3} c_l \sum_n \sum_{\mathcal{C}_l} \mathrm{Tr_c}\left[\mathcal{U}_{\mathcal{C}_l}(n)\right] \tag{F.4}$$

as a generic ansatz for the gauge action and define $\mathcal{U}_{\mathcal{C}_l}(n)$ to be a product l of gauge links $\mathcal{U}_\mu$ that form a closed curve $\mathcal{C}_l$. The notation for the ansatz of Eqn. F.4 differs from that of Ref. [251] to remain consistent with the notation for the Pauli-Villars effective action in Eqn. 5.2 of Chapter 5.2. The coefficients c_l are conventionally normalized as

$$c_0 + 8c_1 + 8c_2 + 16c_3 = 1. \tag{F.5}$$

Note that c_l could depend on g_0^2; however, for the purposes of classical improvement, this does not matter. By expanding each $\sum_{\mathcal{C}_l} \mathrm{Tr_c}\mathcal{U}_{\mathcal{C}_l}(n)$ term in Eqn. F.4 according to Eqns. F.1-F.2 and imposing that all $\mathrm{O}\left(a^2\right)$ terms vanish, one arrives at the following general constraint for c_l

$$\begin{aligned} |x| &< 1/16, \\ c_0 &= \frac{5}{3} - 24x, \\ c_1 &= -\frac{1}{12} + x, \\ c_2 &= 0, \\ c_3 &= x, \end{aligned} \tag{F.6}$$

which represents the most general on-shell improved action that one can construct out of Eqn. F.4 [251]. The choice

$$x = 0 \qquad \text{(tree-level Luescher-Weiss)}$$

is computationally the most convenient; hence, it is often referred to as the *tree-level improved Lüscher-Weiss (Symanzik)* gauge action, as in Chapters 4 and 8.3.

In Chapter 7, I use different Lüscher-Weiss gauge actions for the gauge flow. Because the goal there is not classical improvement, it is not necessary to set c_l according to Eqn. F.6. Rather, I set

$$c_2 = c_3 = 0 \qquad \text{(flow action)}$$

and redefine

$$c_0 \equiv c_\mathrm{p} \text{ and } c_1 \equiv c_\mathrm{r} \qquad \text{(flow action)},$$

which is consistent the constraints imposed by tree-level improvement of the action only when $c_\mathrm{r} = -1/12$. The notation for $c_\mathrm{p}/c_\mathrm{r}$ refers to the fact that each $\mathcal{C}_0$ in Eqn. F.4 is a plaquette, while each $\mathcal{C}_1$ is a planar rectangle.

F.2 Mixed fundamental/adjoint action

Though throughout this thesis I have assumed that the gauge links $\mathcal{U}_\mu(n)$ live in the fundamental representation of $\mathrm{SU}(N)$ by definition, such a restriction is actually not necessary, so long as the gauge action that is constructed from some non-fundamental representation is in the same universality class as the fundamental one. In fact, one can even consider mixing representations within the same action. A natural question, then, is whether using mixed representation actions can reduce lattice artefacts. In Ref. [164], the mixed action

$$\mathcal{S}_\mathrm{G}\left[\mathcal{U}_\mu\right] = \sum_{\alpha=F,A} \frac{\tilde{\beta}_\alpha}{d_\alpha} \sum_n \sum_{\mu<\nu} \Re\mathrm{Tr}_\mathrm{c}\left[\mathbb{1} - \mathcal{U}^{(\alpha)}_{\mu\nu}(n)\right] \tag{F.7}$$

with fundamental (F) and adjoint (A) terms was observed to reduce cutoff effects in spectral measurements of the glueball mass as compared to its pure fundamental (Wilson) counterpart $\tilde{\beta}_A < 0$. In Eqn. F.7, d_α represents is the dimension of the representation, such that

$$d_F = N \text{ and } d_A = N^2 - 1. \tag{F.8}$$

Note that, even though the gauge action has both fundamental and adjoint link variables, we can work exclusively with the fundamental link variables as a consequence of the identity

