Volume 1 Preparation

Part 4 Quantisation

Kiichi, Shiori and Rin · Edition of 2026-11-14 · https://computoergosum.com/en/principia/yang-mills/

> This PDF is a print of the chapters of the web edition. Interactive figures are frozen in the state in which they were printed. Links between chapters lead to the web edition, which holds the latest text.

Chapter		Page
4-01	Canonical quantisation and the Hamiltonian	2
4-02	The path integral and Euclideanisation	13
4-03	Perturbation theory and Feynman diagrams	25
4-04	Divergences and renormalisation	35
4-05	The algebra of divergences	46
4-06	Asymptotic freedom	58
4-07	Dimensional transmutation and Λ	68
4-08	Gauge fixing and Faddeev–Popov	80
4-09	The axioms of quantum field theory	92
4-10	The Osterwalder–Schrader axioms and reconstruction	105
4-11	The definition of the mass gap	117
4-12	Detour: the perturbation series diverges	130

Canonical quantisation and the Hamiltonian — Gauss's law becomes a condition on states

This chapter organises established knowledge (sources carry the “Known” label in the text). This book's own investigation is mainly in Part 15 and Volume 3 (Parts 16–28 and the appendices).

What this chapter shows — The procedure of imposing commutation relations and building the Hamiltonian (canonical quantisation).
That A_0 has no momentum, and Gauss's law becomes a condition on states: "unchanged by small gauge transformations".
That this condition removes the flat motion in gauge directions, and on a circle fixes the levels by the Casimir.

Prerequisite chapters — 2-05 (spectrum, gap), 3-03 (action, Gauss's law). For oscillators, 2-06.

Said first — In the four-dimensional continuum theory, the Hamiltonian of this chapter is a formula in form only and has not been constructed as an operator of this form (what is rigorous is the lattice, 5-03, and low-dimensional models). The construction published in 2026-11 built a Hamiltonian by reconstruction from the Euclidean side (the committee's assessment is that it is proved in Lean; 12-08). It shows nothing about the mass gap.

Order of this chapter

1. The procedure of canonical quantisation
2. A_0 has no momentum — a constraint
3. Quantising in the temporal gauge — wave functionals
4. Gauss's law is a condition on states — charges and large gauge transformations
5. Seen in the free field — removing the flat directions (Figure 1, Table 1)
6. Yang–Mills on a circle — class functions and the Casimir (Figure 2, Table 2, Table 3)
7. Detour: small boxes, and arguments from wave functionals
8. The route of solving Gauss's law outright — Gribov

9. What is formal

01

The procedure of canonical quantisation — making coordinates and momenta into operators

The problem statement defines the mass gap by the spectrum of the Hamiltonian H (the vacuum is $H\Omega = 0$): "A quantum field theory has a mass gap Δ if H has no spectrum in the interval $(0, \Delta)$ for some $\Delta > 0$." (Jaffe–Witten § 4) Known . One route to writing down that H from the classical field is canonical quantisation (the other is 4-02).

With finitely many degrees of freedom there are three steps (2-05). (i) Make the momenta $p_i = \partial L / \partial \dot{q}_i$. (ii) Write $H = \sum_i p_i \dot{q}_i - L$ in terms of q, p. (iii) Replace q, p by operators on a Hilbert space and impose the following.

$$[q_i, p_j] = i\,\delta_{ij} \qquad (\hbar = 1) \tag{1}$$

On wave functions, $p_j = -i\,\partial/\partial q_j$. For fields the index becomes a point $\mathbf{x}$, and δ_{ij} becomes $\delta^3(\mathbf{x} - \mathbf{y})$. **Example.** A scalar field of mass m splits into an oscillator for each wave number, and the gap was m (2-06 § 08). In Yang–Mills, at step (i) a variable appears for which no momentum can be made.

02

A_0 has no momentum — Gauss's law appears as a constraint

The conventions are the same as in 3-03 § 01 ($\hbar = c = 1$, $D_\mu = \partial_\mu - igA_\mu$, $(D_i X)^a = \partial_i X^a + g f^{abc} A_i^b X^c$ on adjoint fields). The colour electric field is

$$E_i^a = F_{i0}^a = -\partial_t A_i^a + (D_i A_0)^a \tag{2}$$

and $L = \frac{1}{2}\int \sum_a (|\mathbf{E}^a|^2 - |\mathbf{B}^a|^2)\, d^3x$ (the magnetic field contains no time derivative). The momentum of A_i^a is $\Pi_i^a = -E_i^a$, but L does not contain $\partial_t A_0$, so the momentum of A_0 is identically 0.

We write step (ii). What corresponds to $\sum_i p_i \dot{q}_i$ is $\int \Pi_i^a\, \partial_t A_i^a\, d^3x$. Reading (2) as $\partial_t A_i = -E_i + D_i A_0$ and substituting,

$$\int \Pi_i^a\, \partial_t A_i^a\, d^3x = \int \left(|\mathbf{E}|^2 - \mathbf{E}\cdot\mathbf{D}A_0\right) d^3x$$

follows. Subtracting L from this leaves $\tfrac{1}{2}(|\mathbf{E}|^2+|\mathbf{B}|^2)$ and $-\int \mathbf{E}\cdot\mathbf{D}A_0$. In the last term, integration by parts moves D onto $\mathbf{E}$ (we take the fields to be 0 far away). The derivative part ∂_i changes sign by ordinary integration by parts. For the $gf^{abc}A_i^b$ part, f^{abc} changes sign under exchange of indices (it is antisymmetric), so relabelling the indices likewise changes the sign once. So $-\int \mathbf{E}\cdot\mathbf{D}A_0\,d^3x=\int A_0^a(\mathbf{D}\cdot\mathbf{E})^a\,d^3x$, and

$$H=\int\Big[\tfrac{1}{2}\big(|\mathbf{E}|^2+|\mathbf{B}|^2\big)+A_0^a\,(\mathbf{D}\cdot\mathbf{E})^a\Big]d^3x \quad (3)$$

follows ($|\mathbf{E}|^2=\sum_a|\mathbf{E}^a|^2$) Known . A_0 is an undetermined multiplier that appears only linearly (the λ of the toy in 2-01 § 08), and its equation $(\mathbf{D}\cdot\mathbf{E})^a=0$ is Gauss's law ((3) of 3-03). Such a condition is called a **constraint**, and its general treatment is said to begin with Dirac (1950).

03

Quantising in the temporal gauge — wave functionals and the electric field operator

We set $A_0=0$ by a gauge transformation (the **temporal gauge**; it is enough to solve $\partial_t\,\Omega=-ig\,\Omega A_0$ at each point). The remaining freedom is time-independent gauge transformations. The coordinates are $A_i^a(\mathbf{x})$ and the momenta are the $\Pi_i^a=-E_i^a$ of § 02. Since $\Pi=-E$, the field version of (1), $[A,\Pi]=i\delta$, is

$$[A_i^a(\mathbf{x}),E_j^b(\mathbf{y})]=-i\,\delta^{ab}\delta_{ij}\,\delta^3(\mathbf{x}-\mathbf{y}) \quad (4)$$

(the sign is negative because $E=-\Pi$). We represent states by functions of configurations (**wave functionals**) $\psi[A]$. For small ϵ and a function $h_i^a(\mathbf{x})$,

$$\psi[A+\epsilon h]-\psi[A]\approx\epsilon\int h_i^a(\mathbf{x})\,\frac{\delta\psi}{\delta A_i^a(\mathbf{x})}\,d^3x$$

the function $\delta\psi/\delta A_i^a(\mathbf{x})$ for which this holds is called the **functional derivative**. It is the finite-dimensional $\psi(q+\epsilon h)-\psi(q)\approx\epsilon\sum_i h_i\,\partial\psi/\partial q_i$ with the index i replaced by the point $\mathbf{x}$ and the sum by an integral. Following the $p=-i\,\partial/\partial q$ of (1), $\Pi_i^a(\mathbf{x})=-i\,\delta/\delta A_i^a(\mathbf{x})$, so it acts as $E_i^a(\mathbf{x})=i\,\delta/\delta A_i^a(\mathbf{x})$, and (3) is formally

$$H=\int\Big[-\tfrac{1}{2}\frac{\delta^2}{\delta A_i^a(\mathbf{x})^2}+\tfrac{1}{2}B_i^aB_i^a\Big]d^3x \quad (5)$$

($\delta^2/\delta A_i^a(\mathbf{x})^2$ is summed over a and i) Known . It can be read as a particle with infinitely many coordinates moving in the potential energy $\tfrac{1}{2}\int|\mathbf{B}|^2$. Strictly, (5) is form only (§ 09). Rescaling as $\hat{A}=gA$,

$$H = \int \Big[-\frac{g^2}{2}\,\frac{\delta^2}{\delta \hat{A}^2} + \frac{1}{2g^2}\,|\hat{\mathbf{B}}|^2 \Big]\, d^3x \tag{6}$$

and when g is large the electric term wins, when small the magnetic term. The coefficients $g^2/2a$ and $2N/g^2 a$ in (16) of 5-03 are the lattice version of these two terms. **Example.** Of the $3(N^2-1)$ components A_i^a at each point, the condition of the next section removes N^2-1 gauge directions, leaving $2(N^2-1)$ (the same number as in 3-03 § 03). The one component that the classical theory removed with two things, the constraint and the remaining gauge transformations, is removed at once by (10) in the quantum theory.

04

Gauss's law is a condition on states — unchanged by small gauge transformations

Since A_0 has been removed, Gauss's law is no longer among the equations of motion. The Gauss operator

$$G^a(\mathbf{x}) = (\mathbf{D}\cdot\mathbf{E})^a = \partial_i E_i^a + g f^{abc} A_i^b E_i^c \tag{7}$$

has three properties Known . (a) **It generates gauge transformations.** Setting $G(\alpha) = \int \alpha^a G^a \, d^3x$, integration by parts and (4) give

$$G(\alpha)\,\psi[A] = -i\,\frac{d}{d\epsilon}\,\psi[A + \epsilon\, D\alpha]\Big|_{\epsilon=0} \tag{8}$$

($A \mapsto A + \epsilon D\alpha$ is (4) of 3-02 with α taken as $g\epsilon\alpha$, that is, an infinitesimal gauge transformation). (b) **It commutes with H.** The two terms of (5) are gauge invariant, so $[H, G^a(\mathbf{x})] = 0$. (c) **It forms a closed algebra.**

$$[G^a(\mathbf{x}), G^b(\mathbf{y})] = -ig\, f^{abc}\, G^c(\mathbf{x})\, \delta^3(\mathbf{x}-\mathbf{y}) \tag{9}$$

$G^a = 0$ cannot be made an operator identity, because the commutator of $G(\alpha)$ and A_j is $-i\,D_j\alpha \;/= 0$. Instead it is imposed on states.

$$G^a(\mathbf{x})\,\psi = 0 \iff \psi[A^\Omega] = \psi[A] \tag{10}$$

The Ω on the right are those continuously connected to the identity transformation. States satisfying (10) are the **physical states**. By (b) time evolution preserves this subspace, and by (c) the conditions do not contradict one another. Observables are restricted to gauge-invariant ones, and the mass gap is the gap of H on the physical states (4-11 § 09).

Example. For a uniform field of $SU(2)$ (a real matrix A_i^a of 3×3; uniform fields are in 3-02 § 07), $G^a = ig\,\varepsilon^{abc}A_i^b\,\partial/\partial A_i^c = -gL^a$. $L^a = -i\varepsilon^{abc}A_i^b\,\partial/\partial A_i^c$ is the angular momentum that turns the colour index, and satisfies $[L^a,L^b] = i\varepsilon^{abc}L^c$. This is the same commutation relation as that of the generators $[T^a,T^b] = if^{abc}T^c$ (3-02 § 02), so $G = -gL$ gives the $-ig$ of (9). Gauge-invariant functions are functions of $A^{\mathsf{T}}A$ and of the sign of $\det A$. With differences we checked that G^a annihilates $\frac{1}{2}|\mathbf{B}|^2$ and $\det A$ to within 2×10^{-10} but does not annihilate $A\frac{1}{1}$, and that (9) holds to relative 2×10^{-11} Computed .

Charges, and large gauge transformations

If there are static colour charges, the condition is $G^a\psi = \rho^a\psi$. The abstract of Kogut–Susskind (1975) describes the gauge-invariant configurations on the lattice as "a collection of strings with quarks at their ends. The strings are lines of non-Abelian electric flux." Known (Table 3). Also, what (8) generates is only transformations connected to the identity. A transformation T that approaches 1 far away and has winding number 1 from $\mathbb{R}^3\cup\{\infty\} = S^3$ to $SU(2)$ (3-05 § 07) is not connected to the identity. So $\psi[A^T] = e^{-i\theta}\psi[A]$ does not contradict (10) (with the convention $(T\psi)[A] = \psi[A^T]$ it has the same direction as (12) of 3-05; with the opposite convention, $e^{+i\theta}$). This is the θ of θ vacua Physics .

05 Seen in the free field — Gauss's law removes the flat directions

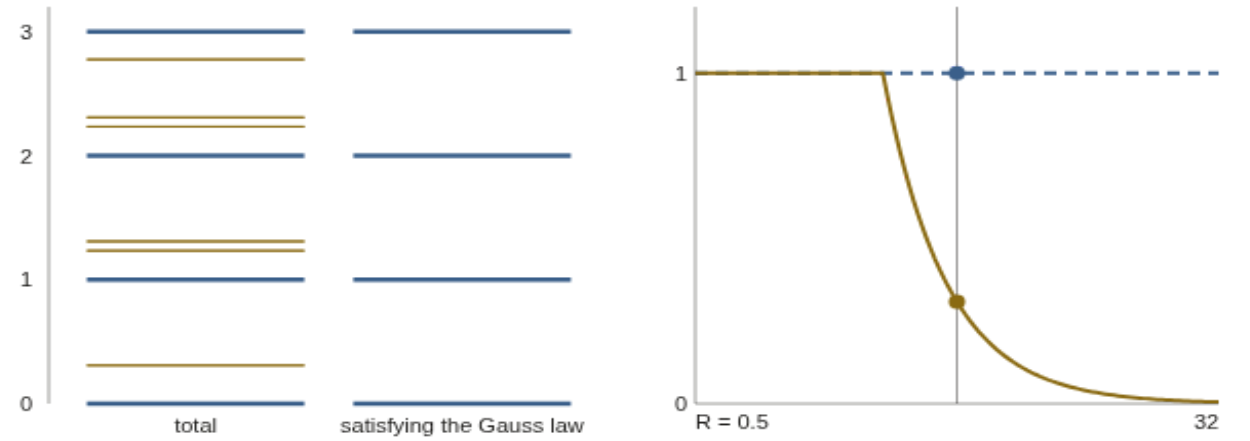
With $g = 0$, split the component of wave number $\mathbf{k}$ into the longitudinal part a_L parallel to $\mathbf{k}$ and the transverse part a_T perpendicular to it. The magnetic field contains only the transverse part, so writing just one transverse direction,

$$H_{\mathbf{k}} = \tfrac{1}{2}(p_T^2 + p_L^2) + \tfrac{1}{2}|\mathbf{k}|^2 a_T^2 \quad (11)$$

and $G_{\mathbf{k}} \propto \mathbf{k}\cdot\mathbf{E}_{\mathbf{k}} \propto p_L$. The longitudinal direction is the direction of the gauge transformation $\mathbf{A}\mapsto\mathbf{A}+\nabla\chi$ and has no restoring force, so in the whole H the energy of longitudinal motion piles up continuously on top of each level. Gauss's law $p_L\psi = 0$ removes this continuum, and the Fock space of the transverse oscillators (photons; 2-06 § 07) remains Known . Figure 1 closes the longitudinal direction into a circle of length $2R$ ($|\mathbf{k}| = 1$) and shows how, of the levels $E(n,m) = n + \frac{1}{2}(\pi m/R)^2$ ($m\in\mathbb{Z}$), $m = 0$ remain.

Figure 1 Changing the length R of the longitudinal direction with the knob changes the levels on the left and the gap on the right. Thick lines are states constant along the longitudinal direction ($m = 0$), thin lines states that

move along it. On the right the solid line is the gap of the whole space, the dashed line the gap of the states satisfying Gauss's law




```
R = 4.00: total gap 0.3084, physical 1, levels with E ≤ 3: 16, physical 4
```

Computed The numbers in this figure were computed on this machine (closed formulas computed on the spot; agreeing with Table 1).

Computed Table 1 The longitudinal length R and the gap ($|\mathbf{k}| = 1$). The number of levels is $E \le 3$ (always 4 for physical ones)

R	Total gap	Physical gap	Number of levels (total)
2	1.00000	1	8
4	0.30843	1	16
32	0.00482	1	120

The total gap $\min(1, \pi^2/2R^2)$ is 0 for $R \to \infty$: the gap of H itself in the temporal gauge is filled by motion in gauge directions. The physical gap $|\mathbf{k}|$ also tends to 0 as the box is enlarged, and the free field has no mass gap (for a field of mass m the gap was m; 2-06 § 08).

06

Yang–Mills on a circle — states are class functions of the holonomy

When space is a circle of length L ($1+1$ dimension) there is no magnetic field, and $H = \frac{1}{2}\int_0^L E^a E^a \, dx$. Gauss's law says that E is carried by parallel transport (1-07), and parallel transport in the adjoint preserves length, so $H = \frac{L}{2} E^a E^a(0)$. All that remains for gauge-invariant functions is the conjugacy class of the holonomy round the circle, $U = P\exp(ig\oint A_1 dx)$. A function unchanged by conjugation $U \mapsto gUg^{-1}$ is called a **class function**, and the trace of a representation matrix a **character** (1-02). Witten (1992) § 4 writes "Ψ must be a class function of U, invariant under conjugation, and so must have an expansion in characters", and expresses H by the Casimir operator Known . $E^a(0)$ acts as g times the derivative that turns U in the direction T^a, and

$$H\,\chi_r(U) = \frac{g^2 L}{2}\,C_2(r)\,\chi_r(U) \tag{12}$$

follows (χ_r is the character of the irreducible representation r, and C_2 the value of $\sum_a T^a T^a$; 1-04). The eigenvalues were found exactly by Rajeev (1988), and the route from the transfer matrix is 7-02 § 07. **Example.** For $SU(2)$, writing the eigenvalues of U as $e^{\pm i\theta}$, $\chi_j = \sin((2j+1)\theta)/\sin\theta$, and the levels are $g^2L/2$ times $j(j+1)$. The gap $\frac{3}{8}g^2L$ can be read as the energy of a loop of electric flux running once round the circle (tension $\frac{3}{8}g^2$).

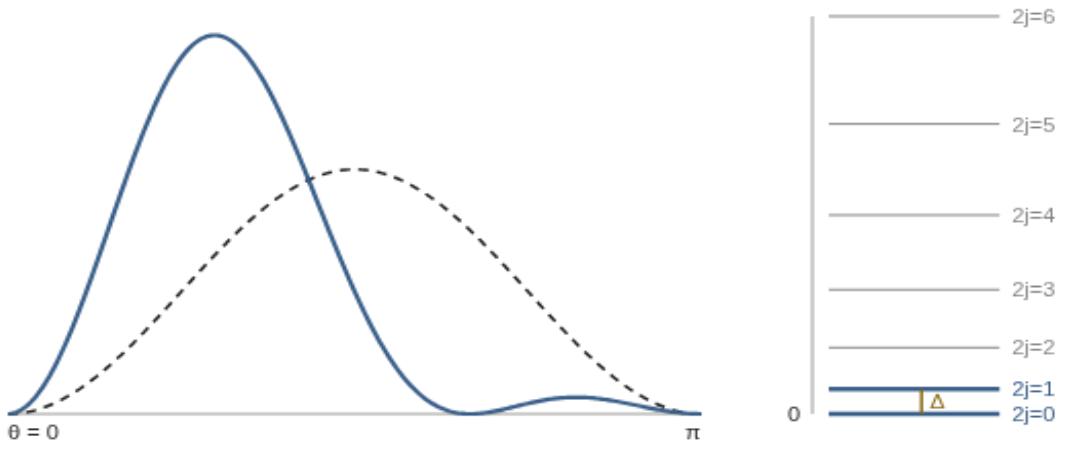
The volume of the orbit fixes the levels

The inner product of class functions is taken with Haar measure $\frac{2}{\pi}\sin^2\theta\,d\theta$. Taking the density $\frac{2}{\pi}\sqrt{1-a_0^2}\,da_0$ (a semicircle) of a_0 in 1-08 § 03 and setting $a_0 = \cos\theta$ gives this form. $\sin^2\theta$ is the area of the conjugacy class (a sphere in S^3), that is, the volume of the gauge orbit (4-02 § 07). Writing (12) with θ,

$$H = \frac{g^2 L}{2}\left(-\frac{1}{4\sin^2\theta}\frac{d}{d\theta}\sin^2\theta\frac{d}{d\theta}\right) \tag{13}$$

and on $u = \sin\theta\, f$, $\frac{g^2L}{8}(-u'' - u)$, that is, a particle in a box on the interval $[0,\pi]$ ($u = \sin n\theta$, $n = 2j+1$) minus 1. If the measure is forgotten and the flat Laplacian is used, the levels change to $k^2/4$ (Table 2). In the three-dimensional Coulomb gauge this volume factor appears as the Faddeev–Popov determinant (§ 08).

Figure 2 A circle of $SU(2)$ ($g = 1$). Choose $2j$ and L and press "Play" to advance time (1 second on screen is time 1); the density (solid line) of θ for the state that superposes χ_0 and χ_j half and half oscillates with angular frequency $\Delta = E_j - E_0$. Making $2j$ or L larger makes the oscillation faster. The dashed line is the Haar weight; on the right, the levels (12)




```
2j = 1, L = 2.0: Δ = 0.750, period 2π/Δ = 8.38, phase Δt = 0.00π
```

Computed The numbers in this figure were computed on this machine (characters and the closed formula of (12) computed on the spot; agreeing with Table 2).

Computed Table 2 Eigenvalues of (13) (in units of $g^2L/2$; differences with 4000 Haar-weighted points), and the $k=2j$-th from the bottom of $-\frac{1}{4}d^2/d\theta^2$ with the measure forgotten (slope 0 at both ends of $[0,\pi]$; a periodic function symmetric under $\theta\mapsto-\theta$)

2j	j(j+1)	Differences (Haar)	k²/4 with the measure forgotten
0	0	0.000000	0
1	0.75	0.750000	0.25
2	2	2.000000	1
3	3.75	3.749998	2.25

Counting on the lattice — states that violate Gauss's law

Cut the circle into N links ($a=L/N$) and set $H=\frac{g^2a}{2}\sum_\ell E_\ell^2$ (the two-dimensional version of 5-03 § 07; there are no plaquettes). As a metaphor, one link is a top spinning on $SU(2)$ (5-03 § 06), with $(2j+1)\times(2j+1)$ states of angular momentum j for left and right rotations. Precisely, this is the Peter–Weyl theorem, which splits the state space $L^2(SU(2))$ of a link by representation Known . So the energy is $\frac{g^2a}{2}\sum_\ell j\ell(j\ell+1)$ and the multiplicity $\prod_\ell(2j\ell+1)^2$. At a vertex the ends of two links meet, and Gauss's law requires that the combination of their two representations be unchanged by rotation (be a scalar). 0 appears in the coupling of angular momenta $j\otimes j'=|j-j'|\oplus\cdots\oplus(j+j')$ only when $j=j'$, and then exactly once (Schur's lemma) Known . So neighbouring j are equal, there is one physical state for each j, and the levels are (12) regardless of N.

Computed Table 3 A lattice circle of $SU(2)$ (in units of $g^2L/2$). The lowest physical excitation is always 0.75, and there are 3 physical states with $E\le 2$

N	Lowest excitation (total)	Number of states with $E\le 2$ (total)
1	0.75	14
4	0.1875	115,504
16	0.046875	2.44×10²²

The lowest excitation of the whole space is a single link with $j=\frac{1}{2}$, that is, a state with a pair of charges at its two ends (§ 04), and its energy $\frac{3}{8}g^2a$ is the tension times the distance a. It tends to 0 as $a\to0$, so $\frac{3}{8}g^2L$ remains only once we restrict to physical states.

07

Detour: small boxes, and arguments from wave functionals

A section you may skip.

