

格点量子色动力学基础

冯旭

暑期学校第一周的课程主要讲的是一些基础的东西，由马建平老师给大家讲微扰 QCD 基础，我负责格点 QCD 基础这一部分。我在准备这次课程讲义的时候，主要有一个想法，就是让大家了解 QCD 的非微扰定义。因为大部分同学学习规范场论的时候都是从微扰论出发的，对于格点 QCD 可能比较陌生一点。大家可能将来并不一定真的要拿格点 QCD 去做各种数值模拟和计算，但即便如此，对于格点 QCD 的了解也能帮助大家更好地理解 QCD。

首先，推荐一些格点 QCD 方面的常用教材。书籍方面主要是

- 1) *Quantum Chromodynamics on the Lattice: An Introductory Presentation*, C. Gattringer & C. Lang
- 2) *Lattice Methods for Quantum Chromodynamics*, Thomas DeGrand & Carleton DeTar
- 3) 格点量子色动力学学导论，刘川

Lecture note 方面，美国核理论所 (INT)2012 年举办过一次暑期学校

- 4) *INT Summer School on Lattice QCD for Nuclear Physics* (Lecture slides 在 INT 网页上都能找到 <http://www.int.washington.edu/PROGRAMS/12-2c/Lectures.html>)

法国 Les Houches 曾经举办过两期格点方面的暑期学校，比较近的一期是

- 5) *Modern Perspectives in Lattice QCD: quantum field theory and high performance computing* (相关 lecture note 已经结集出版)
- 6) 98 年还有一次暑期学校，arXiv 上有一些零星的讲义，比方说 R. Gupta 【hep-lat/9807028】和 M. Lüscher 【hep-lat/9802029】

这些书籍和讲义，对于初学者了解格点 QCD，应该是非常有用的，我在准备这次暑期学校讲义的过程中，也有参考以上文献。

1 引言

谈到 QCD，就不能不说它的三个特殊性质：

- 渐近自由：

$$\alpha_s(\mu) = \frac{1}{4\pi} \frac{1}{\beta_0 \ln \frac{\mu^2}{\Lambda_{\text{QCD}}^2}}, \quad \beta_0 = \left(\frac{11N_c - 2N_f}{3} \right) / 16\pi^2 \quad (1)$$

胶子自相互作用使得 QCD 的性质和 QED 很不一样，在高能区，强相互作用耦合常数变得很小，夸克和胶子渐近自由，而在低能区，相互作用却表现出很强的非微扰特性，使得微扰论失效。

- 色禁闭：

$$V(r) = A + \frac{B}{r} + \sigma r \quad (2)$$

夸克受到 QCD 相互作用的强力束缚，带单个色荷的夸克不可能从核子中单个地分离出来，这给我们研究强子结构造成了困难

- 手征对称性自发破缺：

$$\langle \bar{q}q \rangle = \langle \bar{q}_L q_R + \bar{q}_R q_L \rangle \neq 0 \quad \Rightarrow \quad SU(N_f)_V \times SU(N_f)_A \rightarrow SU(N_f)_V \quad (3)$$

由于真空夸克凝聚，使得手征对称性自发破缺，这也造就了粒子物理世界形形色色的强子谱。首先，Goldstone 定理告诉我们会有 $N_f^2 - 1$ 个无质量 goldstone boson，像 pion、kaon；除此之外，比如矢量介子和轴矢介子的能谱表现出很大的差异

$$m_\rho = 0.770 \text{ GeV}, \quad m_{a_1} = 1.26 \text{ GeV} \quad (4)$$

只有理解了 QCD 在低能区的性质，QCD 的真空结构，才能更好地帮助我们去理解强子的质量来源和相互作用行为。

格点 QCD 的应运而生，就是为了应对 QCD 的几种特殊性质。它能够非微扰地处理低能强相互作用，帮助我们研究 pion 介子、质子甚至一些轻核的强子结构的信息。（以前格点上研究的强子结构主要是与电弱流相关的形状因子、质子电荷半径之谜、质子的质量和自旋分解、以及部分子分布函数的矩等等，主要针对的是不含时的强子矩阵元；随着季向东老师提出的大动量有效理论，把含时的强子矩阵元在大动量框架下和不含时的强子矩阵元关联起来，现在格点也可以用来直接计算部分子分布函数。在这次暑期学校里面，季向东老师会向大家介绍他的大动量有效理论。）另外格点 QCD 也被用来研究色禁闭的机制以及手征对称性自发破缺、手征凝聚等。只要在一个物理过程里面，初态和末态是作为强子态出现的，那么夸克就被禁闭在强子内部，我们计算相关强子矩阵元就需要一种非微扰的处理方式。

随着计算资源的快速增加、以及算法和格点 QCD 研究手段的改进，特别是进入到 2000 年以后，格点 QCD 从淬火近似（也就是说 simulation 里面不含海夸克）到 full QCD 的 simulation，再到 physical quark mass simulation，像轻强子谱的计算以及一些简单强子的强子结构，格点 QCD 计算和实验符合得很好。如果说高能区的深度非弹实验与微扰 QCD 印证的揭示了 QCD 是强相互作用背后最基本的理论，那么低能区的强子谱和强子结构与非微扰格点 QCD 的印证实际上起到了同样的作用。需要指出的是，当前高能物理中的高精度前沿主要通过测量一些高精度的物理，比较它们与标准模型预言的差别来寻找新物理。而标准模型能否给出精确预言，很多时候依赖于我们对低能区 QCD 相互作用贡献的计算，因为这个原因，格点 QCD 在 precision flavor physics 等领域正起着举足轻重的作用。

格点 QCD 处理的对象是一个由强子构成的世界，能量尺度在 $\mu \lesssim 1 \text{ GeV}$ ，这里夸克和胶子禁闭在强子内部，而且强相互作用耦合常数在 $O(1)$ 的量级，微扰论失效。在这个情况下，我们需要一种非微扰的手段来研究夸克的禁闭机制，计算强子谱以及与强子态有关的各种强子矩阵元。为了实现这个目的，一种格点的非微扰的正规化方案被提了出来。应该说，Kenneth G. Wilson 1974 年提出格点规范场论这个概念的时候，并不是直接和计算机联系在一起的。1974 年的超级计算机，运行速度大概是我们现在一台普通笔记本电脑运行速度的千分之一的量级。Wilson 提出格点规范场论，主要是要提供一个解释夸克禁闭的机制。他在强相互作用耦合常数趋于无穷的情况下，得到了静态夸克势随着正反夸克距离变大而变大的结果，从而检验了夸克禁闭。但他把他的这个工作称为模型，原因就在于，他的结论是在耦合常数趋于无穷这个极限下导出来的；另一方面，我们可以看到，因为耦合常数在 $O(1)$ ，所以我们需要非微扰的计算，需要计算机的帮助来进行数值计算；但真当耦合常数趋于无穷时，解析的计算反倒可以做出来。我们后面讲纯规范场 Wilson 圈的时候，还会讲到 Wilson 的这个工作。他文章题目叫 Confinement of quarks, PRD 10 (1974) 2445，目前引用次数 > 4600 次，感兴趣的同学可以找来看一下。

【作业：阅读 K. G. Wilson 这篇开创性的文章，就自己的阅读体会写一个文献阅读报告。】

如果用一句概括格点 QCD，那么格点 QCD 就是构建在分立的欧氏时空格子下的 QCD 理论。我们来看一下这句话里面包含了哪些关键词。

1.1 第一个关键词是分立的时空格子

计算机只能处理有限的自由度，因此，格点 QCD 模拟的系统有一个有限大的体积 L ，有一个不为 0 的格距 a ，那么在一个维度上能容纳的格点数为 $N = L/a \sim 32, 48, 64$ ，考虑到 4 维时空，总的格点数就是 N^4 。

这里，有限体积 L 提供的是一个红外截断，而非零格距 a 提供了一个紫外截断。事实上，格点 QCD 提供一种非微扰的正规化方案，在给定的非零格距下，由于存在着紫外截断 π/a ，因此不会有无穷大发散的存在。另一方面，所有重整化以后的物理量不依赖于截断，在 $a \rightarrow 0$ 时不会发散，或者说它们在 $a \rightarrow 0$ 都有一个好的连续极限值。原则上，我们也可以利用格点正规化来进行微扰的计算，我们通常称为 lattice perturbation theory。只是这种微扰计算比起常用的维数正规化方案来讲，可能更加复杂。但对于非微扰计算来讲，格点正规化方案却是不可或缺的。

由于在量子系统中存在粒子的产生和湮灭，能产生和湮灭的最轻的强子为 pion 介子。因此我们需要格点的体积足够大，至少 pion 介子的康普顿波长 $\sim 1/m_\pi$ ，这样才能保证 pion 介子不会受到有限体积的显著影响；换句话说，这个时候，有限体积误差才会足够小。对于稳定强子来讲，这个误差是随着体积增大而指数衰减的。经验上来讲，当 $m_\pi L \geq 4$ 时，我们认为有效体积效应就不是那么重要了。这个时候，对应的 $L \sim 6$ fm。在格距 a 还没有趋于 0 的时候，如果我们要模拟一个重夸克的系统，比方说 charm quark，质量在 1.3 GeV，那么我们就需要格距 a 足够小，使得 am_c 是个小量，目前格点上模拟的 ultra-fine lattice spacing 在 0.04 fm，对应的能量标度是 5 GeV 左右。如果既想要最大的体积，又想要最小的格距，那么格子的大小为 150^4 ，对于 full QCD 的 simulation 来讲，这依然是个很艰巨的任务。目前各个格点组是根据他们的物理目标进行 simulation。如果研究的物理对象对于轻夸克质量的依赖比较敏感，那么通常物理体积会足够大；这个时候格距可以取得大一点，比方说 0.11, 0.09, 0.07 fm，然后做连续极限外推得到 $a \rightarrow 0$ 的结果。如果对于格距误差 (lattice artifacts) 要求比较高，那么 pion 的质量可以比物理质量重一些，比方说 $m_\pi = 400, 300, 200$ MeV，然后外推到物理的 pion 质量上去。

目前我所知的最大的格点 simulation 来自 M. Lüscher 在 2017 年的格点年会上的汇报，他模拟了一个 $N = 192$ 的淬火格点系统【1707.09758】。它的格距是 0.1fm，物理体积是 19.2fm。这样一个最大的 $SU(3)$ 格点规范场系统的模拟是在一个 64 节点，1536 核，8.2TB 内存的机器上花了 10 天的时间模拟出来的。因为 QCD 系统并不是一个长程的系统，如果格点模拟的体积很大，那么场变量在离得很远的时候的关联很弱，或者可以认为它们的涨落是独立的。这种性质也被称为“stochastic locality”。由于计算机模拟一个 192^4 的格子和模拟 256 个 48^4 的格子所耗费的计算资源是差不多的，那么利用 stochastic locality 以及平移不变性，我们可以在一个 192^4 的大格子上做几百次测量，来得到比几百个 48^4 的规范场组态做平均上相近甚至更好的物理结果。

1.2 第二个关键词是欧氏时空

在 Wick 转动下，我们从闵氏时空转换到欧氏时空。位置和动量空间的 0 分量变为

$$\begin{aligned} x_0 &\equiv t \rightarrow -ix_4 \equiv -i\tau \\ p_0 &\equiv E \rightarrow ip_4 \end{aligned} \quad (5)$$

除了时间分量，欧氏时空中的其它三个分量保持不变，与闵氏时空一样。在欧氏时空下，度规取成 $(+, +, +, +)$ ，因此有

$$\begin{aligned} x_E^2 &= \sum_{i=1}^4 x_i^2 = \vec{x}^2 - t^2 = -x_M^2 \\ p_E^2 &= \sum_{i=1}^4 p_i^2 = \vec{p}^2 - E^2 = -p_M^2 \end{aligned} \quad (6)$$

我们下面来看一个闵氏时空的 2 点格林函数 $\Gamma_M(t, \vec{k}) = \int d^3\vec{x} e^{-i\vec{k}\cdot\vec{x}} \langle 0|T[\hat{O}_M(\vec{x}, t)\hat{O}_M^\dagger(\vec{0}, 0)]|0\rangle$ ，其中算符 $\hat{O}(\vec{x}, t)$ 在海森堡表象下可以写为

$$\hat{O}_M(\vec{x}, t) = e^{i\hat{H}t - i\hat{p}x} \hat{O}_M(\vec{0}, 0) e^{-i\hat{H}t + i\hat{p}x} \quad (7)$$

不妨取 $t > 0$ ，我们可以去掉编时乘积，那么格林函数

$$\Gamma_M(t, \vec{k}) = \int d^3\vec{x} e^{-i\vec{k}\cdot\vec{x}} \langle 0|e^{i\hat{H}t - i\hat{p}x} \hat{O}_M(\vec{0}, 0) e^{-i\hat{H}t + i\hat{p}x} \hat{O}_M^\dagger(\vec{0}, 0)|0\rangle \quad (8)$$

利用 Hilbert 空间的完备性，插入能量本征态 $1 = \sum_n \int \frac{d^3\vec{p}}{(2\pi)^3} |n, \vec{p}\rangle \langle n, \vec{p}|$ ，我们得到

$$\begin{aligned} \Gamma_M(t, \vec{k}) &= \sum_n \langle 0|\hat{O}_M(\vec{0}, 0)|n, \vec{k}\rangle e^{-i\hat{H}t} \langle n, \vec{k}|\hat{O}_M^\dagger(\vec{0}, 0)|0\rangle \\ &= \sum_n \langle 0|\hat{O}_M(\vec{0}, 0)|n, \vec{k}\rangle e^{-iE_n t} \langle n, \vec{k}|\hat{O}_M^\dagger(\vec{0}, 0)|0\rangle \end{aligned} \quad (9)$$

其中 $|n, \vec{k}\rangle$ 代表带动量 \vec{k} 的、量子数对应于 \hat{O}_M 算符的强子态的第 n 个激发态。

对于欧氏时空的 2 点格林函数 $\Gamma_E(\tau, \vec{k}) = \int d^3\vec{x} e^{-i\vec{k}\cdot\vec{x}} \langle 0|T[\hat{O}_E(\vec{x}, \tau)\hat{O}_E^\dagger(\vec{0}, 0)]|0\rangle$ ，其算符写为

$$\hat{O}_E(\vec{x}, \tau) = e^{\hat{H}\tau - i\hat{p}x} \hat{O}_E(\vec{0}, 0) e^{-\hat{H}\tau + i\hat{p}x} \quad (10)$$

在 $t = \tau = 0$ 的时间上，我们有 $\hat{O}_M(\vec{x}, 0) = \hat{O}_E(\vec{x}, 0)$ 。但当 $t = \tau \neq 0$ 时，很明显 $\hat{O}_M(\vec{x}, 0) \neq \hat{O}_E(\vec{x}, 0)$ 。在 Wick 转动底下， $t \rightarrow -i\tau$ ，但哈密顿量算符 \hat{H} 保持不变，对于欧氏时空 2 点函数，我们有

$$\Gamma_E(\tau, \vec{k}) = \sum_n \langle 0|\hat{O}_E(\vec{0}, 0)|n, \vec{k}\rangle e^{-\hat{H}\tau} \langle n, \vec{k}|\hat{O}_E^\dagger(\vec{0}, 0)|0\rangle \quad (11)$$

很明显，欧氏时空的格林函数的时间依赖关系和闵氏时空的格林函数的时间依赖关系是不一样的。因此对于定义在闵氏时空的含时算符，或者说含时矩阵元，它们所描述的物理量，与欧氏时空所定义的物理量，往往是不一样的。解析延拓这件事情，不一定很 trivial。但是这里，希望大家注意到的是，无论在闵氏时空、还是在欧氏时空，哈密顿量是完全一样的，因此对应的能谱都一样，能量本征态也一样。由于欧氏算符和闵氏算符在时间等于 0 的时候是一样，对于不含时的强子矩阵元，比如说 $\langle 0|\hat{O}(\vec{0}, 0)|n, \vec{k}\rangle$ ，欧氏和闵氏时空得到的也是一样的。

比较了相同点，我们下面来说说闵氏时空和欧氏时空物理可能有哪些不一样的地方

- 传统的含时物理量，比如质子的部分子分布函数 (parton distribution function)

$$q(x, \mu^2) = \int \frac{d\xi^-}{4\pi} e^{-ix\xi^- P^+} \langle N(P)|\bar{\psi}(\xi^-)\gamma^+ \exp\left(-ig \int_0^{\xi^-} d\eta^- A^+(\eta^-)\right) \psi(0)|N(P)\rangle \quad (12)$$

其中质子的动量 P^μ 可以取为 z 方向， $P^\mu = (P^0, 0, 0, P^z)$ ， $\xi^\pm = (t \pm z)/\sqrt{2}$ 是光锥坐标， t 是闵氏时间。由于算符明显含时，我们没法在欧氏时空直接计算。解决方案：大动量有效理论（季向东）【1305.1539、1404.6680】等一系列文章。

- 含共振态和多粒子态的物理量。我们前面通过构造关联函数，可以提取出哈密顿量的本征态，对应的是粒子的能谱。但无论共振态还是多粒子态，它们对应的是 branch cut，在能量大于两粒子阈值的时候，能量是连续的，而在格点上，由于整个物理体积是有限的，在有限的体积中能谱必然是分立的。解决方案：有限体积方法 (M. Lüscher) 【Commun.Math.Phys. 105 (1986) 153, NPB 354 (1991) 531】把有限体积中计算得到的能量，转变为无穷体积中的粒子-粒子散射振幅。
- Time-like 和 space-like 区域的差别。我们来看一个真空极化函数，在闵氏时空，它可以定义为

$$\Pi_{\mu\nu}^{(M)} = \int d^4x_M e^{iqx_M} \langle 0 | J_\mu(x_M) J_\nu(0) | 0 \rangle, \quad \Pi_{\mu\nu}^{(M)} = (q_\mu q_\nu - g_{\mu\nu} q^2) \cdot \Pi^{(M)}(q^2) \quad (13)$$