$$\mathrm{Tr}_\mathrm{c}\left[\mathcal{U}^{(A)}\right] = \mathrm{Tr}_\mathrm{c}\left[\mathcal{U}^{(F)}\right]\mathrm{Tr}_\mathrm{c}\left[\mathcal{U}^{(F)\dagger}\right] - 1, \tag{F.9}$$

which reproduces d_A in Eqn. F.8 on the identity. Since the work of Ref. [164], Eqn. F.7 has become a fan favorite action in the beyond Standard Model (BSM) physics community. Modern studies using Eqn. F.7, such as those of Chapters 5 and 7, typically choose

$$\beta_A/\beta_F = -1/4, \tag{F.10}$$

where

$$\beta_A \equiv \tilde{\beta}_F \text{ and } \beta_A \equiv \frac{\tilde{\beta}_A}{1 - 1/N^2}. \tag{F.11}$$

At tree-level, the corresponding "equivalent Wilson bare gauge coupling" is

$$\beta = \beta_F \big(1 + 2\beta_A/\beta_F\big). \tag{F.12}$$

Though the action of Eqn. F.7 has been used for many decades in the BSM literature, very little has been done in the way of systematically exploring its cutoff effects in other observables. Nonetheless, it has been deployed in this thesis for the purposes of remaining consistent with simulations that were already present and available for use. In future studies, the author would prefer to utilize improved gauge actions that are better motivated.

Appendix G

Gauge smearing

Broadly speaking, the RG flow diagram of Fig. 3.1 in Chapter 3.4 suggests that there are three "simple" ways that one can reduce lattice artefacts. One method is to tune g_0^2 closer to the critical surface ($g_0^2 \to 0$) while keeping all relevant scales $l \gg a$ fixed. This is just the continuum limit. Another method is to explicitly cancel off cutoff effects via classical improvement or its extension including logarithmic corrections that arise from quantum effects. This is the classical improvment described in Appendix F. The last method is to average out short-distance fluctuations in such a way that long-distance observables are unaffected by the averaging procedure. This is what an RG transformation does, and it is in essence what gauge smearing does; it brings the action closer to the renormalized trajectory by non-perturbatively smoothing out ultraviolet fluctuations. Note that ideas that underpin smearing are similar, but not strictly equivalent, to those that underpin so-called *perfect actions*, which attempt to utilize RG transformations to construct lattice actions that are so close to the renormalized trajectory that they are devoid of any lattice artefacts [99, 193]. Unfortunately, perfect actions are not practically realizable; see, however, Ref. [204] for some interesting recent work on constructing perfect actions using machine learning. There are many kinds of gauge link smearing, and just about every large-scale lattice collaboration hails the gauge smearing that it uses as "optimal" in some sense. In this appendix, I shall briefly describe only the two types of gauge link smearing that have been deployed in this thesis.

G.1 Normalized hypercubic smearing (nHYP)

Every simulation that utilizes staggered fermions in this thesis deploys hypercubic smearing (nHYP) to smear out the gauge links in the pseudofermion action or the Pauli-Villars action (see Chapter 8.1) [175, 176]. The basic ingredient of nHYP smearing is the APE smearing, which takes the form [3]

$$\mathcal{U}_\mu(n) \to \text{Proj}_G\Big[(1-\alpha)\mathcal{U}_\mu(n) + \frac{\alpha}{N_{\text{stp.}}}\Sigma_\mu\Big], \tag{G.1}$$

where $\text{Proj}_G[U]$ is a projection operator that takes $U \in \text{GL}(N)$ to some group G, Σ_μ is a local sum over "staples" (products of three gauge links that are one link away from forming a closed loop), and $N_{\text{stp.}}$ is the number of staples entering Σ_μ. Note that $\mathcal{P}_G$ does not have to project U back into $\text{SU}(N)$. As long as the resulting action is in the same universality class and the new "fat" link transforms under gauge transformations as $\mathcal{U}_\mu$ does, all is fine and dandy. nHYP smearing is constructed from three levels APE smearing that forces all contributing gauge links to exist only within neighboring hypercubes of the fat link and has