In a three-dimensional periodic box (side ℓ) the uniform components $A_i^a = c_i^a$ are special. Restricting (5) to these components gives $H_0 = -\frac{1}{2\ell^3}\partial_c^2 + \frac{\ell^3}{2}|\mathbf{B}(c)|^2$ (the magnetic field is only the commutator term $g\,c^2$; 3-02 § 07), and rescaling as $c = g^{-1/3}\ell^{-1}y$,

$$H_0 = \frac{g^{2/3}}{\ell}\Big[-\tfrac{1}{2}\frac{\partial^2}{\partial y^2} + \tfrac{1}{2}|\mathbf{B}_1(y)|^2\Big] \quad (14)$$

($\mathbf{B}_1$ is the magnetic field of $g = 1$). The levels are proportional to $g^{2/3}/\ell$, and if g is small they lie below the $2\pi/\ell$ of the non-uniform components. The abstract of Lüscher (1983) writes "When non-abelian gauge fields are enclosed in a box with periodic boundary conditions, the spectrum of the hamiltonian becomes discrete", and expands the levels in powers of $g^{2/3}$ Known . Keeping only two components inside the bracket gives the x^2y^2 well of 2-05. The small box (the side where g is small, by asymptotic freedom) is the opposite end from the four-dimensional question, which asks about infinite volume. Lüscher's g is the renormalised coupling measured at the size of the box, not the classical g of (14) itself.

Feynman (1981), in $2+1$ dimensions, after cautioning in the abstract that "Only qualitative or suggestive discussions are made.", argues that the gauge invariance of the wave functional leads to "a finite energy for any excitation ("glueball") above the ground state" Physics .

08

The route of solving Gauss's law outright — the Coulomb gauge and Gribov

There is also the route of fixing the gauge further and solving the constraint first. Splitting as $\mathbf{E} = \mathbf{E}_T - \nabla\phi$ ($\nabla\cdot\mathbf{E}_T = 0$) in the Coulomb gauge $\nabla\cdot\mathbf{A} = 0$, Gauss's law is a linear equation for ϕ.

$$M\phi = -\rho, \qquad M = -\nabla\cdot\mathbf{D} \quad (15)$$

$\rho^a = g f^{abc} A_i^b E_{T,i}^c$ is the colour charge carried by the transverse field itself. If M can be inverted, then with $K = M^{-1}(-\nabla^2)M^{-1}$,

$$H = \tfrac{1}{2}\int(|\mathbf{E}_T|^2 + |\mathbf{B}|^2) + \tfrac{1}{2}\int \rho\, K\rho \quad (16)$$

(the integral is d^3x) Known . The abstract of Christ–Lee (1980) writes that the change of variables from the temporal gauge produces "an unconventional operator ordering", and that "new nonlocal interactions, even in the familiar Coulomb gauge" appear Known . The volume factor of the orbit corresponding to the $\sin^2\theta$ of § 06 is $\det M$ Known . This is the Jacobian of changing variables to the gauge condition (coordinates on the slice $\nabla\cdot\mathbf{A} = 0$), and is called the Faddeev–Popov determinant (details in 4-08). The difficulty is that when A is large, M has zero eigenvalues. The abstract of Gribov (1978) writes "the fixing of the divergence of the potential in non-Abelian theories does

not fix its gauge” Known . That a slice cannot be chosen globally is Singer's theorem in 3-04. The (10) of the temporal gauge avoids this difficulty.

09

What is formal — the problem of defining the Hamiltonian

In four dimensions (5) lacks meaning in three respects. (i) $E_i^a(\mathbf{x})E_i^a(\mathbf{x})$ is a product of distributions at the same point (1-11 § 09, 1-09 § 09). (ii) The sum of zero-point energies $\frac{1}{2}|\mathbf{k}|$ diverges (2-06 § 09), and with interactions this is paired with renormalisation (4-04). (iii) In the inner product $\int |\psi|^2 \mathcal{D}A$, the absence of Lebesgue measure in infinite dimensions (1-09 § 02) overlaps with the divergence of the orbit volume (the fact that an interacting H cannot be placed on the Fock space of the free field, that is, Haag's theorem, is also involved; 4-09).

On the lattice the state space is fixed, and Gauss's law can be imposed exactly at each vertex (Table 3). What remains is whether the gap stays positive in the limit where the lattice spacing goes to 0 and the box to infinity. The problem statement writes “Establishing a quantum mechanical Hilbert space is part of the solution to this Millennium problem.” (§ 6.5) Known . § 6.1 states that the Osterwalder–Schrader construction links the route of Hilbert spaces and the route of functional integrals. § 6.5 writes that much of the mathematical progress so far has come from understanding functional integration (“Much of the mathematical progress reviewed above results from understanding functional integration”). On this route H is reconstructed from the measure in imaginary time (4-02 § 08, 4-10 § 09). This chapter shows nothing about four-dimensional Yang–Mills.

What this chapter can say

Matter	What is said	What is not said
(1)–(10)	Formal derivation Known . Check on uniform fields Computed	Meaning as operators for fields in four dimensions. θ is at the level of physics
(11)–(13), figures, tables	The free-field model and the levels on a circle Known , differences and counting Computed	The case of space of two or more dimensions
(14)–(16), detour	Counting of scales and classical rewriting Known	Terms from quantum ordering (text not checked). Feynman's argument is qualitative Physics

Matter	What is said	What is not said
The mass gap	—	In four dimensions, H was built by reconstruction in the construction published in 2026-11 (at the level of the committee's assessment; 12-08). Its gap has not been shown (6-04). This chapter shows nothing

Sources and reproduction

Item	Kind	Source / tools
The problem statement (§ 4, § 6.1, § 6.5)	Text	A. Jaffe, E. Witten, *Quantum Yang–Mills Theory*, Clay Math. Inst. (2000)
Class functions and the Casimir on a circle	Text § 4, § 4.2	E. Witten, J. Geom. Phys. 9 (1992) 303–368, arXiv:hep-th/9204083
The ordering terms of § 08 and $\det M$	Abstract (text bibliographic data only; the $\det M$ term not checked)	N. H. Christ, T. D. Lee, Phys. Rev. D 22 (1980) 939
Quotations in § 04 and § 06–§ 08	Abstract	J. Kogut, L. Susskind, Phys. Rev. D 11 (1975) 395 / S. G. Rajeev, Phys. Lett. B 212 (1988) 203 / M. Lüscher, Nucl. Phys. B 219 (1983) 233 / R. P. Feynman, Nucl. Phys. B 188 (1981) 479 / V. N. Gribov, Nucl. Phys. B 139 (1978) 1
General theory of constraints	Bibliographic data only	P. A. M. Dirac, Can. J. Math. 2 (1950) 129–148 (attribution not checked)
(2)–(10), (15)(16)	Known	Standard derivation. Signs recomputed in the conventions of 3-03
The check in § 04, Tables 1–3, figures	Computed	Python 3.12, NumPy, SciPy (`gauss_uniform.py`, `flat_mode.py`, `holonomy.py`, `ring_counts.py`). The figures are plain JavaScript

Next chapter to read: 4-02 The path integral and Euclideanisation. The lattice Hamiltonian is in 5-03, gauge fixing in 4-08.

The path integral and Euclideanisation — make time imaginary, and field theory becomes a problem of probability in four dimensions

> This chapter organises established knowledge (sources carry the “Known” label in the text). This book's own investigation is mainly in Part 15 and Volume 3 (Parts 16–28 and the appendices).

What this chapter shows — That the path integral of a field is a sum over all field configurations.
That making time imaginary turns the weight e^{iS} into e^{-S_E}, and for Yang–Mills, since $S_E \geq 0$, it can be read as a probability weight.
That Lorentz transformations become four-dimensional rotations, and time becomes one of four equal directions.

Prerequisite chapters — 2-07 (the path integral, imaginary time), 2-04 (Lorentz transformations, the action of a field). § 04 uses the non-abelian field strength of 3-02 and 3-03. For the Hamiltonian, 4-01; for Gaussian measures, 1-09.

Said first — Many of the manipulations in this chapter are formal. Wick rotation is a heuristic, and the rigorous correspondence is the Osterwalder–Schrader reconstruction (4-10). The measure of four-dimensional Yang–Mills has not been constructed, and this chapter shows nothing about the mass gap.

Order of this chapter

1. From paths to field configurations — on the lattice, an integral tens of millions of times multiple
2. Making time imaginary — iS becomes $-S_E$
3. The rotation seen in the propagator — turning while avoiding the poles (Figure 1, Table 1)
4. Rotating the Yang–Mills field — $S_E \geq 0$, and the θ term stays imaginary
5. Spacetime becomes space — boosts become rotations (Table 2)
6. Field theory as a probability measure — on the lattice, a genuine measure (Figure 2)

7. What is missing for gauge fields — the volume of the orbits (Table 3)
8. Correlations in imaginary time read the energy — the link to the gap
9. Detour: writing the vacuum wave function on half of spacetime — checked with the oscillator

---

01

From paths to field configurations — what is summed becomes a function

In 2-07 we summed e^{iS} over the paths $x(t)$ of one particle. In field theory, instead of paths, we sum over all functions on spacetime $\varphi(t,\underline{x})$, all histories of the field. The correlation in the vacuum Ω (T is the convention of putting later times to the left) is formally written

$$\langle\Omega|\,\mathrm{T}\,\varphi(x_1)\cdots\varphi(x_n)\,|\Omega\rangle=\frac{\int\mathcal{D}\varphi\;\varphi(x_1)\cdots\varphi(x_n)\,e^{iS[\varphi]}}{\int\mathcal{D}\varphi\;e^{iS[\varphi]}}\qquad(1)$$

Even for one particle $e^{iS}\mathcal{D}x$ was not a measure (2-07 § 03). For fields there are also infinitely many variables at each time (1-09).

Example. Cut into a lattice, (1) is a finite multiple integral. A lattice of space 32^3 and time 64 has 2,097,152 points, and the Yang–Mills field of $SU(3)$ has 4 links per point and 8 real numbers per link, so it is a 67,108,864-fold integral Computed . The difficulties are two: the oscillation of e^{iS}, and the limit of finer steps and a larger box. This chapter removes the first (the second is Parts 5 and 7).

02

Making time imaginary — iS becomes $-S_E$

The substitution $t=-i\tau$ has appeared twice. In 2-04 § 09 we only substituted it into the wave equation and saw oscillation turn into decay. In 2-07 § 05 we rotated the weight of the one-particle path integral. Here we rotate the action and correlations of fields. Putting $t=-i\tau$ ($\hbar=c=1$, signs $(-,+,+,+)$, $x_4=\tau$) into the scalar field action ((7) of 2-04 with the potential energy V added), from $dt=-i\,d\tau$ and $(\partial_t\varphi)^2=-(\partial_\tau\varphi)^2$,

$$iS_M=-S_E,\qquad S_E[\varphi]=\int d^4x\Big[\tfrac12\sum_{\mu=1}^{4}(\partial_\mu\varphi)^2+\tfrac12 m^2\varphi^2+V(\varphi)\Big]\qquad(2)$$

follows. The sign of the time derivative has come into line with the space derivatives, and the four directions are equal. S_E is called the Euclidean action, and if V is

bounded below, the weight e^{-S_E} is positive and bounded. (1) is replaced by the **Schwinger functions** (the name is said to come from Schwinger 1958).

$$S_n(x_1,\dots,x_n)=\frac{1}{Z}\int \mathcal{D}\varphi\ \varphi(x_1)\cdots\varphi(x_n)\,e^{-S_E[\varphi]},\qquad Z=\int \mathcal{D}\varphi\ e^{-S_E[\varphi]} \tag{3}$$

Example. If $V=0$, (3) is a Gaussian integral with covariance $(-\Delta+m^2)^{-1}$ (Δ is the Laplacian $\sum_\mu \partial_\mu^2$; not the Δ of the mass gap), and $S_2(x,y)=G(x-y)$ is the propagator of (10) of 1-11. Note. (1) is not a convergent integral, so putting in $t=-i\tau$ is not a change of variables but a rewriting rule. The precise content is the analytic continuation of correlation functions; seen in a single integral, it takes the form of the next section.

03

The rotation seen in the propagator — turning the contour while avoiding the poles

The two-point function of a free field of mass m, written in momentum, is

$$\langle\Omega|\,\mathrm{T}\,\varphi(x)\varphi(0)\,|\Omega\rangle=\int\frac{d^4p}{(2\pi)^4}\,e^{ip\cdot x}\,\frac{-i}{p^2+m^2-i\varepsilon},\qquad p^2=-(p^0)^2+|\boldsymbol{p}|^2 \tag{4}$$

$\varepsilon\to+0$ fixes on which side the poles are avoided (Feynman's prescription). We first look at the component with time difference $t=0$. Then in $e^{ip\cdot x}$ the factor e^{-ip^0t} from p^0 is 1. Setting $\omega=\sqrt{|\boldsymbol{p}|^2+m^2}$, the integrand in p^0 is $i/((p^0)^2-\omega^2+i\varepsilon)$, and the poles $\pm\sqrt{\omega^2-i\varepsilon}$ are in the fourth and second quadrants.

Here we use just one fact about complex integrals. A **pole** is a point where the denominator is 0. In the complex plane, integrating along a closed path that does not enclose a pole gives 0 (**Cauchy's integral theorem**; precisely, the condition is that the function be complex-differentiable on and inside the path, which a rational function whose denominator does not vanish satisfies) Known . So moving the contour continuously without crossing a pole does not change the value.

Turn the contour anticlockwise as $p^0=e^{i\theta}s$ (s real). While θ is between 0 and 90°, the moving line sweeps only the first and third quadrants and does not cross a pole. Joining the two lines before and after turning by distant arcs of radius R makes a closed path that encloses no pole. The length of the arcs is proportional to R and the integrand decreases as $1/R^2$, so the contribution of the arcs vanishes as $1/R$. So the integral along the real axis equals the integral along the turned line, $\theta=90^\circ$, that is, at $p^0=ip_4$,

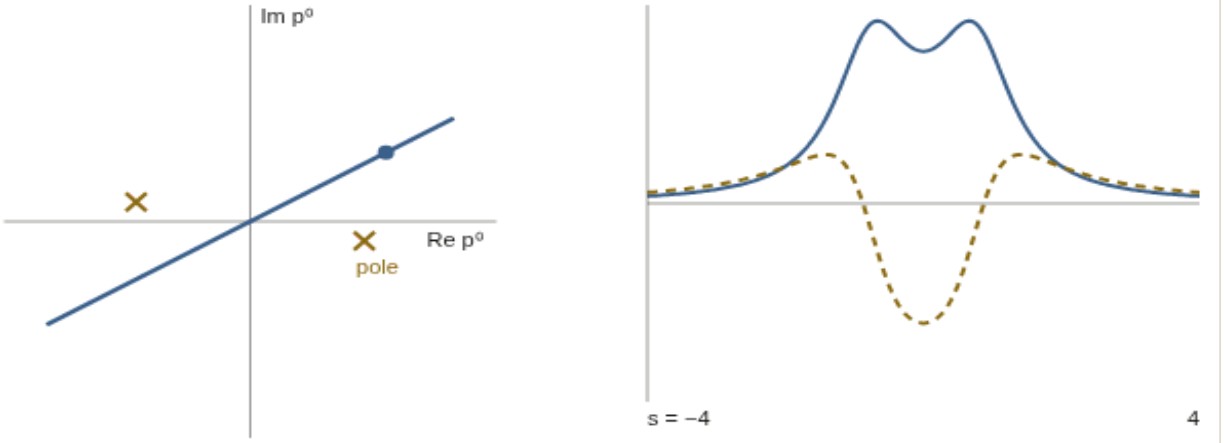
$$\int_{-\infty}^{\infty} \frac{dp^0}{2\pi}\, \frac{i}{(p^0)^2 - \omega^2 + i\varepsilon} = \int_{-\infty}^{\infty} \frac{dp4}{2\pi}\, \frac{1}{p_4^2 + \omega^2} = \frac{1}{2\omega} \qquad (\varepsilon \to 0)$$ (5)

The integrand on the right-hand side is positive and smooth. In four dimensions $-i\,d^4p/(p^2 + m^2 - i\varepsilon)$ is replaced by $d^4p_E/(p_E^2 + m^2)$, and this is the Euclidean propagator [Known] . $1/(2\omega)$ is the $\langle x^2 \rangle$ of the ground state of an oscillator of frequency ω (2-06), and each Fourier component of a free field was an oscillator (2-06 § 08).

The $t \;/= 0$ component carries the factor $e^{-ip^0 t}$. If $p^0 = e^{i\theta} s$ is turned into the first quadrant, on the side of $st > 0$, $|e^{-ip^0 t}| = e^{st \sin\theta}$ grows far away, so the contribution of the arcs does not vanish. So t is also turned by the same angle in the opposite direction, as $t = e^{-i\theta}\tau$. Then at every angle along the way $p^0 t = s\tau$ is real, the factor stays an oscillation of modulus 1, and the argument with arcs goes through. At $\theta = 90°$, $t = -i\tau$, $p^0 = ip4$, $p^0 t = p4\tau$. This is why time turns clockwise and energy anticlockwise.

Turning the other way crosses a pole and changes the sign (Figure 1, Table 1), so the direction is fixed by $i\varepsilon$. The name is said to come from Wick's (1954) treatment of the Bethe–Salpeter equation.

Figure 1 Changing the angle θ changes the contour in the p^0 plane (thick line in the contour plot) and the integrand on it (in the integrand plot, solid line real part, dashed line imaginary part). As long as no pole (×) is crossed, the integral $I(\theta)$ does not change ($\omega = 1$)



θ = 30°, ε = 0.40: I(θ) = 0.4731 + 0.0911i, 1/(2√(ω²−iε)) = 0.4731 + 0.0911i, (above the pole angle -10.9°)

[Computed] The numbers in this figure were computed on this machine (trapezoidal rule with $s = \tan u$).

[Computed] Table 1 $I(\theta)$ of Figure 1 (python3). The angle of the pole for $\varepsilon = 0.4$ is $-10.90°$.

θ	ε	I(θ)
90°, 30°, 0°, −10°	0.4	+0.473094 + 0.091110i
−12°, −45°, −90°	0.4	−0.473094 − 0.091110i
0°	0.001	0.500000 + 0.000250i

The value is the same whatever the angle; turning the other way past the pole only reverses the sign.

04

Rotating the Yang–Mills field — $E^2 - B^2$ becomes $E^2 + B^2$

The notation follows the $\hat{A} = gA$ of 3-03 § 01, written without hats. That is, g is absorbed into A, A_μ are Hermitian matrices, $D_\mu = \partial_\mu - iA_\mu$, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu]$ (the same as (9) of 2-04), $\operatorname{tr}(T^a T^b) = \frac{1}{2}\delta^{ab}$.

The gauge field is a 1-form (1-07), so it is rotated as $A_0\, dt = A_4\, d\tau$, that is, $A_0 = iA_4$. Together with $\partial_0 = i\partial_4$,

$$F_{0j} = i\partial_4 A_j - i\partial_j A_4 - i[iA_4, A_j] = i\,F_{4j}, \qquad F_{\mu\nu}F^{\mu\nu} = -2F_{0j}F_{0j} + F_{jk}F_{jk} = \sum_{\mu,\nu=1}^{4} F$$

In words, only the components with one time index (the components of the electric field) get one i, which squares to -1. In the Minkowski metric the square of the electric field carried -1, so the two cancel, and all four indices line up with the same sign. This is the same thing as the sign of $(\partial_t \varphi)^2$ coming into line with the space derivatives in (2) for the scalar field. Putting this into $S_M = -\frac{1}{2g^2}\int dt\, d^3x \operatorname{tr} F_{\mu\nu}F^{\mu\nu}$ ((1) of 3-03 written in hatted notation) leaves $i \cdot (-i) = 1$, and with $iS_M = -S_E$,

$$S_E[A] = \frac{1}{2g^2}\int d^4x \operatorname{tr} F_{\mu\nu}F_{\mu\nu} = \frac{1}{g^2}\int d^4x \operatorname{tr}\left(E_E^2 + B^2\right) \;\geq\; 0, \qquad E_{E,j} = F_{4j},\; B_j = \frac{1}{2}\varepsilon$$

For Hermitian X, $\operatorname{tr} X^2 = \sum |X_{ij}|^2 \geq 0$, so $S_E \geq 0$, with equality only when $F = 0$. For configurations of topological charge Q, $S_E \geq 8\pi^2|Q|/g^2$ ((12) of 3-03, 3-05) Known .

Example. For configurations with $A_4 = 0$ and A_j independent of τ, $E_E = 0$, so over imaginary time of length T, S_E is $T\mathcal{E}$ ($\mathcal{E} = \frac{1}{g^2}\int \operatorname{tr} B^2 d^3x$ is the magnetic energy), and the weight is $e^{-T\mathcal{E}}$.

The θ term stays imaginary

The topological term θQ is different. Each component contains exactly one F with a time index, so the i from the rotation and the $-i$ of dt cancel, and the weight is $e^{-S_E} e^{i\theta Q}$ (the sign depends on the orientation convention), which for $\theta \neq 0$ is complex. Vicari–Panagopoulos (2009) write 'the complex nature of the θ term in the Euclidean QCD Lagrangian prohibits a direct MC simulation at θ ≠ 0' Known . From here on $\theta = 0$.

05

Spacetime becomes space — boosts become rotations

Putting $t = -i\tau$ and an imaginary rapidity $\phi = i\alpha$ into the boost $t' = t\cosh\phi - x\sinh\phi$, $x' = x\cosh\phi - t\sinh\phi$ of 2-04, from $\cosh i\alpha = \cos\alpha$ and $\sinh i\alpha = i\sin\alpha$,

$$\tau' = \tau\cos\alpha + x\sin\alpha, \qquad x' = x\cos\alpha - \tau\sin\alpha \ \text{(8)}$$

which is a rotation of the (τ, x) plane. The Lorentz group $SO(3,1)$ is replaced by the rotation group $SO(4)$ Known . The interval becomes $\tau^2 + |\underline{x}|^2 \geq 0$, and the light cone and the distinction between past and future disappear. Which direction to call time is chosen afterwards. If reflection positivity (4-10) also holds, then by rotation invariance the same theory comes back whichever direction is chosen as time. The method of carrying each into the other is an achievement of axiomatic field theory Known (§ 3 of the exposition by Jaffe–Witten [2000, 0-03], the official problem statement of the Clay prize problem; hereafter "the exposition").

The lattice reduces rotations to 90°

The rotations left on the lattice are only steps of 90° and reflections, and that $SO(4)$ comes back for $a \to 0$ also has to be shown. For the free field on a four-dimensional lattice ($m = 1$, $a = 1/2n$), we compare the correlations at distance-1 points $(2n, 0, 0, 0)$ and (n, n, n, n). The continuum correlation is $G(r) = \dfrac{m\,K_1(mr)}{4\pi^2 r}$. K_1 is a special function called the modified Bessel function, which decays as e^{-mr} far away (near the origin it behaves like $1/(mr)$, $G \approx 1/(4\pi^2 r^2)$).

Computed Table 2 Ratio of the lattice two-point function to the continuum $G(1) = K_1(1)/4\pi^2$. The lattice side is computed from a single-integral representation using the family of modified Bessel functions I_ν.

a	1/2	1/4	1/8	1/16	1/32
Axis	1.4329	1.1297	1.0247	1.0058	1.0014
Diagonal	0.8126	0.9475	0.9869	0.9967	0.9992

On a coarse lattice the correlation differs by a factor of about 1.8 depending on direction, and the difference vanishes in proportion to a^2. Wilson's lattice (5-01) keeps reflection positivity (4-10) exactly at the price of breaking rotations. The exposition writes that if a regularisation loses reflection positivity, there are few ways to recover it (Jaffe–Witten § 6.5).

```
06
```

Field theory as a probability measure — reading e^{-S_E} as a probability

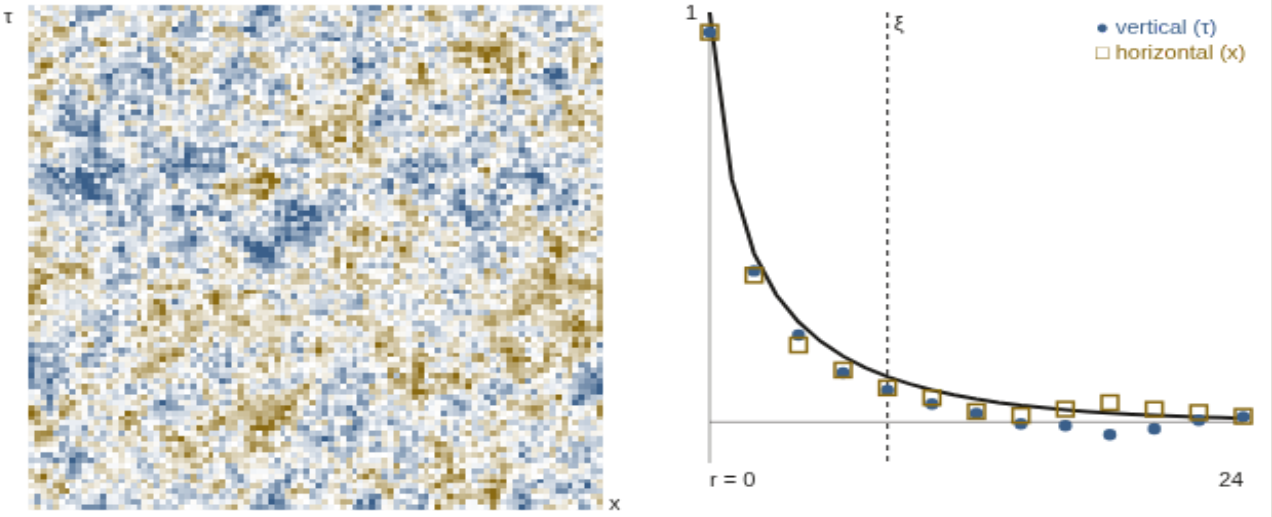
If S_E is bounded below, e^{-S_E}/Z can be read like a probability density on field configurations, and (3) are its moments. Field theory becomes a theory of random functions on four-dimensional space. The exposition attributes this view to Symanzik (Jaffe–Witten § 3) Known .