这个真空极化函数对应于 $e^+e^- \rightarrow \gamma^* \rightarrow \text{hadrons}$ 的过程。 $s = q^2 > 0$ ，对应 time-like 区域，这里 s 可以认为是末态强子的不变质量平方。根据光学定理 (从么正性导出)，这个真空极化函数的虚部，与强子谱密度 (spectral density) 成正比，可以由 BESIII 实验的 R 值测量来给出

$$\text{Im } \Pi^{(M)}(s) = 2\pi\rho(s) = \frac{R(s)}{6\pi}, \quad R(s) \equiv \frac{\sigma(e^+e^- \rightarrow \text{hadrons})}{4\pi\alpha(s)^2/(3s)} \quad (14)$$

而对于欧氏时空，做完傅立叶变换以后，我们得到

$$\Pi_{\mu\nu}^{(E)} = \int d^4x_E e^{iQx_E} \langle 0 | J_\mu(x_E) J_\nu(0) | 0 \rangle, \quad \Pi_{\mu\nu}^{(E)}(Q^2) = (Q_\mu Q_\nu - \delta_{\mu\nu} Q^2) \cdot \Pi^{(E)}(Q^2) \quad (15)$$

这里 $q^2 = -Q^2 < 0$ ，对应 space-like 区域。根据真空极化函数的解析性 (可以从因果率导出)，两个区域的物理可以通过 once-subtracted dispersion relation 关联起来

$$\Pi^{(E)}(Q^2) - \Pi^{(E)}(0) = Q^2 \int_0^\infty \frac{ds}{2\pi} \frac{\text{Im } \Pi^{(M)}(s)}{s(s+Q^2)} \quad (16)$$

在 time-like 区域，通过实验 R 值测量，我们可以看到各个共振态的峰，而 space-like 区域，真空极化函数随能量的变化是平滑的，两者的物理现象很不一样。要在格点上处理 time-like 区域的物理，也不是说没有办法，对于末态是两体散射或者两个散射道的情况，可以用 Lellouch-Lüscher 方法 【hep-lat/0003023】去处理，但对于深度非弹性的过程，末态粒子包含各种强子，目前还没有好的解决方案。

从路径积分的角度来看一个 2 点格林函数，在闵氏时空下，它可以写成

$$\langle \hat{O}_1 \hat{O}_2 \rangle_M = \frac{1}{Z_M} \int \mathcal{D}[A_\mu] \mathcal{D}[\psi] \mathcal{D}[\bar{\psi}] O_1 O_2 e^{-iS_M} \quad (17)$$

这里，配分函数 Z 由路径积分给出

$$Z_M = \int \mathcal{D}[A_\mu] \mathcal{D}[\psi] \mathcal{D}[\bar{\psi}] e^{-iS_M} \quad (18)$$

它在场论里面代表真空到真空的振幅。等式 (17) 的左边用的是量子场论场算符的语言，而等式右边与场算符没有任何关系，这里的 $O_{1,2}$ 代表的是以规范场和费米子场为变量的泛函。注意到，在闵氏时空，路径积分带有一个权重因子 e^{-iS_M} 。

如果用计算机对路径积分进行 Monte Carlo 数值模拟，那么我们需要一个高斯型的权重因子，而不是一个震荡型的分布函数。这就是为什么我们在欧氏时空去处理量子场论问题。在 Wick 转动下

$t \rightarrow -i\tau$, 权重因子变为 $e^{-iS_M} \rightarrow e^{-S_E}$ 。这个时候的权重因子实际上相当于统计力学系统的玻尔兹曼因子。2 点格林函数可以写成

$$\langle \hat{O}_1 \hat{O}_2 \rangle_E = \frac{1}{Z_E} \int \mathcal{D}[A_\mu] \mathcal{D}[\psi] D\bar{\psi} O_1 O_2 e^{-S_E} \quad (19)$$

这里

$$Z_E = \int \mathcal{D}[A_\mu] \mathcal{D}[\psi] D\bar{\psi} e^{-S_E} \quad (20)$$

事实上配分函数 Z_M 和 Z_E 并没有太大区别。对于欧氏时空, 它告诉我们作用量最小的路径对于路径积分的贡献最大, 其它路径的贡献被指数压低了; 而对于闵氏时空, 大作用量对应的路径有一个剧烈震荡的因子, 在积分时贡献也会抵消掉, 出来的效果是一致的。这里我要强调一点, 把闵氏空间和欧氏空间的关联函数和路径积分分别进行比较, 不仅仅是为了说明欧氏时空场论给出的物理和闵氏时空是一致的, 更为根本的, 我们可以把欧氏时空的场论看成是量子场论的一种非微扰定义。

把与场算符相关的格林函数用路径积分的形式表达出来, 可以说是整个格点 QCD simulation 的基础。在这个基础上, 我们可以把欧氏时空路径积分系统类比为 一个统计力学系统。对格点 QCD 做数值模拟和对一个统计力学系统做数值模拟从原理上来讲是一样的。

表 1: 欧氏时空场论 vs 经典统计力学

欧氏时空场论	经典统计力学
作用量: $S_E[\psi, \bar{\psi}, A]$	哈密顿量: H
路径积分权重因子: e^{-S_E}	玻尔兹曼因子: $e^{-\beta H}$
真空到真空的振幅: $\int \mathcal{D}[A_\mu] \mathcal{D}[\psi] D\bar{\psi} e^{-S_E}$	配分函数: $\sum_{conf.} e^{-\beta H}$
真空能量	自由能
真空期望值: $\langle 0 \hat{O} 0 \rangle$	正则系综平均: $\langle O \rangle$
格林函数: $\langle 0 T [O_1 \cdots O_n] 0 \rangle$	关联函数: $\langle O_1 \cdots O_n \rangle$

1.3 第三个关键词: QCD 理论——格点 QCD 构建的必须是个 QCD 理论

同学们可以想一下, 假如你是 Wilson, 要在格点上构建 QCD 理论, 你需要做些什么事情: 第一、你要构建格点离散化的费米子场和规范场, 这是格点正规化所必须的; 第二、用这些场来构建作用量, 在构建作用量的时候要尽可能保证原有的对称性不被破坏, 在格距趋于 0 的时候要保证能回到连续时空的 QCD 理论; 第三、我们需要做路径积分, 那么还需要明确定义费米子场和规范场积分测度。或者, 换句话说, 我们要给路径积分一个严格的非微扰的定义。

我们先来做第一件事情, 把费米子场放在欧氏 4 维时空格子的每个格点上

$$\psi(n), \quad \bar{\psi}(n), \quad n_\mu = 0, 1, \dots, L-1, \quad \mu = 1, 2, 3, 4 \quad (21)$$

连续时空的自由费米子场作用量的表达式给出如下

$$S_F^0[\psi, \bar{\psi}] = \int d^4x \bar{\psi}(x) (\gamma_\mu \partial_\mu + m) \psi(x) \quad (22)$$

在分立的时空里面, 我们把求导替换为差分算符, 可以有向前、向后差分

$$\partial_\mu \psi(n) = \frac{1}{a} (\psi(n + \hat{\mu}) - \psi(n)), \quad \partial_\mu^* \psi(n) = \frac{1}{a} (\psi(n) - \psi(n - \hat{\mu})) \quad (23)$$

或者两者求平均

$$\bar{\partial}_\mu \psi(n) = \frac{\psi(n + \hat{\mu}) - \psi(n - \hat{\mu})}{2a} \quad (24)$$

这样, 格点上给出的自由费米子场作用量为

$$S_F^0[\psi, \bar{\psi}] = a^4 \sum_n \bar{\psi}(n) \left(\sum_{\mu=1}^4 \gamma_\mu \frac{\psi(n + \hat{\mu}) - \psi(n - \hat{\mu})}{2a} + m\psi(n) \right) \quad (25)$$

计算机处理的都是无量纲的数, 我们也可以把格距 a 吸收到费米子场和质量中, 构成无量纲的量 $a^{\frac{3}{2}}\psi(n)$ 、 am 。对于费米子场的路径积分来讲, 格点上的费米子场依然可以用 Grassmann 数来表示, 对费米子场的积分依然满足 Grassmann 数积分

$$\begin{aligned} \int d\psi d\bar{\psi} &= \int d\psi d\bar{\psi} \psi = \int d\psi d\bar{\psi} \bar{\psi} = 0 \\ \int d\psi \psi &= \int d\bar{\psi} \bar{\psi} = 1 \\ \int d\psi d\bar{\psi} \bar{\psi} \psi &= 1 \end{aligned} \quad (26)$$

还有一点需要指出的是, 由于我们在欧氏时空选择了度规 $(+, +, +, +)$ 。那么我们的 γ 矩阵选择的也是欧氏时空的 γ 矩阵, 其中一种手征表示是

$$\vec{\gamma} = \begin{pmatrix} 0 & i\vec{\sigma} \\ -i\vec{\sigma} & 0 \end{pmatrix}, \quad \gamma_4 = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} \quad (27)$$

它们具有厄米性 $\gamma_\mu^\dagger = \gamma_\mu$, 并且满足反对易关系

$$\{\gamma_\mu, \gamma_\nu\} = 2\delta_{\mu\nu} I \quad (28)$$

这里, 我们得到的是 $2\delta_{\mu\nu}$, 而不是 $2g_{\mu\nu}$ 。在欧氏时空中, γ_5 矩阵定义为

$$\gamma_5 = \gamma_1 \gamma_2 \gamma_3 \gamma_4 = \begin{pmatrix} -I & 0 \\ 0 & I \end{pmatrix} \quad (29)$$

有了费米子场, 下面要加入规范场。很明显, 这里规范对称性是我们首先要保证的对称性。对费米子场做规范变换, 我们有

$$\psi(n) \rightarrow V(n)\psi(n), \quad \bar{\psi}(m) \rightarrow \bar{\psi}(m)V^\dagger(m) \quad (30)$$

我们要使得正反夸克构成的算符规范不变, 最好的做法是像 Wilson 一样引进一个规范场链接 (link), 连接正反夸克场

$$\bar{\psi}(m)U(m, n)\psi(n) \equiv \bar{\psi}(m) \left(\mathcal{P} e^{\int_n^m igA_\mu(x)dx_\mu} \right) \psi(n) \quad (31)$$

这里 \mathcal{P} 表示积分和路径有关。链接 $U(m, n)$ 满足规范变换

$$U(m, n) \rightarrow V(m)U(m, n)V^\dagger(n) \quad (32)$$

有了这个链接以后, 整个夸克 bilinear 算符是规范不变的。在我们之前构造的费米子作用量里, 只有相邻的正反夸克场发生耦合, 这个时候我们可以引进长度为 1 的链接, 刚好连接这样的两个夸克场

$$U_\mu(n) \equiv U(n, n + \hat{\mu}) = e^{iagA_\mu(n + \frac{\hat{\mu}}{2})} \quad (33)$$

从这个定义可以看出, $U_\mu(n)$ 并非定义在 n 上, 而是定义在从 n 出发向四个方向的链接上。这个定义其实来得很自然。因为规范场描述的就是夸克之间的相互作用, 如果把夸克放在格点上, 那么把规范场放在连接格点的 link 上是恰如其分的。

有了 $U_\mu(n)$ 以后, 我们也可以定义一个反向的 link

$$U_{-\mu}(n) \equiv U(n, n - \hat{\mu}) = U(n - \hat{\mu}, n)^\dagger = U_\mu(n - \hat{\mu})^\dagger = e^{-iagA_\mu(n - \frac{\hat{\mu}}{2})} \quad (34)$$

它满足规范变换

$$U_{-\mu}(n) \rightarrow V(n)U_{-\mu}(n)V(n - \hat{\mu})^\dagger \quad (35)$$

把规范场加入到费米子场作用量里面去, 我们得到

$$S_F[\psi, \bar{\psi}] = a^4 \sum_n \bar{\psi}(n) \left(\sum_{\mu=1}^4 \gamma_\mu \frac{U_\mu(n)\psi(n + \hat{\mu}) - U_{-\mu}(n)\psi(n - \hat{\mu})}{2a} + m\psi(n) \right) \quad (36)$$

这种构造很简单, 只需要给不在同一位置上的夸克场引入连接它们的链接即可。这里我们顺便可以给出向前、向后协变差分的定义

$$D_\mu\psi(n) = \frac{1}{a}(U_\mu(n)\psi(n + \hat{\mu}) - \psi(n)), \quad D_\mu^*\psi(n) = \frac{1}{a}(\psi(n) - U_{-\mu}(n)\psi(n - \hat{\mu})) \quad (37)$$

【作业：对费米子场作用量进行格距 a 的微扰展开, 给出 $O(a)$ 和 $O(a^2)$ 项的表达式, 并确认, $O(1)$ 这一项和连续时空的形式是一致的。】

需要指出的是, 这里所用的费米子场叫做 naive fermion, 它会有 doubling 的问题, 也就是说费米子传播子除了一个物理的极点之外, 还存在非物理的极点, 因此, 并不是我们真正在格点中使用的费米子场作用量。我们后面还会讲到这件事, 现在先略过去。

在格点上, 构造规范不变的算符有两种方式, 一种是正反夸克场之间用 gauge link 来连接, 形如

$$\bar{\psi}(m)U_\mu(m)U_\nu(m + \hat{\mu}) \cdots U_\rho(n - \hat{\rho})\psi(n) \quad (38)$$

还有一种是由纯规范场构成, 由 gauge linke 形成一个闭合的圈, 最简单的例子是小方块

$$P_{\mu\nu}(n) = \text{Re Tr}(U_\mu(n)U_\nu(n + \hat{\mu})U_\mu^\dagger(n + \hat{\nu})U_\nu^\dagger(n)) \quad (39)$$

由于 $U_\mu(n)$ 是个 SU(3) 的矩阵, 小方块求迹之后是个复数。可以证明, 如果小方块的 4 个链接反着走, 与正着走刚好形成复共轭。取实部实际上是把这两种路径做了平均。

【作业：请确认对于规范场的闭合圈, 正着走和反着走刚好形成复共轭。另外, 请确认对于规范不变的算符, 只有上述两种构造方式。】

在连续极限下, 我们可以把一个闭合圈上的规范场乘积写为

$$G(x, y) = \mathcal{P} \exp \left(ig \int_{\mathcal{C}_{x,y}} A \cdot dr \right) = \exp \left(ig \int_{\mathcal{S}_{x,y}} \nabla \times A \cdot ds \right) \quad (40)$$

第二个式子, 由 Stokes 公式得到。如果我们现在处理的是一个 μ, ν 方向的小方块, 那么我们有

$$P_{\mu\nu}(n) = \exp (ia^2 g F_{\mu\nu} + O(a^4)) \quad (41)$$

【作业：请从小方块 $P_{\mu\nu}(n)$ 的定义出发，验证这个式子，先考虑 $U(1)$ 规范场，再考虑更为复杂的 $SU(3)$ 非阿贝尔规范场。提示：在 $n + \frac{\mu+\nu}{2}$ 处做展开】

下一步， $P_{\mu\nu}(n)$ 对 a 做 Taylor 展开，得到

$$P_{\mu\nu}(n) = 1 + ia^2 g F_{\mu\nu} - \frac{1}{2} a^4 g^2 F_{\mu\nu} F_{\mu\nu} + O(g^6) \quad (42)$$

为了消去第一项和第二项，我们将规范场作用量定义为

$$S_G[U] = \frac{2}{g^2} \sum_n \sum_{\mu < \nu} \text{Re Tr}[1 - P_{\mu\nu}(n)] = \frac{a^4}{2g^2} \sum_n \sum_{\mu, \nu} \text{Tr}[F_{\mu\nu} F_{\mu\nu}] + O(a^2) \quad (43)$$

在 $a \rightarrow 0$ 的极限下， $a^4 \sum_n \rightarrow \int d^4x$ ，我们得到 $S_G[U] = S_G[A]$ 。

有了作用量以后，我们来看一下格点系统中的对称性。首先，在欧氏时空下 Lorentz 对称性变成了 $SO(4)$ 对称性。由于格点离散化，这个对称性会进一步破缺为 hypercubic 群。原本任意的旋转变换现在只能允许转 90° 。平移不变性这件事情，只能以格距为单位进行平移，才能保证。这也意味着，在周期性边界条件下，格点上的动量只能允许取

$$k = \frac{2\pi n}{La}, \quad n = 0, 1, 2, \dots, L-1 \quad (44)$$

或者等价写成

$$k = \pm \frac{2\pi n}{La}, \quad n = 0, 1, 2, \dots, L/2 \quad (45)$$

这些对称性的破缺，在格点正规化这个框架下，几乎是不可避免的。但是除此之外，格点 QCD 尽可能地保留了其它的对称性，比方说宇称 (\mathcal{P})、电荷共轭 (\mathcal{C}) 和时间反演 (\mathcal{T})。费米子场和规范场在 \mathcal{P} 、 \mathcal{C} 、 \mathcal{T} 变换下满足