$$\text{Proj}_{\text{U}(N)}(U) = U\big(U^\dagger U\big)^{-1/2}. \tag{G.2}$$

Explicitly,

$$\begin{aligned}
\overline{\mathcal{V}}_{\mu;\nu\rho} &= \text{Proj}_{\text{U}(N)}\left[(1-\alpha_1)\mathcal{U}_\mu(n) + \frac{\alpha_1}{2}\sum_{\pm\eta\neq\mu,\nu,\rho}\mathcal{U}_\eta(n)\mathcal{U}_\mu(n+\hat{\eta})\mathcal{U}_\eta(n+\hat{\mu})^\dagger\right] && (1),\\
\widetilde{\mathcal{V}}_{\mu;\nu} &= \text{Proj}_{\text{U}(N)}\left[(1-\alpha_2)\mathcal{U}_\mu(n) + \frac{\alpha_2}{4}\sum_{\pm\rho\neq\mu,\nu}\overline{\mathcal{V}}_{\rho;\nu\mu}(n)\overline{\mathcal{V}}_{\mu;\rho\nu}(n+\hat{\rho})\overline{\mathcal{V}}_{\rho;\nu\mu}(n+\hat{\mu})^\dagger\right] && (2),\\
\mathcal{V}_\mu(n) &= \text{Proj}_{\text{U}(N)}\left[(1-\alpha_3)\mathcal{U}_\mu(n) + \frac{\alpha_3}{6}\sum_{\pm\nu\neq\mu}\widetilde{\mathcal{V}}_{\nu;\mu}(n)\widetilde{\mathcal{V}}_{\mu;\nu}(n+\hat{\nu})\widetilde{\mathcal{V}}_{\nu;\mu}(n+\hat{\mu})^\dagger\right] && (3).
\end{aligned} \tag{G.3}$$

Because the projection onto $\text{U}(N)$ is differentiable, the full collection of smearing steps comprising the nHYP smearing transformation of Eqn. G.1 can be implemented in HMC. Note that the ordering on $\alpha \equiv (\alpha_1, \alpha_2, \alpha_3)$ in Eqn. G.1 is the convention in `QEX` [284], which differs from Ref. [175] by an exchange of α_1 with α_3. Details of calculating the force from the fermion sector when using nHYP-smeared links can be found in Ref. [175]. The force is heroically implemented in `QEX` [284].

G.2 Stout smearing

The gradient flow transformation (see Chapter 4.1) is an infinitesimal stout smearing. To see how, note that the first-order solution to the gradient flow equation for the gauge fields (Eqn. 8.26, alternatively Eqn. 4.2) by Euler integration is

$$\mathcal{U}_\mu(n,t+\epsilon) \approx \exp\big[\epsilon Z_\mu(n,t)\big]\mathcal{U}_\mu(n,t) \tag{G.4}$$

for $\epsilon \ll 1$ (in lattice units). Eqn. G.4 is equivalent to the stout smearing introduced in Ref. [266] with the identification [227]

$$\begin{aligned} \epsilon &\equiv \rho, \\ \Sigma_\mu(n,t) &\equiv \sum_{\pm\nu\neq\mu} \mathcal{U}_\nu(n,t)\mathcal{U}_\mu(n+\hat{\nu},t)\mathcal{U}_\nu(n+\hat{\mu},t)^\dagger, \\ \Omega_\mu(n,t) &\equiv \Sigma_\mu(n,t)\mathcal{U}_\mu(n,t)^\dagger - \mathcal{U}_\mu(n,t)\Sigma_\mu(n,t)^\dagger, \\ Z_\mu(n,t) &= \frac{1}{2}\Omega_\mu(n,t) - \frac{1}{N}\mathrm{Tr}_\mathrm{c}\big[\Omega_\mu(n,t)\big] \in \mathfrak{su}(N). \end{aligned} \tag{G.5}$$