Precisely, there is no Lebesgue measure $\mathcal{D}\varphi$ in infinite dimensions (1-09), so it is correct as it stands only in finite dimensions. On a periodic lattice $\Lambda_{a,L}$ with spacing a and side L,

$$d\mu_{a,L}(\varphi) = \frac{1}{Z_{a,L}}\, e^{-S_E^{(a)}(\varphi)} \prod_{x\in\Lambda_{a,L}} d\varphi(x) \qquad (9)$$

is a genuine probability measure ($S_E^{(a)}$ is (2) with derivatives replaced by differences and integrals by sums). Making a field theory means showing the limits of the moments as $a \to 0$, $L \to \infty$ and their properties (7-01). The limit of the free field is a Gaussian measure on a space of distributions, and the exposition writes that nonlinear fields need non-Gaussian measures ('one requires non-Gaussian dμ to obtain nonlinear fields') (Jaffe–Witten § 6.2, 1-09).

Figure 2 Changing the correlation length ξ changes a sample of the free field on a two-dimensional periodic lattice (96×96) (left; colour is $\varphi/\sqrt{G(0)}$) and its vertical and horizontal correlations (right, ● and □; the solid line is the exact $G(r)/G(0)$). It is a single sample, so the correlations at points fluctuate ("Redraw" gives another sample; the average over 200 samples is in the text)



ξ = 8 : G(0) = 0.6056、G(8)/G(0) = 0.111 (exact), this sample: vertical 0.078, horizontal 0.083 (a single sample fluctuates; the 200-sample average is in the text)

Computed The numbers in this figure were computed on this machine (random numbers with a fixed seed).

Looking at the samples, one cannot tell which axis is time. In 200 samples made the same way (python3), with $\xi = 8$ (the mass coefficient of the action $M = 2\sinh\frac{1}{2\xi}$,

chosen so that the correlation decays exactly as $e^{-r/\xi}$), the correlation at distance 8 is 0.1120 ± 0.0032 vertically and 0.1112 ± 0.0030 horizontally, against 0.1107 exactly, and for $\xi = 2, 4, 16$ too they agree within twice the standard error Computed . That $G(0)$ grows logarithmically with ξ is a manifestation of the divergence of point values in two dimensions (2-07 § 09).

07

What is missing for gauge fields — the volume of the orbits and compact groups

$S_E[A]$ does not change under gauge transformations (3-02), so e^{-SE} is constant on gauge orbits (the collections of connections carried into one another by gauge transformations, 3-04). An orbit is infinite-dimensional, with one group attached at each point, and the integral in its direction is infinite even formally.

(A) The weight $e^{-(x-y)^2/2}$ does not change under the shift $(x, y) \to (x + c, y + c)$, and the orbits are infinitely long lines. The integral Z_R over the square $[-R, R]^2$ grows in proportion to R, but the average of $(x - y)^2$, which is unchanged by the shift, tends to 1. (B) The orbits of the weight $e^{\beta \cos(\theta_1 - \theta_2)}$ are circles of length 2π, and integrating with the uniform measure on the circle (Haar measure), $Z = (2\pi)^2 I_0(\beta)$ is finite for every β (checked by double integrals and Bessel functions) Computed . Here I_0 is the $\nu = 0$ of the I_ν of Table 2, and

$$I_0(\beta) = \frac{1}{2\pi} \int_0^{2\pi} e^{\beta \cos\theta}\, d\theta$$

Setting $\theta = \theta_1 - \theta_2$, the remaining integral over θ_2 just produces one factor of the orbit length 2π.

Computed Table 3 Toy (A). Z_R agrees with the closed formula to 10 digits.

R	$Z_R/2R$	$\langle (x-y)^2 \rangle_R$
1	1.528	0.4341
10	2.407	0.9584
100	2.497	0.9960
1000	2.506	0.9996

In (A) the denominator and numerator diverge at the same rate, and only the ratio remains ($Z_R/2R \to \sqrt{2\pi} = 2.5066$). Wilson's lattice (5-01) has the form of (B). On a finite lattice the link variables $U \in G$ are integrated with the Haar measure of a compact group, so the integral is finite without fixing the gauge ('the lattice approximation reduces the functional integration to a finite-dimensional integral', Jaffe–Witten § 6.5)

Known . Treating it in the continuum needs gauge fixing (4-08), and there is the obstruction that representatives cannot be chosen globally and continuously (3-04). Either way, only gauge-invariant observables (Wilson loops, $\operatorname{tr} F^2$ and so on) have meaning.

08

Correlations in imaginary time read the energy — the link to the mass gap

If there are a Hamiltonian H (with vacuum energy 0) and a vacuum Ω, the correlation of observables O separated by imaginary time τ, expanded in eigenstates $|n\rangle$, is

$$\langle O(\tau)O(0)\rangle = \langle\Omega|\, O\, e^{-\tau H}\, O\,|\Omega\rangle = \sum_n |\langle n|O|\Omega\rangle|^2\, e^{-E_n\tau} \quad (10)$$

(an integral if the spectrum is continuous). If the vacuum is unique (the only state of energy 0 is Ω), $\langle\Omega|O|\Omega\rangle = 0$, and the spectrum of H has nothing in $(0,\Delta)$, the correlation decays at least as $e^{-\Delta\tau}$. **Example.** Integrating the two-point function of the free field over space leaves only one particle with momentum 0, and the decay rate is exactly the mass Known .

$$\int d^3x\; S_2\big((\tau,\vec{x}),0\big) = \int \frac{dp_4}{2\pi}\,\frac{e^{ip_4\tau}}{p_4^2+m^2} = \frac{e^{-m|\tau|}}{2m} \quad (11)$$

This is how the gap shows on the Euclidean side; the definition is set out in 4-11. On the lattice the weight joining layers at neighbouring times is the transfer matrix $T = e^{-aH}$ (2-07 § 08), and in gauge theory, in the limit where the step in the time direction goes to 0, the Kogut–Susskind Hamiltonian comes out of the gauge transfer matrix of $A_0 = 0$ (abstract of Creutz 1977) Known . The lattice Hamiltonian is treated in 5-03.

Mind the direction

(10) is a formula for when H and the Hilbert space are there first. Building H from the Euclidean measure alone needs positivity with respect to the reflection $\tau \to -\tau$ (reflection positivity) (4-10). The exposition writes 'Establishing a quantum mechanical Hilbert space is part of the solution to this Millennium problem' (Jaffe–Witten § 6.5). "Gap $\Longleftrightarrow$ exponential decay of correlations" is a correspondence after reconstruction. For four-dimensional Yang–Mills, that Hilbert space was built by the construction published in 2026-11 (the committee's assessment is that it is proved in Lean; 12-08). That correlations decay exponentially there has not been shown (6-04).

09

Detour: writing the vacuum wave function as a path integral over half of spacetime

A section you may skip.

The $e^{-\tau H}$ of (10) has another use. Expanding a state in eigenstates of H, $e^{-\tau H}$ multiplies the E_n component by $e^{-E_n \tau}$. Over a long imaginary time, every component other than the lowest-energy one dies away relatively faster. That is, $e^{-\tau H}$ is a filter that extracts only the vacuum from any state that has a vacuum component (2-07 § 05). Precisely, the conditions are that the initial state has a vacuum component and that the vacuum is unique.

Writing this as a path integral, the vacuum wave function itself becomes a path integral. We look at one particle. $\langle x | e^{-TH} | x_i \rangle$ is the sum of e^{-S_E} over all paths that leave x_i at imaginary time $\tau = -T$ and arrive at x at $\tau = 0$. For $T \to \infty$, $\psi_0(x)\,\psi_0(x_i)\,e^{-E_0 T}$ dominates, and as a function of x only the vacuum wave function $\psi_0(x)$ remains.

$$\psi_0(x) \;\propto\; \int_{x(0)=x} \mathcal{D}x\; e^{-S_E[x]}, \qquad S_E[x] = \int_{-\infty}^{0} \left[\tfrac{1}{2}\dot{x}^2 + \tfrac{1}{2}\omega^2 x^2\right] d\tau \quad (12)$$

What is summed is the paths on the half-line of imaginary time $\tau \le 0$ with only the value at the end $\tau = 0$ fixed.

Example. For the oscillator, the exponent in (12) is quadratic in $x(\cdot)$, so the dependence of the answer on x is fixed by the path of least action (the contribution of fluctuations around it is a constant independent of x). The path that satisfies $\ddot{x} = \omega^2 x$, is x at $\tau = 0$ and is 0 at $\tau \to -\infty$ is $x(\tau) = x\, e^{\omega\tau}$. Its action is

$$S_E = \int_{-\infty}^{0} \omega^2 x^2 e^{2\omega\tau}\, d\tau = \frac{\omega x^2}{2}, \qquad \psi_0(x) \propto e^{-\omega x^2/2}$$

which is exactly the ground state of 2-06 Known .

It can also be checked on a lattice. On an imaginary-time lattice with spacing a, fix $x_0 = x$ and integrate $x_{-1}, x_{-2}, \ldots$ in turn. Each integration preserves the form $e^{-c\,x^2/2}$, and the coefficient c moves by a recurrence and settles to one value. Counting the potential energy at the end point x_0 by the convention of splitting it half and half with the neighbouring half of spacetime, the value it settles to is $c(a) = \omega\sqrt{1 + a^2\omega^2/4}$. At $\omega = 1$, for $a = 1/2$ it is 1.0308 and for $a = 1/10$ 1.00125, and the difference from the continuum $\omega = 1$ vanishes in proportion to a^2 Computed .

For fields

For fields the form is the same. The vacuum wave functional $\Psi_0[\varphi_0]$ (the wave functional of 4-01 § 03) is the sum over all field configurations on the half of spacetime $\tau < 0$ with $\varphi = \varphi_0$ at $\tau = 0$. For the free field each Fourier component is an oscillator of frequency $\omega_p = \sqrt{|p|^2 + m^2}$ (2-06 § 08), so multiplying the oscillator answers component by component,

$$\Psi_0[\varphi_0] \;\propto\; \exp\Big(-\frac{1}{2}\int \frac{d^3p}{(2\pi)^3}\,\omega_p\,|\tilde{\varphi}_0(p)|^2\Big), \qquad \tilde{\varphi}_0(p) = \int d^3x\, e^{-ip\cdot x}\varphi_0(x)$$

Known . The Osterwalder–Schrader reconstruction (4-10) follows this view in reverse. It builds states from quantities fixed by the fields on the half of spacetime $\tau > 0$ alone, and defines the inner product with the reflection $\tau \to -\tau$. What guarantees that the inner product is positive is reflection positivity. Precisely, (12) is a formula for when the Hilbert space and H are there first. For four-dimensional Yang–Mills, the Hilbert space and the Hamiltonian were built by reconstruction in the construction published in 2026-11 (at the level of the committee's assessment; § 08), but a measure on field configurations has not been constructed, and (12) as a path integral over half of spacetime remains formal.

What this chapter can say

	Content
What is said	The manipulations of (2), (5)–(8) and (11), the rotation of the contour in (5), that the lattice (9) is a probability measure, that for compact groups the lattice integral is finite Known . Tables 1–3, Figures 1–2, the numbers in § 01, § 06 and § 09 Computed
Conditional	(10) and (12) are formulas for when the Hilbert space and the Hamiltonian are there first and the vacuum is unique (the conditions for going back are in 4-10)
What is not said	The $\mathcal{D}\varphi$ of (1) and (3) is formal and not a measure. The existence of the measure of four-dimensional Yang–Mills has not been shown (6-04). The restoration of rotational symmetry and a continuum limit preserving reflection positivity were shown at the level of correlation functions by the construction published in 2026-11 (at the level of the committee's assessment; 12-01, 12-06). Nothing is shown about the mass gap

Sources and reproduction

Item	Kind	Source / tools
Quotations from the exposition (§ 3, § 6.2, § 6.5)	Text	A. Jaffe, E. Witten, *Quantum Yang–Mills Theory*, Clay Math. Inst. (2000)
That the θ term is complex	Text § 1	E. Vicari, H. Panagopoulos, Phys. Rept. 470 (2009) 93, arXiv:0803.1593
The transfer matrix and $A_0 = 0$	Abstract	M. Creutz, Phys. Rev. D 15 (1977) 1128–1136
The names of Wick and Schwinger	Bibliographic data only	G. C. Wick, Phys. Rev. 96 (1954) 1124–1134 / J. Schwinger, PNAS 44 (1958) 956–965 (origin of the names not checked)

Item	Kind	Source / tools
The correspondence between Euclidean and Minkowski	Bibliographic data only	K. Osterwalder, R. Schrader, CMP 31 (1973) 83–112, 42 (1975) 281–305 / K. Osterwalder, E. Seiler, Ann. Phys. 110 (1978) 440–471
The definition of the gap and the Hilbert space	Text	Jaffe–Witten § 4, § 6.5
Table 1, Figure 1	Computed	python3 (SciPy; agreeing with the closed form and 3×10^{-13}). The figure is JavaScript and agrees with the table
Table 2	Computed	python3 (SciPy). Checked against the value at the origin 0.154933 (the same as 1-09) and the FFT of a periodic lattice of 40^4 (agreeing to 7 digits at $a = 1/2$)
$c(a)$ of § 09	Computed	python3 (the recurrence and direct computation of the 400-point quadratic form agree to 10 digits; numerical integration of the action of the classical solution)
Figure 2 and § 06, Table 3 and § 01	Computed	python3 (NumPy FFT, 200 samples with a fixed seed / SciPy / counting). The figure is one sample made the same way

Next chapter to read: 4-03 Perturbation theory and Feynman diagrams. The Euclidean propagator of (5) becomes the lines. The conditions for going back to physical spacetime are in 4-10.

Perturbation theory and Feynman diagrams — expand around a Gaussian integral, and the terms become diagrams

> This chapter organises established knowledge (sources carry the “Known” label in the text). This book's own investigation is mainly in Part 15 and Volume 3 (Parts 16–28 and the appendices).

What this chapter shows — That each term of the expansion becomes a sum over the ways of pairing in a Gaussian integral, which can be counted with diagrams of lines and vertices.
Symmetry factors, connected diagrams, loops, and that the number of loops counts powers of $\hbar$.
That Yang–Mills needs gauge fixing, and three-point and four-point vertices and ghosts appear.

Prerequisite chapters — 4-02. For counting pairings, 1-08; for the propagator, 1-11; for the cubic and quartic terms, 3-02. $\hbar = c = 1$ (restored only in § 06).

Said first — Perturbation theory is a method of building a power series term by term, and gives neither its sum nor the existence of the theory. The integrals for the coefficients diverge (4-04), and the series is also thought to diverge (4-12). The mass gap is thought not to show at any order (4-07), and this chapter shows nothing about it.

Order of this chapter

1. The idea of perturbation theory — expanding around a Gaussian
2. Drawing pairings as lines — symmetry factors (Figure 1, Table 1) / Detour: 1949
3. Connected diagrams — the logarithm is connected diagrams (Table 2)
4. To fields — lines are propagators
5. Writing in momentum — conservation laws and loops
6. The number of loops is the power of ħ — trees are classical
7. The Feynman rules of Yang–Mills — gauge fixing is needed
8. The perturbation series — an asymptotic series (Table 3)

9. What perturbation theory says, and what it does not

01

The idea of perturbation theory — expanding around a Gaussian integral

What can be computed exactly in the Euclidean path integral ((3) of 4-02) is the case where the action is quadratic in the field, that is, a Gaussian integral (2-07 § 02). So we split the action into the quadratic part S_0 and the interaction S_{int}, and expand $e^{-S_{\mathrm{int}}}$ ($\langle\cdot\rangle_0$ is the expectation with S_0 alone).

$$\langle F\rangle = \frac{\langle F\, e^{-S_{\mathrm{int}}}\rangle_0}{\langle e^{-S_{\mathrm{int}}}\rangle_0}, \qquad \langle F\, e^{-S_{\mathrm{int}}}\rangle_0 \overset{\text{formal}}{=} \sum_{n\geq 0} \frac{(-1)^n}{n!}\,\langle F\, S_{\mathrm{int}}^n\rangle_0 \quad (1)$$

Exchanging the sum and the integral is formal (§ 08). **Example.** In place of a field take one real number x, and let $S = \frac{1}{2}x^2 + \frac{u}{4!}x^4$ ($u \geq 0$; the same convention as 4-04). It is a "zero-dimensional field theory" whose spacetime is a single point. From $\langle x^{2k}\rangle_0 = (2k-1)!!$ (1-08 § 06),

$$Z(u) = \int \frac{dx}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2 - \frac{u}{4!}x^4} \;\sim\; \sum_{n\geq 0} \frac{(4n-1)!!}{n!}\left(-\frac{u}{4!}\right)^n = 1 - \frac{u}{8} + \frac{35}{384}u^2 - \frac{385}{3072}u^3 + \cdots$$

(the meaning of $\sim$ is in § 08; the coefficients computed as fractions Computed). All the values of the diagrams are numbers, so the bookkeeping can be checked by hand, and in four dimensions the shapes of the diagrams and the bookkeeping of weights are the same. Each term of the expansion can be written only with averages of polynomials under the Gaussian distribution, that is, moments of the normal distribution. Not one hard integral remains; all that remains is counting.

02

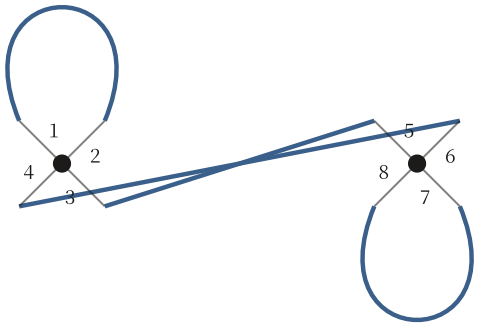
Drawing pairings as lines — from Wick's theorem to diagrams

The term of order n in (2) was the number of ways of pairing up $4n$ copies of x two by two (Isserlis' formula, in physics Wick's theorem; (9) of 1-08). Draw each $\frac{u}{4!}x^4$ as a point (vertex) with 4 legs, and join paired legs by lines. Pairings that give the same shape after joining have the same value, so they are grouped by shape. These are Feynman diagrams.

$$\frac{\#\{\text{pairings giving the shape } \Gamma\}}{n!\,(4!)^n} = \frac{1}{S(\Gamma)} \quad (3)$$

The denominator is the number of permutations of vertices and legs, and usually exactly this many pairings make one shape. If the shape has a symmetry, some of the permutations give back the same pairing and the number drops; that ratio is the symmetry factor $S(\Gamma)$, precisely the number of automorphisms of the diagram (exchanges of vertices and reattachments of legs that do not change how the lines are connected) Known . **Example.** With one vertex there are 3 pairings, all of them a "figure eight". $3/4! = 1/8$ is the coefficient of u in (2), and with the exchange of the loops and the reversal of each loop, $S = 2 \cdot 2 \cdot 2$.

Figure 1 Choose a case and move the pairing number with the knob, and the pairing of legs (blue lines) and the shape of the diagram it makes change. "Count all" counts by shape



`Pairing 4 / 105 : 「bubble with a loop on each side", pairings of this shape: 72, 1/S = 1/16`

Computed The counting in this figure is done in the browser and has been checked to agree with Table 1. The numbers are the legs of the vertices; x, y and 1–4 are external points (the x put in from outside).

Computed Table 1 All pairings generated and sorted into shapes (isomorphism classes) (python3). In brackets: (number of pairings, $1/S$)

Case	Pairings	Shapes
Vacuum, 1 vertex	3	figure eight (3, 1/8)
Vacuum, 2 vertices	105	bubble with a loop on each side (72, 1/16), joined by four lines (24, 1/48), two figure eights (9, 1/128)
Two-point, 1 vertex	15	line with a loop (12, 1/2), line and a figure eight (3, 1/8)
Two-point, 2 vertices	945	loop on a loop (288, 1/4), two loops in a row (288, 1/4), joined by three lines (192, 1/6), and those containing bubbles: line and bubble with a loop on each side (72, 1/16), line with a loop and a figure eight (72, 1/16), line and joined by four lines (24, 1/48), line and two figure eights (9, 1/128)
Four-point, 1 vertex	105	$\times$ (24, 1), line and line with a loop (6 ways $\times$ 12, each 1/2), two lines and a figure eight (3 ways $\times$ 3, each 1/8)

Detour: the 1949 papers and a theorem the following year

A subsection you may skip.

The abstract of Feynman (1949) says it presents a viewpoint from which matrix elements can be 'written down directly for any specific problem', and in the same year Dyson (1949) showed the equivalence of the theories of Feynman and Schwinger. The next year Wick (1950) proved an algebraic theorem that simplifies the reduction to 'a sum of "graph" terms'. The formula itself is found in Isserlis (1918), in statistics Known .

03

Connected diagrams — bubbles vanish by division, and the logarithm counts connected diagrams

Table 1 also includes diagrams containing parts not connected to the external points (vacuum bubbles). The value of a diagram made of separate parts is the product of the values of the parts, and the denominator Z of (1) is the sum of diagrams made only of bubbles, so bubbles vanish by division. Moreover, the sum of all diagrams is the exponential of the sum of connected diagrams. When k copies of the same connected diagram C appear side by side, the exchanges $k!$ are added to S, giving $\sum_k C^k/k! = e^C$ Known . In zero dimensions the value of a line is 1, so the value of a diagram is $(-u)^n/S$.

$$\log Z = \sum_{\text{connected vacuum diagrams}} \frac{(-u)^n}{S}, \qquad \langle x^2 \rangle = \sum_{\text{connected two-point diagrams}} \frac{(-u)}{S}$$

Example. For u^2, the $\frac{1}{16} + \frac{1}{48} = \frac{1}{12}$ of the connected vacuum diagrams agrees with the coefficient $\frac{35}{384} - \frac{1}{2}\cdot\frac{1}{64} = \frac{1}{12}$ of $\log Z$, and the $\frac{1}{4} + \frac{1}{4} + \frac{1}{6} = \frac{2}{3}$ of the connected two-point diagrams is the coefficient of $\langle x^2 \rangle$. Read the other way, a single integral is a generating function that counts diagrams, which leads to the counting of diagrams on surfaces (Bessis–Itzykson–Zuber 1980, 17-04).