	\mathcal{P}	\mathcal{C}	\mathcal{T}
$U_4(\vec{x}, \tau)$	$U_4(-\vec{x}, \tau)$	$U_4^*(\vec{x}, \tau)$	$U_{-4}(\vec{x}, -\tau)$
$U_i(\vec{x}, \tau)$	$U_{-i}(-\vec{x}, \tau)$	$U_i^*(\vec{x}, \tau)$	$U_i(\vec{x}, -\tau)$
$\psi(\vec{x}, \tau)$	$\gamma_4 \psi(-\vec{x}, \tau)$	$C \bar{\psi}^T(\vec{x}, \tau)$	$\gamma_4 \gamma_5 \psi(\vec{x}, -\tau)$
$\bar{\psi}(\vec{x}, \tau)$	$\bar{\psi}(-\vec{x}, \tau) \gamma_4$	$-\psi^T(\vec{x}, \tau) C^{-1}$	$\bar{\psi}(\vec{x}, -\tau) \gamma_5 \gamma_4$

这里电荷共轭矩阵 C 可以写成 $C = \gamma_4 \gamma_2$ 的形式，它满足关系式

$$C \gamma_\mu C^{-1} = -\gamma_\mu^T = -\gamma_\mu^* \quad (46)$$

还有一个保存下来的对称性的是规范不变性，这在前面作用量的构造里面可以看到。对于 naive 费米子来讲，手征对称性也是满足的，但是 naive 费米子会有 doubling 的问题，Wilson 的解决方案是引进了 Wilson 项，但这一引进项破坏了手征对称性。我们在后面讲到费米子章节的时候会具体讲到这个问题。

既然格点 QCD 给出的是 QCD 的非微扰定义，那么所有 QCD 依赖的参数我们这里也需要：一个是 α_s ，在我们的格点作用量里面由裸的参数 g^2 来给出；还有一个是夸克质量，包括 up, down, strange

三个较轻的夸克，以及比较重的 charm 和 bottom quark, top 夸克寿命太短，也不形成强子态，所以不在格点 QCD 的研究范畴。charm 夸克目前是直接可以像三个较轻的夸克一样进行 simulation 的, bottom quark 的引入一般要借助于重夸克有效理论、或者 Non-relativistic QCD 等手段。格点作用量里面作为参数引入的夸克质量也是裸的夸克质量。之所以说这些参数是裸的，是因为它们是在格点正规化下引入的，会依赖于截断，也就是非零格距 a 。我们要调节裸参数 $g(a), m_i(a)$ ，使得格点计算出来的物理量和实验符合

$$O(g(a), m_i(a), a) = O_{\text{phys}} \quad (47)$$

比方说，我们调节裸夸克的质量，使得格点 QCD 计算出来的 π 介子质量等于实验的测量值，那么这个时候的 u/d 夸克参数可以认为是物理的参数，我们让 kaon 或者 Ω 重子的质量也等于实验测量值，那么这个时候的 s 夸克也调到了物理值。一旦用几个物理量把 QCD 的参数定下来以后，其它的物理量全是格点 QCD 的 prediction。

这里其实还有一个问题，就是计算机上得到的数全是无量纲的数，以 pion 介子质量为例，我们计算得到的实际是无量纲量 am_π 的值。要与 PDG 上物理的值进行比较，我们首先要知道，格距 a 到底是多大？而且，由于格距 a 这个量的存在，粗看起来，感觉我们好像在格点 QCD 里面多引进了一个参数，这是怎么回事？我们下面来回答这两个问题。我们先考虑一个没有相互作用的系统，假如格距 $a = 0.1$ fm, 物理尺寸 $L = 1$ fm, 那个放在计算机上，它对应于一个 $(L/a)^4 = 10^4$ 的立方格子；如果我们再定义一个格距 $a = 0.2$ fm, 物理尺寸 $L = 2$ fm 的系统，放在计算机上，它还是对应于一个 $(L/a)^4 = 10^4$ 的立方格子；在无相互作用的时候，两者完全没有区别。但当引进了相互作用以后，实际上低能 QCD 是存在一个非微扰的 haronic scale, 这个 scale 也可以认为是我们之前提到的 Λ_{QCD} 。裸的耦合常数 g^2 和截断 a 之间满足重整化群耦合常数跑动

$$g^2(a) = \frac{1}{\beta_0 \ln \frac{1}{a^2 \Lambda^2}} \Rightarrow \Lambda = \frac{1}{a} \exp\left(-\frac{1}{2\beta_0 g^2}\right) \quad (48)$$

由于 Λ 这个 scale 的存在，使得我们输入的裸耦合常数和格距 a 是不独立的。输入了裸耦合常数，相当于我们已经设定了格距 a 。当然，格点上最后定格距 a 并不是通过上面这个式子，原因在于， Λ 是个非微扰的 scale，而等式的左边却是从微扰论里来得。我们后面会具体讲到如何给格距定标。

因为格点 QCD 的参数可调，事实它不仅给出我们感兴趣的物理量的标准模型预言值，也可以通过研究物理量对夸克质量的依赖关系，帮助我们研究手征微扰论，确定低能有效常数；另外也可以用到重夸克有效理论中去。

2 格点上的纯规范场

在纯规范场下，物理量 O 的期望值可以由路径积分

$$\langle O \rangle = \frac{1}{Z} \int \mathcal{D}[U] O[U] e^{-S_G[U]} \quad (49)$$

给出。这里配分函数 Z 定义为

$$Z = \int \mathcal{D}[U] e^{-S_G[U]}, \quad (50)$$

积分测度定义为

$$\mathcal{D}[U] = \prod_{x \in L^{3T}} \prod_{\mu=1}^4 \int dU_\mu(x) \quad (51)$$

规范场的作用量定义为

$$S_G[U] = \frac{\beta}{N_c} \sum_{x \in L^3 T} \sum_{\mu < \nu} \text{Re Tr}[1 - P_{\mu\nu}(x)] \quad (52)$$

这里 $\beta = \frac{2N_c}{g^2}$ 叫做 inverse coupling, 之所以这么定义是有历史的原因, 我们后面会看到。

因为所有的变量 $U_\mu(x)$ 的取值都是 $SU(3)$ 群的群元, 所以我们实际上是在一个连续的紧致群 (compact group) 上进行积分, 这时候的积分测度在数学上被称为 Haar 测度。在规范变化下, $U_\mu(x)$ 变为

$$U_\mu(x) \rightarrow U'_\mu(x) = V(x)U_\mu(x)V(x + \hat{\mu})^\dagger \quad (53)$$

要满足规范不变性, 我们要求积分测度

$$dU_\mu(x) = dU'_\mu(x) = d(V(x)U_\mu(x)V(x + \hat{\mu})^\dagger) \quad (54)$$

因为群元 $V(x)$ 的选取具有任意性, 因此, 我们实际上要求无论左乘还是右乘一个群元 $W \in SU(3)$, 都有

$$dU = d(UW) = d(WU) \quad (55)$$

因为 U_μ 是从 $SU(3)$ 群群元里面取值, 如果用矩阵表示, 矩阵元素的取值范围在 0 到 1 之间, 我们管这种规范场叫 compact gauge field。对于我们在微扰论里面经常用的规范场 A_μ , 它是通过 $SU(3)$ 群的生成元来构造, 这种规范场称为 non-compact gauge field。在连续极限下, 规范场的自由度从 compact 变成 non-compact。但是在格点正规化的情况, 由于非零格距的引入, 规范场自然而然的是 compact 的, 所有的积分测度都不会发散, 因此也不需要做 gauge fixing。由于在做路径积分的时候, 有限的积分测度会在分子、分母里抵消掉一个因子, 因此, 我们不妨引入归一化条件

$$\int dU 1 = 1 \quad (56)$$

利用这两个条件, 我们可以得到一系列 $SU(3)$ 积分

$$\begin{aligned} \int_{SU(3)} dU U_{ab} &= 0 \\ \int_{SU(3)} dU U_{ab}U_{cd} &= 0 \\ \int_{SU(3)} dU U_{ab}(U^\dagger)_{cd} &= \frac{1}{3}\delta_{ad}\delta_{bc} \\ \int_{SU(3)} dU U_{ab}U_{cd}U_{ef} &= \frac{1}{6}\epsilon_{ace}\epsilon_{bdf} \end{aligned} \quad (57)$$

导出这几个积分的思路如下: 由条件 (55) 我们可以推出

$$\int dU f(U) = \int d(VU) f(VU) = \int dU f(VU) \quad (58)$$

这其实是个很霸道的条件, 我们在数学上有个类似的例子, 就是如果 $x = cx$, 且 c 可以取任意实数, 那么必然有 $x = 0$ 。类似的, 对于被积函数 $f(U)$, 在大部分情况下, 积分给出的结果都是 0。比方说 $f(U)$ 的形式是 U_{ab} 或者 $U_{ab}U_{cd}$, 我们必然有

$$\int dU U_{ab} = \int dU (VU)_{ab} = V_{ac} \int dU U_{cb} \quad (59)$$

由于 V_{ac} 可以取任意 $SU(3)$ 群元, 我们只能让 $\int dU U_{ab} = 0$ 。但是当 $f(U)$ 的形式是 UU^\dagger 的时候, 根据归一化条件, 我们可以有

$$\sum_b \int dU U_{ab} U_{bd}^\dagger = \delta_{ab} \quad \Rightarrow \quad \int dU U_{ab} U_{cd}^\dagger = \frac{1}{N_c} \delta_{ac} \delta_{bd} \quad (60)$$

也就是积分不为 0 的情况对应于 $f(U)$ 给出的值是 1, 或者 $f(U)$ 函数给出的是 U 的平庸表示。如果 $f(U)$ 给出的是显式依赖于 U 的表达式, 那么积分就为 0。除了 $UU^\dagger = 1$ 之外, 我们知道 U 的行列式也等于 1, $\det(U) = 1$, 这个时候对应于

$$f(U) = \det(U) = \epsilon_{abc} U_{a1} U_{b2} U_{c3} \quad (61)$$

或者, 我们写出更一般的表达式来, $f(U)$ 可以取

$$f(U) = \epsilon_{abc} U_{ad} U_{be} U_{cf} = \epsilon_{def} \quad (62)$$

$$\int dU \epsilon_{abc} U_{ad} U_{be} U_{cf} = \epsilon_{def}, \quad \int dU U_{ad} U_{be} U_{cf} = \frac{1}{6} \epsilon_{abc} \epsilon_{def} \quad (63)$$

这几个积分公式里面, 最重要的是第三个, 从它我们可以得到

$$\int dU \text{Tr}[VU] \text{Tr}[U^\dagger W] = \frac{1}{3} \text{Tr}[VW]. \quad (64)$$



2.1 夸克禁闭与格距的确定

对于由纯规范场构成的规范不变量来讲, 它是由规范场链接构成的闭合圈来给出的

$$L(U) = \text{Tr} \left[\prod_{(x,\mu) \in \mathcal{L}} U_\mu(x) \right] \quad (65)$$

这里 \mathcal{L} 是个闭合圈。Wilson 圈是其中最著名的一种闭合圈, 它的特点是由四部分构成, 在时间片 t_0 和 t_1 处, 有两条空间方向的 Wilson 线 $S(\vec{x}, \vec{y}; t_0)$ 和 $S(\vec{x}, \vec{y}; t_1)$; 另外, 给定空间点 \vec{x} 和 \vec{y} , 在时间方向上有两条线 $T(\vec{x}; t_0, t_1)$ 和 $T(\vec{y}; t_0, t_1)$ 。这里空间线 $S(\vec{x}, \vec{y}; t)$ 定义为

$$S(\vec{x}, \vec{y}; t) = \prod_{(\vec{z}, j) \in \mathcal{C}_{\vec{x}, \vec{y}}} U_j(\vec{z}, t) \quad (66)$$

时间线 $T(\vec{x}; t_0, t_1)$ 定义为

$$T(\vec{x}; t_0, t_1) = \prod_{t=t_0}^{t_1-1} U_4(\vec{x}, t). \quad (67)$$

Wilson 圈的值通过求迹来得到

$$W_{\mathcal{L}}[U] = \text{Tr}[S(\vec{x}, \vec{y}; t_1) T(\vec{y}; t_0, t_1)^\dagger S(\vec{x}, \vec{y}; t_0)^\dagger T(\vec{x}; t_0, t_1)]. \quad (68)$$

最简单的 Wilson 圈是沿着时间方向为 $t = t_1 - t_0$ ，沿着空间方向间隔为 $r = |\vec{y} - \vec{x}|$ 的长方形。为了理解 Wilson 圈的物理含义，我们选取一种时间规范，这种规范固定是将所有时间方向上的规范场链接都设成单位矩阵

$$U_0(x) = 1, \quad \forall x \quad (69)$$

由于 $U_0(x) = \exp(iaA_0(x))$ ，这个规范等价于令 $A_0(x) = 0$ 。于是 Wilson 圈在路径积分下的期望值可以写成

$$\langle W_{\mathcal{L}}[U] \rangle = \langle \text{Tr}[S(\vec{x}, \vec{y}; t_1) S(\vec{x}, \vec{y}; t_0)^\dagger] \rangle \quad (70)$$

我们后面在费米子章节会证明，在无穷重夸克极限下，如果 x 和 y 两点存在一对静态正反夸克的源，那么这对静态正反夸克之间的传播子可以由最短距离连接 x 、 y 点的 Wilson 线给出。如果用数学的形式写出来

$$Q(\vec{x}, t; \vec{y}, t) \equiv \Psi(\vec{x}, t)_{\alpha, a} \bar{\Psi}(\vec{y}, t)_{\beta, b} \propto S(\vec{x}, \vec{y}; t)_{a, b} \quad (71)$$

这里我们忽略了传播子中的 spinor 指标 α, β ，主要考察规范场的作用。于是从 Wilson 圈，我们可以得到

$$\langle W_{\mathcal{L}}(r, t) \rangle \rightarrow \langle \text{Tr}[Q(\vec{x}, t_1; \vec{y}, t_1) Q(\vec{x}, t_0; \vec{y}, t_0)^\dagger] \rangle \rightarrow \langle [\Psi(\vec{x}, t_1) \bar{\Psi}(\vec{y}, t_1)] [\Psi(\vec{x}, t_0)_{\alpha, a} \bar{\Psi}(\vec{y}, t_0)]^\dagger \rangle \quad (72)$$

这实际上告诉我们，Wilson 圈对应于一个格林函数，这个格林函数在 t_0 这个时间片产生一对距离为 r 的正反夸克和在 t_1 这个时间片上湮灭一对正反夸克。如果我们在两个 Wilson 线之间插入能量的本征态，我们得到

$$\langle W_{\mathcal{L}}(r, t) \rangle = \langle [S(\vec{x}, \vec{y}; 0)]_{ab} |n\rangle \langle n| [S(\vec{x}, \vec{y}; 0)^\dagger]_{ba} \rangle e^{-E_n t} \quad (73)$$

这里的基态能量对应的是正反夸克对的能量，可以表述为空间距离为 r 的静态夸克势

$$E_1 = V(r) = \lim_{t \rightarrow \infty} -\frac{1}{t} \log \langle W_{\mathcal{L}}(r, t) \rangle \quad (74)$$

更高能量的激发态里除了我们感兴趣的夸克对之外，还有其它正反夸克的组合，这些组合的量子数与真空一致。事实上，只要时间间隔 $t = t_1 - t_0$ 取得足够大，Wilson 圈的贡献主要来自基态。

要判断夸克禁闭可以有两种方式，一种是给定一对带色荷的正反夸克，看它们之间的相互作用势能是否随着距离增大，如果一直增大的话，意味着正反两个色荷的夸克离得无穷远时，能量无穷大。还有一种判断方式是，看分离出来的单个的色荷的自由能是否是无穷大，如果是的话，意味着我们需要加入无穷大的能量，才能把单个色荷从系统中分离出来。我们先来看第一种情况，也就是看带色荷的正反夸克之间的势能 $V(r)$ 随着距离是如何变化的。事实上，Wilson 在他的那篇开创性的文章里，就证明了当强相互作用耦合常数趋于无穷时， $V(r)$ 随着 r 增加是线性增长的。另一方面，在小耦合常数下，QCD 两个色荷之间的势很类似于 QED 的库伦势，都是 $1/r$ 的形式。所以我们可以把 $V(r)$ 表述为

$$V(r) = A + \frac{B}{r} + \sigma r \quad (75)$$

这种静态夸克势的表达方式，也被称为 Cornell 势，其中 σ 称为弦张量。

我们下面来看一下，强耦合极限下， σ 的计算。用路径积分来表达 Wilson 圈的期望值

$$\langle W_{\mathcal{L}}[U] \rangle = \frac{1}{Z} \int \mathcal{D}[U] \exp \left(-\frac{\beta}{N_c} \sum_P \text{Re Tr}[1 - U_P] \right) \text{Tr} \left[\prod_{l \in \mathcal{L}} U_l \right] \quad (76)$$

其中分子、分母可以消去一个 $\exp(-\beta/N_c \sum_P \text{Re Tr}[1])$ 。我们得到

$$\begin{aligned} \langle W_{\mathcal{L}}[U] \rangle &= \frac{1}{Z'} \int \mathcal{D}[U] \exp\left(\frac{\beta}{N_c} \sum_P \text{Re Tr}[U_P]\right) \text{Tr} \left[\prod_{l \in \mathcal{L}} U_l \right] \\ &= \frac{1}{Z'} \int \mathcal{D}[U] \exp\left(\frac{\beta}{2N_c} \sum_P \left(\text{Tr}[U_P] + \text{Tr}[U_P^\dagger]\right)\right) \text{Tr} \left[\prod_{l \in \mathcal{L}} U_l \right] \end{aligned} \quad (77)$$