The staple sum Σ_μ is the same as we already encountered in Sec. G.1 for APE smearing and nHYP smearing [3, 175]. Because $Z_\mu(n,t) \in \mathfrak{su}(N)$, the stout-smeared link stays in SU(N), as is the case for gradient flowed links. The exponential can be accurately evaluated numerically using the Cayley-Hamilton theorem. Moreover, the stout smearing transformation is differentiable; as such, it can be implemented in HMC. See Ref. [266] for details regarding the force from pseudofermion actions involving stout-smeared links. Note, also, that future implementations of stout smearing could benefit from a version of automatic differentiation adapted to group derivatives; see the appendix of Ref. [204] and the beautiful `Julia`-based `Gaugefields.jl` library.[1] Usually, many levels

---

[1] `Gaugefields.jl` can be found at `https://github.com/akio-tomiya/Gaugefields.jl`.

of stout smearing are performed on the gauge links. For example, the simulations of Chapter 8.3 are performed with six levels of stout smearing with $\rho = 0.1$.

Appendix H

Symplectic Integrators

`QEX`, specifically Xiaoyong Jin's `MDevolve`,[1] uses a symplectic Omelyan integrator with various options for including higher-order terms, which I shall now describe. Following Ref. [353], let q and p be generalized coordinates and let f stand for either q or p. Furthermore, define a linear operator $L(H)$ (with H the Hamiltonian) acting on f via the Poisson bracket as [330, 353]

$$L(H)f \equiv \{f, H\} = \dot{f}. \tag{H.1}$$

Written explicitly in terms of q or p, Eqn. H.1 gives Hamilton's equations of motion, with the formal (approximate) solution

$$f(t+\delta t) \approx \exp\big(\delta t L(H)\big) f(t). \tag{H.2}$$

Assuming that $L(H)$ can be decomposed as $L(H) = T + V$, where T is the kinetic part (involving only p) and V is the interaction part (involving only q), we have

$$\exp\big(\delta t L(H)\big) = \exp\left[\frac{1}{2}\delta t T\right] \exp(\delta t V) \exp\left[\frac{1}{2}\delta t T\right] + \mathcal{O}\big(\delta t^3\big) \equiv G_2(\Delta t) + \mathcal{O}\big(\delta t^3\big). \tag{H.3}$$

Eqn. H.3 gives the well-known leapfrog (2LF) integrator. The mapping $G_2(\delta t) : \mathcal{M} \to \mathcal{M}$, where $\mathcal{M}$ is the symplectic manifold that p and q live on, is symplectic and time reversible. Hence, $G_2(\delta t)$

---

[1] `MDevolve` is available at https://github.com/jxy/MDevolve.

may be used to update the fields in an HMC simulation.[2]

Higher-order integrators $G_{2k+2}(\delta t)$ may be generated recursively from lower-order integrators $G_{2k}(\delta t)$; however, the performance of such integrators may be sub-optimal [353]. An alternative to the recursive approach, famously explored by Omelyan *et al.* in Ref. [282], starts by decomposing $\exp\big(\delta t L(H)\big)$ as

$$\exp\big(\delta t L(H)\big) = \prod_{i=1}^{k} \exp\big(c_i \delta t T\big) \exp\big(d_i \delta t V\big) + \mathcal{O}\big(\delta t^{n+1}\big), \tag{H.4}$$

where c_i, d_i are constrained to obey $\sum_i c_i = \sum_i d_i = 1$. Time-reversibility imposes additional constraints on c_i and d_i [353]. For a fixed choice of n (the order of the integrator), the error depends on the choice of k and additional details regarding the choice of c_i and d_i; in particular, it may be parameterized by a set of free variables that are constrained to depend on c_i, d_i and k in such a way that the error is minimized.