Computed Table 2 Sums of $(-1)^n/S$ for connected shapes counted from all pairings (python3). All agree with the coefficients of the series. In brackets, the coefficient from the series (the two-point $n = 4$ was not counted)

n	Vacuum pairings	Shapes (connected)	$\log Z$	Two-point pairings	$\langle x^2 \rangle$
1	3	1 (1)	-1/8	15	-1/2
2	105	3 (2)	1/12	945	2/3
3	10,395	7 (4)	-11/96	135,135	-11/8
4	2,027,025	20 (10)	17/72	—	(34/9)

04

To fields — lines are propagators, vertices are integrals over position

For a four-dimensional field the procedure is the same. The quadratic part of the d-dimensional action

$$S_E[\phi] = \int d^d x \left[\tfrac{1}{2} (\partial_\mu \phi)^2 + \tfrac{1}{2} m^2 \phi^2 + \tfrac{u}{4!} \phi^4 \right] \quad (5)$$

is a Gaussian measure with covariance $G = (-\Delta + m^2)^{-1}$ (Δ is the Laplacian $\sum_\mu \partial_\mu^2$; not the Δ of the mass gap) (4-02 § 06), and Wick's theorem can be used as it is. Two things change. The value of a line becomes the propagator $G(x-y)$ (the degree to which the fluctuation of the field at the point x reaches the point y; it becomes smaller with distance), and vertices can be placed anywhere in spacetime, so their positions z are integrated over.

$$\langle \phi(x) \phi(y) \rangle = G(x-y) - \frac{u}{2} \int d^d z \, G(x-z) \, G(0) \, G(z-y) + O(u^2) \quad (6)$$

The second term is the "line with a loop"; the number 1 for $\langle x^2 \rangle = 1 - \frac{u}{2} + \cdots$ in zero dimensions has simply been replaced by a propagator. The lowest-order four-point "$\times$" is $-u \int d^d z \prod_i G(x_i - z)$, which can be read as four particles meeting at one point. Precisely, it means integrating the position of that point over all space. **Note.** $G(0)$ is infinite at $d \geq 2$ (1-11), and it diverges from first order. Its treatment is 4-04, and the $T(\Lambda)$ there is this loop.

05

Writing in momentum — conservation laws, and the loop integrals that remain

Putting in $G(x) = \int \frac{d^d p}{(2\pi)^d} \frac{e^{ip \cdot x}}{p^2 + m^2}$ (the inverse Fourier transform of the $\hat{G}$ of (10) in 1-11; the rotation is (5) of 4-02) and integrating over the positions of the vertices gives $\int d^d z \, e^{i(p_1 + \cdots + p_4) \cdot z} = (2\pi)^d \delta^{(d)}(p_1 + \cdots + p_4)$: momentum is conserved at each vertex. The δ here is the relation of 1-11 that integrating a plane wave over all space concentrates the value only where the wave number is 0. The integral over positions turns into a condition on adding momenta. There are V conditions for I internal lines, and one of them is overall conservation, involving only the external lines, so the number of momenta left free in a connected diagram is

$$L = I - V + 1 \quad (7)$$

which is the number of independent cycles (loops) of the diagram Known . The rules are $1/(p^2 + m^2)$ for each line, $-u$ for each vertex, $\int d^d k / (2\pi)^d$ for each loop, and $1/S$

overall. **Example.** The one-loop four-point "bubble" has $I = V = 2$, $L = 1$, and with momentum P flowing in,

$$\frac{u^2}{2}\int \frac{d^dk}{(2\pi)^d}\,\frac{1}{(k^2+m^2)\big((k+P)^2+m^2\big)} \quad (8)$$

The two-point "joined by three lines" diagram has $I = 3$, $V = 2$, $L = 2$, and is a double integral. (8) behaves like $k^{d-1}dk/k^4$ for large $|k|$, and diverges logarithmically at $d = 4$. Adding the three bubbles and changing the sign gives the one-loop term of (4) in 4-04.

06

The number of loops is the power of $\hbar$ — trees are classical, loops are quantum corrections

Restoring the weight to $e^{-SE/\hbar}$, propagators are multiplied by $\hbar$ and vertices by $1/\hbar$. For a connected diagram with I' lines including those going to external points, V vertices and E external points, counting external points as points too, $L = I' - V - E + 1$, so

$$\hbar^{I'-V} = \hbar^{L+E-1} \quad (9)$$

Fixing E, the power of $\hbar$ is fixed by the number of loops (Srednicki § 21) Known . The lowest order is diagrams without loops (trees), and the abstract of Boulware–Brown (1968) writes that the classical field made by a source is 'the generating functional of the tree-graph approximation' Known .

Example. The classical solution of $\frac{1}{2}x^2 + \frac{u}{4!}x^4 - Jx$, zero dimensions with a source added, is $x_{\mathrm{cl}} = J - \frac{u}{6}J^3 + \frac{u^2}{12}J^5 - \frac{u^3}{18}J^7 + \cdots$. In general it can be written $x_{\mathrm{cl}} = \sum_k (-1)^k T_k\,(u/6)^k J^{2k+1}$, where $T_k = 1, 1, 3, 12, 55, \ldots$ is the number of ternary trees. Integrating numerically at $u = 1$, $J = 0.8$, $(\langle x\rangle - x_{\mathrm{cl}})/\hbar$ approaches -0.2167, -0.2266, -0.2277 at $\hbar = 0.1,\ 0.01,\ 0.001$, the value -0.2278 of the one-loop diagram (a line with a loop around x_{cl}) Computed . In Yang–Mills, rescaling as $\hat{A} = gA$ makes the action $\frac{1}{2g^2}\int \mathrm{tr}\,\hat{F}^2$ ((7) of 4-02), which enters the weight only in the combination $g^2\hbar$, so the loop expansion is an expansion in g^2 (3-03).

07

The Feynman rules of Yang–Mills — a quadratic form that cannot be inverted, and three-point and four-point vertices

The conventions are the same as in 3-02: g is written split, $F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu$, $S_E = \frac{1}{4}\int F^a_{\mu\nu} F^a_{\mu\nu}\, d^4x$. Split into quadratic, cubic and quartic parts ((11) of 3-02 read in Euclidean form as in 4-02 § 04; all indices are lower, and repeated indices are summed over 1–4),

$$S_E = \frac{1}{2}\int A^a_\mu\left(-\partial^2\delta_{\mu\nu} + \partial_\mu\partial_\nu\right)A^a_\nu + g f^{abc}\int (\partial_\mu A^a_\nu) A^b_\mu A^c_\nu + \frac{g^2}{4} f^{abe} f^{cde}\int A^a_\mu A^b_\nu A^c_\mu A^d_\nu$$

The quadratic kernel is $K_{\mu\nu} = p^2\delta_{\mu\nu} - p_\mu p_\nu$ in momentum, and sends p_μ to 0. This is because $A_\mu = \partial_\mu\lambda$ is the direction of infinitesimal gauge transformations, which do not change the action (4-02 § 07). There is no inverse matrix, so no propagator can be made. Adding the gauge-fixing term $\frac{1}{2\alpha}\int(\partial_\mu A^a_\mu)^2$ (justified in 4-08),

$$K^{(\alpha)}_{\mu\nu} = p^2\delta_{\mu\nu} - \left(1 - \frac{1}{\alpha}\right)p_\mu p_\nu, \qquad D^{ab}_{\mu\nu}(p) = \frac{\delta^{ab}}{p^2}\left(\delta_{\mu\nu} - (1-\alpha)\frac{p_\mu p_\nu}{p^2}\right) \quad (11)$$

and for $\alpha = 1$ (the Feynman gauge), $\delta^{ab}\delta_{\mu\nu}/p^2$ (the same letter as in 4-08). The passage to momentum is $A^a_\mu(x) = \int \frac{d^4p}{(2\pi)^4} e^{ip\cdot x}\tilde{A}^a_\mu(p)$ (derivatives become ip_μ). A vertex is the term of order n of the action differentiated n times with respect to $\tilde{A}$, with the sign changed (excluding $(2\pi)^4\delta(\sum p)$); the sign of e^{-S} enters here (all momenta flowing in). Taking the opposite direction of Fourier transform reverses the sign of i in (12).

$$V^{abc}_{\mu\nu\lambda}(p,q,r) = ig\, f^{abc}\left[\delta_{\mu\nu}(p-q)_\lambda + \delta_{\nu\lambda}(q-r)_\mu + \delta_{\lambda\mu}(r-p)_\nu\right] \quad (12)$$

$$V^{abcd}_{\mu\nu\lambda\rho} = -g^2\left[f^{abe}f^{cde}(\delta_{\mu\lambda}\delta_{\nu\rho} - \delta_{\mu\rho}\delta_{\nu\lambda}) + f^{ace}f^{bde}(\delta_{\mu\nu}\delta_{\lambda\rho} - \delta_{\mu\rho}\delta_{\nu\lambda}) + f^{ade}f^{bce}(\delta_{\mu}\right.$$

The three-point vertex is proportional to momentum, the four-point vertex does not depend on momentum, and both contain f^{abc}. **Example.** Putting plane waves into (10), extracting the coefficients directly, and comparing with (11)–(13) for 20 random sets of momenta, polarisations and colours in SU(2) and SU(3), the relative error was at most 8×10^{-13} Computed .

Ghosts. The abstract of Faddeev–Popov (1967) writes 'Feynman and De Witt showed, that the rules must be changed for ... diagrams with closed loops', and proposed a method of obtaining the rules for arbitrary diagrams from the path integral. The result is the rule of adding ghosts with propagator δ^{ab}/p^2 and a vertex with two ghost lines and one gauge field line, and multiplying by -1 for each closed ghost loop (Srednicki § 71) Known . -1 comes from ghosts being anticommuting numbers (4-08). The derivation, and the limits of the local gauge fixing it presupposes (3-04), are in 4-08.

08

The perturbation series — the coefficients grow factorially, and it must be cut off somewhere

The ratio of the coefficients a_n of (2) is $|a_{n+1}/a_n| = \frac{(4n+1)(4n+3)}{24(n+1)} \approx \frac{2n}{3}$, so a_n grows at a rate of roughly $n!\,(2/3)^n$, and the series does not converge for any $u > 0$. That the integral (2) diverges for $u < 0$ shows the same circumstance. The argument of this form in quantum electrodynamics was made by Dyson (1952), whose abstract qualifies the conclusion as 'tentatively' (4-12). Even so, putting $s = \frac{u}{4!}x^4$ into the Taylor remainder $\left|e^{-s} - \sum_{n<N}(-s)^n/n!\right| \le s^N/N!$ $(s \ge 0)$ and integrating gives

$$\left| Z(u) - \sum_{n<N} a_n u^n \right| \le |a_N|\, u^N \qquad (u \ge 0) \tag{14}$$

Known . Fixing N and letting $u \to 0$, the error vanishes at the rate u^N. This is an asymptotic series, and it is the meaning of the $\sim$ in (2). Fixing u and increasing N, the error first falls and then rises, and the best is around $N \approx 3/(2u)$, where the terms are smallest (the valley of term sizes can be moved in Figure 1 of 4-12).

Computed Table 3 Truncation of $\langle x^2\rangle = 1 - \frac{1}{2}u + \frac{2}{3}u^2 - \frac{11}{8}u^3 + \frac{34}{9}u^4 - \cdots$ (python3, $N \le 16$)

u	Exact $\langle x^2\rangle$	Best N	Least error	Error at N=10
0.1	0.955577	14	1.3e−6	2.3e−6
0.2	0.919366	6	1.6e−3	3.2e−3
0.5	0.838595	2	7.8e−2	41
1.0	0.750511	0	0.25	4.7e4

For $u = 1$, adding nothing, $N = 0$, is best. (14) also held numerically for $u \le 1$, $N \le 16$ Computed . For field theory, the problem statement writes that in many known examples ('In most known examples') perturbation series are 'divergent expansions' (Jaffe–Witten § 6.1), and that in the two-dimensional models that have been constructed the scattering matrix can be expanded as 'an asymptotic series in the coupling constants' (§ 6.2). For four-dimensional Yang–Mills, the construction published in 2026-11 built the theory (at the level of the committee's assessment; 6-00), but agreement with perturbation theory is claimed only within a range such as the one-loop running of the coupling, and it has not been shown that the series is an asymptotic expansion of that theory (6-04).

09

What perturbation theory says, and what it does not

The coefficient at each order is a finite sum of diagram integrals, and that each order is finite after renormalisation is a theorem, gauge theories included (4-04 § 08, 4-05) Known . **The sum of the series** has no meaning in general. There are also **things that do not show in perturbation theory**. The propagator (11) has a pole at $p = 0$ and represents a massless particle. If there is a mass gap, it is thought to come in a form that looks 0 at every order in g^2, like $\Lambda \propto e^{-1/(2b_0 g^2)}$ (4-07) Physics . § 6.6 of the problem statement also writes 'the mass gap is not classically visible'.

Even so, perturbation theory is part of the conditions of the problem statement. § 4 requires that the correlation functions of the constructed theory agree at short distances with asymptotic freedom and the predictions of perturbative renormalisation theory (quoted in 0-03 § 07). Perturbation theory is not a method of making the theory but a mark that the theory made must meet. Asymptotic freedom is 4-06, and large N is 17-04.

What this chapter can say

	Content
What is said	The expansion of (1)(2)(4) and how to count diagrams, the symmetry factor of (3), the loop counting of (7)(9), the form of the Feynman rules (11)–(13), and (14) in zero dimensions Known . Tables 1–3, Figure 1, the numbers in § 06 and § 07 Computed
Conditional	The values of field diagrams such as (6)(8) have meaning only once a cutoff is put in (4-04). The ghost rules are rules under local gauge fixing (4-08). The view that the mass gap comes in the form e^{-1/g^2} is an expectation of physics Physics
What is not said	The sum of the perturbation series, an estimate corresponding to (14) in field theory, the existence of four-dimensional Yang–Mills, the mass gap (this site shows none of them)

Sources and reproduction

Item	Kind	Source / tools
Quotations from the problem statement (§ 4, § 6.1, § 6.2, § 6.6)	Text	A. Jaffe, E. Witten, *Quantum Yang–Mills Theory*, Clay Math. Inst. (2000)
Counting of $\hbar$, the -1 of ghosts	Text § 21, § 71	M. Srednicki, *Quantum Field Theory*, Cambridge UP (2007) (the author's public draft)
The beginnings of the diagram method	Abstract	R. P. Feynman, Phys. Rev. 76 (1949) 769 / F. J. Dyson, Phys. Rev. 75 (1949) 486 / G. C. Wick, Phys. Rev. 80 (1950) 268

Item	Kind	Source / tools
The formula for Gaussian moments	Bibliographic data only	L. Isserlis, Biometrika 12 (1918) 134–139
The ghost rules	Abstract	L. D. Faddeev, V. N. Popov, Phys. Lett. B 25 (1967) 29–30. The earlier Feynman, Acta Phys. Polon. 24 (1963) 697 and DeWitt, Phys. Rev. 162 (1967) 1195, 1239 bibliographic data only
Tree diagrams and the classical / counting diagrams / divergence of the series	Abstract	D. G. Boulware, L. S. Brown, Phys. Rev. 172 (1968) 1628 / D. Bessis, C. Itzykson, J.-B. Zuber, Adv. Appl. Math. 1 (1980) 109 / F. J. Dyson, Phys. Rev. 85 (1952) 631
Tables 1 and 2, Figure 1	Computed	python3 (all pairings generated and sorted into isomorphism classes by permutations; vacuum up to the 2,027,025 pairings at order 4). The figure's JS uses the same classification, checked to agree with node
The comparison in § 07	Computed	python3 (NumPy; coefficients extracted by putting in plane waves)
Table 3, the numbers in § 06	Computed	python3 (coefficients as fractions; exact values by SciPy numerical integration, Z also agreeing with the closed formula with the Bessel function $K_{1/4}$)

Next chapter to read: 4-04 Divergences and renormalisation. The $G(0)$ of (6) and the logarithmic divergence of (8) are rewritten in terms of the measured coupling.

Divergences and renormalisation — rewriting infinities in terms of measured quantities

This chapter organises established knowledge (sources carry the “Known” label in the text). This book's own investigation is mainly in Part 15 and Volume 3 (Parts 16–28 and the appendices).

What this chapter shows — That loop integrals diverge because they add up too many short-distance fluctuations. That the types of divergence can be found by counting dimensions, and are finitely many for ϕ^4 in four dimensions and for Yang–Mills. That absorbing the divergences into measured couplings makes predictions independent of the cutoff, and that those couplings run with the scale.

Prerequisite chapters — 4-03 (propagators, vertices, loops). Helpful: 1-11, 2-08. $\hbar = c = 1$, Euclidean momenta. β is the β function (not the inverse coupling of the lattice).

Said first — This is a chapter of perturbative calculations and how to read them. Renormalisability is a theorem at the level of each order of perturbation theory (formal power series), and gives neither the sum of the series nor the existence of the four-dimensional theory. This site shows neither the construction of the continuum limit nor the mass gap.

Order of this chapter

1. Where the infinities come from — adding up too many short-distance fluctuations
2. Putting in a cutoff — regularisation, and its price (Table 1)
3. Counting dimensions — the superficial degree of divergence (Table 2)
4. Counterterms — rewriting the bare coupling in terms of the measured coupling (Figure 1)
5. The running coupling and the renormalisation group equation
6. Wilson's view — relevant, irrelevant, marginal (Figure 2) / Detour: the Fermi interaction
7. In gauge theories — subtracting without breaking the symmetry

8. Renormalisation seen from mathematics — where the theorems are, and where they are not yet (Table 3)

01

Where the infinities come from — adding up too many short-distance fluctuations

When a perturbative diagram (4-03) has a loop, an integral over a momentum k not fixed from outside remains. Large $|k|$ are short-distance fluctuations, and in four dimensions there are too many of them for convergence. We cut at $|k| < \Lambda$ (Λ is the **cutoff**) and look at how it grows.

$$T(\Lambda) = \int_{|k|<\Lambda} \frac{d^4k}{(2\pi)^4} \frac{1}{k^2+m^2} = \frac{1}{16\pi^2}\left[\Lambda^2 - m^2 \ln\left(1+\frac{\Lambda^2}{m^2}\right)\right]$$
$$I(\Lambda) = \int_{|k|<\Lambda} \frac{d^4k}{(2\pi)^4} \frac{1}{(k^2+m^2)^2} = \frac{1}{16\pi^2}\left[\ln\left(1+\frac{\Lambda^2}{m^2}\right) - \frac{\Lambda^2}{\Lambda^2+m^2}\right] \quad (1)$$

T is the diagram closing one propagator, and I the diagram closed with two (a bubble) ($d^4k = 2\pi^2 k^3 dk$). Multiplying the volume $k^3 dk$ by k^{-2} for each propagator, T is $k\,dk$ (quadratic divergence) and I is $k^{-1}dk$ (logarithmic divergence). On the position side this corresponds to $G(x)^2 \sim |x|^{-4}$ not being integrable at the origin (1-11).

Example. With $m = 1$ and $\Lambda = 10, 10^3, 10^6$, $16\pi^2 T$ is 95.4, 1.0×10^6, 1.0×10^{12}, and $16\pi^2 I$ is 3.625, 12.816, 26.631. Each time Λ is multiplied by 10, T grows about 100-fold and $16\pi^2 I$ grows by $2\ln 10 = 4.605$ Computed .

02

Putting in a cutoff — regularisation, and its price

Putting in an artificial scale to make things finite is called **regularisation**. Regularisation is scaffolding, and the goal is that as $\Lambda \to \infty$ (on a lattice, $a \to 0$) the answer does not depend on the scaffolding. Each of the four in Table 1 loses something.

Table 1 Typical regularisations

Regularisation	Keeps	Loses
Momentum cutoff	Rotations	Gauge invariance (§ 07), reflection positivity
Lattice (5-01, $\Lambda \sim \pi/a$)	Gauge invariance, reflection positivity (5-03)	Continuous rotations (4-02)

Regularisation	Keeps	Loses
Pauli–Villars (subtracting heavy fields)	Rotations	Reflection positivity (4-10) (in the non-abelian case also gauge invariance; § 07)
Dimensional regularisation ($d = 4 - \epsilon$)	Rotations, gauge invariance	Reflection positivity (4-05)

Dimensional regularisation (Bollini–Giambiagi 1972, 't Hooft–Veltman 1972) is the standard tool of perturbation theory, Yang–Mills included. On the other hand, in Table 1 only the lattice row keeps reflection positivity. Reflection positivity is the property needed to assemble the Hilbert space of quantum mechanics (the space of states) from Euclidean correlation functions (4-10). The problem statement writes that there are few ways to recover reflection positivity once lost by regularisation, giving dimensional regularisation and Pauli–Villars as examples (§ 6.5; quoted in 0-03 § 03). Being able to keep from the start a property that is hard to recover once lost is one of the reasons the lattice is chosen on the side of construction.

03

Counting dimensions — the superficial degree of divergence

With $\hbar = c = 1$, the dimension $[\cdot]$ of a quantity is a power of mass (length is -1). The action is dimensionless, $d^d x$ is $-d$, and derivatives are $+1$, so the dimensions of fields and couplings are fixed.

$$\int d^d x \Big[\tfrac{1}{2}(\partial\phi)^2 + \tfrac{u}{4!}\phi^4\Big] : \ [\phi] = \tfrac{d-2}{2}, \ [u] = 4 - d; \qquad \frac{1}{2g^2}\int d^d x \, \mathrm{tr}\, F_{\mu\nu}F_{\mu\nu} : \ [A] = 1, \ [g^2]$$

In four dimensions both u and g^2 are dimensionless. This is what "four dimensions is marginal" means (0-03, 1-12; how to derive the counting of dimensions is (10) of 2-04 § 08). The power of momentum of a diagram with L loops and I internal lines is $\omega = dL - 2I$ (adding the number of vertices that carry derivatives), and if $\omega \geq 0$ it superficially diverges. For n vertices of ϕ^4 and E external lines, $4n = 2I + E$, and from $L = I - n + 1$,

$$\omega_{\phi^4} = d - \frac{d-2}{2}E - [u]\,n, \qquad \omega_{\mathrm{YM}} = d - \frac{d-2}{2}E - \frac{[g^2]}{2}(V_3 + 2V_4) \quad (3)$$

(V_3, V_4 is the number of three-point and four-point vertices of Yang–Mills, and the external lines are gluons). In four dimensions $\omega_{\mathrm{YM}} = 4 - E$. The dimension of the coupling appears in the coefficient of the order, so there are three cases Known . If positive, ω falls with the order, and only finitely many diagrams diverge (**super-renormalisable**). If 0, the types of divergence (values of E) are finitely many but appear at all orders (**renormalisable**). If negative, the types increase without bound (**non-renormalisable**).

Computed Table 2 Types of loop diagrams with $\omega \geq 0$ (excluding $E = 0$, up to order 30; enumerated from combinations of vertices and external lines, agreeing with (3); for Yang–Mills $E = 1$ is 0 by Lorentz symmetry and colour symmetry)

Theory	Dimension of the coupling	E (ω)	Orders	Classification
ϕ^4, $d = 2$	2	2 (0)	1 only	Super-renormalisable
ϕ^4, $d = 3$	1	2 (1, 0)	1, 2	Super-renormalisable
ϕ^4, $d = 4$	0	2 (2), 4 (0)	All	Renormalisable
ϕ^4, $d = 6$	−2	2, 4, 6, 8, ⋯	All	Non-renormalisable
Yang–Mills, $d = 4$	0	1 (3), 2 (2), 3 (1), 4 (0)	All	Renormalisable

The two-dimensional type has the same form as T and can be removed by Wick ordering (7-03; for three-dimensional Yang–Mills, 7-06). We say "superficial" because even with $\omega < 0$ a diagram diverges if it has a $\omega \geq 0$ subdiagram inside. The theorem that states convergence by a power condition including subdiagrams is Weinberg's 1960 theorem, and the method of subtracting subdiagrams in turn is treated in 4-05.

04

Counterterms — rewriting the bare coupling in terms of the measured coupling

We follow the procedure at one loop for ϕ^4 in four dimensions. The coefficients m_0, u_0 of the action are called the **bare** mass and coupling (coefficients under the cutoff, not directly measurable). The four-point strength Γ^4 (not the diagram symbol Γ of 4-03; the four-point function with the external propagators removed, summing only diagrams that do not split in two when one line is cut; u_0 at tree level) shifts by the amount of one inserted bubble.

$$\Gamma^4 = u_0 - \frac{u_0^2}{2}\sum_{j=1}^{3} B(P_j;\Lambda) + O(u_0^3), \qquad P_1 = p_1 + p_2,\ P_2 = p_1 + p_3,\ P_3 = p_1 + p_4 \tag{4}$$

The three terms are the three ways of dividing the external lines into pairs, and $\frac{1}{2}$ is the symmetry factor (counted the same way as 7-08). The bubble with momentum P flowing in reduces to the I of (1) by $\frac{1}{AB} = \int_0^1 \frac{dx}{[xA+(1-x)B]^2}$ (a Feynman parameter) and a shift of variable.

$$B(P;\Lambda) = \int \frac{d^4k}{(2\pi)^4}\,\frac{1}{(k^2+m^2)\big((k+P)^2+m^2\big)} = \int_0^1 I\big(\Lambda;\, m^2 \to m^2 + x(1-x)P^2\big)\,dx \tag{5}$$

The difference between the value cut in the shifted variable and the value cut in the original k is 0 as $\Lambda \to \infty$ ($P = m$, $\Lambda = 10^3 m$, multiplied by $16\pi^2$, 3×10^{-7}) Computed . So we redefine the coupling by a measured quantity. The Γ^4 measured in the configuration where all three $|P_j|$ are μ is taken as the **renormalised coupling** $u_R(\mu)$ at the scale μ.