强耦合极限 $g^2 \rightarrow \infty$ 对应于 $\beta \rightarrow 0$ 极限, 对于配分函数, 我们有

$$Z' = \int \mathcal{D}[U] \exp\left(\frac{\beta}{2N_c} \sum_P \left(\text{Tr}[U_P] + \text{Tr}[U_P^\dagger]\right)\right) = \int \mathcal{D}[U] (1 + O(\beta)) = 1 + O(\beta) \quad (78)$$

对玻尔兹曼因子 $\exp\left(\frac{\beta}{2N_c} \sum_P \left(\text{Tr}[U_P] + \text{Tr}[U_P^\dagger]\right)\right)$ 这项做 Taylor 展开, 只保留与 $\text{Tr}[U_P^\dagger]$ 有关的项, 我们得到

$$\begin{aligned} &\int \mathcal{D}[U] \frac{1}{n_A!} \left(\frac{1}{2N_c}\right)^{n_A} \left(\sum_P \text{Tr}[U_P^\dagger]\right)^{n_A} \text{Tr} \left[\prod_{l \in \mathcal{L}} U_l \right] \\ &= \left(\frac{1}{2N_c}\right)^{n_A} \int \mathcal{D}[U] \prod_{P \in \mathcal{A}_{\mathcal{L}}} \text{Tr}[U_P^\dagger] \text{Tr} \left[\prod_{l \in \mathcal{L}} U_l \right] \\ &= \left(\frac{1}{2N_c}\right)^{n_A} \left(\frac{1}{N_c}\right)^{n_A} \text{Tr}[1] \\ &= 3 \exp\left(n_A \log \frac{\beta}{2N_c^2}\right) \end{aligned} \quad (79)$$

这里 n_A 对应的是铺满整个 Wilson 圈所需的小方格子的个数: $n_A = \frac{r^4}{a^2}$ 。所以, 我们得到

$$V(r) \sim -\frac{r}{a^2} \log \frac{\beta}{2N_c^2} (1 + O(\beta)) \quad \Rightarrow \quad \sigma = -\frac{1}{a^2} \log \frac{\beta}{2N_c^2} (1 + O(\beta)) \quad (80)$$

我们发现在强耦合展开中, 我们实际得到了所谓的面积法则, 也就是说 Wilson 圈的期望值随着面积增大而指数衰减。这种行为即意味着禁闭的出现。我们看到, 面积法则这件事情实际并不依赖于具体的规范群, 也就是说, 无论 $SU(3)$ 的 QCD 还是 $U(1)$ 的 QED, 在强耦合下都会出现面积法则, 也就意味着禁闭的发生。尽管前者具有很强的胶子自相互作用, 而后者是阿贝尔规范理论, 没有自相互作用项。我们知道 QED 是不禁闭的, 那么我们如何从场论的角度理解 QED 中出现面积法则这件事情呢? 事实上, 即使像 QED 这样的规范理论, 在强耦合极限下, 它也是禁闭的。但是真正的物理理论对应的耦合常数很小, 这个时候系统会解禁闭, 解禁闭的相我们称之为库伦相。如果我们不考虑真空极化和粒子的产生、湮灭, 那么这个时候一对正反电子之间的势能也就是我们所熟悉的库伦势。当耦合常数很大时, 即进入禁闭相。随着耦合常数的变化, 规范理论存在着从库伦相到禁闭相的转变。前者可以有弱耦合极限下研究, 后者在强耦合极限下研究, 因此两个极端区域的行为都可以解析地得到。但在中间区域, 我们需要用计算机非微扰地处理。早期的格点研究表明, 库伦相到禁闭相的转变是一种 crossover, 它在一个以耦合常数为参量的区域内光滑地从一个相过渡到另一个相的, 而不是在某个临界耦合常数值发生一级相变。

这里, 我们需要注意的是, 我们的讨论是基于静态夸克势。真正的 full QCD 理论, 我们需要考虑夸克的动力学效应, 包括正反夸克对的产生和湮灭。事实上, 当一对正反夸克被拉得足够远, 能量大到足以产生一对正反夸克, 就会与初始的一对夸克重新组合, 形成两个介子。我们把这种现象称之为 string breaking。

从静态夸克势中，我们可以提取一个和距离有关的量，叫做 Sommer parameter，它定义为

$$r^2 \frac{\partial V(r)}{\partial r} \Big|_{r=r_0} = 1.65 \quad (81)$$

这个量可以认为是个物理量。因为 Cornell 势可以很成功地用于预言 charmonium 和 Upsilon 系统的能级，通过和实验数据做对比，可以得到

$$r_0 = 0.49 \text{ fm} \quad (82)$$

因此 r_0 这个量可以帮助我们来确定格距 a 的大小。我们用格点上计算出来的 r_0/a 的值；通过和 r_0 的物理值比较，马上可以定出格距大小。有了格距大小，我们可以给其它所有的物理量来定标。采用 r_0 来定标有几个好处，第一是它其实对于势模型的选取很不敏感，所以模型依赖性可忽略；第二， r_0 这个距离取得比较适中，它保证了这个时候静态夸克势可以很好地由 $c\bar{c}$ 或者 $b\bar{b}$ 的谱来确定下来；第三，格点 QCD 能够给出很精确的 r_0 的值；第四，它对海夸克的依赖很不敏感，不需要我们先把 u/d 夸克参数调到物理值，或者说， r_0 这个量对于 full QCD 和淬火 QCD 都适用。

当然，确定格距的方法不只一种，比方说可以用某个强子 H 的质量。我们有

$$a = \frac{(am_H)^{\text{lat}}}{m_H^{\text{exp}}} \quad (83)$$

这里常用的物理量有 Ω^- 重子的质量以及 pion 介子衰变常数等。两个最基本的要求是：一、该物理量在格点上的计算结果很精确（当然实验值也要足够精确）；二、该物理量对于轻夸克的依赖关系比较小。因为夸克质量也是 QCD 的参数，我们不希望同时调两个参数来和实验进行比对。最后，无论采取哪种方法，我们得到的格距的值应该是自洽的，这是对于格点 QCD simulation 结果的一个 highly non-trivial 的检验。

除了 Wilson 圈之外，还有一种很有用的纯规范场物理量叫做 Polyakov 圈。它是对规范场链接在时间方向上取周期性边界条件，然后构造时间线，时间线的长度和整个格点的时间长度一样大。由于周期性边界条件，那么这条时间线在时间方向上实际是走了个圈，我们把这个圈称为 Polyakov 圈。它可以写成

$$P(\vec{x}) = \text{Tr} \left[\prod_{j=0}^{T-1} U_4(\vec{x}, j) \right] \quad (84)$$

很显然，这样一个首尾相接的 Polyakov 圈是规范不变的。我们可以计算两个距离为 $r = |\vec{x} - \vec{y}|$ 的 Polyakov 圈相乘的关联函数

$$\langle P(\vec{x})P(\vec{y})^\dagger \rangle \quad (85)$$

这实际上等价于一个时间长度为 $t = T$ ，空间长度为 r 的 Wilson 圈，只是这个 Wilson 圈中两个空间线都取成了 $SU(3)$ 的单位群元，因此所有的贡献都来自于两条走向相反的 Polyakov 圈。从这个关联函数中，我们同样可以印证面积法则以及计算静态夸克势 $\langle P(\vec{x})P(\vec{y})^\dagger \rangle \propto e^{-TV(r)}$ 。引入 Polyakov 圈的另外一个用途是，单个的 Polyakov 圈的真空期望值实际上对应了单个孤立的带色荷的夸克的自由能

$$\langle P(\vec{x}) \rangle \sim e^{-F_q T} \quad (86)$$

如果我们能提取出单个孤立的夸克来， F_q 就只能是有有限值，也就是说 Polyakov 圈的真空期望值不为 0。因此我们可以把 $\langle P \rangle$ 当成是一个序参量

$$\begin{aligned} \langle P \rangle \neq 0 &: \text{退禁闭相 (库伦相) } (F_q \text{ 是有限值}) \\ \langle P \rangle = 0 &: \text{禁闭相 } (F_q \text{ 无穷大}) \end{aligned} \quad (87)$$

还有一个比较有意思的物理量是胶球谱。比方说我们可以考虑 0^{++} 的胶球。它对应的最简单的关联函数可以由两个小方格给出

$$\langle P_{\mu\nu}(t)P_{\mu\nu}^\dagger(0) \rangle_{\text{conn}} = \langle P_{\mu\nu}(t)P_{\mu\nu}^\dagger(0) \rangle - \langle P_{\mu\nu}(t) \rangle \langle P_{\mu\nu}^\dagger(0) \rangle \sim e^{-M_G t} \quad (88)$$

前面的 P 指的 Polyakov 圈，这里的 P 指的是 1×1 小方格 (plaquette)。在强耦合极限极限下

$$\langle P_{\mu\nu}(t)P_{\mu\nu}^\dagger(0) \rangle_{\text{conn}} \sim e^{-M_G t} = \left(\frac{\beta}{2N_c^2} \right)^{4t} (1 + O(\beta)) \Rightarrow M_G = -4 \ln \frac{\beta}{2N_c^2} + \dots \quad (89)$$

$4t$ 这个因子是由 4 个面上的面积为 $1 \times t$ 的流管来给出。

【作业：思考一下如果把 1×1 的小方格替换为一个 $R \times R$ 的 Wilson 圈，在强耦合极限下，我们应该如何来构造流管。请证明 M_G 的表达式与 R 的大小无关。】

2.2 重整化群与连续极限

在格点 QCD 中，我们很自然地引进了紫外截断，由非零格距来给出 $\Lambda \sim \frac{\pi}{a}$ 。截断这件事情，其实是很自然的，因为在 $\Lambda \rightarrow \infty$ 的区域内，我们并不知道那个地方的物理，实验也达不到那么高的能量。换个观点，我们可以认为，很高能量区域 (从 ∞ 到 π/a) 的物理都被积掉，留给我们的是一个有效理论，这个有效理论就是我们能获知的关于相互作用的所有信息。当然这个有效理论的裸参量是依赖于截断的，因为所有高能区的信息都在裸参量里面。而且裸参量中的截断依赖抵消掉了有效理论中的截断依赖，因此最终格点 QCD 计算得到的物理量是与截断无关的。我们知道，这些裸参量随着截断的变化是由重整化群来描述。为了简化讨论，我们考虑一个纯规范理论，它只依赖一个裸参量，就是耦合常数 $g(a)$ ，或者说是 β ，这是我们做格点 simulation 需要放入的一个参数。令 $O(g(a), a)$ 是一个物理量，在 $a \rightarrow 0$ 时，能给我们物理的值

$$\lim_{a \rightarrow 0} O(g(a), a) = O_{\text{phys}} \quad (90)$$

在非零格距 $a \neq 0$ 下，我们可以写成

$$O(g(a), a) = O_{\text{phys}} + \sum_n a^n O_n \quad (91)$$

这些 $a^n O_n$ 的项用重整化群的语言来讲，是由 irrelevant operators 引起的，当 $a \rightarrow 0$ 时会消失。格点里面一般把它们叫做 lattice artifacts。我们先忽略这些 lattice artifacts 的项，那么截断所引起的非物理效应，就完全由 $g(a)$ 来抵消掉了，因此，我们可以把 $O(g(a), a)$ 看成是与截断 a 无关的量，于是得到 Callan-Symanzik 重整化群方程

$$\frac{dO(g, a)}{d \ln a} = 0 \quad \text{或者} \quad \left(\frac{\partial}{\partial \ln a} + \frac{\partial g}{\partial \ln a} \frac{\partial}{\partial g} \right) O(g, a) = 0 \quad (92)$$

其中第二项的系数被称为 β 函数

$$\beta(g) \equiv -\frac{\partial g}{\partial \ln a} \quad (93)$$

它描述了耦合常数是如何随着截断 a 变化的。渐近自由告诉我们，当 $a \rightarrow 0$ 时， $g \rightarrow 0$ 。通过微扰论，我们可以得到

$$\beta(g) = -\beta_0 g^3 - \beta_1 g^5 + \dots \quad (94)$$

这里

$$\begin{aligned}\beta_0 &= \left(\frac{11N_c - 2N_f}{3}\right)/16\pi^2 \\ \beta_1 &= \left(\frac{34N_c^2}{3} - \frac{10N_cN_f}{3} - \frac{N_f(N_c^2 - 1)}{N_c}\right)/(16\pi^2)^2\end{aligned}\quad (95)$$

β_0 和 β_1 这两个参数是和正规化方案无关的。如果我们仅考虑 β_0 的贡献，那么在截断从 a_1 变到 a_2 的过程中，耦合常数的变化为

$$\frac{1}{2\beta_0g^2(a_2)} - \frac{1}{2\beta_0g^2(a_1)} = \ln \frac{a_2}{a_1} \quad \text{或者} \quad \frac{1}{2\beta_0g^2(a)} - \log a = \text{constant} = \log \Lambda \quad (96)$$

这里我们不妨引入一个 scale，叫 Λ ，那么

$$a = \frac{1}{\Lambda} e^{-\frac{1}{2\beta_0g^2}} \quad (97)$$

如果考虑 β_1 的贡献，那么

$$a(g) = \frac{1}{\Lambda} e^{-\frac{1}{2\beta_0g^2}} (\beta_0g^2)^{-\frac{\beta_1}{2\beta_0^2}} (1 + O(g^2)) \quad (98)$$

我们通过输入裸参数 g^2 来做格点 QCD simulation，从而计算出 Wilson 圈和静态夸克势，并进一步确定格距 a 。有了 a 和 g^2 的对应关系，我们就能确定出 Λ

$$\Lambda = \frac{1}{a} e^{-\frac{1}{2\beta_0g^2}} (\beta_0g^2)^{-\frac{\beta_1}{2\beta_0^2}} (1 + O(g^2)) \quad (99)$$

也就是说，我们从一个完全没有量纲的系统出发，得到了一个有量纲的量 Λ 。究其原因，是因为由于强相互作用的存在，会自然地产生一个 scale。

公式 (98) 告诉我们，格距 a 随着耦合常数 g 减小 (β 增加) 而减小，那么在 β 趋于 ∞ 时，我们自然而然地得到了连续极限的结果。但实际我们无法在 β 趋于 ∞ 的极限下去做 simulation，因为这个时候如果格点数目给定，比方说 N^4 ，那么对应的物理体积是 $(Na)^4$ 会趋于 0。要保持物理体积不变，那么格点数就会反比于 a 增加。我们一般的做法是用几个不同的 β 值做参数，来定出不同的格距值，然后在每个格距上计算我们感兴趣的物理量。再对这个物理量做连续极限外推。

无论是格点 QCD，还是微扰 QCD，我们最终的结果都是要给出物理的预言值。区别在于，格点 QCD 调节的是裸参数随着截断 $1/a$ 的变化。格点 QCD 做 simulation 的时候是没法直接让截断 $1/a$ 趋于无穷的。因此非零格距 a 作为系统误差存在，要得到物理预言，我们需要做连续极限外推。好处是这种系统误差是可以被控制的。而对于微扰 QCD 来讲，它利用的是重整化以后的拉氏量，重整化以后的参数不依赖于截断，而是随着重整化能标变化。加入高圈修正以后，计算得到的物理量对重整化能标的依赖会越来越小，最终得到非常可靠物理的预言值。但这件事情，在非微扰区域就变得很困难，因为高圈修正的收敛性会很差。

2.3 耦合常数跑动

反解公式 (98)，我们得到裸耦合常数作为格距 a 的函数

$$g(a)^{-2} = \beta_0 \ln(a^{-2}\Lambda^{-2}) + \frac{\beta_1}{\beta_0} \ln(\ln(a^{-2}\Lambda^{-2})) + O(1/\ln(a^2\Lambda^2)). \quad (100)$$

但这个耦合常数不是很有用，物理上大家更感兴趣的应该是重整化以后的耦合常数随着能标 μ 是如何跑动的。我们希望将格点得到的 coupling constant running 和微扰计算结合起来：由格点计算得到 $\alpha_s(\mu)$ 在低能区的值，然后通过重整化群演化，得到 α_s 在能标 $\mu = M_Z$ 的值。 α_s 的精确确定，对于标准模型的精确预言意义重大。如果不用格点 QCD 的数据做输入，那么通过实验得到的 global average 是

$$\alpha_{\overline{\text{MS}}}^{(5)}(M_Z) = 0.1175(17) \quad \text{PDG} \quad (101)$$

这里主要采用了四种实验数据： $\tau \rightarrow \nu_\tau + \text{hadrons}$, $e^+e^- \rightarrow \text{hadrons}$, 深度非弹以及高精度电弱衰变数据。格点计算得到的结果是

$$\alpha_{\overline{\text{MS}}}^{(5)}(M_Z) = 0.1183(12) \quad \text{FLAG} \quad (102)$$

格点 QCD 定 α_s 的方法有好些种。格点 QCD 有一个组织，叫做 Flavour lattice averaging group, 他们在做一件类似 PDG 的事情。就是统计各个格点组计算得到的结果，分析计算中可能的各种系统误差，然后再来做 global average。在它们 2017 年的 review 里面列了 6 种不同的计算 α_s 的方法。我这里举其中一种方法为例，让大家有个直观的印象。我们知道无论是 tau 衰变、还是 e^+e^- 湮灭的过程，其强子部分都是由夸克矢量流或者轴矢流构成的真空极化函数来给出。在格点上，我们不妨去计算这样一个真空极化函数

$$\langle J_\mu^a J_\nu^b \rangle(Q) = \delta^{ab} \left[(\delta_{\mu\nu} Q^2 - Q_\mu Q_\nu) \Pi_J^{(1)}(Q) - Q_\mu Q_\nu \Pi_J^{(0)}(Q) \right] \quad (103)$$

ab 是味道指标， J 可以是矢量流或者轴矢流， $\Pi_J^{(1)}(Q)$ 代表真空极化函数的横向分量， $\Pi_J^{(0)}(Q)$ 代表纵向分量。我们可以定义

$$\Pi_J(Q) = \Pi_J^{(1)}(Q) + \Pi_J^{(0)}(Q), \quad \Pi_{V+A}(Q) = \Pi_V(Q) + \Pi_A(Q) \quad (104)$$