A popular choice of integrator, first explored by Omelyan *et. al* in Ref. [282], is the 2nd-order minimum-norm (2MN) integrator $I_{\mathrm{2MN}}(\delta t) : \mathcal{M} \to \mathcal{M}$, defined by

$$I_{\mathrm{2MN}}(\delta t) \equiv \exp\big(\lambda \delta t T\big) \exp\left[\frac{1}{2}\delta t V\right] \exp\Big[(1-2\lambda)\delta t T\Big] \exp\left[\frac{1}{2}\delta t V\right] \exp\big(\lambda \delta t T\big), \tag{H.5}$$

where λ is obtained by minimizing the error, resulting in

$$\lambda = \frac{1}{2} - \frac{(2\sqrt{326}+36)^{1/3}}{12} + \frac{1}{(6\sqrt{326}+36)^{1/3}} \approx 0.1931833... \tag{H.6}$$

The 2MN integrator requires twice as many force evaluations as the 2LF integrator; however, the error on the 2MN integrator is approximately eleven times smaller, meaning that the total number integration steps is reduced by a factor of the square root of eleven such that the total computational cost of the 2MN integrator is approximately half as expensive as the 2LF integrator.

In implementing the 2MN integration, one must make a choice of starting variables (either first update q, the gauge fields, or p, the momenta). Different choices of starting variables can affect

---

[2] Symplecticity and time-reversibility are sufficient to ensure that the molecular dynamics update in an HMC simulation obeys the detailed balance condition.

the performance of the integrator; however, the 2MN integrator is typically insensitive to this choice. We choose to update q (that is, $\mathcal{U}_\mu$) first, such that

$$\mathcal{U}_\mu^{(1)}(n) = \exp\left[\delta t \mathfrak{P}_\mu^{(0)}(n)\right]\mathcal{U}_\mu^{(0)}(n), \tag{H.7}$$

where $\mathfrak{P}_\mu$ is the momentum of the gauge variable. The momentum is then updated as

$$\mathfrak{P}_\mu^{(1)}(n) = \mathfrak{P}_\mu^{(0)}(n) - \delta t \sum_{\mathrm{f}=\mathrm{G,F,PV}} \mathfrak{F}_\mu^{(\mathrm{f})}(n), \tag{H.8}$$

where $\mathfrak{F}_\mu^{(\mathrm{G})}(n)$ is the force from the gauge fields, $\mathfrak{F}_\mu^{(\mathrm{F})}(n)$ the pseudofermion fields, and $\mathfrak{F}_\mu^{(\mathrm{PV})}(n)$ the Pauli-Villars fields (see Chapter 8). The latter sequence of updates is repeated once more with the addition of a final $\mathcal{U}_\mu$ update at each step of the trajectory.

Higher-order minimum norm integrators often exhibit a higher degree of sensitivity to the starting variables. Therefore, higher-order integrators are typically differentiated by their starting variables (ending with "FP" if the $\mathcal{U}$ update is done first and "FV" if the $\mathfrak{P}$ update is done first) and number of force evaluations. For example, the fourth-order minimum norm integrator starting with the $\mathcal{U}_\mu$ update and being composed of five force updates is given the name "4MN5FP". I shall not delve further into these higher-order integration schemes. Nonetheless, I have included options in `qex_staghmc` based off of the example codes contained in `QEX` to utilize these higher-order integrators as the user desires.

H.1 Tuning the 2MN integrator

In this subsection, I follow Ref. [353]. Take a unit time trajectory. If we split such a trajectory into N steps, such that $\delta t = 1/N$, then the error in the Hamiltonian produced by an nth-order integrator is $\mathcal{O}(\delta t^n)$. That is, $\delta\mathcal{H} \sim \delta t^n$. From *Creutz's equality* $\langle \exp(\delta\mathcal{H})\rangle = 1$ (or *Jansen's inequality*) [87],

$$\langle \delta\mathcal{H}\rangle = \frac{1}{2}\langle \delta\mathcal{H}^2\rangle + \mathcal{O}(\delta\mathcal{H}^3) \sim C_n V \delta t^{2n}, \tag{H.9}$$

where C_n is a coefficient that depends upon the order of the integrator and details of the Hamiltonian and V is the volume of the simulation. The acceptance $\langle P_{\mathrm{acc.}} \rangle$ of an HMC algorithm in a sufficiently large volume is given by

$$\langle P_{\mathrm{acc.}} \rangle \underbrace{\approx}_{V/a^4 \gg 1} \mathrm{erfc}\left[\frac{1}{8}\left\langle \delta\mathcal{H}^2 \right\rangle^{1/2}\right] \underbrace{\approx}_{\langle \delta\mathcal{H}^2 \rangle \ll 1} \exp\left[-\frac{1}{\sqrt{2\pi}}\left\langle \delta\mathcal{H}^2 \right\rangle^{1/2}\right]. \tag{H.10}$$