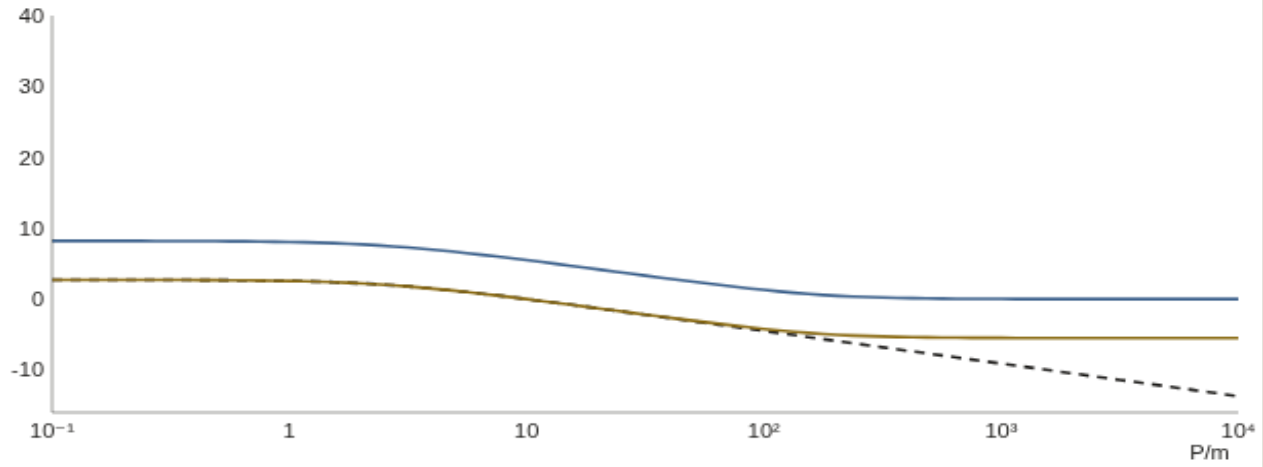
$$u_R(\mu) = u_0 - \tfrac{3}{2} u_0^2 B(\mu;\Lambda) + O(u_0^3), \qquad u_0 = u_R + \underbrace{\tfrac{3}{2} u_R^2 B(\mu;\Lambda)}_{\text{counterterm } \delta u} + O(u_R^3) \tag{6}$$

$\delta u \approx \frac{3u_R^2}{32\pi^2} \ln \frac{\Lambda^2}{\mu^2}$ diverges. Putting this back into (4), the divergence is shut up in the form of a difference.

$$\Gamma^4 = u_R - \frac{u_R^2}{2} \sum_j \big[B(P_j;\Lambda) - B(\mu;\Lambda) \big] + O(u_R^3) \xrightarrow{\Lambda\to\infty} u_R - \frac{u_R^2}{32\pi^2} \sum_j \int_0^1 \ln \frac{m^2 + x}{m^2 + x}$$

No Λ remains on the right-hand side. The content of renormalisation is that **predictions written in terms of the measured coupling are finite**, and the infinity is pushed into the unmeasurable u_0. The divergence of the mass (T) is pushed into m_0, and from two loops on also into the field normalisation Z; for ϕ^4 in four dimensions these three suffice at all orders. This is the practical meaning of "renormalisable" Known .

Figure 1 Raising the cutoff Λ with the knob, the bare bubble $16\pi^2 B(P;\Lambda)$ (solid line, upper) keeps rising, while the difference with the value at $\mu = 10m$ subtracted (solid line, lower) settles to the limit shown dashed. The horizontal axis is P (log). "Play" moves Λ back and forth



Λ = $10^{2.0}$ m: at P = 100m, bare 1.289, difference −4.208 (limit −4.496)

Computed The numbers in this figure were computed on this machine ((5)(7) integrated numerically in python3; the figure integrates the same formulas on the spot by Simpson's rule, agreeing to 4 digits). Static values ($16\pi^2$ times, $\mu = 10m$): the difference at $P = 1000m$ is −5.477 at $\Lambda = 10^2 m$, −9.096 at $10^4 m$, −9.099 at $10^8 m$ (limit −9.099). At the same Λ the bare $B(100m)$ keeps growing: 1.289, 10.208, 28.629.

05

The running coupling and the renormalisation group equation

The reference scale μ can be chosen freely. How u_R changes when u_0, Λ is held fixed and μ is moved is the **β function**.

$$\beta(u_R) = \mu \frac{du_R}{d\mu} = -\frac{3}{2} u_R^2 \, \mu \frac{\partial B(\mu;\Lambda)}{\partial \mu} + O(u_R^3) = \frac{3u_R^2}{16\pi^2} + O(u_R^3) \qquad (\mu \gg m) \quad (8)$$

Λ vanishes under differentiation. Differentiating (5) numerically, the coefficient approaches $3/16\pi^2 = 0.018998$: 0.01728 at $\mu = 10m$, 0.018997 at $10^3 m$ Computed . Integrating at one loop,

$$\frac{1}{u_R(\mu)} = \frac{1}{u_R(\mu_0)} - \frac{3}{16\pi^2} \ln \frac{\mu}{\mu_0} \quad (9)$$

and the coupling becomes stronger at shorter distances. For $u_R(\mu_0) = 1$, it is 1.78 at $\mu/\mu_0 = 10^{10}$, and the denominator is 0 at $10^{22.86}$ (the Landau pole; 7-08) Computed Physics . Yang–Mills has the opposite sign (4-06).

Example. At one loop in quantum electrodynamics, each Dirac particle of charge 1 (charged lepton) gives $d(1/\alpha)/d\ln\mu = -2/3\pi$. Extending from $1/\alpha = 137.036$ to $M_Z = 91.188$ GeV with the electron, μ and τ gives 132.20 (only the leading logarithm at one loop) Computed . The PDG 2024 value is $\hat{\alpha}^{(5)}(M_Z)^{-1} = 127.930$. MS is the scheme that, instead of a cutoff, subtracts only the poles of dimensional regularisation (4-05). The superscript (5) indicates the value with the five kinds of quark other than top included. The difference from 132.20 is due to the quarks and the constants of the scheme (3-07).

It can also be said for whole functions. The n-point function written in bare quantities knows nothing of μ, and the renormalised $\Gamma_R^{(n)}$ differs by the field normalisation Z, so

$$\left[\mu \frac{\partial}{\partial \mu} + \beta(u) \frac{\partial}{\partial u} - n\gamma(u)\right] \Gamma_R^{(n)}(P; u, \mu) = 0, \qquad \gamma = \tfrac{1}{2} \mu \, \partial_\mu \ln Z \quad (10)$$

(γ is the **anomalous dimension**; for ϕ^4 from two loops on). Where masses do not matter, multiplying momenta by s while moving the coupling along with them brings the left-hand side of (10) into a single ordinary derivative in s (the method of characteristics). Ignoring γ, $\Gamma^4(sP; u, \mu) = \Gamma^4(P; \bar{u}(s), \mu)$ ($s\,d\bar{u}/ds = \beta(\bar{u})$, $\bar{u}(1) = u$). Looking at momenta multiplied by s is replacing the coupling by its value at the scale $s\mu$ Known . Gell-Mann–Low 1954 defined the effective charge of the electron $e(\mu)$ as a function of scale (lecture notes: D. Tong, *Statistical Field Theory* § 3.5.3). Stueckelberg–Petermann 1953 are said to have noticed that the transformations of renormalisation form a group. The form of the equation that moves masses is due to Callan 1970 and Symanzik 1970 (the Callan–Symanzik equation), and in textbooks the

form (10) is also called by this name. All are statements at each order of perturbation theory.

06

Wilson's view — integrating out short scales

Regarding Λ as the genuine smallest scale, integrating only the fluctuations in $\Lambda/b < |k| < \Lambda$ and remeasuring lengths b times, all the coefficients of the action move a little. Repeating this is Wilson's 1971 renormalisation group, which implements Kadanoff's idea of blocking (2-08) as integration shell by shell in momentum (a rigorous version on the lattice is 7-05). As a metaphor, it is the operation of lowering the resolution of a photograph one step at a time. Precisely, it is a dynamical system in which the sequence of coefficients of the effective action is carried by a map at each step.

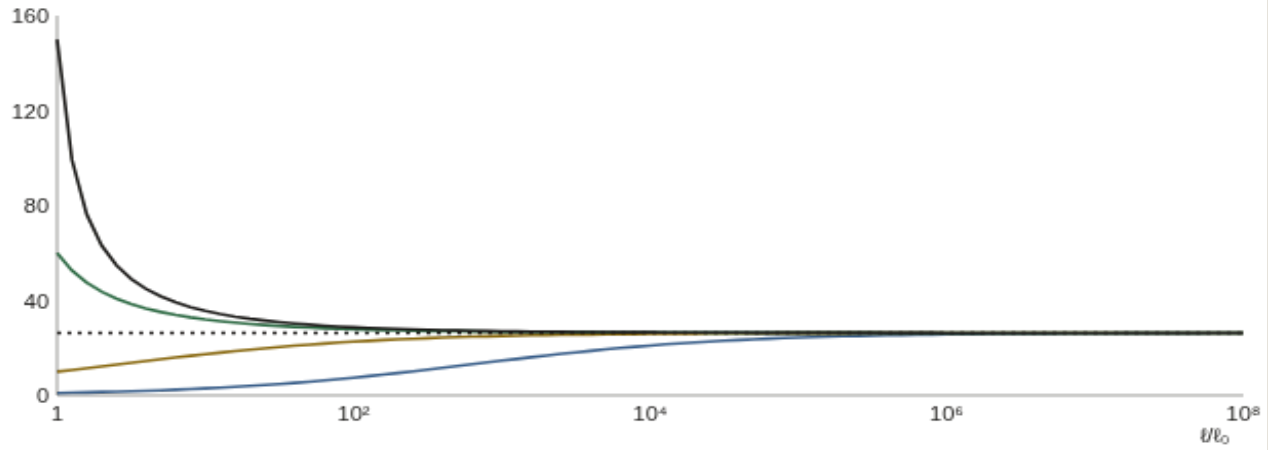
The coefficient of a local term of dimension $\Delta_{\mathcal{O}}$ has dimension $d - \Delta_{\mathcal{O}}$, and the dimensionless coefficient is multiplied by $b^{\,d-\Delta_{\mathcal{O}}}$ at each step towards long distances. The lecture notes (§ 3.2.2) divide them as 'Relevant if Δ_O < d', 'Irrelevant if Δ_O > d', 'Marginal if Δ_O = d' Known . Irrelevant terms are diluted to about $(E/\Lambda)^{\Delta_{\mathcal{O}}-d}$ at energy E, and at long distances only relevant and marginal terms remain. In four dimensions $\mathrm{tr}F^2$ ($\Delta = 4$) is marginal and $\mathrm{tr}F^4$ ($\Delta = 8$) is irrelevant (the question of 3-01). The expectation that the same continuum limit is reached regardless of the details of the lattice action (universality; 18-04) also comes from here Physics .

The direction of a marginal coupling is decided by quantum corrections. The lecture notes write 'most couplings which naively appear marginal will end up becoming either marginally relevant or marginally irrelevant due to such corrections', and give 'the Kondo effect and non-Abelian gauge theories' as examples that turn relevant (§ 3.4). ϕ^4_4 is on the side that weakens at long distances, Yang–Mills on the side that strengthens. With $t = \ln(\ell/\ell_0)$ towards long distances and the dimensionless $\hat{u} = u\,\ell^{\,4-d}$, extending d to real values,

$$\frac{d\hat{u}}{dt} = (4-d)\,\hat{u} - \frac{3}{16\pi^2}\,\hat{u}^2 \qquad (11)$$

(the formula of the lecture notes § 3.5 converted to the normalisation of $u/4!$; opposite in direction and in sign to (8)). For $d < 4$ there is a fixed point at $\hat{u}_* = 16\pi^2(4-d)/3$, and every positive initial value flows into it. This is the Wilson–Fisher 1972 fixed point, the smallest model of universality. For $d = 4$ it goes to 0 slowly, for $d > 4$ fast (7-08) Physics . A figure placing the $[g^2]$ of (2) and the fixed point of (11) next to other quantities that switch at four is the dimension knob of 1-12 § 10.

Figure 2 Moving the dimension d with the knob changes the flow of (11). The lines are $\hat{u}_0 = 1, 10, 60, 150$, the dotted line is the fixed point. For d<4 they gather to one point; for d≥4 they fall to 0. "Play" moves d back and forth between 3 and 5



`d = 3.50: û* = 26.32, at length ×10⁸: û = 26.25, 26.31, 26.32, 26.32`

Computed Physics The numbers in this figure were computed on this machine (closed-form solution; agreeing with RK4). Static values (length $\times 10^8$, $\hat{u}_0 = 1$ / 100): at $d = 3$, 52.6 / 52.6 ($\hat{u}_* = 52.64$); at $d = 4$, 0.741 / 2.78; at $d = 5$, 9.8×10^{-9} / 3.5×10^{-7}. The formula is a one-loop approximation, reliable when $4 - d$ is small.

Detour: non-renormalisable theories are useful too — the Fermi interaction

A subsection you may skip.

In the low-energy description of the weak interaction, the term in which four fermions meet at one point ($\Delta = 6$) carries $G_F = 1.1663788(6) \times 10^{-5}\ \mathrm{GeV}^{-2}$ (PDG 2024), and it is non-renormalisable. Even so, $G_F E^2$ is 1.2×10^{-5} at 1 GeV and reaches 1 only at $E = 292.8$ GeV Computed . Before that, the gauge theory of the W and Z (3-07) takes over. Such theories are useful as effective theories below some scale Known .

07

In gauge theories — subtracting without breaking the symmetry

One would like the counterterms of Yang–Mills, once added, to stay within the Yang–Mills form. The gauge-invariant local terms of dimension at most 4 are only $\mathrm{tr}F^2$ and $\mathrm{tr}F\tilde{F}$ (the θ term; a total derivative that does not enter perturbation theory; 3-05), and the mass term $\mathrm{tr}A_\mu A_\mu$ cannot be written (3-02). However, perturbation theory is done with the gauge-fixed action. There, instead of gauge invariance, BRST symmetry (the symmetry that remains after gauge fixing) constrains the form of the counterterms (4-08). As a result, the gauge field part is kept in the form $\mathrm{tr}F^2$. The divergences are absorbed into g and the field normalisations (including those of the ghost and gauge-fixing terms), and since there is one parameter, the integration constant of the running becomes the only scale Λ_{YM} (a different quantity from the cutoff Λ; 4-07). Gauge invariance also reduces divergences. The two-point correction of the gluon is

restricted to transverse form by the Ward (in the non-abelian case, Slavnov–Taylor) identities Known .

$$\Pi_{\mu\nu}(P) = \left(P^2\delta_{\mu\nu} - P_\mu P_\nu\right)\Pi(P^2) \quad (12)$$

Transverse means $P_\mu \Pi_{\mu\nu} = 0$, that is, there is no component along the momentum. Since the second power of P factors out, of the $\omega = 2$ in Table 2 only the logarithmically divergent Π remains. $\Lambda^2\delta_{\mu\nu}$, corresponding to a mass term, does not appear in gauge-invariant regularisations. A momentum cutoff breaks this, and the construction of Magnen–Rivasseau–Sénéor 1993 also says it 'breaks gauge invariance and requires gauge restoring counterterms'. The abstract of 't Hooft 1971 says that regulator fields that are not gauge invariant can also be used if gauge invariance can be restored by finitely many counterterms in the limit of large regulator mass, and that massless Yang–Mills fields 'appear to be renormalizable' in the usual sense. Gauge fixing and ghosts are in 4-08. The lattice keeps gauge invariance exactly at the stage of regularisation (5-01).

08

Renormalisation seen from mathematics — where the theorems are, and where they are not yet

The key is not to mix up three levels.

Table 3 Three levels of renormalisation

Level	Status	Chapter
That each order of perturbation theory is finite	Known Theorem. BPHZ (Bogoliubov–Parasiuk 1957, Hepp 1966, Zimmermann 1969), Epstein–Glaser 1973. For gauge theories, 't Hooft 1971, 't Hooft–Veltman 1972, Lee–Zinn-Justin 1972, Becchi–Rouet–Stora 1976. Another proof by the BV formalism, Costello 2011	4-05
Summing the series (four dimensions)	Physics For both ϕ^4 and Yang–Mills the series is thought to diverge, and no meaning is given to its sum	4-12
Summing the series (two and three dimensions)	Known For ϕ^4 in two dimensions, divergence of the series (Jaffe 1965) and Borel summability (Eckmann–Magnen–Sénéor 1975); for ϕ^4 in three dimensions, Borel summability (Magnen–Sénéor 1977); for the massive two-dimensional Gross–Neveu model, the Borel sum (Feldman–Magnen–Rivasseau–Sénéor 1986)	4-12

Level	Status	Chapter
The theory itself with the cutoff removed	Known ϕ^4 in two and three dimensions has been constructed. Four-dimensional Yang–Mills: uniform bounds on the effective action in finite volume, on the lattice, at small coupling (Bałaban's series of papers in the 1980s; the 1989 theorem carries an assumption on the effective coupling; what can be taken from uniform bounds is a convergent subsequence by compactness, and the limit is not necessarily unique). Foundations of a continuum construction with an infrared cutoff (Magnen–Rivasseau–Sénéor 1993). The continuum theory as the scaling limit of Wilson's lattice was built by the construction published in 2026-11 (at the level of the committee's assessment; 6-00). The uniqueness of the infinite-volume limit has not been shown (6-04)	7-01, 7-07

An early example of the first way of thinking is Dyson 1949 for quantum electrodynamics (an argument removing the divergences at each order by renormalising mass and charge; the complete proof came later with Weinberg 1960 and BPHZ; convergence of the series was not shown). On the third, the problem statement writes that for ϕ^4 in two and three dimensions the coefficients are chosen 'to depend on κ in the way that perturbation theory suggests' on the cutoff κ, and 'One then proves that the expectations converge as $\kappa \to \infty$, even though the specified constants α, . . . diverge.' (§ 6.2). This is the constructive version of (7) being finite even though the δu of (6) diverges.

For four-dimensional Yang–Mills, it is required that the correlation functions agree at short distances with the textbook predictions of asymptotic freedom and perturbative renormalisation theory (§ 4; quoted in 0-03 § 07). § 2 writes 'Asymptotic freedom ensures a decisive regularity in the case when classical Sobolev inequalities are borderline'. In this article's reading, "borderline" corresponds to "the coupling is dimensionless in four dimensions" of § 03 (18-04), and the sentence is an outlook, not a theorem. That handling the infinities is not itself the centre of the difficulty is in 16-08.

What this chapter can say

	Content
What is said	The calculations (1)–(11), the properties of the regularisations in Table 1, the threefold classification of renormalisability — as calculations and theorems of perturbation theory Known . Table 2, Figures 1 and 2, the numerical differentiation of the β coefficient, the running of QED and the numbers for G_F Computed
Conditional	(11) and Figure 2 are an approximation extending the one-loop formula to real dimensions. Universality remains a premise for four-dimensional Yang–Mills Physics . (12) is quoted at textbook level, and the source is bibliographic data only

	Content
What is not said	The sum of the perturbation series. The existence of the four-dimensional theory with the cutoff removed, the continuum limit of Yang–Mills, the mass gap (this site shows none of them)

Sources and reproduction

Item	Kind	Source / tools
Read in full text	Known	A. Jaffe, E. Witten, *Quantum Yang–Mills Theory* (the Clay problem statement) § 2, § 4, § 6.2, § 6.5. D. Tong, *Statistical Field Theory* (lecture notes) Chapter 3 § 3.2.2, § 3.4, § 3.5, § 3.5.3. J. Erler, A. Freitas, Electroweak Model…, in S. Navas et al. (PDG), Phys. Rev. D 110 (2024) 030001 (eqs. (10.8)(10.11), $\hat{\alpha}(M_Z)$) and Summary Tables
Read in abstract only	Known	Gell-Mann–Low, Phys. Rev. 95 (1954) 1300; Weinberg, Phys. Rev. 118 (1960) 838; Callan, Phys. Rev. D 2 (1970) 1541; Symanzik, CMP 18 (1970) 227; 't Hooft, Nucl. Phys. B 33 (1971) 173; 't Hooft–Veltman, Nucl. Phys. B 44 (1972) 189; 't Hooft–Veltman, Nucl. Phys. B 50 (1972) 318; Bollini–Giambiagi, Nuovo Cim. B 12 (1972) 20; Lee–Zinn-Justin, Phys. Rev. D 5 (1972) 3121; Jaffe, CMP 1 (1965) 127; Eckmann–Magnen–Sénéor, CMP 39 (1975) 251; Magnen–Sénéor, CMP 56 (1977) 237; Wilson, Phys. Rev. B 4 (1971) 3174; Wilson–Fisher, Phys. Rev. Lett. 28 (1972) 240 (INSPIRE)
Checked against the survey material	Known	Dyson, Phys. Rev. 75 (1949) 486, 1736 (abstracts); Magnen–Rivasseau–Sénéor, CMP 155 (1993) 325; Bogoliubov–Parasiuk, Acta Math. 97 (1957) 227; Hepp, CMP 2 (1966) 301; Zimmermann, CMP 15 (1969) 208; Epstein–Glaser, Ann. IHP A 19 (1973) 211; Costello, AMS (2011); Bałaban, CMP 122 (1989) 175, 355 (CMP 109 (1987) 249 bibliographic data only); Feldman–Magnen–Rivasseau–Sénéor, CMP 103 (1986) 67 (abstract)
Bibliographic data only	Known	Becchi–Rouet–Stora, Ann. Phys. 98 (1976) 287 (from the bibliography of the problem statement); Stueckelberg–Petermann, Helv. Phys. Acta 26 (1953) 499; Peskin–Schroeder, *An Introduction to Quantum Field Theory* (1995; the form of (12))
Table 2, Figures 1 and 2, numerical examples	Computed	python3 (scipy, enumeration, RK4; each under a minute). The figures' JS checked with node

Next chapter to read: 4-05 The algebra of divergences (BPHZ, dimensional regularisation, Connes–Kreimer). The sign of the β function is in 4-06.

The algebra of divergences — nested subtraction becomes a decomposition in a group, yet still does not leave formal power series

> This chapter organises established knowledge (sources carry the “Known” label in the text). This book's own investigation is mainly in Part 15 and Volume 3 (Parts 16–28 and the appendices).

What this chapter shows — Three prescriptions for subtracting nested divergences (BPHZ, Epstein–Glaser, dimensional regularisation). That subtraction becomes a Birkhoff decomposition of a group in the Hopf algebra of Connes–Kreimer (computed in a small model). Costello's families of effective actions, and where all this falls short of the mass gap.

Prerequisite chapters — 4-04 Divergences and renormalisation. Helpful: 1-11 (distributions), 4-03 (Feynman diagrams).

Said first — This is a guide to known results at the level of formal power series, and shows nothing about the existence of Yang–Mills theory or the mass gap. The small model is for seeing the algebraic structure, not a four-dimensional calculation.

Order of this chapter

1. Nested divergences — seen in a small model: subtracting only the whole leaves the outer scale
2. The forest formula — BPHZ: subtract from the inside, and count the sets to subtract by forests
3. Extension in position space — Epstein–Glaser: the freedom is δ and its derivatives
4. Moving the dimension continuously — dimensional regularisation: divergences become poles 1/ε. The MS scheme
5. The algebra of diagrams — the Hopf algebra of Connes–Kreimer: the coproduct writes the extraction
6. Birkhoff decomposition — subtraction becomes a decomposition in a group: counterterms are one factor. Detour: rooted trees and Runge–Kutta methods

7. Families of effective actions — Costello: a family of actions, one for each scale
8. Where the algebra reaches and where it does not: the measure and $\exp(-1/g^2)$ lie outside

---

01

Nested divergences — seen in a small model

In 4-04 we removed one-loop divergences by subtracting the value at a reference scale μ. What was subtracted was a number independent of the external momenta (a counterterm). From two loops on, a part of a diagram (a **subdiagram**) diverges on its own (a **subdivergence**). Two subdivergences are either **nested**, one containing the other, **overlapping**, sharing a part without either containing the other, or disjoint.