当 Q 比较大的时候，我们可以对真空极化函数 $\Pi_{V+A}(Q)$ 做 OPE 展开。

$$\begin{aligned} \Pi_{V+A}(Q) &= c + C_0(Q^2, \mu^2, \alpha_s) \\ &+ C_m^{V+A}(Q^2, \mu^2, \alpha_s) \frac{\bar{m}^2(Q)}{Q^2} \\ &+ \sum_{q=u,d,s} C_{\bar{q}q}^{V+A}(Q^2, \alpha_s) \frac{\langle m_q \bar{q}q \rangle}{Q^4} \\ &+ C_{GG}(Q^2, \alpha_s) \frac{\langle (\alpha_s/\pi) GG \rangle}{Q^4} \\ &+ \dots \end{aligned} \quad (105)$$

其中 c 这一项是个 contact term, 与非零格距导有关，可以很容易通过减除去掉非物理的这一项。这些展开的 Wilson 系数 C_X , ($X = 0, \bar{q}q, GG$), 微扰可算。 $\bar{m}^2(Q)$ 是 running quark mass, 和格点输入的参数有关。夸克凝聚 $\langle m_q \bar{q}q \rangle$ 这一项格点可以计算，在给定 scale 的情况下，我们也可以把它当作是一个待定参数；胶子凝聚这一项 $\langle (\alpha_s/\pi) GG \rangle$ 也可以当成是一个待定参数。我们可以选取重整化 scale $\mu = 2 \text{ GeV}$, 这样 α_s 也作为一个未知的参数，研究 $\Pi_{V+A}(Q)$ 随着 Q 的变化行为，我们就可以定出 α_s 在 $\mu = 2 \text{ GeV}$ 时的值。

尽管格点上计算 α_s 的方法有好些种，但它们背后的原理是类似的。通过这个具体的例子我们可以看到：首先我们要选定一个 short-distance 物理量 $\langle O \rangle$, 这个物理量要求既能够被微扰地计算，又能够通过格点进行非微扰计算，然后把格点数据和微扰计算相结合，我们就能定出 α_s ；这和把实验数据和微扰计算相结合来定 α_s 是类似的。

2.4 Symanzik 作用量改进

我们把格点 QCD 看成是一个带有截断 π/a 的有效理论。之前提到过, 对于一个物理量 $O(g(a), a) \equiv O_L(\pi/a)$, 在 $a \rightarrow 0$ 时, 能给我们物理的值

$$\lim_{a \rightarrow 0} O(g(a), a) = O_{\text{phys}} = O_C \quad (106)$$

在非零格距 $a \neq 0$ 下, 我们可以写成

$$O_C = O_L(\pi/a) + \sum_n \sum_i a^n D_i^{(n)} O_i^{(n)}(\pi/a) \quad (107)$$

由于在 $a \rightarrow 0$ 的极限下, $O_i^{(n)}(\pi/a)$ 这些项会趋于 0, 在我们之前的讨论里面把这些项忽略掉了。但如果我们考虑如何让格点计算出来的结果更快地趋于物理的连续极限, 那么就需要考虑这些 lattice artifacts 项的贡献。

同样的, 对于作用量, 我们也可以在连续极限下的作用量 ($a \rightarrow 0$) 和格点作用量之间就可以构建一个关系式

$$S_C = S_L(\pi/a) + \sum_n \sum_i a^n C_i^{(n)} A_i^{(n)}(\pi/a) \quad (108)$$

其中, $A_i^{(n)}(\pi/a)$ 是 higher dimensional 算符。这里, 系数 C_i 和 D_i 都是耦合常数 α_s 的函数。

下面我们来考察一个给定算符的期望值

$$\begin{aligned} \langle O_C \rangle &= \int \mathcal{D}U O_C e^{-S_C} \\ &= \int \mathcal{D}U \left\{ O_L + \sum_n \sum_i a^n D_i^{(n)} O_i^{(n)} \right\} e^{-S_L - \sum_m \sum_i a^m C_i^{(m)} A_i^{(m)}} \\ &= \int \mathcal{D}U \left\{ O_L + \sum_n \sum_i a^n D_i^{(n)} O_i^{(n)} \right\} \left(1 - \sum_m \sum_i a^m C_i^{(m)} A_i^{(m)} + \dots \right) e^{-S_L} \\ &= \int \mathcal{D}U O_L e^{-S_L} + \text{correction terms} \end{aligned} \quad (109)$$

很明显, 格点 QCD 由于是在非零格距下模拟出来的, 它所有的计算结果都带有非零格距导致的系统误差。但需要指出的是, 这些系统误差到底有多大, 是可以通过连续极限外推来进行判断和计算的, 因此这个误差是可控的。代价是我们需要在多个不同的格距上进行 simulation, 而小格距往往意味着大的计算资源消耗。如果我们能在比较大的格距上也能保证 lattice artifacts 比较小, 岂不是更好? 这就是作用量改进的意义。

在讲作用量改进之前, 我们先来看一个简单的 toy model。假设我们本来是要实现 $f(x)$ 求一阶导数这件事情。在格点上我们用差分来代替一阶导数, 那么就是

$$\frac{f(x+a) - f(x-a)}{2a} = f'(x) + a^2 C^{(2)}(x) + a^4 C^{(4)}(x) + \dots \quad (110)$$

根据 Taylor 展开, 我们很容易知道 $C^{(2)}(x) = \frac{1}{6} f'''(x)$, 这一项由 $f(x)$ 的高阶导数给出, 实际上对应于一个 high dimensional 算符。

我们要做改进, 把 $a^2 C^{(2)}(x)$ 这项去掉, 该怎么做? 我们还是只能在分立的格点场上进行构造, 利用 $f(x), f(x \pm a), f(x \pm 2a), \dots$ 进行组合来得到 $f'''(x)$ 项

$$D^{(3)}[f](x) = \frac{f(x+2a) - 2f(x+a) + 2f(x-a) - f(x-2a)}{2a^3} = f'''(x) + O(a^2) \quad (111)$$

于是

$$\frac{f(x+a) - f(x-a)}{2a} - \frac{1}{6}a^2 D^{(3)}[f](x) = f'(x) + O(a^4) \quad (112)$$

就是我们要得到的改进表达式。我们可以看到，在构造 $D^{(3)}[f](x)$ ，我们用到了 $f(x \pm 2a)$ ，原则上我们也可以使用 $f(x \pm 3a)$ 来代替，因此构造改进作用量的方式并不唯一。

从这个 Toy model 出发我们来总结一下规律：首先，我们要知道对哪一项做修正，这个修正项通常是由连续极限下的表达式来给出，我们这里是 $f'''(x)$ ；其次，需要修正的表达式最好有一定的对称性，在我们的情况里是 $\pm na$ 成对出现，它的好处是修正项成 a^2 压低，换句话说，对称性的存在使得 high dimensional 算符的数目减少。再次，我们需要用格点的表达式来构造出我们想要的修正项。最后这种构造并不是唯一的。这样一种改进方案通常被称为 Symanzik improvement program。

我们这个 toy model 里面修正项的系数就等于 $1/6$ ，比较简单。当我们处理 QCD 时，这个系数相当于是一个裸参量，重整化的效应会影响系数的大小，我们通常需要做格点微扰论计算或者非微扰的 matching，才能把这个系数确定下来。

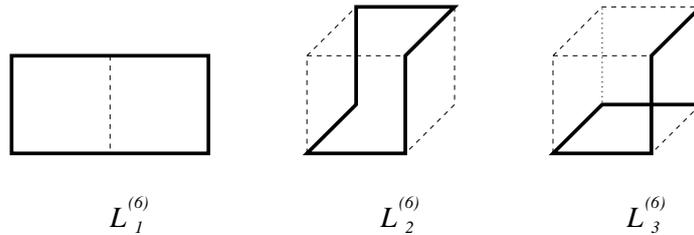
下面我们来看规范场。对于纯规范场来讲，所有 Wilson 圈的领头阶展开都是 $\sum_{\mu\nu} \text{Tr}(F_{\mu\nu}F_{\mu\nu})$ 的形式，这是一个 dimension 4 的算符。因为对于纯规范场算符，我们没有 dimension 5 的算符，因此可能的修正来源于 dimension 6 的算符，它们贡献的是 $O(a^2)$ 的修正。把所有 dimension 4 和 dimension 6 的算符都找出来，我们有

$$\begin{aligned} \mathcal{O}^{(4)} &= \sum_{\mu\nu} \text{Tr}(F_{\mu\nu}F_{\mu\nu}) \\ \mathcal{O}_1^{(6)} &= \sum_{\mu\nu} \text{Tr}(D_\mu F_{\mu\nu} D_\mu F_{\mu\nu}) \\ \mathcal{O}_2^{(6)} &= \sum_{\mu\nu\rho} \text{Tr}(D_\mu F_{\nu\rho} D_\mu F_{\nu\rho}) \\ \mathcal{O}_3^{(6)} &= \sum_{\mu\nu\rho} \text{Tr}(D_\mu F_{\mu\rho} D_\nu F_{\nu\rho}) \end{aligned} \quad (113)$$

给定一个 Wilson 圈 \mathcal{L} ，展开以后可以写成

$$\mathcal{L} = r^{(4)} \mathcal{O}^{(4)} + a^2 r_1^{(6)} \mathcal{O}_1^{(6)} + a^2 r_2^{(6)} \mathcal{O}_2^{(6)} + a^2 r_3^{(6)} \mathcal{O}_3^{(6)} + O(a^4) \quad (114)$$

我们下面来找最简单的 Wilson 圈，除了 4 个链接的小方块，就是 6 个链接的 Wilson 圈，有下面 3 种结构



这 3 种 Wilson 圈，正好可以帮助我们去掉 3 个 dimension 6 算符的贡献。把 Wilson 圈对格距做 Taylor 展开以后，得到的系数为

$Loop$	$r^{(4)}$	$r_1^{(6)}$	$r_2^{(6)}$	$r_3^{(6)}$
$\mathcal{L}^{(4)}$	$-\frac{1}{4}$	$\frac{1}{24}$	0	0
$\mathcal{L}_1^{(6)}$	-2	$\frac{5}{6}$	0	0
$\mathcal{L}_2^{(6)}$	-2	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$
$\mathcal{L}_3^{(6)}$	-4	$\frac{1}{6}$	0	$\frac{1}{2}$

我们把规范场作用量写成一个带参数的普遍的形式

$$S_g = \frac{2N_c}{g^2} \left(c^{(4)}(g^2)\mathcal{L}^{(4)} + \sum_{i=1,2,3} c_i^{(6)}(g^2)\mathcal{L}_i^{(6)} \right) \quad (115)$$

归一化条件要求

$$-\frac{1}{4}c^{(4)}(g^2) - 2c_1^{(6)}(g^2) - 2c_2^{(6)}(g^2) - 4c_3^{(6)}(g^2) = -\frac{1}{4} \quad (116)$$

从表格里面我们可以看到，在树图水平，不考虑高阶微扰论的修正的话，那么要消除 a^2 项，我们只需取

$$\frac{1}{24}c^{(4)}(0) + \frac{5}{6}c_1^{(6)}(0) = 0, \quad c_2^{(6)}(0) = 0, \quad c_3^{(6)}(0) = 0. \quad (117)$$

也就是说

$$c^{(4)}(0) = \frac{5}{3}, \quad c_1^{(6)}(0) = -\frac{1}{12}, \quad c_2^{(6)}(0) = 0, \quad c_3^{(6)}(0) = 0. \quad (118)$$

这被称为 tree-level Lüscher-Weisz action。如果考虑 1 圈图修正，那么可以把系数写为 $c_i(g^2) = c_i(0) + g^2 d_i + \dots$ ，Lüscher 和 Weisz 计算了 Polyakov 圈的关联函数，在要求 $O(a^2)$ 改进的情况下定出了系数

$$\begin{aligned} c^{(4)}(g^2) &= \frac{5}{3} + 0.2370g^2 \\ c_1^{(6)}(g^2) &= -\frac{1}{12} - 0.02521g^2 \\ c_2^{(6)}(g^2) &= -0.00441g^2 \\ c_3^{(6)}(g^2) &= 0 \end{aligned} \quad (119)$$

这个结果也可以看成是在考虑重整化的情况下，原则上所有 dimension 6 的算符都会混合进来。但是，这里有一个例外， $c_3^{(6)}$ 取成 0 是个约定俗成，主要是为了简化作用量。如果我们回过去看算符 $\mathcal{O}_3^{(6)}$ ，它实际上可以由规范场的重新定义 $A_\mu \rightarrow A_\mu + a^2 g^2 d(g) D_\nu F_{\nu\mu}$ 来产生，因此 4 个 c_i 参数，其中一个多余的，我们可以直接把 $c_3^{(6)}$ 设成 0。

需要指出的是，格点作用量的改进，往往意味着作用量会变得更加复杂，这会增加计算的复杂性以及消耗更多的 simulation 时间。所以减小格距误差往往是在作用量改进和连续极限外推之间找一个平衡点。

3 格点上的夸克场

我们来回顾一下之前讲过的 naive 费米子

$$\begin{aligned} S_F[\psi, \bar{\psi}] &= a^4 \sum_n \bar{\psi}(n) \left(\sum_{\mu=1}^4 \gamma_\mu \frac{U_\mu(n)\psi(n+\hat{\mu}) - U_{-\mu}(n)\psi(n-\hat{\mu})}{2a} + m_0\psi(n) \right) \\ &= a^4 \sum_{n,m} \bar{\psi}(n) D(n,m) \psi(m) \end{aligned} \quad (120)$$

其中 Dirac 矩阵可以写为

$$D(n,m) = \sum_{\mu=1}^4 \gamma_\mu \frac{U_\mu(n)\delta_{n+\hat{\mu},m} - U_{-\mu}(n)\delta_{n-\hat{\mu},m}}{2a} + m_0\delta_{n,m} \quad (121)$$

为了简单起见，我们可以把规范场链接设为 1，这样就得到了自由费米子的 Dirac 矩阵。我们对自由 Dirac 矩阵做傅立叶变换

$$\begin{aligned} D(p,q) &= \frac{1}{N_L^3 N_T} \sum_{n,m} e^{-ip \cdot na} e^{iq \cdot ma} D(n,m) \\ &= \frac{1}{N_L^3 N_T} \sum_n e^{-i(p-q) \cdot na} \left(\sum_{\mu=1}^4 \gamma_\mu \frac{e^{iq_\mu a} - e^{-iq_\mu a}}{2a} + m_0\delta_{n,m} \right) \\ &= \delta_{p,q} \tilde{D}(p) \end{aligned} \quad (122)$$

其中，动量空间的 Dirac 矩阵形式要比位置空间的形式简单不少。

$$\tilde{D}(p) = \frac{i}{a} \sum_{\mu=1}^4 \gamma_\mu \sin(p_\mu a) + m_0 \quad (123)$$

我们不妨取质量为 $m_0 = 0$ ，我们可以看到传播子 $\tilde{D}^{-1}(p)$ 除了有一个物理的 $p = (0, 0, 0, 0)$ 的极点，还有 15 个非物理的极点，我们管它们叫加倍子

$$p = (\pi/a, 0, 0, 0), \quad (0, \pi/a, 0, 0), \quad \dots \quad (\pi/a, \pi/a, \pi/a, \pi/a) \quad (124)$$

如果我们采用 naive 费米子，那么对于任何物理过程的计算，我们都没办法去掉这 15 个加倍子导致的非物理的效应。

我们怎么来理解加倍子这件事情呢？首先我们可以看到，在 massless 极限下 $m_0 = 0$ ，naive 费米子作用量在手征变换下

$$\psi(n) \rightarrow e^{i\alpha\gamma_5}\psi(n), \quad \bar{\psi}(n) \rightarrow \bar{\psi}(n)e^{i\alpha\gamma_5} \quad (125)$$

是不变的。这一点可以通过 γ_5 矩阵与 Dirac 矩阵反对易关系来判断 $\gamma_5 D + D\gamma_5 = 0$ 。所以 naive 费米子保证了作用量的手征对称性，也保证了 axial Ward identity 的存在。但不同于连续时空场论，格点中是存在着 lattice spacing 硬截断的。由于硬截断的存在，就违背了 ABJ 手征反常成立的条件。所以 naive 费米子有手征性，却没有手征反常。加倍子的出现，就是为了抵消手征反常。事实上，Nielsen-Ninomiya 证明过一个 No-go 定理，它说的是，一个定义在 4 维格点上的费米子场作用量无法同时满足以下 4 个条件 1) 平移不变性 2) 相互作用的局域性 3) 手征对称性 4) 无加倍子。