By considering the *work per accepted trajectory*, the optimal acceptance turns out to be dependent only upon the order n of the integrator, such that

$$\langle P_{\mathrm{acc.}} \rangle \sim \exp(-1/n), \tag{H.11}$$

which yields an optimal acceptance of $\sim 61\%$ for the 2MN integrator. In practice, we do not typically concern ourselves if the acceptance is $\sim 80\% - 95\%$, as it does not affect the simulation and merely indicates that our simulations are not being performed as efficiently as they could be. Simulations monitoring $\left\langle \delta\mathcal{H}^2 \right\rangle$ tend to find that the optimal λ from Eqn. H.6 is slightly higher than ~ 0.193 [353]. More information about tuning the HMC integrators can be found in Refs. [61, 83].

Appendix I

Conjugate gradient

Multiple stages of a single Hamiltonian Monte Carlo step require the solution to the Dirac equation, with

$$\mathfrak{D}\varphi = \phi. \tag{I.1}$$

In Sec. 8.1.1, I discussed the even/odd preconditioning procedure, which cuts the size of the inversion problem down considerably. In this appendix, I shall gather the algorithmic details of the inversion process together.

Let us begin by describing the beautiful method by which the solution x to the generic linear system of equations of the form

$$Ax = b \tag{I.2}$$

may be acquired, assuming $A : \mathbb{C}^N \to \mathbb{C}^N$ is a Hermitian (and positive-definite) linear operator and $x, b \in \mathbb{C}^N$ (as implicitly stated in the definition of A). Solving Eqn. I.2 is equivalent to finding the unique minimum of the functional $f : \mathbb{C}^N \to \mathbb{C}$ defined by

$$f(x) \equiv \langle x, Ax\rangle - \langle x, b\rangle, \tag{I.3}$$

where $\langle\cdot,\cdot\rangle : \mathbb{C}^N \times \mathbb{C}^N \to \mathbb{R}$ is the familiar Hermitian inner product on $\mathbb{C}^N$ defined by

$$\langle y, z\rangle \equiv y^\dagger z, \tag{I.4}$$

for any $x, y, z \in \mathbb{C}^N$. Furthermore, the solution x may be decomposed into a basis $\{p_k\}$ as

$$x = \sum_{k=0}^{N-1} \alpha_k p_k, \tag{I.5}$$

which, by imposing the orthogonality constraint

$$\langle p_i, A p_j \rangle = 0 \text{ for any } i \neq j, \tag{I.6}$$

further allows the α_k coefficients to be easily solved for using Eqn. I.2.[1] With some preliminary details out of the way, I am able to state the overarching goal of the conjugate gradient method. The conjugate gradient method for finding the solution x to the linear system of Eqn. I.2 consists of iteratively finding the minimum of the functional defined by Eqn. I.3 by updating the solution x expressed in the basis $\{p_k\}$ defined by the constraint of Eqn. I.6 in order of decreasing significance. What I mean by "order of decreasing significance" is that the error (which we must quantify) of the solution projected onto the subspace defined by the basis $\{p_k\}_J \equiv \{p_k | k < J-1 \text{ and } J < N\}$ is smaller in magnitude than the error of the solution projected onto the subspace defined by the basis $\{p_k\}_K$ if $J > K$. That is, as we update the solution x with more $\alpha_k p_k$ (subject to the constraint of Eqn. I.6), the approximation error decreases.