We use a small model with nesting only (Kreimer's toy model; the definition in § 3.2 of Panzer 2012). "One loop" is $\int_0^\infty dy/(y+s)$, where $s>0$ is the outer scale and y the loop variable. Like the four-dimensional bubble it diverges logarithmically, so we **regularise** by multiplying by y^{-z} (z is a small positive number, corresponding to the ε of § 04), and build nesting by putting the inner value into the outer integrand.

$$\varphi(\bullet)(s)=\int_0^\infty \frac{y^{-z}\,dy}{y+s}=s^{-z}F(z),\qquad F(z)=\frac{\pi}{\sin\pi z}=\frac{1}{z}+\frac{\pi^2}{6}z+\cdots \tag{1}$$

$$\varphi(\ell_2)(s)=\int_0^\infty \frac{y^{-z}}{y+s}\,\varphi(\bullet)(y)\,dy=s^{-2z}F(z)\,F(2z) \tag{2}$$

Writing the part with negative powers of z (the **pole part**) as pole, with $L=\ln(s/\mu)$ and $\mu=1$,

$$\operatorname{pole}\varphi(\bullet)=\frac{1}{z},\qquad \operatorname{pole}\varphi(\ell_2)=\frac{1}{2z^2}-\frac{L}{z} \tag{3}$$

Computed The pole of the double loop contains L. Subtracting only the pole of the whole, the subtracted quantity contains $\ln s$ and is not a local term that can be added to the action. **Example:** at $z=0.1$, $L=1$, $\varphi(\ell_2)=44.49$, and its pole part is $50-10=40$.

02

The forest formula — BPHZ

The cure is to subtract from the inside. In Bogoliubov's prescription, one first adds the term in which the subdivergence is replaced by its counterterm $S(\bullet) = -1/z$, and then subtracts the whole.

$$\bar{R}(\ell_2) = \varphi(\ell_2) + S(\bullet)\,\varphi(\bullet), \qquad \operatorname{pole}\bar{R}(\ell_2) = -\frac{1}{2z^2} \tag{4}$$

Computed L/z disappears, and the overall counterterm $S(\ell_2) = 1/(2z^2)$ does not depend on the outer scale ($\bar{R}$: Bogoliubov's map). For general diagrams the sum is over **forests** — sets of divergent subdiagrams any two of which are nested or disjoint (including the empty set and the whole diagram). Zimmermann's forest formula is

$$R\,\Gamma = \sum_{\mathcal{F}:\ \text{forests}} \prod_{\gamma\in\mathcal{F}} (-T_\gamma)\,\Gamma \tag{5}$$

where T_γ is the operation extracting the divergent part of γ (in momentum space, the low orders of the Taylor expansion in the external momenta), and the product acts from the inside (a form with details omitted). **Example:** the forests of ℓ_2 are $\varnothing, \{\bullet\}, \{\ell_2\}, \{\bullet, \ell_2\}$, and (5) is the expansion $(1 - T_{\ell_2})(1 - T_\bullet)\,\ell_2$. The $(1 - T_\bullet)$ part corresponds to the $\bar{R}$ of (4), and the remaining $-T_{\ell_2}$ to adding the overall counterterm $S(\ell_2) = 1/(2z^2)$, leaving a value without poles (the φ_+ of § 05). It is called BPHZ after the initials of Bogoliubov–Parasiuk (1957), Hepp (1966) and Zimmermann (1969), and the gaps in the first proof are said to have been filled by the work of the latter two (§ 1 of Ebrahimi-Fard–Kreimer 2005). The forest formula with overlaps also follows from the Hopf algebra of rooted trees (§ 05) (abstract of Kreimer 1999). Known

03

Extension in position space — Epstein–Glaser

This section uses the distributions of 1-11 § 04 and § 08. We give only the essentials. A **test function** g is a function that can be differentiated any number of times and goes to 0 quickly far away, and a **distribution** t is a rule assigning linearly a number $\langle t, g\rangle$ to each test function. $\langle \delta, g\rangle = g(0)$, and its derivative is defined by $\langle \partial^\alpha \delta, g\rangle = (-1)^{|\alpha|}\partial^\alpha g(0)$. A sequence of distributions "converges as distributions" if $\langle t_n, g\rangle$ converges for every test function g.

In position space, a divergence is the impossibility of extending a distribution to a singular point. We take examples in Euclidean $\mathbb{R}^4$. The square of the four-dimensional massless propagator $G = 1/(4\pi^2|x|^2)$ (the bubble diagram) behaves as $|x|^{-4}$ near the

origin, and its integral diverges logarithmically. The method of Epstein–Glaser (1973) was set up in Minkowski space (spacetime with time); it builds time-ordered products (whose expectations are the correlation functions) inductively in the number of vertices (§ 4 of Brunetti–Fredenhagen 2000), and **extends** values fixed off the **diagonal**, where points coincide, onto the diagonal. Renormalisation is exactly the extension of expectations to the diagonal (abstract of the same paper). Since it is formulated as extension, there is no need to write divergent integrals along the way. The extension problem itself has the same form in Euclidean $\mathbb{R}^d$, as below. The freedom is measured by the **scaling degree** ((33) of the same paper; numbering of the arXiv version; convergence as distributions).

$$\operatorname{sd}(t) = \inf\{\omega : \lambda^\omega\, t(\lambda x) \to 0 \ (\lambda \downarrow 0)\}, \qquad \operatorname{sd}(|x|^{-a}) = a, \quad \operatorname{sd}(\delta) = d \tag{6}$$

For t_0 on $\mathbb{R}^d \setminus \{0\}$ (Theorems 5.2 and 5.3 of the same paper) Known : if $\operatorname{sd}(t_0) < d$, the extension preserving the scaling degree is unique. If $d \le \operatorname{sd}(t_0) < \infty$, extensions exist, and the difference of two extensions is supported at the origin, of the form

$$t - t' = \sum_{|\alpha| \le \rho} c_\alpha\, \partial^\alpha \delta, \qquad \rho = \operatorname{sd}(t_0) - d \tag{7}$$

ρ corresponds to the superficial degree of divergence of 4-04, and the constants c_α are the freedom of local counterterms. **Example:** for $t_0 = |x|^{-4}$ ($d = 4$, $\rho = 0$), the extension that subtracts $g(0)$ only inside the ball of radius $1/\mu$,

$$\langle u_\mu, g\rangle = \int_{\mathbb{R}^4} \frac{g(x) - g(0)\,\mathbf{1}\{|x| < 1/\mu\}}{|x|^4}\, d^4x \tag{8}$$

agrees with t_0 for $x \neq 0$, and converges because the numerator is $O(|x|)$. The difference on changing μ is $2\pi^2 \ln(\mu'/\mu)\, g(0)$ ($2\pi^2$ is the area of S^3), a constant multiple of δ.

Example: for $g_s = e^{-|x|^2/s^2}$, $\langle u_\mu, g_s\rangle = 2\pi^2(\ln(\mu s) - \gamma/2)$ (γ is Euler's constant). At $\mu = 1$, for $s = 0.1, 1, 10$ it is −51.148, −5.697, 39.754, and the difference from $\mu = 2$ is $2\pi^2 \ln 2 = 13.682$ in each case. Removing $|x| < \eta$ without extending gives 85.21, 357.91 at $\eta = 10^{-2}, 10^{-8}$ ($s = 1$), growing. Computed

04

Moving the dimension continuously — dimensional regularisation

We regard momentum integrals as functions of the dimension d. The bubble of 4-04 (mass m) converges for $d < 4$, giving

$$I(d)=\int \frac{d^dk}{(2\pi)^d}\,\frac{1}{(k^2+m^2)^2}=\frac{\Gamma(2-d/2)}{(4\pi)^{d/2}}\,(m^2)^{d/2-2} \quad (9)$$

($1/(8\pi m)$ at $d=3$). The right-hand side extends **meromorphically** to complex d (as a function complex-differentiable except at isolated points called poles), and has a pole at $d=4,6,\dots$. Using this is **dimensional regularisation**. Here we write $d=4-2\varepsilon$ (half the ϵ of the $d=4-\epsilon$ of 4-04). Then

$$16\pi^2\,I=\frac{1}{\varepsilon}-\gamma+\ln 4\pi-\ln m^2+O(\varepsilon) \quad (10)$$

and the divergence becomes the pole $1/\varepsilon$. At $\varepsilon=0.1, 0.01, 0.001$, $16\pi^2 I$ ($m=1$) is 12.2535, 101.9814, 1001.9565, and the differences from the approximation (10) are 0.2997, 0.0276, 0.0027. The role of $\ln k_{\max}^2$ in the $\ln(k_{\max}^2/m^2)-1$ cut at $|k|<k_{\max}$ in momentum is played by $1/\varepsilon$. Computed The **MS scheme** subtracts only $1/\varepsilon$, and the **$\overline{\text{MS}}$ scheme** subtracts $1/\varepsilon-\gamma+\ln 4\pi$. Since a scale μ is introduced (the coupling is multiplied by $\mu^{2\varepsilon}$), $-\ln(m^2/\mu^2)$ remains at $\overline{\text{MS}}$. The $\Lambda_{\overline{\text{MS}}}$ of 4-06 is the Λ of this scheme.

In 1972 there were papers by Bollini–Giambiagi and by 't Hooft–Veltman (titled "Regularization and renormalization of gauge fields"), and it is said to be used as standard in Yang–Mills perturbation theory as a regularisation that does not break gauge symmetry. The price is that (9) is an analytic continuation diagram by diagram, and gives no measure for non-integer d. The problem statement too writes that there are few ways to recover the reflection positivity lost by dimensional regularisation and the like (p.11; quoted in 0-03 § 03; 4-10).

05

The algebra of diagrams — the Hopf algebra of Connes–Kreimer

Subtraction is the repetition of "extracting a subdiagram γ" and "the remainder Γ/γ with γ collapsed to a point". Before making this an operation, we prepare three words.

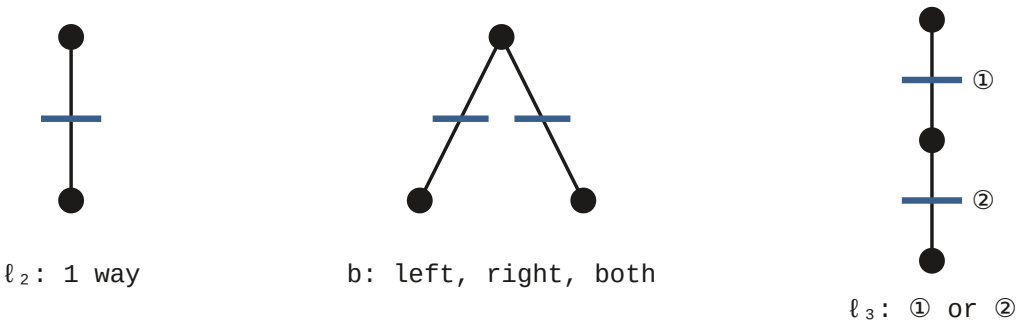
- **One-particle irreducible**: a diagram that does not split in two when any one internal line is cut.
- $\otimes$ (tensor product): $\gamma\otimes\Gamma/\gamma$ is a symbol that formally sets a pair of diagrams side by side. It is linear in each of left and right (bilinear), and does not multiply left and right. When values are put in later, it is multiplied as $\varphi(\gamma)\,\varphi(\Gamma/\gamma)$.
- **Hopf algebra**: an algebra with a product, a **coproduct** (a map splitting one element into a sum of pairs $\otimes$), and auxiliary maps called the counit and the antipode. This chapter uses only the coproduct.

Making subtraction an operation is the Hopf algebra of Kreimer (1998) and Connes–Kreimer (1998, 2000): the generators are divergent one-particle irreducible diagrams, the product is setting diagrams side by side, and extraction is the coproduct ((15) of Ebrahimi-Fard–Kreimer 2005). The sum in (11) is over divergent one-particle irreducible proper subdiagrams and their disjoint sets (set side by side as a product). For a set, Γ/γ is the remainder with all of them collapsed to points. Writing the coproduct as Δ is a convention of the literature, unrelated to the Δ of the mass gap.

$$\Delta(\Gamma) = \Gamma \otimes 1 + 1 \otimes \Gamma + \sum_{\gamma \subsetneq \Gamma} \gamma \otimes \Gamma/\gamma \quad (11)$$

With nesting only, **rooted trees** suffice in place of diagrams. Vertices are divergent subdiagrams, and children are inner subdivergences. The sum in (11) becomes a sum over **admissible cuts** (choices of edges that cut every path from the root to a leaf at most once), with the cut-off side going on the left and the root side on the right (Figure 1).

Figure 1 Rooted trees and admissible cuts. The top is the root (the whole diagram), lower is further inside. Coloured lines are cuts; the lower side cut off is the left factor of (12), the root side the right factor. In ℓ_3, ① and ② are on the same path, so they are not cut together



$$\Delta(\ell_2) = \ell_2 \otimes 1 + 1 \otimes \ell_2 + \bullet \otimes \bullet$$
$$\Delta(b) = b \otimes 1 + 1 \otimes b + 2\ \bullet \otimes \ell 2 + \bullet\bullet \otimes \bullet \quad (12)$$

b is the tree with two children at the root. Regularised values φ (maps taking values in Laurent series in z and carrying products to products; **characters**) form a group under the convolution $\varphi_1 \star \varphi_2 = m \circ (\varphi_1 \otimes \varphi_2) \circ \Delta$ (m is the product of values). The unit is the character that gives 1 to the empty diagram and 0 to other diagrams. Since there are inverses, the problem can be posed as multiplying φ by a factor S_R that cancels its divergences to make $S_R \star \varphi = \varphi_+$ finite. The counterterm S_R plays the role of the inverse of φ (§ 06). Fixing a scheme R (for MS the pole part, for MOM the value at $s = \mu$), the counterterm S_R and the renormalised value φ_+ are fixed in order of tree size ((21)(25)(26) of Ebrahimi-Fard–Kreimer 2005).

$$S_R(\Gamma) = -R\big[\bar{R}(\Gamma)\big], \qquad \varphi_+(\Gamma) = (1-R)\big[\bar{R}(\Gamma)\big]$$
$$\bar{R}(\Gamma) = \varphi(\Gamma) + \sum_{\gamma} S_R(\gamma)\,\varphi(\Gamma/\gamma) \tag{13}$$

In the toy model, writing the subtree rooted at vertex v as t_v and the number of vertices as $|t|$, the values take closed form (Proposition 3.2.2 of Panzer 2012).

$$\varphi(t)(s) = s^{-z|t|} \prod_{v} F\big(z\,|t_v|\big) \tag{14}$$

Computing (13) for the 8 trees with up to 4 vertices, (i) none of the counterterms contains L (locality), (ii) the renormalised values are polynomials of degree $|t|$ in L with top-degree term $(-L)^{|t|}/t!$ ($t! = \prod_v |t_v|$ is the **tree factorial**), and (iii) the MOM values agreed with Panzer's (3.4.2)–(3.4.5). The difference between MS and MOM is finite. Computed

Computed Table 1 Renormalisation in the toy model (computed exactly in rationals). The right-hand column is the value at $z \to 0$. Subtracting by MS changes the constants (ℓ_2 becomes $\frac{L^2}{2} + \frac{\pi^2}{4}$, ℓ_3 becomes $-\frac{L^3}{6} - \frac{5\pi^2 L}{12}$).

Tree	$t!$	Pole part of the bare value	Counterterm (MS)	φ_+ (MOM)
$\bullet$	1	$\frac{1}{z}$	$-\frac{1}{z}$	$-L$
ℓ_2	2	$\cancel{\frac{1_2}{2z^2}} - \frac{L}{z}$	$\cancel{\frac{1_2}{2z^2}}$	$\frac{L^2}{2}$
b	3	$\cancel{\frac{1_3}{3z^3}} - \cancel{\frac{L_2}{z^2}} + \frac{27L^2+11\pi^2}{18z}$	$-\cancel{\frac{1_3}{3z^3}} + \frac{\pi^2}{18z}$	$-\frac{L^3}{3} - \frac{\pi^2 L}{3}$
ℓ_3	6	$\cancel{\frac{1_3}{6z^3}} - \cancel{\frac{L_2}{2z^2}} + \frac{27L^2+14\pi^2}{36z}$	$-\cancel{\frac{1_3}{6z^3}} - \frac{\pi^2}{18z}$	$-\frac{L^3}{6} - \frac{\pi^2 L}{6}$

06

Birkhoff decomposition — subtraction becomes a decomposition in a group

The main theorem of Connes–Kreimer (2000) (abstract) is that the dimensionally regularised bare theory gives a loop $\gamma(z)$ in the group of characters on a small circle around the integer spacetime dimension D, and 'the renormalized theory is just the evaluation at z = D of the holomorphic part γ+ of the Birkhoff decomposition of γ'. $z = D$ is the integer dimension of spacetime, γ is the loop given by the regularised theory (the family of characters as the dimension z moves on the circle), and γ_+ is the holomorphic factor of its decomposition. Known Below we write z as the deviation from D (the same role as the z of the toy model, with $z = 0$ the integer dimension).

$$\varphi(z) = \varphi_-(z)^{\star -1} \star \varphi_+(z) \tag{15}$$

φ_- is a polynomial in $1/z$ (constant term the unit) and is the MS counterterm; φ_+ is holomorphic at $z = 0$, and $\varphi_+(0)$ is the renormalised value. As an analogy, it is the

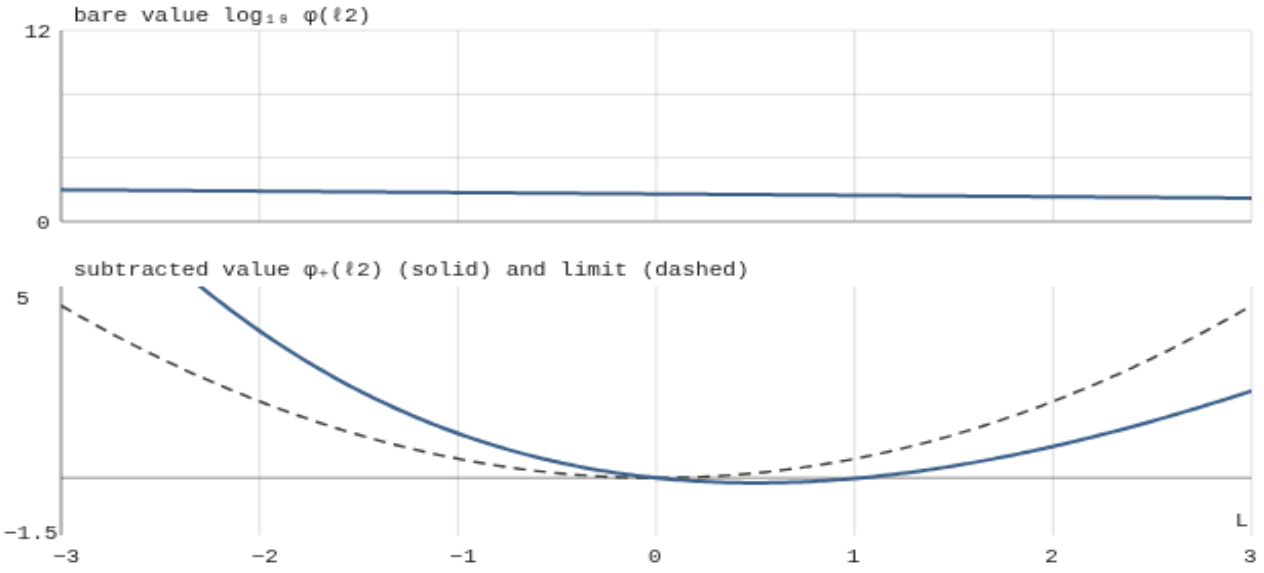
multiplicative version of splitting a Laurent series into its pole part and its holomorphic part. Precisely, since the group is not commutative, they are fixed in order of tree size by the induction of (13). That S_R and φ_+ are again characters is because R satisfies $R^2 = R$ and the **Rota–Baxter identity**

$$R(x)\,R(y) + R(xy) = R\big(R(x)\,y + x\,R(y)\big) \quad (16)$$

((20) of Ebrahimi-Fard–Kreimer 2005). Known **Example:** for the pole part R and $x = 1/z + a$, $y = 1/z + b$, both sides are $2/z^2 + (a+b)/z$.

Figure 2 is the MOM of the ladder ℓ_n (n vertices in a row). The sum in (13) becomes $\bar{R}(\ell_n) = \varphi(\ell_n) + \sum_{j<n} S_R(\ell_j)\,\varphi(\ell_{n-j})$; as $z \to 0$ the bare value runs away at the rate z^{-n}, while the subtracted value settles to a polynomial. Computed

Figure 2 Making the knob z smaller, the bare value $\log_{10}\varphi(\ell_n)$ in the upper panel runs away upwards, and the subtracted value $\varphi_+(\ell_n)$ in the lower panel (solid line) approaches a polynomial in $z \to 0$ (dashed line). The horizontal axis is L, and n is the number of rungs of the ladder. "Play" lowers z



n = 2, z = 0.100, at L = 1, bare value 44.49, subtracted value −0.0141, limit 0.5000

Computed The numbers in this figure were computed on this machine. The closed forms of (13)(14) are evaluated in the figure and checked against python3.

If JavaScript is disabled: at $L = 1$, the subtracted values at $z = 0.02,\ 0.005$ are, for $n = 1$: −0.991, −0.998 (limit −1); $n = 2$: 0.394, 0.473 (0.5); $n = 3$: −1.696, −1.776 (−1.812); $n = 4$: 0.655, 1.419 (1.687). The bare values at $z = 0.02$ are 49, 1205, 19802, 245195. The difference from the limit shrinks roughly in proportion to z. Computed

Detour: rooted trees and Runge–Kutta methods

A subsection you may skip.

Expanding the solution of $y' = f(y)$ in h, reading the derivative of order k as "a vertex with k children", $f,\ f'f,\ f''(f,f),\ f'f'f$ becomes $\bullet,\ \ell_2,\ b,\ \ell_3$.

$$y(h) = y + hf + \frac{h^2}{2} f'f + \frac{h^3}{6}\big(f''(f,f) + f'f'f\big) + O(h^4) \tag{17}$$

The coefficient of h^3 is $1/t!$ for ℓ_3, and for b the reciprocal of the product of 2, from exchanging left and right, and $t! = 3$; this is the same tree factorial as in Table 1. Runge–Kutta methods are built to reproduce these coefficients up to a given order, and the group handling the expansion is called Butcher's group (Butcher 1972). Brouder (2000), in his abstract, points out the connection between the algebra of rooted trees in renormalisation and Runge–Kutta methods. Known

07

Families of effective actions — Costello

Another route is to regard the family of actions, one for each scale, as the theory (Wilson's renormalisation group; 7-05). Costello (a 2007 paper, a 2011 book) restricts the t (the square of a length) of the propagator $1/k^2 = \int_0^\infty e^{-tk^2}\,dt$ from ε to T.

$$P(\varepsilon, T)(k) = \int_\varepsilon^T e^{-tk^2}\,dt = \frac{e^{-\varepsilon k^2} - e^{-Tk^2}}{k^2} \tag{18}$$

Precisely, the paper is written with the heat kernel on a compact manifold, which on flat space becomes e^{-tk^2}. **Example:** at $\varepsilon = 0.01$, $T = 100$, $k^2 P$ is 0.990, 0.368, 1.2×10^{-4} at $k = 1, 10, 30$. Computed Actions at different scales are linked by the renormalisation group equation ($\Gamma(P, S)$ is the sum of diagrams joining vertices of S by P).

$$S_{\mathrm{eff}}(T_2) = \Gamma\big(P(T_1, T_2),\, S_{\mathrm{eff}}(T_1)\big) \tag{19}$$

The Counterterm Uniqueness Theorem of the paper (Theorem B of Costello) is that, once a scheme is fixed (a choice of which functions of ε to call the singular part), there is a unique counterterm $S^{\mathrm{CT}}(\varepsilon)$ such that $\lim_{\varepsilon\to 0} \Gamma(P(\varepsilon, T), S - S^{\mathrm{CT}})$ exists, and the counterterms are local and independent of T. The counterterms are power series in $\hbar$ (the letter counting loops). The **effective action** at scale T is the limit as $\varepsilon \to 0$ after subtracting these counterterms,

$$\Gamma^R\big(P(0,T), S\big) = \lim_{\varepsilon\to 0} \Gamma\big(P(\varepsilon, T),\, S - S^{\mathrm{CT}}(\varepsilon)\big) \tag{20}$$

and building diagrams with $P(0,T)$ as it is diverges. The effective action too is a formal power series in $\hbar$ (§ 08). The Classification Theorem for Families of Effective Actions (Theorem C of Costello) is that, once a scheme is fixed, local actions correspond one-to-one to families of effective actions satisfying (19) and locality as $T \to 0$. Changing the scheme amounts to putting different coordinates on the same family. Known

Gauge symmetry is expressed by the quantum master equation of the Batalin–Vilkovisky (BV) formalism (here only the name; it gathers the condition of keeping gauge symmetry into one equation with auxiliary fields added). The 2007 paper distinguishes renormalisation (constructing counterterms) from quantisation (constructing solutions of the quantum master equation), and gives pure Yang–Mills on compact four-dimensional manifolds as an example of the former, and Chern–Simons theory as a non-trivial example of the latter. The 2011 book is described as containing a proof of the perturbative renormalisability of Yang–Mills by cohomology (an algebraic tool that looks at $\ker Q / \operatorname{im} Q$ for maps Q with $Q^2 = 0$; here only the name). Renormalisability itself is said to have been established by 't Hooft (1971), 't Hooft–Veltman (1972), Lee–Zinn-Justin (1972), Becchi–Rouet–Stora (1976) and others (7-12 § 06). Known

08

Where the algebra reaches and where it does not

All the tools so far work on **formal power series** in the coupling constant, make each order finite, and confine the freedom to finitely many local terms. Known Three things lie out of reach. **(i) The sum of the series:** the perturbation series of four-dimensional Yang–Mills is thought to diverge (4-12). Physics **(ii) Quantities that are 0 at every order:** the one-loop scale (4-07; for pure $SU(N)$, $\beta_0 = 11N/3$; conversion between four ways of writing the coefficient in Table 1 of 4-06)

$$\Lambda = \mu \, \exp\!\Big(-\frac{8\pi^2}{\beta_0 \, g^2}\Big) \quad (21)$$

(with $b_0 = \beta_0/(16\pi^2)$, the same formula as (3) of 4-07) has all its Taylor coefficients in g equal to 0. The mass gap is expected to be proportional to Λ Physics , and such quantities cannot be distinguished from 0 within formal power series (7-12 § 07). **(iii) The measure and positivity:** what the problem statement asks for is not a sequence of coefficients but a theory satisfying the axioms (0-03), whose correlation functions should agree at short distances with asymptotic freedom and the predictions of perturbative renormalisation theory (p.6; quoted in 0-03 § 07). The algebra of this chapter is on the side of the answers that a construction must reproduce.