为了去除加倍子，Wilson 选择放弃手征对称性。后来，人们经过二十多年的努力，终于在格点上把手征对称性放了回去。其中一种方案叫 Domain wall fermion 【D. B. Kaplan, PLB288 (1992) 342】。它的想法是引进一个额外的第 5 维，左手和右手费米子定义在第 5 维的两个 4 维超平面上。在第 5 维长度足够长的时候，可以认为左右手费米子不发生耦合。因为引进了第 5 维，所以不必遵守 No-go 定理，就可以保证 4 个条件同时满足。当然在第 5 维无法去到无穷大的时候，还是会有微小的手征对称性破缺。还有一种方案更绝，叫做 overlap fermion 【H. Neuberger, PLB417 (1998) 141】。它还是在一个 4 维的时空里面进行作用量的构造，所以它不满足连续时空下的手征对称性。但如果我们把手征变换的形式做了微小的变动，定义一个格点上的手征变换（这种变换在连续极限下会回到我们所熟悉的手征变换），那么在格点手征变换下，overlap fermion 是具有严格手征性的，并且，它也不破坏 ABJ 手征反常定理。保护手征对称性的好处是，在有些物理过程中，手征对称性起着重要的作用；如果手征性被破坏，在重整化的时候会有很复杂的算符混合。这个时候，具有手征性的费米子就显示出它们的优势。需要指出的是，Domain wall 和 overlap fermion 的格点 simulation 和计算要比 Wilson 费米子更为复杂，耗得计算时间也更多，所以采用什么的费米子作用量，通常是和我们要研究的物理目的相关。除了 Wilson, Domain wall, overlap 费米子，其实还有其它的费米子作用量，在这个课上我们不可能一一涉及，所以我主要讲两种，一种是最简单的 Wilson 费米子，还有一种是具有严格手征性的 overlap 费米子。

3.1 Wilson 费米子

Wilson 解决加倍子的方案是，在原有的 naive 费米子作用量之上，加入新的一项，叫做 Wilson 项

$$\tilde{D}_W(p) = \frac{i}{a} \sum_{\mu=1}^4 \gamma_{\mu} \sin(p_{\mu}a) + m_0 + \frac{r}{a} \sum_{\mu=1}^4 (1 - \cos p_{\mu}a) \quad (126)$$

很明显，除了物理的极点对应于 $\tilde{D}_W(p)$ 的零点，其它的加倍子所对应的 4 动量都不再是 $\tilde{D}_W(p)$ 的零点。因此加倍子问题被解决了。但是这个 Wilson 项和 explicit 的质量项一样，都是破坏 γ_5 反对易关系的。因此，作用量的手征对称性被破坏了。如果在位置空间写出来，并且加入规范场，Wilson 项的形式为

$$ar \sum_{\mu=1}^4 \frac{2\delta_{n,m} - U_{\mu}(n)\delta_{n+\hat{\mu},m} - U_{-\mu}(n)\delta_{n-\hat{\mu},m}}{2a^2} \quad (127)$$

另外，如果我们对 a 做 Taylor 展开的话，会发现，naive 费米子的 lattice artifacts 是在 $O(a^2)$ 平方阶的，而 Wilson 费米子的 lattice artifacts 是在 $O(a)$ 阶。所以，实际应用的过程当中，我们还要做作用量改进，加入修正项，得到所谓的 Clover Wilson 费米子。一般在格点上，如果不加特殊说明，Wilson 费米子指的就是 Clover Wilson 费米子。

我们现在写出 Wilson 费米子的 Dirac 矩阵

$$D(n, m) = \left(-\frac{1}{2a} \sum_{\mu=\pm 1}^{\pm 4} (r - \gamma_{\mu}) U_{\mu}(n) \delta_{n+\hat{\mu}, m} \right) + \left(m_0 + \frac{4r}{a} \right) \delta_{n, m} \quad (128)$$

这里，我们定义 $\gamma_{-\mu} = -\gamma_{\mu}$ 。

下面我们来看一下，静态夸克极限下，传播子的形式。我们引进参量 $\kappa = \frac{1}{2m_0a+8r}$ ，把 Dirac 矩阵

改写为

$$D(n, m) = C(\delta_{n, m} - \kappa H(n, m)), \quad C = m_0 + \frac{4r}{a}, \quad H(n, m) = \sum_{\mu=\pm 1}^{\pm 4} (r - \gamma_\mu) U_\mu(n) \delta_{n+\hat{\mu}, m} \quad (129)$$

然后求传播子

$$[D^{-1}](n, m) \propto [(1 - \kappa H)^{-1}](n, m) = \sum_{j=0}^{\infty} [\kappa^j H^j]_{n, m} \quad (130)$$

我们先忽略 H 中的 γ 矩阵, 单纯研究规范场。那么 H^j 项实际上对应着 j 个规范场链接相乘所构成的 Wilson 线。只有那些从 m 出发, 走到 n 的 Wilson 线会对夸克传播子有贡献。当然这种走法是多种多样的。但是, 如果我们取重夸克极限, 那么只有最短的那条 Wilson 线的贡献是主要的, 更长长度的 Wilson 线都被 κ 高阶压低了。我们之前用长方形的 Wilson 圈来计算得到面积法则, 那里用的 Wilson 线就是直线, 刚好对应的是一对静态无穷重正反夸克的传播子。因此, 我们可以用长方形的 Wilson 圈来研究静态夸克势。

格点上定义的费米子作用有一个重要的性质叫做 γ_5 -厄米性, 也即 $D^\dagger = \gamma_5 D \gamma_5$ 。

【作业：证明 Wilson 费米子 Dirac 矩阵的 γ_5 -厄米性。】

从 γ_5 -厄米性出发, 我们马上能得到 Dirac 矩阵的本征值要么是实数, 要么是成对的复共轭数

$$\det(D - \lambda) = \det(\gamma_5^2(D - \lambda)) = \det(\gamma_5(D - \lambda)\gamma_5) = \det(D^\dagger - \lambda) = \det(D - \lambda^*)^* \quad (131)$$

所以我们看到, 如果 λ 是 D 的本征值, λ^* 也必然是 D 的本征值。除此之外, 我们还可以论证, 如果 D 的本征值 λ 为复数, 那么该本征值对应的本征矢量 ϕ 关于 γ_5 的内积为 0

$$\lambda(\phi, \gamma_5 \phi) = (\phi, \gamma_5 D \phi) = (\phi, D^\dagger \gamma_5 \phi) = (D \phi, \gamma_5 \phi) = \lambda^*(\phi, \gamma_5 \phi) \quad (132)$$

这里, 内积定义为 $\chi^\dagger \phi = (\chi, \phi)$ 。如果 $\text{Im}(\lambda) \neq 0$, 那么 $(\phi, \gamma_5 \phi) = 0$ 。

3.2 手征费米子

关于手征费米子, 故事还要从 Wilson 重整化群谈起。Wilson 重整化群讲的是我们把格点上定义的作用量看作是有效作用量, 所有的尺度小于 a 的场都被积分积掉了, 留给我们的的是一个有截断 a 依赖关系的有效理论。我们下面来看对于自由费米子场, 如果把尺度小于 a 的场都积掉, 会怎么样?

首先, 引入连续时空费米子作用量

$$S_F[\phi, \bar{\phi}] = \int d^4x \bar{\phi}(x) \gamma_\mu \partial_\mu \phi(x) = \int d^4p \bar{\phi}(-p) (i\gamma_\mu p_\mu) \phi(p) \quad (133)$$

这里, 为了区别格点上的费米子场, 我们用 ϕ 来代表连续时空费米子场。那么格点上的费米子场的定义方式为

$$\psi(n) = \int_{c_{na}} d^4x \frac{x}{a} \phi(x) = \left(\prod_{\mu=1}^4 \int_{n_\mu - \frac{1}{2}}^{n_\mu + \frac{1}{2}} d\frac{x_\mu}{a} \right) \phi(x) = \int_{-\infty}^{+\infty} \frac{d^4p}{(2\pi)^4} \phi(p) \pi(p) e^{ip \cdot na} \quad (134)$$

这里, $\pi(p)$ 的表达式为

$$\pi(p) = \left(\prod_{\mu=1}^4 \int_{-\frac{1}{2}}^{+\frac{1}{2}} d\frac{x_\mu}{a} \right) e^{ipx} = \prod_{\mu=1}^4 \frac{2 \sin(p_\mu a/2)}{p_\mu a} \quad (135)$$

格点费米子场的傅立叶变换为

$$\psi(q) = a^4 \sum_n \psi(n) e^{-iq \cdot na} = \sum_m \int_{-\infty}^{+\infty} \frac{d^4 p}{(2\pi)^4} \phi(p) \pi(p) \delta_{(p-q)a/2\pi, m} = \sum_m \phi\left(q + \frac{2\pi m}{a}\right) \pi\left(q + \frac{2\pi m}{a}\right) \quad (136)$$

我们把连续场的自由度积掉，来得到格点场

$$e^{-S_F[\psi, \bar{\psi}]} = \int \mathcal{D}\phi \mathcal{D}\bar{\phi} e^{-S_F[\phi, \bar{\phi}]} \delta\left(\psi(n) - \int_{c_{na}} d^4 \frac{x}{a} \phi(x)\right) \delta\left(\bar{\psi}(n) - \int_{c_{na}} d^4 \frac{x}{a} \bar{\phi}(x)\right) \quad (137)$$

等式的右边可以看成是把 c_{na} 这个小区域中的场自由度积掉。等式的左边是我们定义的格点费米子作用量。作用量 $S_F[\psi, \bar{\psi}]$ 的具体计算可以在文献【hep-lat/9510026】中找到，这里只给出最后结果

$$S_F[\psi, \bar{\psi}] = \int_{-\frac{\pi}{a}}^{+\frac{\pi}{a}} \frac{d^4 p}{(2\pi)^4} \bar{\psi}(-q) D(q) \psi(q) \quad (138)$$

这里 $D(q)$ 的表达式为

$$D(q) = \left(a - i \sum_m \frac{\pi(q + 2\pi m/a)^2}{\gamma_\mu(q + 2\pi m/a)_\mu} \right)^{-1} \quad (139)$$

这里，我们是通过 block 连续时空场论自由度的方法来得到一个格点上的有效理论。block 的方法不一样，最后得到的格点作用量也不一样。怎么处理高能自由度，对应于不同的格点正规化方案。尽管 $D(q)$ 的形式并不唯一，但它有一些共同点。以我们现在得到的 $D(q)$ 为例，首先，在连续极限下，它能回到正确的连续极限形式；其次，它能够把加倍子给去除掉。我们下面来检验这两件事情：在连续极限下，我们有

$$\pi(q + 2\pi m/a)^2 = \begin{cases} \delta_{m0} & \text{对于所有 } q_\mu \approx 0 \\ O(\frac{1}{|m|^8}) & \text{对于任意 } q_\mu \approx \pi/a \end{cases} \Rightarrow \sum_m \frac{\pi(q + 2\pi m/a)^2}{\gamma_\mu(q + 2\pi m/a)_\mu} = \begin{cases} \frac{1}{\gamma_\mu q_\mu} \\ O(a) \end{cases} \quad (140)$$

于是我们马上可以导出

$$D(q) = \begin{cases} i\gamma_\mu q_\mu + O(a) & \text{对于所有 } q_\mu \approx 0 \\ O(1/a) & \text{对于任意 } q_\mu \approx \pi/a \end{cases} \quad (141)$$

另外，我们可以验证，我们构造的这个无加倍子的费米子作用量是破坏手征对称性的。我们有

$$\gamma_5 D^{-1} + D^{-1} \gamma_5 = 2a\gamma_5 \Rightarrow D\gamma_5 + \gamma_5 D = 2aD\gamma_5 D \quad (142)$$

这个关系式叫做 Ginsparg-Wilson relation。我们知道，手征对称性要求 $\{\gamma_5, D\} = 0$ 。很明显 Ginsparg-Wilson relation 违反这个反对易性。从重整化群和有效理论的角度，我们同样得到了无加倍子和保留手征对称性之间的矛盾。

既然这件事情是在格点正规化下是无法调和的矛盾，要解决这个问题，我们就要打破原来的思维模式。我们之所以要求反对易关系式 $\{\gamma_5, D\} = 0$ ，是因为我们采用的手征变换形式是连续极限下的手征变换。如果我们把手征变换的形式换一下，会怎么样？M. Lüscher 提出来，我们可以采用定义在格点上的手征变换。我们下面改变一下手征变换的形式 (modified chiral rotation)

$$\delta\psi = i\epsilon\gamma_5 \left(1 - \frac{a}{2}D\right) \psi, \quad \delta\bar{\psi} = i\epsilon\bar{\psi} \left(1 - \frac{a}{2}D\right) \gamma_5 \quad (143)$$

【作业：请验证，在这种新的手征变换下，如果作用量保持不变，那么 Dirac 矩阵 D 满足 Ginsparg-Wilson relation。】

虽然我们是在自由费米子场下导出的 Ginsparg-Wilson relation, 但在有规范场存在的情况下, 事实上也可以得到同样的关系式。我们把满足 Ginsparg-Wilson relation 的费米子称为 Ginsparg-Wilson 费米子。利用该费米子 Dirac 矩阵的 γ_5 -厄米性: $D^\dagger = \gamma_5 D \gamma_5$, 以及 $D + D^\dagger = a D^\dagger D$, 我们发现, 矩阵 D 的本征值 $\lambda = x + iy$ 的分布是位于复平面上一个圆心为 $(1/a, 0)$, 半径为 $1/a$ 的圆上

$$(x - 1/a)^2 + y^2 = (1/a)^2 \quad \Rightarrow \quad \lambda = \frac{1}{a}(1 - e^{i\varphi}) \quad (144)$$

如果矩阵 D 的本征矢量同时还是 γ_5 的本征矢量, 我们知道 γ_5 的本征值只可能为 ± 1 : $\gamma_5 \phi = \pm \phi$ 。那么, 不难得到 D 的本征值

$$(\gamma_5 D + D \gamma_5 - a D \gamma_5 D) \phi = (\pm 2\lambda \phi \mp a \lambda^2) \phi = 0 \quad (145)$$

公共本征矢量所对应的 D 的本征值只可能是 $\lambda = 0$ 或者 $\lambda = \frac{2}{a}$ 。反过来, 如果 D 的本征值 λ 是实数, 那么我们一定可以把 λ 对应的本征矢量构造成为既是矩阵 D 又是 γ_5 的公共本征矢量。首先, 如果 ϕ 是 D 的本征矢量, 则 $\gamma_5 \phi$ 必然也是。

$$(D \gamma_5 + \gamma_5 D - a D \gamma_5 D) \phi = D \gamma_5 \phi + \lambda \gamma_5 \phi - a \lambda D \gamma_5 \phi = 0, \quad \Rightarrow \quad D(\gamma_5 \phi) = \frac{\lambda}{a\lambda - 1}(\gamma_5 \phi) \quad (146)$$

假如 ϕ 不对应 γ_5 的本征矢量, 也即 $\gamma_5 \phi \neq \pm \phi$, 我们总可以将 ϕ 和 $\gamma_5 \phi$ 重新组合成 γ_5 的本征矢量 $\frac{1 \pm \gamma_5}{2} \phi$ 。

在 Dirac 矩阵 D 的所有本征矢量里面, 我们把本征值为 0 的本征矢量称为零模, 记为 ϕ_0 。如果 ϕ_0 还满足 $\gamma_5 \phi_0 = +\phi_0$, 我们称这种零模的手征性为正; 如果 $\gamma_5 \phi_0 = -\phi_0$, 我们称这种零模的手征性为负。我们把手征性为正的零模的总个数记为 n_+ , 把手征性为负的零模的总个数记为 n_- , 事实上, 这两个值和 $\text{Tr}[\gamma_5 D]$ 之间有个简单的关系

$$Q \equiv \frac{a}{2} \text{Tr}[\gamma_5 D] = n_- - n_+$$

这个关系叫做指标定理 (index theorem)。根据我们之前得到的关于 Dirac 矩阵的性质, 我们容易验证

$$\begin{aligned} \frac{a}{2} \text{Tr}[\gamma_5 D] &= -\frac{1}{2} \text{Tr}[\gamma_5 (2 - aD)] = -\frac{1}{2} \sum_{\lambda} (\phi_{\lambda}, \gamma_5 (2 - aD) \phi_{\lambda}) \\ &= -\frac{1}{2} \sum_{\lambda} (2 - a\lambda) (\phi_{\lambda}, \gamma_5 \phi_{\lambda}) = n_- - n_+ \end{aligned} \quad (147)$$

对于复本征值, 我们有 $(\phi_{\lambda}, \gamma_5 \phi_{\lambda}) = 0$; 对于 $\lambda = \frac{2}{a}$, 前面系数为 0。所以只有零模有贡献, 其贡献为 $n_- - n_+$ 。

在连续时空, 也有一个 Atiyah-Singer 指标定理, 说的是无质量 Dirac 算符的正负手性零模的数量差等于规范场的拓扑荷 (topological charge): $Q^{\text{cont}} = n_- - n_+$ 。这里的拓扑荷与规范场张量之间的关系为

$$Q^{\text{cont}} = \int d^4x q(x)^{\text{cont}}, \quad q(x)^{\text{cont}} = \frac{1}{32\pi^2} \epsilon_{\mu\nu\rho\sigma} \text{Tr}[F_{\mu\nu}(x) F_{\rho\sigma}(x)] \quad (148)$$

其中 $q(x)^{\text{cont}}$ 为拓扑荷密度。在格点上, 我们可以把 $Q = \frac{a}{2} \text{Tr}[\gamma_5 D]$ 当作是拓扑荷, 那么拓扑荷密度可以定义为

$$Q = a^4 \sum_n q(n), \quad q(n) = \frac{1}{2a^3} \text{Tr}_{sc}[\gamma_5 D(n, n)] \quad (149)$$