Now that we have an idea of the objective of conjugate gradient, let's fill in the details of its execution. Assume first that we have some initial guess for the solution x_0 and define the *residual* $r(x_0)$ as

$$r(x_0) \equiv b - Ax_0 = A(x - x_0). \tag{I.7}$$

The magnitude of the residual quantifies the error in x_0, since, roughly, $|r(x_0)| \sim |x - x_0|$. We can update our guess x_0 to a more accurate estimate x_1 of x by following the gradient of the functional

---

[1] In the jargon of the literature, the vectors $p_i, p_j \in \mathbb{C}^N$ for $i, j \leq N-1$ are said to be *conjugate* to one another with respect to A if they satisfy the orthogonality constraint of Eqn. I.6.

defined in Eqn. I.3, such that

$$x_1 = x_0 - \alpha_0 \nabla f(x)\big|_{x=x_0} = x_0 + \alpha_1 r(x_0) \equiv x_0 + \alpha_0 p_0.$$

At this point, the coefficient α_0 is a free parameter. We can constrain it completely simply by minimizing $f(x_1)$; that is, by imposing

$$\frac{\partial}{\partial \alpha_0} f(x_1) = 0,$$

which yields

$$\alpha_0 = \frac{\langle r_0, p_0 \rangle}{\langle p_0, A p_0 \rangle}.$$

With α_0 completely determined by known quantities, we can then further update the residual

$$r(x_1) = r_0 - A x_1 = r_0 - \alpha_0 A p_0.$$

To summarize the first step, we started out with an initial guess x_0. Due to the equivalence of solving the linear system defined by Eqn. I.2 and minimizing the functional defined in Eqn. I.3, we were able to update our guess by following the gradient of f and enforcing that our new solution x_1 is as close to the minimum of f as is possible. This process gives the following set of updates

$$p_0 \equiv r(x_0) \tag{I.8}$$

$$\alpha_0 = \frac{\langle r_0, p_0 \rangle}{\langle p_0, A p_0 \rangle} = \frac{\langle r_0, r_0 \rangle}{\langle p_0, A p_0 \rangle} \tag{I.9}$$

$$x_0 \to x_1 = x_0 + \alpha_0 p_0 \tag{I.10}$$

$$r(x_0) \to r(x_1) = r(x_0) - \alpha_0 A p_0. \tag{I.11}$$

The next update will essentially follow the same sequence of updates; however, at this point (and in subsequent steps), we must now satisfy the constraint imposed by Eqn. I.6. Taking

$$p_n = r_{n+1} + \beta_n p_n$$

as an initially mysterious ansatz [98], Eqn. I.6 is satisfied if, in all subsequent updates,

$$\beta_n = \frac{\langle r_{n+1}, r_{n+1} \rangle}{\langle r_n, r_n \rangle}.$$

One can then repeat the procedure leading to α_0 in the first step to get α_n in all subsequent steps using $r_n = b - Ax_n$ for the residual and p_n from the previous update. The solution is updated in exactly the same fashion as in the first step, and the same is true for the residual.

To summarize, once we have updated our initial guess x_0 to x_1, all subsequent solutions x_n are obtained via the following chain of updates [98]

$$\beta_n = \frac{\langle r_{n+1}, r_{n+1} \rangle}{\langle r_n, r_n \rangle} \tag{I.12}$$

$$p_n \to p_{n+1} = r_{n+1} + \beta_n p_n \tag{I.13}$$

$$\alpha_n = \frac{\langle r_n, r_n \rangle}{\langle p_n, Ap_n \rangle} \tag{I.14}$$

$$x_n \to x_{n+1} = x_n + \alpha_n p_n \tag{I.15}$$

$$r(x_n) \to r(x_{n+1}) = r(x_n) - \alpha_n A p_n. \tag{I.16}$$

As n increases, each $\alpha_n p_n$ contributes less to x_{n+1}, since each p_n is constructed out of the residual r_n of the previous step. The latter sequence of updates is repeated until

$$\delta(x_n) \equiv \langle x - x_n, x - x_n \rangle_A \equiv \langle x - x_n, r(x_n) \rangle \tag{I.17}$$

reaches some desired precision, at which point the updating procedure is terminated and the solution x_{n+1} is returned. Note that, because A is positive-definite, so is $\delta(x_n)$. Moreover, because A is