An example of the same algebra being used inside a tool outside formal power series is the renormalisation of Hairer's regularity structures. Regularity structures are a framework for solving the equations of **stochastic quantisation** (7-09; here only the name; a method that solves a time-evolution equation shaken by noise and tries to obtain the measure as its distribution after a long time), and their renormalisation is written with a pair of Hopf algebras and a variant of the Birkhoff decomposition (abstract of Bruned–Hairer–Zambotti 2019). Regularity structures have been used to

construct local-in-time solutions (solutions only for a short time) of the stochastic quantisation equation of three-dimensional Yang–Mills–Higgs (abstract of Chandra–Chevyrev–Hairer–Shen 2024). Known What they can handle is subcritical equations (the non-linear term weakens at shorter scales), and four-dimensional Yang–Mills is **critical** (the strength of the non-linear term does not depend on scale), so it lies outside.

What this chapter can say

	Content
What is said	The five formulations and the theorems quoted (classification of extensions, the main theorem on Birkhoff decomposition, Costello's Counterterm Uniqueness Theorem and Classification Theorem for Families of Effective Actions) Known . In the toy model: that subtracting only the whole leaves poles containing L and that Bogoliubov's prescription removes them; that for 8 trees the counterterms are local and the values are polynomials in L. That the difference of extensions of $\lvert x\rvert^{-4}$ is a constant multiple of δ. The pole of the bubble Computed
Conditional	The toy model has nesting only, and locality is checked within it (in general it rests on the theorems in the literature). The divergence of the perturbation series and $\Delta \propto \Lambda$ are expectations Physics
What is not said	Multi-loop calculations in four-dimensional Yang–Mills. The sum of the series. The non-perturbative measure, reflection positivity, the mass gap. The content of the proofs in literature cited by bibliographic data only

Sources and reproduction

Item	Kind	Source / tools
Read in full text	Known	Jaffe–Witten, Quantum Yang–Mills Theory (the Clay problem statement) p.6, p.11. Panzer, arXiv:1202.3552 (2012) § 3.2–3.4. Ebrahimi-Fard–Kreimer, J. Phys. A 38 (2005) R385 § 1–2. Brunetti–Fredenhagen, arXiv:math-ph/9903028 (Commun. Math. Phys. 208 (2000) 623), abstract, § 4, § 5.1–5.2 (numbering of formulas and theorems from the arXiv version). Costello, arXiv:0706.1533 (2007) § 1
Read in abstract only	Known	Kreimer, Adv. Theor. Math. Phys. 2 (1998) 303. Connes–Kreimer, Commun. Math. Phys. 199 (1998) 203; 210 (2000) 249. Kreimer, Commun. Math. Phys. 204 (1999) 669. Brouder, Eur. Phys. J. C 12 (2000) 521. Bruned–Hairer–Zambotti, Invent. Math. 215 (2019) 1039. Chandra–Chevyrev–Hairer–Shen, Invent. Math. 237 (2024) 541 (arXiv:2201.03487)

Item	Kind	Source / tools
Bibliographic data only	Known	Bogoliubov–Parasiuk, Acta Math. 97 (1957) 227. Hepp, Commun. Math. Phys. 2 (1966) 301. Zimmermann, ibid. 15 (1969) 208. Epstein–Glaser, Ann. Inst. H. Poincaré A 19 (1973) 211. Bollini–Giambiagi, Nuovo Cim. B 12 (1972) 20. 't Hooft–Veltman, Nucl. Phys. B 44 (1972) 189 (standard use in gauge theories not checked). 't Hooft, ibid. 33 (1971) 173. Lee–Zinn-Justin, Phys. Rev. D 5 (1972) 3121. Becchi–Rouet–Stora, Ann. Phys. 98 (1976) 287. Costello, Renormalization and Effective Field Theory, AMS (2011) (from the publisher's description). Butcher, Math. Comp. 26 (1972) 79
Equations (3)(4), Table 1	Computed	python3 (Laurent series in z computed exactly with the standard library's rationals; 8 trees, MS and MOM)
Figure 2, the numbers in § 03, § 04, § 06 and § 07	Computed	python3 (SciPy numerical integration and the Γ function, closed formulas). The JavaScript of Figure 2 was checked against the python3 values with node

Next chapter to read: 4-06 Asymptotic freedom. Detours: 7-12, 4-12.

Asymptotic freedom — the sign that weakens the coupling at short distances, and what is inside 11N/3

This chapter organises established knowledge (sources carry the “Known” label in the text). This book's own investigation is mainly in Part 15 and Volume 3 (Parts 16–28 and the appendices).

What this chapter shows — The β function of Yang–Mills is negative, and the coupling weakens logarithmically at shorter distances. The coefficient 11/3 splits into 4 from spin and −1/3 from orbital motion. b_0, b_1 do not depend on the definition of the coupling, and on that footing lattice numbers can be compared with the prediction.

Prerequisite chapters — 4-04 (the running coupling). Helpful: 17-01, 3-02, 5-01. $\hbar = c = 1$, $\mathrm{tr}(T^a T^b) = \frac{1}{2}\delta^{ab}$. We write the β function as $\beta(g)$ and the inverse lattice coupling as $\beta_{\mathrm{L}} = 2N/g_0^2$, to keep them apart.

Said first — The coefficients are perturbative calculations (Known here means "known as a perturbative calculation"). The four-dimensional continuum theory was built by the construction published in 2026-11, and it includes the one-loop running of the coupling and several short-distance checks as statements (the committee's assessment is that it is proved in Lean; 10-06, 6-01). Agreement with perturbation theory at all orders has not been shown (6-04). The comparison with the lattice in this chapter is a comparison of numbers, and does not in itself show that the continuum limit exists. This site shows neither the mass gap nor the construction of the continuum theory.

Order of this chapter

1. The running coupling — the sign of the β function
2. The one-loop coefficient — four ways of writing it (Table 1)
3. What is inside 11/3 — spin and orbital motion (Figure 1, Table 2) / Detour: the vacuum is a paramagnet
4. Adding matter — Coleman–Gross
5. Two loops and universality — b_0, b_1 do not depend on the definition
6. How it was discovered — a timeline (Table 3)
7. Agreement with the lattice — the ratio of Λs and r_0/a (Figure 2, Table 4)

8. Asymptotic freedom in the problem statement — what it says and what it does not (Table 5)

01

The running coupling — the sign of the β function

The renormalised coupling g has a value only once the scale μ at which it is measured is fixed (a momentum; as a distance, roughly $1/\mu$) (4-04). How it changes is the **β function**, and the right-hand side is an expansion in small g.

$$\beta(g) = \mu \frac{dg}{d\mu} = -b_0\, g^3 - b_1\, g^5 - \cdots \quad (1)$$

If $\beta > 0$, the coupling grows at short distances (large μ); QED and four-dimensional ϕ^4 are on this side (7-08). If $b_0 > 0$, it weakens at short distances. This is **asymptotic freedom**. Dropping b_1 and higher and integrating, $1/g^2$ becomes a linear function of $\ln \mu$ (the same formula as in 17-01).

$$\frac{1}{g^2(\mu)} = \frac{1}{g^2(\mu_0)} + 2b_0 \ln \frac{\mu}{\mu_0} \quad (2)$$

Formula (2) means that $1/g^2$ grows in proportion to the logarithm of the factor by which μ has grown. Whether μ is multiplied by 10 or by 100, it grows in fixed steps, and the coupling weakens only slowly.

Example. Writing α for the square of the coupling divided by 4π, one-loop QED with electrons only gives $d(1/\alpha)/d\ln\mu = -2/(3\pi) \approx -0.212$ (a rewriting of the formula in § 2.4.1 of Tong's lecture notes), and pure SU(3) gives $+1.751$. In the latter, $1/\alpha$ grows by about 4.03 each time μ grows by a factor of 10 (0-03) Computed .

02

The one-loop coefficient — four ways of writing 11N/3

For a gauge group with n_f Dirac particles (representation R_f) and n_s complex scalars (representation R_s) added, the one-loop coefficient is as follows (§ 2.4 and § 2.7 of Tong's lecture notes; formula (5.3) of 't Hooft 1999 is the same formula with a different normalisation) Known .

$$16\pi^2\, b_0 = \frac{11}{3}\, C_A - \frac{4}{3}\, T(R_f)\, n_f - \frac{1}{3}\, T(R_s)\, n_s \quad (3)$$

$T(R)$ is the number fixed by $\mathrm{tr}_R(T^a T^b) = T(R)\,\delta^{ab}$: $\frac{1}{2}$ for the fundamental representation and C_A for the adjoint representation (N for SU(N); 1-04). For pure SU(N), $b_0 = \frac{11N}{48\pi^2}$. The apparent numbers change with the choice of variable, so we list them in Table 1 (the four rows say the same thing).

Table 1 Four ways of writing the one-loop coefficient of pure SU(N)

Variable	Flow equation	Coefficient	SU(2)	SU(3)
g	$\mu\, dg/d\mu = -b_0 g^3$	$\frac{11N}{48\pi^2}$	0.04644	0.06966
$1/g^2$	$d(1/g^2)/d\ln\mu = 2b_0$	$\frac{11N}{24\pi^2}$	0.09288	0.1393
$\alpha = g^2/(4\pi)$	$\mu^2 d\alpha/d\mu^2 = -b_0^{(\alpha)}\alpha^2$	$\frac{11N}{12\pi}$	0.5836	0.8754
$\beta_{\mathrm{L}} = 2N/g^2$	$d\beta_{\mathrm{L}}/d\ln\mu = \kappa$	$\kappa = 4Nb_0 = \frac{11N^2}{12\pi^2}$	0.3715	0.8359

Computed Evaluated with python3. $b_0^{(\alpha)} = 33/(12\pi)$ for SU(3) agrees with formula (9.3) of the PDG and the formula for b_0 right after it.

The last row is in the language of the lattice (5-01). The one-loop prediction is that the same physics is kept if β_{L} is raised by $\kappa \ln 2$ (0.258 for SU(2), 0.579 for SU(3)) each time the lattice spacing a is halved Computed . $\kappa = 11/(3\pi^2)$ for SU(2) is the same number as the slope κ in the record of this book's investigation, and is the known one-loop coefficient written in lattice conventions (23-02) Known .

03

What is inside 11/3 — 4 from spin and −1/3 from orbital motion

Why is the sign reversed? The **background field method** places a slowly varying external field (the background field), integrates out the quantum fluctuations around it at one loop, and looks at how the coefficient $1/g^2$ of the action of the background field changes. Calculated this way (§ 2.4.2 of Tong's lecture notes), the gluon contribution splits in two. The term that counts the components of the field like scalars gives, after the excess is subtracted with ghosts, $-\frac{1}{3}$ in the screening direction, the same as in electromagnetism. The term that comes from spin 1 (the lecture notes' 'the paramagnetic piece') gives $+4$ in the antiscreening direction. The overall minus sign comes from here ('This is where the overall minus sign comes from') Known .

$$\frac{11}{3} = 4 - \frac{1}{3} \quad (4)$$

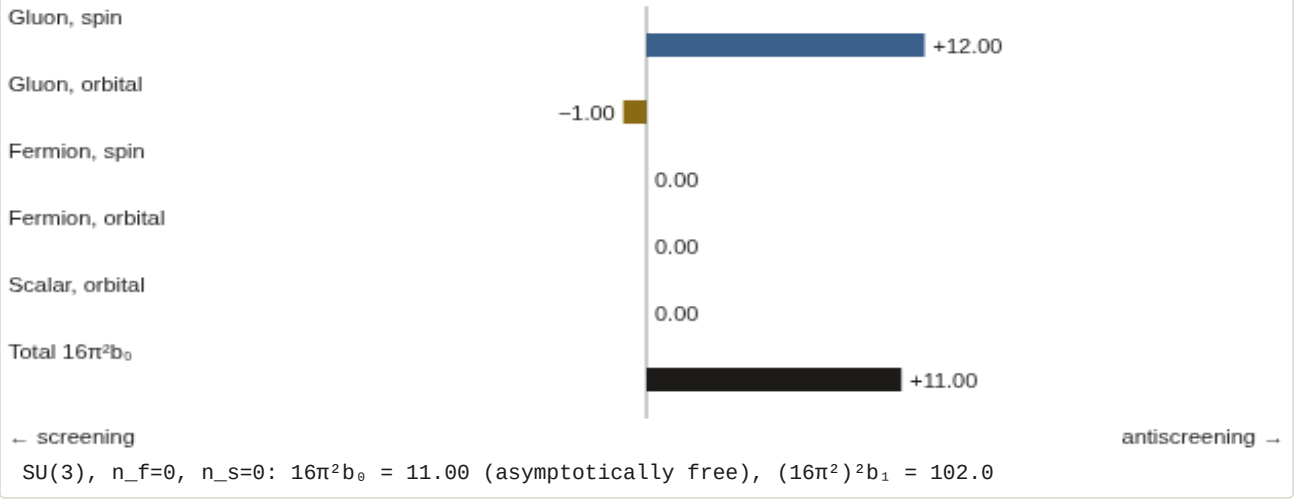
The photon carries no charge and takes no part in the vacuum polarisation of QED, but the gluon carries colour charge, so it does take part (3-02). Wilczek's Nobel lecture also writes 'it is the effect of virtual gluons that is responsible for antiscreening'.

The same split can be used for matter. We count each state of a massless particle (spin s, component s_z along the magnetic field) as contributing to $16\pi^2 b_0$ the amount

$$\tfrac{1}{2}\,(-1)^{2s}\big((2s_z)^2-\tfrac{1}{3}\big)\,T(R) \quad (5)$$

each. $(2s_z)^2$ is the spin (paramagnetic) part, $-\frac{1}{3}$ the orbital (diamagnetic) part, and $(-1)^{2s}$ the sign of a fermion loop. A gluon ($s_z=\pm1$) gives $\frac{11}{3}C_A$, a Dirac particle $-\frac{4}{3}T$, and a complex scalar $-\frac{1}{3}T$, and all of formula (3) comes out (Table 2). In this chapter we set down formula (5) as a way of counting that reproduces the known coefficients. That this rule reproduces all the coefficients of formula (3) was checked on this machine Computed . The derivation of the rule itself is not treated in this chapter. The analogy that views asymptotic freedom as paramagnetism is said to go back to Nielsen 1981 (Tong's note on the literature; Tong also mentions a private attribution to 't Hooft).

Figure 1 Moving the sliders for N, the number n_f of Dirac particles in the fundamental representation and the number n_s of complex scalars changes the contributions to $16\pi^2 b_0$ and their total. The right (positive) is the asymptotically free side. "Play" moves n_f back and forth



SU(3), n_f=0, n_s=0: $16\pi^2 b_0$ = 11.00 (asymptotically free), $(16\pi^2)^2 b_1$ = 102.0

Table 2 Contribution per particle by formula (5) (before multiplying by $T(R)$)

Particle	States	Spin	Orbital	Total = known coefficient
Gluon ($\times C_A$)	2	+4	−1/3	+11/3
Dirac particle	4	−2	+2/3	−4/3
Complex scalar	2	0	−1/3	−1/3

Computed The numbers in this figure and table were computed on this machine (formula (5) evaluated with python3 rationals, agreeing with the known coefficients; the figure adds up the same formula on the spot). If JavaScript is disabled, this table can be read instead.

Detour: the vacuum is a paramagnet

A subsection you may skip.

In the relativistic vacuum the speed of light is 1, so the product of permittivity and permeability is $\varepsilon\mu_{\mathrm{m}} = 1$. If it is paramagnetic ($\mu_{\mathrm{m}} > 1$), then $\varepsilon < 1$, and charge is antiscreened (Tong § 2.4.1). In a free electron gas, Pauli paramagnetism from spin and Landau diamagnetism from orbital motion are in the ratio $1 : -\frac{1}{3}$ (Tong's statistical physics § 3.6), the same ratio as the bracket $1 - \frac{1}{3}$ in formula (5). For fermions in the vacuum, however, the loop sign $(-1)^{2s}$ flips the whole, so the contribution of the electron is on the screening side ($-\frac{4}{3}$ in Table 2). For the gluon, $4 - \frac{1}{3}$. Speaking in metaphor, the vacuum of gluons is a strong paramagnet. Strictly, "spin" and "orbital" in formula (5) are names of terms in the calculation, not an observable such as the magnetisation of the vacuum.

04

Adding matter — $n_f < 11N/2$ and Coleman–Gross

Matter lowers b_0. With n_f Dirac particles in the fundamental representation, formula (3) is $16\pi^2 b_0 = \frac{11N}{3} - \frac{2n_f}{3}$, and the theory is asymptotically free only when $n_f < 11N/2$ (less than 16.5 for SU(3)) Computed . 't Hooft's recollection (1999) has the same number. There are 6 real quarks, and QCD is on the asymptotically free side.

For an abelian group $C_A = 0$, so only the matter terms remain in (3), and $b_0 < 0$ (the direction of QED in § 01). For $U(1)_Y$ of the Standard Model (normalised by $Q = T_3 + Y$; the hypercharges are in Table 1 of 3-07), replacing $T(R)$ by Y^2 and counting one left-handed Weyl component as half a Dirac particle, the three generations of fermions ($\sum Y^2 = 10/3$ per generation) and the Higgs doublet give

$$16\pi^2\, b_0\big|_{U(1)_Y} = -\frac{2}{3}\cdot 3\cdot\frac{10}{3} - \frac{1}{3}\cdot 2\cdot\frac{1}{4} = -\frac{41}{6}$$

and at one loop the coupling grows at short distances Computed . This is a different number from b_1 in § 05 (the two-loop coefficient).

For theories without Yang–Mills fields, the abstract of Coleman and Gross (1973) states 'No renormalizable field theory without non-Abelian gauge fields can be asymptotically free.' (a claim at the level of calculating β functions) Known . § 6.2 of the problem statement also places the fact that four-dimensional ϕ^4 is not asymptotically free next to triviality (7-08), and calls Yang–Mills 'the natural candidate for four-dimensional CQFT'. This is one reason why Yang–Mills is the protagonist of the problem.

05

Two loops and universality — b_0, b_1 do not depend on the definition

The two-loop coefficient was found by Caswell and by Jones, each in a 1974 paper [Known] . For pure SU(N) it is positive, and rewriting the PDG formula for α in terms of g gives the following form (checked for SU(3) [Computed]).

$$b_1 = \frac{34}{3}\,\frac{N^2}{(16\pi^2)^2} \quad (6)$$

b_0 and b_1 do not change when the definition of the coupling (the renormalisation scheme) is changed. Changing to another coupling $\tilde{g} = g + c\,g^3 + d\,g^5 + \cdots$ and rewriting $\mu\, d\tilde{g}/d\mu = (1 + 3cg^2 + 5dg^4)\beta(g)$ in terms of $\tilde{g}$ gives

$$\tilde{\beta}(\tilde{g}) = -b_0\tilde{g}^3 - b_1\tilde{g}^5 - \tilde{b}_2\tilde{g}^7 - \cdots$$
$$\tilde{b}_2 = b_2 - 2c\,b_1 + (2d - 3c^2)\,b_0 \quad (7)$$

and c, d enter only from $\tilde{b}_2$ on (checked with series of rationals [Computed]). The PDG also writes that b_2 and beyond are 'renormalization-scheme-dependent', and Capitani et al. 1999 call b_0, b_1 'The universal coefficients' [Known] . This universality is the footing for § 07.

Example. For pure SU(3), starting from the same integration constant Λ (4-07), α at $\mu/\Lambda = 10$ is 0.248 at one loop and 0.186 at two loops, and even at 10^{10} it is 0.0248 and 0.0232, a difference of 7% [Computed] . The correction is a "log of a log" and is slow to disappear. Adding matter to make $b_1 < 0$, the β function vanishes at $g_*^2 = -b_0/b_1$, and for SU(3) with $n_f = 16$ an infrared fixed point with $\alpha_* \approx 0.042$ appears [Computed] (Banks–Zaks 1982 [Physics]). The coefficients have been computed up to five loops (the PDG's list of references; for the meaning of the series as a whole, 4-12).

06

How it was discovered — the sign had been seen several times

The negative sign had appeared in calculations before 1973. According to 't Hooft's recollection (1999), at the time it was thought to be universal that the β function is positive, and quantum field theory was thought unable to explain Bjorken scaling (the regularity, in high-energy scattering, that the constituents inside a hadron appear to move almost freely at short distances).

Table 3 A timeline of asymptotic freedom

Year	Event	Source
1964	Vanyashin–Terentyev: a negative sign for charged vector mesons, put down to non-renormalisability (paper from 1965)	't Hooft 1999

Year	Event	Source
1969	Khriplovich: the opposite sign for Yang–Mills in Coulomb gauge, not connected with asymptotic freedom	Same
1972	At the Marseille conference 't Hooft writes the one-loop formula on the blackboard (not made into a paper)	Same
1973	Gross–Wilczek, Politzer, Coleman–Gross	Abstract
1974	Caswell, Jones: two loops	Abstract
1980	Creutz: lattice numbers. Hasenfratz–Hasenfratz: lattice and continuum Λ	Abstract
2004	Nobel Prize in Physics to Gross, Politzer and Wilczek	nobelprize.org

Politzer's title carried a question mark: 'Reliable Perturbative Results for Strong Interactions?' Known . 't Hooft writes of the 1973 papers 'I was surprised, not about the result, which I had known all along, but about the stir they caused', and goes on to say that the real problem is confinement. The problem of this book also lies beyond that (17-01).