可以证明, 如果格距 a 足够小, 规范场足够光滑, 那么格点上定义的 $q(n)$ 在连续极限下趋于 $q(x)^{\text{cont}}$: $q(n) = q(x)^{\text{cont}} + O(a^2)$ 。

在格点上对不同的规范场组态做平均, 不同的组态具有不同的拓扑荷, 正反拓扑荷抵消, 平均之后应该为 0。但是拓扑荷平方的真空期望值不为 0, 我们可以定义拓扑荷磁化率 (topological susceptibility)

$$\chi_{top} = \frac{1}{V} \langle Q^2 \rangle = \frac{a^4}{N_L^3 N_T} \sum_{m,n} \langle q(m)q(n) \rangle = a^4 \sum_m \langle q(m)q(0) \rangle \quad (150)$$

下面我们讲一下手征反常。从路径积分的角度去理解手征反常, 它说的是在手征变换下, 尽管作用量是手征不变的, 但积分测度并不是手征不变的, 因此会导致手征反常的出现。但在格点上, 在引入非零格距硬截断的情况下, 手征反常就不应该存在了, 其中一个例子就是 naive fermion, 加倍子的出现就是为了抵消手征反常。但如果我们采用的是 Ginsparg-Wilson 费米子, 实际上我们可以恢复手征反常。这里关键的因素是, 我们定义了新的手征变换

$$\psi' = \left[1 + i\epsilon\gamma_5 \left(1 - \frac{a}{2}D \right) \right] \psi, \quad \bar{\psi}' = \bar{\psi} \left[1 + i\epsilon \left(1 - \frac{a}{2}D \right) \gamma_5 \right] \quad (151)$$

这种手征变换即使在硬截断的情况下, 也可以使得积分测度破坏手征不变性。

$$\mathcal{D}[\psi, \bar{\psi}] = \mathcal{D}[\psi', \bar{\psi}'] \det \left[1 + i\epsilon\gamma_5 \left(1 - \frac{a}{2}D \right) \right]^2 \quad (152)$$

其中行列式这一项可以表达为

$$\begin{aligned} \det \left[1 + i\epsilon\gamma_5 \left(1 - \frac{a}{2}D \right) \right]^2 &= \exp \left(2 \text{Tr} \left[\ln \left(1 + i\epsilon\gamma_5 \left(1 - \frac{a}{2}D \right) \right) \right] \right) \\ &= \exp \left(2i\epsilon \text{Tr} \left[\gamma_5 \left(1 - \frac{a}{2}D \right) \right] + O(\epsilon^2) \right) \\ &= 1 + 2i\epsilon \text{Tr} \left[\gamma_5 \left(1 - \frac{a}{2}D \right) \right] + O(\epsilon^2) = 1 - 2i\epsilon Q + O(\epsilon^2) \\ &= 1 - 2i \int d^4x \frac{1}{32\pi^2} \epsilon_{\mu\nu\rho\sigma} \text{Tr} [F_{\mu\nu}(x)F_{\rho\sigma}(x)] + O(\epsilon^2) + O(a^2) \end{aligned} \quad (153)$$

因此手征反常得以恢复。

还有一点值得说明的是, 人们发现在作用量中加入任何 dimension-5 的算符都是破坏 Ginsparg-Wilson relation 的。这句话换个方式理解, 就是说手征对称性保证了 Ginsparg-Wilson 费米子算符不会和 dimension-5 算符混合。这从侧面反映 Ginsparg-Wilson 费米子的格距误差来自于 dimension-6 算符, 是被 $O(a^2)$ 压低的。从这一点上来看, Ginsparg-Wilson 费米子的性质也比 Wilson 费米子要更好。

以上是关于 Ginsparg-Wilson 费米子的一系列性质的描述。下面我们讨论一下如何给出 Ginsparg-Wilson 费米子的具体的形式, 或者说要给 Ginsparg-Wilson relation 找到一个解。这个解就是 overlap 费米子, 它的定义是

$$D_{ov} = \frac{1}{a} (1 + \gamma_5 \text{sign}[H]), \quad H = \gamma_5 D_W \quad (154)$$

因为 H 是个矩阵, 这里的符号函数定义为

$$\text{sign}[H] = \frac{H}{\sqrt{H^2}} = \text{sign} \left[\sum_i \lambda_i |i\rangle\langle i| \right] = \sum_i \text{sign}(\lambda_i) |i\rangle\langle i| \quad (155)$$

在实际的数值模拟中, 我们可以用 Zolotarev 近似, 用多项式的比值去逼近 $1/\sqrt{x^2}$

$$\frac{1}{\sqrt{x^2}} \approx d \prod_{n=1}^m \frac{x^2 + c_{2n}}{x^2 + c_{2n-1}} \quad (156)$$

这也意味着，具有完美手征性的 overlap 费米子，在数值计算的时候所耗费的计算量也更大。

到目前为止，我们讨论了 Ginsparg-Wilson relation，以及它的解，但我们并没有给出 Ginsparg-Wilson 费米子的 explicit 质量项。质量项代表的是左右手费米子的耦合，所以我们要先定义左右手费米子。根据 Ginsparg-Wilson relation，我们有

$$[\gamma_5(1 - aD)]^2 = 1 \quad \Rightarrow \quad \hat{\gamma}_5 = \gamma_5(1 - aD) \quad (157)$$

这可以帮助我们定义新的左右手费米子投影算符

$$\hat{P}_{R,L} = \frac{1 + \hat{\gamma}_5}{2}, \quad P_{R,L} = \frac{1 + \gamma_5}{2}, \quad \psi_{R,L} = \hat{P}_{R,L}\psi, \quad \bar{\psi}_{R,L} = \bar{\psi}P_{L,R} \quad (158)$$

我们很容易证明

$$\bar{\psi}D\psi = \bar{\psi}_L D\psi_L + \bar{\psi}_R D\psi_R \quad (159)$$

对于质量项，我们有

$$m(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R) = m\bar{\psi}(P_L\hat{P}_L + P_R\hat{P}_R)\psi = m\bar{\psi}\left(1 - \frac{a}{2}D\right)\psi \quad (160)$$

最后，带质量的 Ginsparg-Wilson 费米子的 Dirac 矩阵可以写为

$$D_m = D + m\bar{\psi}\left(1 - \frac{a}{2}D\right)\psi = \left(1 - \frac{am}{2}\right)D + m \quad (161)$$

含有质量的 Dirac 矩阵的本征矢量和不含质量的本征矢量一样，但本征值有个变化

$$D_m\phi = \lambda_m\phi, \quad \lambda_m = \left(1 - \frac{am}{2}\right)\lambda + m = \left(1 - \frac{am}{2}\right)\frac{1}{a}(1 - e^{i\varphi}) + m \quad (162)$$

由于 $\pm\varphi$ 成对出现，而质量项的平移又保证了 $|\lambda| > 0$ ，所以矩阵 D_m 是正定的，它的行列式 $\det(D_m) > 0$ 。

4 路径积分和 Monte Carlo 模拟

一旦格点作用量被选定以后，下一步就是做路径积分，配分函数的形式为

$$Z = \int \mathcal{D}[U] \int \mathcal{D}[\psi, \bar{\psi}] e^{-S_G[U] - \sum_f \bar{\psi}_f(D[U] + m_f)\psi_f} \quad (163)$$

利用 Grassmann 数积分，先把费米子场积掉，我们得到配分函数

$$Z = \int \mathcal{D}[U] e^{-S_G[U]} \prod_f \det(D[U] + m_f) \quad (164)$$

我们可以计算多个 local 的规范不变算符的真空期望值，也就是它们的关联函数

$$\langle O(U, \psi, \bar{\psi}) \rangle = \frac{1}{Z} \int \mathcal{D}[U] \int \mathcal{D}[\psi, \bar{\psi}] O(U, \psi, \bar{\psi}) e^{-S_G[U] - \sum_f \bar{\psi}_f(D[U] + m_f)\psi_f} \quad (165)$$

如果算符依赖于正反夸克场，那么积掉这些场不光会给我们行列式 $\prod_f \det(D[U] + m_f)$ ，而且根据 Wick 定理，我们会得到一系列的夸克传播子 $(D[U] + m_f)^{-1}$ 。传播子其实就是规范场的函数，所以路径积分的最终形式可以写成

$$\langle O(U, \psi, \bar{\psi}) \rangle = \frac{1}{Z} \int \mathcal{D}[U] O[U, m_f] e^{-S_G[U]} \prod_f \det(D[U] + m_f) \quad (166)$$

我们把算符中出现的夸克叫做价夸克,把作用量中出现的叫做海夸克。价夸克的效应体现在传播子中,海夸克的效应体现在行列式中。如果海夸克和价夸克的质量不一样,我们称之为部分淬火 (partial quenching); 如果把整个行列式忽略掉, 设成 1, 就是我们平常所说的淬火近似。

这个路径积分, 即使在我们引入了格点正规化和有限体积之后, 依然很难直接计算。我们不妨考虑一个更简单的 Ising 统计模型, 假设我们一共有 16^4 个时空点, 每个点上有自旋向上和向下这两种可能, 如果我们的路径积分要把各种自旋组态都做加权平均, 总的组态数是 $2^{16^4} = 2^{65536} = 10^{19728}$ 次, 这是一个不可能的任务。所以我们必须要做 Monte Carlo 模拟, 也就是我们生成 N 个规范场组态, 这些组态按照几率 $\frac{1}{Z} \exp(-S_G[U]) \prod_f \det(D[U] + m_f)$ 进行分布。我们只需要对物理量进行组态求平均, 就能得到我们想要的物理结果

$$\langle O(U, \psi, \bar{\psi}) \rangle = \frac{1}{N} \sum_{n=1}^N O[U_n, m_f] \quad (167)$$

当然, 我们所得的物理结果是带统计误差的, 这个误差随着 N 增大成 $1/\sqrt{N}$ 减小。这里的 U_n 指的是第 n 个规范场组态, 每个规范场组态是所有格点时空点上 4 个方向的 $SU(3)$ 矩阵的集合 $U = \{U_\mu(x)\}$ 。

Monte Carlo 归根结底是个取样的思想。如果我们想要计算某个函数的积分, 我们可以在积分区间上随机的撒点, 要求这些随机点 x_n 出现的几率密度是 $\rho(x_n) = \frac{1}{b-a}$ 。这是一个简单的取样, 然后我们计算这些随机点上的函数值 $f(x_n)$, 再对函数值求平均, 就得到我们想要的积分结果

$$\frac{1}{b-a} \int_a^b dx f(x) = \langle f \rangle_\rho = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N f(x_n) \quad (168)$$

对于一维的积分, 我们当然有各种方法可以做, 但是我们的路径积分至少是个 N^4 维的积分, 这个时候取样求和的思想就显然特别重要。当然, 我们这里的取样指的是重点取样, 因为我们的积分有个权重因子, 我们要求对路径积分贡献大的组态的分布几率高一点, 贡献低的组态分布几率低一些。如果在一维情况下, 有个权重因子 $\rho(x)$, 那么对函数的加权积分可以表达为

$$\langle f \rangle_\rho = \frac{\int_a^b dx \rho(x) f(x)}{\int_a^b dx \rho(x)} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N f(x_n) \quad (169)$$

这里 x_n 的取样满足几率分布密度

$$dP(x) = \frac{\rho(x) dx}{\int_a^b dx \rho(x)} \quad (170)$$

类比过来, 对于 $SU(3)$ 规范场组态, 我们希望它满足几率分布密度

$$dP(U) = \frac{e^{-S[U]} \prod_f \det(D_f) \mathcal{D}[U]}{\int \mathcal{D}[U] e^{-S[U]} \prod_f \det(D_f)} \quad (171)$$

为简单起见, 我们先不考虑几率密度函数中的行列式。现在的问题是, 我们怎么样得到具有给定几率分布密度的规范场组态。一个组态很难被凭空直接构造出来, 但它可以从另外一个组态被“修改”出来。基于这种想法, 我们就可以从一个初始组态出发, 按照一定规则以及通过随机演化构造出下一个组态, 然后继续这个过程, 直到构造出满足分布几率的组态。在这个过程中, 每个组态只依赖于它之前的那个组态, 这个过程被称为马尔可夫过程, 而这些组态构成了一个马尔可夫链。

$$U_0 \rightarrow U_1 \rightarrow U_2 \rightarrow \cdots U_n \rightarrow \cdots \quad (172)$$

因为在马尔可夫链的演化过程是通过随机演化进行的，初始规范场组态取什么样并不重要，一种取法是把所有的链接都取成单位矩阵。

我们所寻求的算法，实际上就是把一个组态随机变成另外一个组态的规则。在这个规则下，组态作为系综空间上的一个点，随机游走在系综空间上。在经历了一定的计算机演化时间以后，我们统计这个点出现在系综空间的各个位置上的几率，如果这个几率符合路径积分中规定的几率分布函数，我们就认为这个算法是正确的。当然，因为初始的组态是人为选定的，所以在组态演化的早期要有一个“预热”的过程，在产生了一定的不可用组态之后，再产生的组态才是我们希望的。

一个组态对应于系综空间上的一个点，就如同一维积分里的坐标一样，其实我们并不太关心这个组态是什么。我们真正关心的是组态的几率密度。我们把组态 U 出现的几率密度叫 $P[U]$ 。在马尔可夫过程中，我们定义 U 跃迁到 U' 的跃迁几率密度为 $W(U \rightarrow U')$

$$P^{(1)}[U'] = \int \mathcal{D}U W(U \rightarrow U') P^{(0)}[U] \quad (173)$$

由于初始组态是给定的，它对应的几率分布密度可以定义为 $P^{(0)}[U] = \delta(U - U_0)$ 。然后通过跃迁规则进行演化，不停的更新可以产生一系列的几率密度

$$P^{(0)}[U] \xrightarrow{W} P^{(1)}[U] \xrightarrow{W} \dots P^{(n)}[U] \xrightarrow{W} \dots \quad (174)$$

既然是跃迁几率密度，我们要求 W 满足

$$0 \leq W(U \rightarrow U') \leq 1, \quad \int \mathcal{D}[U'] W(U \rightarrow U') = 1 \quad (175)$$

第二个条件表示，从组态 U 出发跃迁到其它组态（也包括 U 自身）的几率总和是 1。除了这两个最基本的条件外，我们还要求跃迁几率密度 W 能够保证在足够长时间的演化以后，得到我们想要的几率分布

$$\lim_{n \rightarrow \infty} P^{(n)}[U] = \frac{1}{Z} e^{-S[U]} \quad (176)$$

这需要跃迁规则 W 满足两个条件：“各态历经”和“稳定性条件”。各态历经条件很容易理解，即在足够长时间的随机游走中，系综空间的每一个点都会被走到，不存在死角。稳定性条件则要求我们最后达到的几率分布（我们可以叫它玻尔兹曼分布）在跃迁几率 W 的作用下是稳定的，

$$e^{-S[U']} = \int \mathcal{D}[U] W(U \rightarrow U') e^{-S[U]} \quad (177)$$

这个稳定性条件告诉我们，玻尔兹曼分布几率 $e^{-S[U]}$ 是整个马尔可夫过程几率密度演化的不动点。还有一个比稳定性条件更强的条件叫做“细致平衡条件”，它表述为

$$e^{-S[U']} W[U' \rightarrow U] = W[U \rightarrow U'] e^{-S[U]} \quad (178)$$

即在系综空间的任何两个点之间的“概率流”都必须平衡的。将上面这个式子对 $\mathcal{D}[U]$ 进行积分，我们马上可以得到稳定性条件。

下面我们介绍 Metropolis 算法来实现马尔可夫过程，Metropolis 算法是二十世纪十大算法之一。对于我们要处理的路径积分，Metropolis 算法可以分三步。

- 第一步：预选一种几率 $W_0(U \rightarrow U')$ ，使得组态 U 能够以几率 W_0 的形式跃迁到另一个组态 U'

- 第二步：设置一定几率决定接受 U' 作为新的组态，还是放弃 U' 。这里接收的几率为

$$W_A(U \rightarrow U') = \min \left(1, \frac{W_0(U \rightarrow U') \exp(-S[U'])}{W_0(U' \rightarrow U) \exp(-S[U])} \right) \quad (179)$$

- 第三步：重新回到第一步

结合前两步，从组态 U 到 U' 的跃迁几率为 $W(U \rightarrow U') = W_A(U \rightarrow U')W_0(U \rightarrow U')$ 。我们要求对于预选几率 W_0 满足对称性

$$W_0(U \rightarrow U') = W_0(U' \rightarrow U) \quad (180)$$

我们下面可以验证一下，现在构造的跃迁几率满足细致平衡条件

$$\begin{aligned} W(U \rightarrow U')e^{-S[U]} &= W_0(U \rightarrow U') \min \left(1, \frac{W_0(U \rightarrow U') \exp(-S[U'])}{W_0(U' \rightarrow U) \exp(-S[U])} \right) \exp(-S[U]) \\ &= \min (W_0(U' \rightarrow U) \exp(-S[U]), W_0(U \rightarrow U') \exp(-S[U'])) \\ &= W(U' \rightarrow U)e^{-S[U']} \end{aligned} \quad (181)$$