Hermitian,

$$\delta(x_n) = \langle A^{-1} r(x_n), r(x_n)\rangle \underbrace{\sim}_{\text{roughly}} \langle r(x_n), r(x_n)\rangle. \tag{I.18}$$

Strictly speaking, $\delta(x_n)$ determines the size of the approximation error; however, because A^{-1} is constant, Eqn. I.18 implies that one can also use the norm of the residual, as one might have guessed intuitively. In practice, it is the norm of the residual $|r(x_n)|$ that is often used to set the precision of the conjugate gradient solve.

Appendix J

Monte Carlo algorithms for spin systems

In Chapter 6, I explore various spin systems for the purposes of testing out the neural-network-based finite size scaling method of that chapter. The data for the spin systems of that chapter is generated from Markov Chain Monte Carlo algorithms that work especially well for classical spin systems. This appendix briefly summarizes the two Monte Carlo algorithms used in that chapter.

J.1 The cluster algorithm

Cluster algorithms have been around for some time now. The first widely-used cluster algorithm was the Swendsen-Wang algorithm, which updates configurations by finding many clusters of spins and flipping them [348]. The Wolff cluster algorithm that is applied to the 2-/3-state Potts model and the 4-state clock model in Chapter 6 works similarly; however, instead of grabbing multiple clusters, it grabs one big cluster and flips it [377]. Take the reduced Hamiltonian of the spin system under consideration to be

$$\mathcal{H} = -K \sum_{\langle ij \rangle} \boldsymbol{s}_i \cdot \boldsymbol{s}_j, \tag{J.1}$$

where $\langle ij \rangle$ denotes a sum over all sites and their nearest neighbors j. All of the Hamiltonians in Chapter 6 can be written as Eqn. J.1, even if the spin degrees of freedom are discrete. The first step of each iteration of the Wolff cluster algorithm is to pick a random site i and a random direction $\hat{\boldsymbol{n}}$, then flip $\boldsymbol{s}_i$ along $\hat{\boldsymbol{n}}$ as

$$\boldsymbol{s}_i \to \boldsymbol{s}_i - 2\big(\boldsymbol{s}_i \cdot \hat{\boldsymbol{n}}\big)\hat{\boldsymbol{n}}. \tag{J.2}$$

With $\boldsymbol{s}_i$ flipped, a cluster is constructed recursively by visiting all neighboring sites, flipping them along $\hat{\boldsymbol{n}}$ with some probability, then visiting the neighbors of the flipped site until there are no more sites to update. The probability of flipping a neighboring site j is

$$\alpha = 1 - \exp\left[\min\bigl(0, 2K(\boldsymbol{s}_i \cdot \hat{\boldsymbol{n}})(\boldsymbol{s}_j \cdot \hat{\boldsymbol{n}})\bigr)\right]. \tag{J.3}$$

These steps are repeated until the desired precision is met. Because cluster algorithms flip entire clusters of spins, they suffer far less from critical slowing down than local update algorithms do. Hence, cluster algorithms are to this day considered the state-of-the-art when it comes to simulating spin systems close to criticality.

J.2 Heatbath algorithms

Heatbath algorithms operate by selecting individual spins, flipping them, then accepting the new spin-flipped configuration with probability $\alpha = (1 + \exp(-\delta\mathcal{H}))^{-1}$. Note that this is different than the standard Metropolis-Hastings update, which has a different acceptance probability. As the updates are local, heatbath algorithms tend to be inefficient within the vicinity of a 2nd-order phase transition without the help of additional overrelaxation. The heatbath algorithm is utilized to simulate the XY model in Chapter 6 for historical reasons that are discussed in that chapter. For the XY model, it is much more optimal to utilize the cluster algorithm discussed in Chapter J.1.