下面，我们来讨论一下针对 $SU(3)$ 规范场，Metropolis 算法应该如何实现。我们注意到规范场作用量是个局域作用量，如果我们将其中某个链接从 $U_\mu(n)$ 改为 $U_\mu(n)'$ ，对作用量的影响仅仅是那些包含了 $U_\mu(n)$ 链接的小方块。这样的小方块一共有 6 个，我们只需计算这部分小方块对作用量产生的影响

$$\Delta S = S[U_\mu(n)']_{\text{loc}} - S[U_\mu(n)]_{\text{loc}} \quad (182)$$

我们要求对于新组态的接收率为 $\min(1, e^{-\Delta S})$ 。要做到这一点，只需在 $\Delta S < 0$ 时，全盘接收；在 $\Delta S > 0$ 时，要在 $[0, 1)$ 之间随机产生一个数 r ，在 $r \leq e^{-\Delta S}$ 时接收即可。另外，对于规范场链接 $U_\mu(n)$ 变到 $U_\mu(n)'$ ，我们只需要随机生成一个 $SU(3)$ 群元 X 即可

$$U_\mu(n)' = XU_\mu(n) \quad (183)$$

当然，这里 X 的选取和预选几率 $W_0(U \rightarrow U')$ 有关。如果要求 W_0 是对称的，那么选取 X 的几率要保证和选取 X^{-1} 的几率是一致的。

对于一个 $SU(3)$ 矩阵 X ，我们可以把它写成三个 $SU(2)$ 型矩阵的乘积

$$X = RST, \quad R = \begin{pmatrix} r_{11} & r_{12} & 0 \\ r_{21} & r_{22} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad S = \begin{pmatrix} s_{11} & 0 & s_{12} \\ 0 & 1 & 0 \\ s_{21} & 0 & s_{22} \end{pmatrix}, \quad T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & t_{11} & t_{12} \\ 0 & t_{21} & t_{22} \end{pmatrix} \quad (184)$$

对于每个 $SU(2)$ 型的矩阵，它可以参数化为

$$U = x_0 + i\vec{x} \cdot \vec{\sigma}, \quad \text{这里要求 } \det(U) = |x|^2 = x_0^2 + \vec{x}^2 = 1 \quad (185)$$

参数化以后，我们只需要设定一定的概率去取 X 即可。 X 一般在单位矩阵 1 附近取值，如果取得太偏离 1，接收率会低；如果太接近 1，那么组态之间关联性太强。所以需要找一个平衡点。

Metropolis 算法是利用了规范场作用量的局域性，对每个链接逐个进行更新。还有一种算法是对全局的规范场变量进行更新，称为 Hybrid Monte Carlo，它采用了分子动力学的算法。假设我们要产生的几率分布为

$$P[\phi] \propto e^{-S[\phi]} \quad (186)$$

其中 $e^{-S[\phi]}$ 是几率分布函数。我们知道自然界中的空气分子的分布自动按照玻尔兹曼分布，只不过其中 $e^{-S[\phi]}$ 表示的是空气分子的重力势而已。从任意的初态开始，空气分子可以靠着分子的运动和碰撞达成玻尔兹曼分布。我们可以模拟空气分子演化的这一过程。由于 $S[\phi]$ 代表的是势能，我们在系统中引入广义动量 π ，构造哈密顿量

$$H[\pi, \phi] = \sum_x \frac{\pi(x)^2}{2} + S[\phi] \quad (187)$$

我们要求在虚拟时间 (计算机时间) τ 中进行经典力学的演化

$$\begin{cases} \dot{\pi} \equiv \frac{d\pi}{d\tau} = -\frac{\partial H[\pi, \phi]}{\partial \phi} = -\frac{\partial S}{\partial \phi} \\ \dot{\phi} \equiv \frac{d\phi}{d\tau} = \frac{\partial H}{\partial \pi} = \pi \end{cases} \quad (188)$$

经过足够长的时间演化以后，我们会得到玻尔兹曼分布

$$P[\pi, \phi] = e^{-H[\pi, \phi]} = \exp\left(-\sum_x \frac{\pi(x)^2}{2}\right) \exp(-S[\phi]) \quad (189)$$

由于哈密顿量是动能部分和势能部分之和，因此体系关于动量和坐标的几率分布是完全独立的。也就是说，当系统达到平衡时，它的势能部分，或者说 ϕ 场满足的分布，自动就是我们需要的 $e^{-S[\phi]}$ 分布。

尽管这个分子动量学的图像比较清楚，但这个微分方程的求解却并不见得容易。如果不能获得它的解析解，我们可以引入一个积分步长 $\delta\tau$ ，求一个近似积分。近似求解微分方程的一个后果是 $\Delta H = H(\tau) - H(0) \neq 0$ 。我们可以引入一个 Monte Carlo 判据，以概率 $\min(1, e^{-\Delta H})$ 来接收新的组态。

对于规范场来讲，场变量是定义在 $SU(3)$ 群元上面，而与之对应的广义动量是定义在相应的李代数中。我们一样可以写出体系的哈密顿量，然后模拟它的演化，只是具体过程更加复杂。我们这里只是提一下，不再做具体介绍，感兴趣的同学可以阅读相关的格点书籍。

对于含费米子行列式的路径积分，

$$Z = \int \mathcal{D}[U] e^{-S_G[U]} \prod_f \det(D[U, m_f]) \quad (190)$$

这里面有一个费米子行列式正定性的问题。对于 Ginsparg-Wilson 费米子，我们论证过费米子行列式一定是正的。但对于其它的费米子，比方说 Wilson 费米子，其行列式就不一定是正的。其后果就是被积函数就无法解释成几率密度从而破坏了 Monte Carlo 算法。好在我们有一个 Dirac 矩阵的 γ_5 -厄米性，它保证了

$$\det(D^\dagger) = \det(\gamma_5 D \gamma_5) = \det(D) \Rightarrow \det(D)^2 = \det(DD^\dagger) \quad (191)$$

那么对于模拟两味道 u 、 d 夸克的系统，如果我们不考虑 u -夸克和 d -夸克之间微小的质量差，那么行列式的平方自动可以写成 $\det(DD^\dagger)$ 的形式，因此被积函数是正定的。我们可以引入玻色型的赝费米子场 ϕ ，并对 ϕ 做路径积分

$$Z = \int \mathcal{D}[U] \mathcal{D}[\phi^*] \mathcal{D}[\phi] e^{-S_G[U] - \phi^\dagger (DD^\dagger)^{-1} \phi} = \int \mathcal{D}[U] e^{-S_G[U]} \det(D[U, m_{u,d}])^2 \quad (192)$$

如果是 $2+1$ 味的格点模拟， s -夸克不是成对出现，那就有可能出现单个非正定的 Dirac 矩阵。但我们知道非正定的出现完全是由于我们做格点正规化的缘故。如果我们选取性质比较好的 overlap 费米子，那么行列式一定为正。或者我们也可以认为，行列式非正定这件事情是一种 lattice artifacts，在格距足够小的时候，行列式应该是正定的。基于这种考虑，我们在模拟的时候取 s -夸克的行列式的绝对值

$$|\det(D(m_s))| = |\det(D(m_s))| = \det(DD^\dagger)^{\frac{1}{2}} \quad (193)$$

于是我们需要模拟的系统为

$$Z = \int \mathcal{D}[U] \mathcal{D}[\phi^*] \mathcal{D}[\phi] e^{-S_G[U] - \phi^\dagger (DD^\dagger)^{-\frac{1}{2}} \phi} \quad (194)$$

其中 $\frac{1}{\sqrt{x^2}}$ 这件事情，可以由有理分式来逼近，所以 Monte Carlo 模拟也可以做到。真正的问题出现在有限密度 QCD。

我们之前讲的路径积分，它的配分函数是真空到真空的振幅。在真空中，正反重子或者正反夸克的总和抵消，净重子数或者说净夸克数为 0。（重子数和夸克数的关系为 $N_B = N_q/3$ 。）但如果我们考察的是以物质场为背景的 QCD 相互作用呢？比方说重离子碰撞实验，我们就必须要考虑一个夸克数或者说重子数不为 0 的系统，这样的系统我们也称有限密度系统。于是我们需要在作用量中引入夸克数算符

$$N_q = \int d^3x \bar{\psi}(x) \gamma_4 \psi(x) \quad (195)$$

其中 $\bar{\psi}(x) \gamma_4 \psi(x)$ 是守恒矢量流的时间分量。有了夸克数算符，配分函数函数改写为

$$Z(T, \mu) = \text{Tr} \left[e^{-(H - \mu N_q)T} \right] \quad (196)$$

这里的 T 指的是格点欧氏时间长度，它在有限温度、有限密度理论里面的物理含义是温度的倒数。系数 μ 叫做夸克化学势，它表征的是背景场中非零夸克数或者重子数的大小。加入 N_q 这一项，原来的费米子作用量就变化为

$$\int d^4x \bar{\psi} D \psi \rightarrow \int d^4x \bar{\psi} D \psi + \mu \bar{\psi} \gamma_4 \psi \quad (197)$$

很明显 γ_4 矩阵存在关系式 $\gamma_5 \gamma_4 \gamma_5 = -\gamma_4^\dagger$ ，破坏了原来 Dirac 矩阵的 γ_5 -厄米性。它导致的后果是费米子行列式在路径积分中有正、有负，这在统计物理里面被称为符号问题 (sign problem)。符号问题就其字面意思上来讲，是指高维度积分过程中由于被积函数的剧烈震荡行为，正的贡献和负的贡献又接近抵消，使得数值的方法几乎没办法得到可靠的结果。从物理上来讲，符号问题通常对应于具有大量强相互作用费米子的量子力学系统或者对应于具有非零密度的强相互作用费米子的场论系统，当两个费米子交换的时候它们的波函数会改变符号。因此，当我们对所有的多粒子态求和的时候，就会出现被求和函数剧烈震荡的情况。目前，符号问题还是 QCD 悬而未决的问题之一。

4.1 一个简单的数值路径积分案例：谐振子 (G. P. Lepage [hep-lat/0506036])

我们考虑一个一维的量子力学系统，哈密顿量为 H 。我们用路径积分来计算在 t_i 时刻从初态 $|x_i\rangle$ 跃迁到末态 $|x_f\rangle$ 的振幅

$$\langle x_f | e^{-H(t_f - t_i)} | x_i \rangle = \int \mathcal{D}[x(t)] e^{-S[x]} \quad (198)$$

等式的左边是哈密顿量算符，右边是路径积分的作用量

$$S[x] \equiv \int_{t_i}^{t_f} dt L(x, \dot{x}) \equiv \int_{t_i}^{t_f} dt \left[\frac{m\dot{x}(t)^2}{2} + V(x(t)) \right] \quad (199)$$

这里我们已经做了 Wick 转动，选取了欧氏时空。 $\int \mathcal{D}[x(t)]$ 要求对所有可能的路径求和

$$\{x(t) \text{ for } t = t_i \rightarrow t_f\} \quad (200)$$

初始和最后时刻给定

$$x(t_i) = x_i, \quad x(t_f) = x_f \quad (201)$$

如果我们选取

$$x_i = x_f \equiv x, \quad t_f - t_i \equiv T \quad (202)$$

那么振幅可以写成

$$\langle x|e^{-HT}|x\rangle = \sum_n \langle x|E_n\rangle e^{-E_n T} \langle E_n|x\rangle \quad (203)$$

在 T 很大的情况下, 只有基态能贡献

$$\langle x|e^{-HT}|x\rangle \xrightarrow{T \rightarrow \infty} e^{-E_0 T} |\langle x|E_0\rangle|^2 \quad (204)$$

我们可以得到基态能量 E_0 和基态波函数 $\langle x|E_0\rangle$ 。

下面进行格点分立化

$$t_j = t_i + ja, \quad \text{for } j = 0, 1, 2, \dots, N \quad (205)$$

其中格距 $a = (t_f - t_i)/N$ 。我们想要的组态由一个 $N + 1$ 维矢量来给出

$$x = \{x(t_0), x(t_1), \dots, x(t_N)\} \quad (206)$$

路径积分在格点离散的情况下有严格的数学定义

$$\int \mathcal{D}[x(t)] \rightarrow A \int_{-\infty}^{\infty} dx_1 dx_2 \cdots dx_{N-1} \quad (207)$$

其中 x_0 和 x_N 满足边界条件

$$x_0 = x_N = x \quad (208)$$

这里归一化因子 $A = (\frac{m}{2\pi a})^{\frac{N}{2}}$ 由关系式

$$\langle x_{j+1}|e^{-ap^2/2m}|x_j\rangle = \int \frac{dp}{2\pi} e^{ip(x_{j+1}-x_j)} e^{-ap^2/2m} = \sqrt{\frac{m}{2\pi a}} e^{-(x_{j+1}-x_j)^2 m/(2a)} \quad (209)$$

这里我们用到了平面波的定义

$$\langle x|p\rangle = \frac{1}{\sqrt{2\pi}} e^{ipx} \quad (210)$$

在分立的时空下, 跃迁振幅等于

$$\langle x|e^{-HT}|x\rangle \approx A \int_{-\infty}^{\infty} dx_1 \cdots dx_{N-1} e^{-S_{\text{lat}}[x]} \quad (211)$$

格点上的作用量给出如下

$$S_{\text{lat}}[x] \equiv \sum_{j=0}^{N-1} \left[\frac{m}{2a} (x_{j+1} - x_j)^2 + aV(x_j) \right] \quad (212)$$

这里 $x_0 = x_N = x$, $a = T/N$ 。

【作业：取 $a = 1/2$ 和 $N = 8$, $V(x)$ 取谐振子势 $V(x) = \frac{x^2}{2}$, 质量 $m = 1$ 。请通过 7 维积分, 求出振幅 $\langle x|e^{-HT}|x\rangle$ 在若干个不同的 x 处的值, 比方说 x 可以从 0 取到 2, 然后验证

$$\langle x|e^{-HT}|x\rangle \approx |\langle x|E_0\rangle|^2 e^{-E_0 T} \quad (213)$$

这里 $E_0 = 1/2$, $\langle x|E_0\rangle = \frac{e^{-x^2/2}}{\pi^{1/4}}$ **】**

下一步我们计算关联函数

$$\langle x(t_2)x(t_1) \rangle \equiv \frac{\int \mathcal{D}[x(t)] x(t_2)x(t_1)e^{-S[x]}}{\int \mathcal{D}[x(t)] e^{-S[x]}}. \quad (214)$$

这里我们对 $x_i = x_f = x$ 进行积分。等式左边可以写为

$$\int dx \langle x | e^{-H(t_f-t_2)} x e^{-H(t_2-t_1)} x e^{-H(t_1-t_i)} | x \rangle. \quad (215)$$

令 $T = t_f - t_i$, $t = t_2 - t_1$, 我们可以把关联函数写成

$$\langle x(t_2)x(t_1) \rangle = \frac{\sum e^{-E_n T} \langle E_n | x e^{-(H-E_n)t} x | E_n \rangle}{\sum e^{-E_n T}}. \quad (216)$$

如果 $T \gg t$, 那么基态 $|E_0\rangle$ 占有所有态求和中的主导贡献。

$$G(t) = \langle x(t_2)x(t_1) \rangle \rightarrow \langle E_0 | x e^{-(H-E_0)t} x | E_0 \rangle \xrightarrow{t \gg 0} |\langle E_0 | x | E_1 \rangle|^2 e^{-(E_1-E_0)t} \quad (217)$$

在 t 很大时, 主要是第一激发态 $|E_1\rangle$ 的贡献。因为算符 x 具有奇宇称, 所以它会混合基态和第一激发态。通过 $G(t)$, 我们可以求出

$$(E_1 - E_0)a = \lim_{t \rightarrow \infty} \log(G(t)/G(t+a)) \quad (218)$$

对于一个算符 O , 它的期望值为组态平均

$$\langle O \rangle = \sum_{k=1}^{N_{conf}} O[x^{(k)}] \quad (219)$$

误差 σ_O 为

$$\sigma_O^2 \approx \frac{1}{N_{conf}} \left\{ \frac{1}{N_{conf}} \sum_{k=1}^{N_{conf}} O^2[x^{(k)}] - \langle O \rangle^2 \right\} \quad (220)$$

作业：1、用 Metropolis 算法生成组态。先给定一组初始的组态 $x = \{x_0, x_1, \dots, x_{N-1}\}$, $x_0 = x_N$ 满足周期性边界条件。这里 N 取为 20。依次从 x_0 到 x_{N-1} 进行更新。对于第 j 个元素 x_j 的更新, 我们在区间 $[-\epsilon, +\epsilon]$ 之间产生一个随机数 ξ , 然后更新 $x_j \rightarrow x_j + \xi$, 并计算 ΔS 。如果 $\Delta S < 0$, 接收新的 x_j ; 如果 $\Delta S > 0$, 在 $[0, 1]$ 之间产生一个随机数 η , 如果 $e^{-\Delta S} > \eta$, 接收新的值, 否则放弃。这里 ϵ 不能取得太大, 否则接收率很低; 也不能太小, 否则每次更新组态变动很小。我们这里取 $\epsilon = 1.4$ 。整体而言, 更新前后的组态并不是独立的, 我们认为更新一定的次数 N_{cor} 以后才能产生一个独立的组态, 这里 N_{cor} 取为 20。另外, 从初始组态开始, 需要进行一定的预定, 也能产生满足我们分布要求的组态, 这里预热的更新次数取为 $10N_{cor}$ 。请用 Metropolis 算法产生一万个组态。

2、通过组态平均来计算 $G(t)$, 可以分别采用 $N_{conf} = 25, 100, 1000, 10000$ 个组态

$$G(t) = \frac{1}{N} \sum_j \langle x(t_j + t)x(t_j) \rangle \Rightarrow G_n = \frac{1}{N} \sum_j \langle x_{(n+j) \bmod N} x_j \rangle \quad (221)$$

$$\Delta E_n \equiv \log(G_n/G_{n+1}) \xrightarrow{n \rightarrow \infty} (E_1 - E_0)a \quad (222)$$