07

Agreement with the lattice — the running of the bare coupling and the ratio of Λs

On the lattice (5-01) the lattice spacing a is the cutoff, and the bare coupling g_0 runs with a. By the universality of § 05 its β function also begins with the same b_0, b_1, and integrating to two loops gives

$$a\Lambda_{\mathrm{L}} = (b_0 g_0^2)^{-b_1/(2b_0^2)}\, e^{-1/(2b_0 g_0^2)}\, \left(1 + O(g_0^2)\right) \quad (8)$$

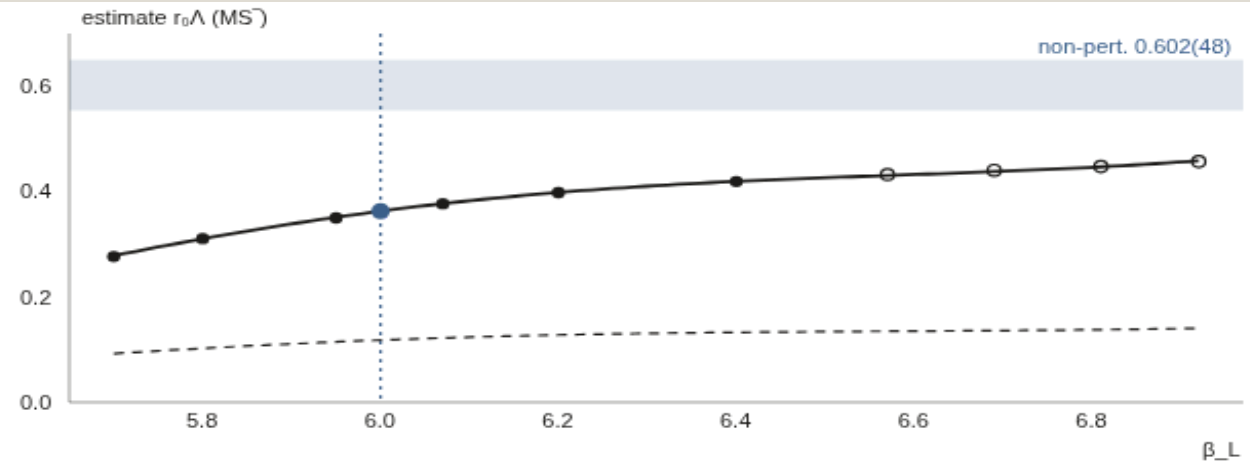
(the form of formula (2.7) of Capitani et al. 1999 with $\mu = 1/a$) Known . We compare the lattice Λ_{L} with the Λ of the continuum MS scheme. MS is the standard definition of the coupling in continuum theory, which in dimensional regularisation subtracts only the poles of the divergences and the constants that come with them (4-05). The two Λ differ only by a constant factor fixed at one loop; for the Wilson action $\Lambda_{\overline{\mathrm{MS}}}/\Lambda_{\mathrm{L}} = 28.81$ (SU(3)) and 19.82 (SU(2)). Computed from the coefficient k_1 of Lüscher–Weisz 1995, the SU(3) value agreed with 28.80934 in formula (4.1) of the same paper Computed . That formula (8) holds for lattice numbers is called **asymptotic scaling** (5-05; an early test is Creutz 1980).

At today's precision — r_0/a for pure SU(3)

The length standard r_0 is fixed from the force F between a pair of static colour charges by $r_0^2 F(r_0) = 1.65$ (17-01). The naive estimate of $r_0\Lambda_{\overline{\mathrm{MS}}}$, obtained by multiplying r_0/a

of Necco–Sommer 2002 by formula (8) and the ratio of Λs, is a quantity that should be constant if asymptotic scaling holds.

Figure 2 Moving the slider for β_L shows below the r_0/a at that point and the step in β_L that halves a (data and two loops). The curves are the estimate of $r_0\Lambda_{\overline{MS}}$ (solid two loops, dashed one loop), the points are lattice values (filled: direct values of r_0/a; open: converted from r_c/a, the * in Table 4), and the band is the non-perturbative value



β_L = 6.00: r₀/a = 5.37, step: data 0.475, two loops 0.614, estimate r₀Λ = 0.363

Table 4 Pure SU(3), Wilson action: r_0/a, the estimate of $r_0\Lambda_{\overline{MS}}$, and the step that halves a

β_L	r_0/a	Estimate (two loops)	Step (data)	Step (two loops)
5.70	2.922	0.277	0.354	0.616
5.95	4.898	0.350	0.456	0.614
6.20	7.380	0.398	0.524	0.613
6.40	9.74	0.419	—	0.612
6.57	12.18*	0.432	—	0.611
6.92	19.13*	0.457	—	0.609

Computed The numbers in this figure and table were computed on this machine. r_0/a is from Table 1 of Necco–Sommer 2002 (* converts r_c/a with $r_c/r_0 = 0.5133$), and the step (data) is from the interpolation formula (2.6) of the same paper (— is out of range). The estimate is $(r_0/a) \times a\Lambda_L \times 28.81$. The values in the figure agree with python3. If JavaScript is disabled, this table can be read instead.

The estimate keeps growing from 0.28 to 0.46 and has not yet reached the non-perturbative value 0.602(48). The expansion in the bare coupling converges slowly (with one loop alone, 0.09 to 0.14). Still, the direction and the order of magnitude agree, and the higher β_L, the closer to the prediction. In the method that measures the running coupling non-perturbatively with the size of the box as the scale (Lüscher–Weisz–Wolff 1991), the measured coupling is 'closely matched by the perturbative evolution', and 0.602(48) comes from there (Capitani et al. 1999) Known . All of these, however, are comparisons of finite-lattice numbers with perturbation theory, and show neither the existence nor the uniqueness of the continuum limit (5-09, 16-08).

08

Asymptotic freedom in the problem statement — what it says and what it does not

The problem statement first introduces asymptotic freedom as a physical explanation (§ 1), and gives it two roles. One is a requirement: the correlation functions must agree at short distances with the predictions of asymptotic freedom and perturbative renormalisation theory (§ 4; quoted in 0-03 § 07). Its role is to check at short distances that the theory built is "Yang–Mills". The other is a tool: 'Asymptotic freedom ensures a decisive regularity in the case when classical Sobolev inequalities are borderline' (§ 2) Known . This can be read as saying that four is the borderline dimension in which classical estimates fall just short, and asymptotic freedom makes up the shortfall (1-12; for Bałaban's work, 7-07).

Table 5 What asymptotic freedom says, and what it does not

What it says	What it does not say
The coupling weakens logarithmically at short distances (as a perturbative calculation). A handle on ultraviolet control (7-05)	That this holds in a constructed theory. Behaviour at long distances (as $\mu \to \Lambda$, g becomes large and formula (2) cannot be used)
An integration constant Λ appears and does not depend on μ (4-07)	The mass gap. A quantity expected to be proportional to Λ has the form $e^{-1/(2b_0 g^2)}$ and shows up at no order of perturbation theory (4-12)
Asymptotic freedom in four dimensions needs non-abelian gauge fields	Confinement (17-02)

Asymptotic freedom says "short distances are easy", and at the same time says "at long distances there is no handle for an expansion". The difficulty of the problem lies on the latter side (7-10, 7-11).

What this chapter can say

	Content
What is said	Formulas (3)(6)(7)(8), the threshold on the amount of matter, Coleman–Gross, as perturbative calculations Known . The checks of formulas (5)(7), the numbers in the tables and figures, the ratio of Λs Computed
Conditional	The agreement of the lattice with the prediction, as a comparison of finite-lattice numbers with perturbative formulas. The infrared fixed point within the range of the two-loop formula Physics

	Content
What is not said	That asymptotic freedom agrees with perturbation theory at all orders in the constructed four-dimensional theory (the statements of the construction published in 2026-11 go up to the one-loop running; 10-06). The uniqueness of the infinite-volume limit (6-04). "Spin" and "orbital" in formula (5) are names of terms. The events of 1964 to 1969 are taken from recollections (the original papers not read). Confinement and the mass gap

Sources and reproduction

Item	Kind	Source / tools
Read in full text	Known	PDG: J. Huston, K. Rabbertz, G. Zanderighi, in S. Navas et al., Phys. Rev. D 110 (2024) 030001, § 9.1.1. G. 't Hooft, Nucl. Phys. B Proc. Suppl. 74 (1999) 413, arXiv:hep-th/9808154. F. Wilczek 2005, arXiv:hep-ph/0502113. S. Capitani, M. Lüscher, R. Sommer, H. Wittig 1999, arXiv:hep-lat/9810063. S. Necco, R. Sommer 2002, arXiv:hep-lat/0108008. M. Lüscher, P. Weisz 1995, arXiv:hep-lat/9502001, hep-lat/9505011. D. Tong's lecture notes *Gauge Theory* § 2.4 and *Statistical Physics* § 3.6. A. Jaffe, E. Witten (the Clay problem statement)
Read in abstract only	Known	Gross–Wilczek, PRL 30 (1973) 1343; Politzer, PRL 30 (1973) 1346; Coleman–Gross, PRL 31 (1973) 851; Caswell, PRL 33 (1974) 244; Jones, Nucl. Phys. B 75 (1974) 531; Banks–Zaks, Nucl. Phys. B 196 (1982) 189; Creutz, Phys. Rev. D 21 (1980) 2308; Hasenfratz–Hasenfratz, Phys. Lett. B 93 (1980) 165; Lüscher–Weisz–Wolff, Nucl. Phys. B 359 (1991) 221. The prize citation is on nobelprize.org
Bibliographic data only	Known	Vanyashin–Terentyev, Zh. Eksp. Teor. Fiz. 48 (1965) 565; Khriplovich, Yad. Fiz. 10 (1969) 409 (from 't Hooft's list of references); N. K. Nielsen, Am. J. Phys. 49 (1981) 1171; the literature on higher coefficients (from the PDG's list of references)
Formulas (5)(7), Tables 1, 2 and 4, Figures 1 and 2, the ratio of Λs, the numerical examples, $-41/6$ for $U(1)_Y$ in § 04	Computed	python3 (series of rationals, bisection; each under 1 second). The JavaScript values of the figures were checked against the python3 values with node

Next chapter to read: 4-07 Dimensional transmutation and Λ (the Λ that appears when formula (2) is extended to the long-distance side). The continuation on the lattice side is 5-05.

Dimensional transmutation and Λ — quantum theory brings one scale into a classical theory without scales

This chapter organises established knowledge (sources carry the “Known” label in the text). This book's own investigation is mainly in Part 15 and Volume 3 (Parts 16–28 and the appendices).

What this chapter shows — From a classical theory without scales, one scale Λ appears in quantum theory as the integration constant of the running coupling. Masses are expected to be constant multiples of Λ, and their form $e^{-1/(2b_0 g^2)}$ does not show up in perturbation theory. The two-dimensional delta well as an exact example.

Prerequisite chapters — 4-06 (the β function and b_0, b_1). Helpful: 1-12, 2-05, 3-05. The notation is the same as in 4-06 ($\hbar = c = 1$, $\beta(g) = \mu\, dg/d\mu$, $b_0 = 11N/(48\pi^2)$).

Said first — Dimensional transmutation is a physical argument read off from the perturbative renormalisation group, and in four dimensions it is not a theorem. The renormalisation group says only "if there is a mass, it is a constant multiple of Λ", and does not say that the coefficient is non-zero (the mass gap). This site shows neither the mass gap nor the construction of the continuum theory.

Order of this chapter

1. The classical theory has no scale
2. Λ is the integration constant of the running (Table 2)
3. One parameter disappears (Figure 1)
4. Physical quantities are constant multiples of Λ (Table 3)
5. $\exp(-1/g^2)$ does not show up in perturbation theory (Table 4)
6. The two-dimensional delta well (Figure 2, Table 5; Detour: Cooper pairs in a metal)
7. Λ changes with the convention (Table 6)
8. How scale symmetry is broken
9. What can be said and what cannot (Table 7)

01

The classical theory has no scale

With $\hbar = c = 1$, length is the inverse of mass, and every quantity can be measured in powers of mass. The dimension of the coupling in the action $S = \frac{1}{2g^2}\int \operatorname{tr} F_{\mu\nu}F_{\mu\nu}\, d^dx$ is

$$[g^2] = (\text{mass})^{4-d} \quad (1)$$

(2-04, 1-12). In three dimensions g^2 is itself a mass, and one can write $m = c\,g^2$ (7-06). In four dimensions g is dimensionless, so no mass can be made from the classical ingredients. Classical waves have the dispersion $\omega = |\mathbf{k}|$, and the problem statement also writes 'massless waves that travel at the speed of light' (§ 1) Known . There are no static finite-energy solutions (lumps) either (3-03; Deser 1976, Coleman 1977) Known .

Example. The action $8\pi^2/g^2$ of the BPST instanton does not depend on its size ρ (3-05). Classical solutions have no preferred size. If $\Delta > 0$ in the quantum theory, that scale must come from outside the classical action. This is what 'the mass gap is not classically visible' in § 6.6 of the problem statement is about (§ 05).

02

Λ is the integration constant of the running

The $\beta(g) = \mu\, dg/d\mu$ of 4-06 is a first-order ordinary differential equation. Separating the variables and integrating gives

$$\ln\frac{\mu}{\Lambda} = \int^{g(\mu)} \frac{dg'}{\beta(g')} \quad (2)$$

The freedom in the lower limit on the right-hand side, taken to the left-hand side as a constant, is Λ; being an integration constant, it has the same value whichever point of the flow it is computed from. At one loop ($\beta = -b_0 g^3$), choosing the constant to be 0 gives

$$\Lambda = \mu\, \exp\!\Big(-\frac{1}{2b_0\, g^2(\mu)}\Big) \quad (3)$$

and differentiating $\ln\Lambda$ with respect to $\ln\mu$ gives $1 + \beta/(b_0 g^3) = 0$. Rewritten for SU(2) on the lattice ($\beta = 4/g^2$) it is $1/(a\Lambda) \propto e^{3\pi^2\beta/11}$, the same formula as the one-loop part of the bookkeeping length ξ_{RG} of Part 23 (23-01 The line of construction and the

diffusive-relaxation-time assumption) Known . From two loops on, a convention for the lower limit is needed. The one widely used is formula (2.7) of Capitani et al. 1999.

$$\Lambda = \mu\,(b_0 g^2)^{-b_1/(2b_0^2)}\, e^{-1/(2b_0 g^2)} \exp\!\Big(-\int_0^{g} \Big[\frac{1}{\beta(x)} + \frac{1}{b_0 x^3} - \frac{b_1}{b_0^2 x}\Big]\, dx \Big) \quad (4)$$

The bracket is finite as $x \to 0$, so 0 can be used as the lower limit. The PDG review explains Λ as 'a constant of integration, which corresponds to the scale where the perturbatively-defined coupling would diverge' Known .

Example (Table 2). Treating the two-loop β as exact, we integrated the flow numerically for SU(3) from $g^2(\mu_0) = 2$. Formula (4) is constant to 9 digits. Formula (3) drifts by a factor of 2.7 while μ grows by a factor of 10^{16}, and the formula with only the prefactor power attached merely shrinks the difference from 4.6% to 0.43%. The numbers are given with their defining formulas.

Table 2 Three "Λ/μ_0" along the two-loop flow of pure SU(3)

μ/μ_0	$g^2(\mu)$	Formula (3)	With the power only	Formula (4)
1	2.000	0.02763	0.06341	0.066444376
10^{16}	0.1733	0.01028	0.06616	0.066444376

Computed python3 (RK4). The integral in formula (4) agrees between the closed form and numerical integration to 9 digits.

For how this book uses this, see 24-07 § 01 (reading the dimensionless coupling as the first source of the difficulty of four dimensions).

03 One parameter disappears — Λ is a unit

Turning formula (3) around, the whole flow is fixed by Λ alone.

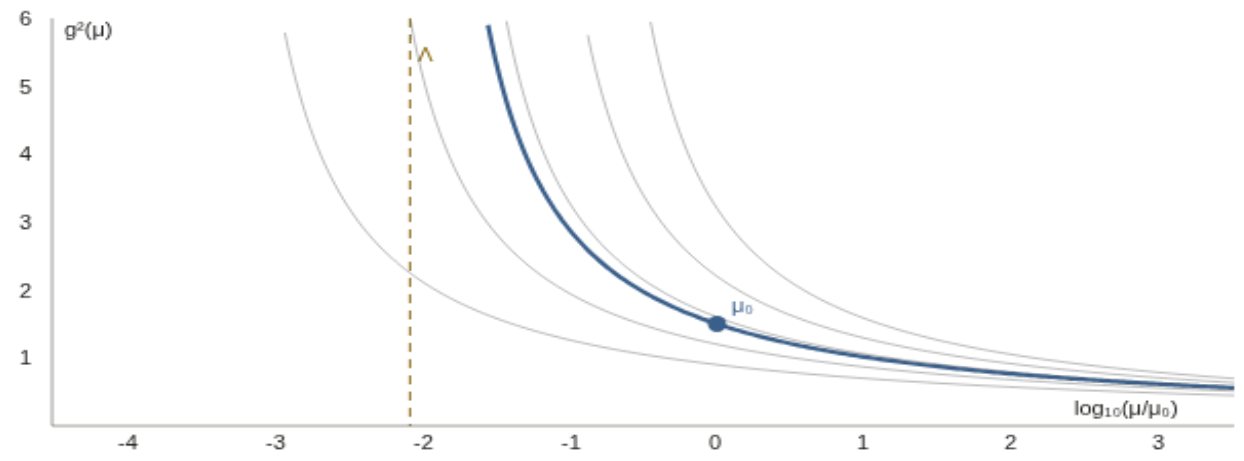
$$g^2(\mu) = \frac{1}{2b_0 \ln(\mu/\Lambda)} \qquad (\mu > \Lambda,\ \text{one loop}) \quad (5)$$

Λ enters only in the form μ/Λ. Doubling Λ is the same as halving the unit of mass, and the one-parameter family "choose the coupling at μ_0" is, up to units, a single theory. The dimensionless g has been replaced by the dimensionful Λ. This is called **dimensional transmutation** Physics .

Tong writes that this idea was known from the 1973 work of Coleman and E. Weinberg (Tong § 2.9). E. Weinberg's thesis of the same year writes, for the electrodynamics of a massless scalar field, that changing the vacuum expectation value $\langle\phi\rangle$ generated by quantum effects (the value of the scalar field in the vacuum, which serves as the scale of mass) is 'completely equivalent to changing the scale in which masses are measured'

Known . Pure Yang–Mills has no dimensionless coupling left, and the only remaining inputs are the group (N) and the θ angle (3-05).

Figure 1 Moving the slider for $g^2(\mu_0)$ shifts the one-loop flow of pure SU(3) (thick line) and Λ (dashed line) sideways (the thin lines are $g^2(\mu_0)$ = 0.9 to 3.2). "Horizontal axis μ/Λ" makes the six coincide. "Play" moves back and forth



g²(μ₀) = 1.50 → Λ/μ₀ = 8.35e-3

Computed The numbers in this figure were computed on this machine (the figure evaluates formulas (3)(5) on the spot). Λ/μ_0 of the thin lines ranges from 3.44×10^{-4} to 0.106, but g^2 at $\mu = 10\Lambda$, $10^3\Lambda$ and $10^6\Lambda$ is always 3.117, 1.039 and 0.520.

The scale comes from the cutoff. On the lattice Λ is made from the pair (a, g_0) (formula (8) of 4-06).

04

Physical quantities are constant multiples of Λ — nothing is said about the sign of the coefficient

Computing a physical quantity with the dimension of mass, dimensional analysis gives the form $m = \mu\, F(g(\mu))$. m does not depend on the choice of μ, so

$$\mu \frac{d}{d\mu}\big(\mu\, F(g(\mu))\big) = \mu\big(F(g) + \beta(g)\, F'(g)\big) = 0 \quad (6)$$

and the solution is $F(g) = C \exp\big(-\int^g dg'/\beta(g')\big)$. Comparing with formula (2),

$$m = C\,\Lambda \quad (7)$$

where C is a pure number. Formula (6) is the renormalisation group equation (4-04) for a physical quantity without anomalous dimension. For the string tension $\sigma = C'\Lambda^2$, and the ratio $m/\sqrt{\sigma}$ becomes a pure number fixed by N and θ alone, with no value left to adjust Physics . This is what "one scale" means (17-01, 17-12).

The solutions of formula (6) include $C = 0$. What can be said is only "if there is a mass, it is a constant multiple of Λ", and $C > 0$ cannot be said. The mass gap problem can be read as the problem of constructing the theory and showing $C > 0$ for the state next to the vacuum (0-03, 4-11).

Example (Table 3). Lattice numbers come out as dimensionless ratios (r_0 is the length of 4-06). $m(0^{++})$ is the mass of the lightest glueball with spin 0, positive parity and positive charge conjugation (5-08).

Table 3 Dimensionless ratios of pure SU(3)

Ratio	Value	Source
$r_0 \Lambda_{\overline{\text{MS}}}$	0.602(48)	Capitani et al. 1999
$r_0\, m(0^{++})$	4.21(11)(4)	Morningstar–Peardon 1999
$m(0^{++})/\sqrt{\sigma}$	3.405(21)	Athenodorou–Teper 2020
$m(0^{++})/\Lambda_{\overline{\text{MS}}}$	6.99(59)	Division of the two rows above (a rough guide combining values from different papers)

Computed The last row was divided with python3, propagating the errors as independent. Physics All are extrapolations from the lattice, to be read on the assumption that the continuum theory exists (17-01).

05

exp(−1/g²) does not show up in perturbation theory

Rewriting formula (7) at one loop gives

$$\frac{m}{\mu} = C\,\exp\!\Big(-\frac{1}{2b_0\, g^2(\mu)}\Big) \quad (8)$$

$f(x) = e^{-1/x}$ $(x > 0)$, $f(x) = 0$ $(x \le 0)$ can be differentiated any number of times at $x = 0$, and all its derivatives are 0. This is because $f^{(n)}(x) = P_n(1/x)\, e^{-1/x}$ ($P_1(u) = u^2$, $P_2(u) = u^4 - 2u^3$), and polynomials cannot beat e^{-u} Known . Expanding (8) in powers of g^2, all the coefficients are 0.

Perturbation theory (4-03) computes a power series in g^2, so a quantity of the form (8) looks like 0 at every order. § 6.6 of the problem statement writes that, since the gap is not classically visible, showing it may need a non-classical change of variables ('change of variables or “duality transformation.”') Known . The abstract of Gross–Neveu 1974 also says that a mass generated dynamically in an asymptotically free theory 'must depend on the coupling constants in a nonanalytic fashion, vanishing exponentially when these vanish' (at the level of the renormalisation group). The tunnelling splitting of the double well is of the same type (2-05).

Example (Table 4). As a rough guide with coefficients set to 1, the order $n^* = \ln(\Lambda/\mu)/\ln\alpha$ at which the n-th power of $\alpha = g^2/4\pi$ drops to Λ/μ is 2.8 at $g^2 = 1$, and grows without bound as $g \to 0$. The actual coefficients grow with the order, and the series is thought to diverge (4-12).

Table 4 Pure SU(3), one loop

g^2	Λ/μ	n^*	$\log_{10} e^{-8\pi^2/g^2}$
2	2.76×10⁻²	1.95	−17.15
1	7.63×10⁻⁴	2.84	−34.29
0.25	3.39×10⁻¹³	7.33	−137.16
0.1	6.71×10⁻³²	14.85	−342.91

Computed python3. The rightmost column (§08) agrees with $(\Lambda/\mu)^{11}$ in all digits. P_n was checked up to $n = 6$.

06

The two-dimensional delta well — exact dimensional transmutation

An exact example exists in two-dimensional quantum mechanics. In units with $\hbar^2/2m = 1$, consider the delta-function well in d dimensions

$$H = -\nabla^2 - g\,\delta^d(\mathbf{x}) \tag{9}$$

δ^d has the dimension of length to the power $-d$, and ∇^2 to the power -2, so g has the dimension of length to the power $d-2$, and is dimensionless at $d = 2$. This is the same situation as four-dimensional Yang–Mills. Taking the Fourier transform as $\hat\psi(\mathbf{k}) = \int \psi(\mathbf{x})\, e^{-i\mathbf{k}\cdot\mathbf{x}}\, d^dx$ (1-11), the transform of δ^d is 1. A bound state $E = -\kappa^2$ satisfies $(k^2 + \kappa^2)\,\hat\psi(\mathbf{k}) = g\,\psi(0)$, so dividing by $k^2 + \kappa^2$ and integrating gives

$$1 = g \int_{|\mathbf{k}|<K} \frac{d^dk}{(2\pi)^d}\,\frac{1}{k^2+\kappa^2} \tag{10}$$

(K is the momentum cutoff). For $d = 2$ the right-hand side is $(g/4\pi)\ln(1 + K^2/\kappa^2)$, which diverges logarithmically, as does one-loop four-dimensional Yang–Mills (4-04). Solving,

$$E_B = \kappa^2 = \frac{K^2}{e^{4\pi/g} - 1} \approx K^2\, e^{-4\pi/g} \tag{11}$$

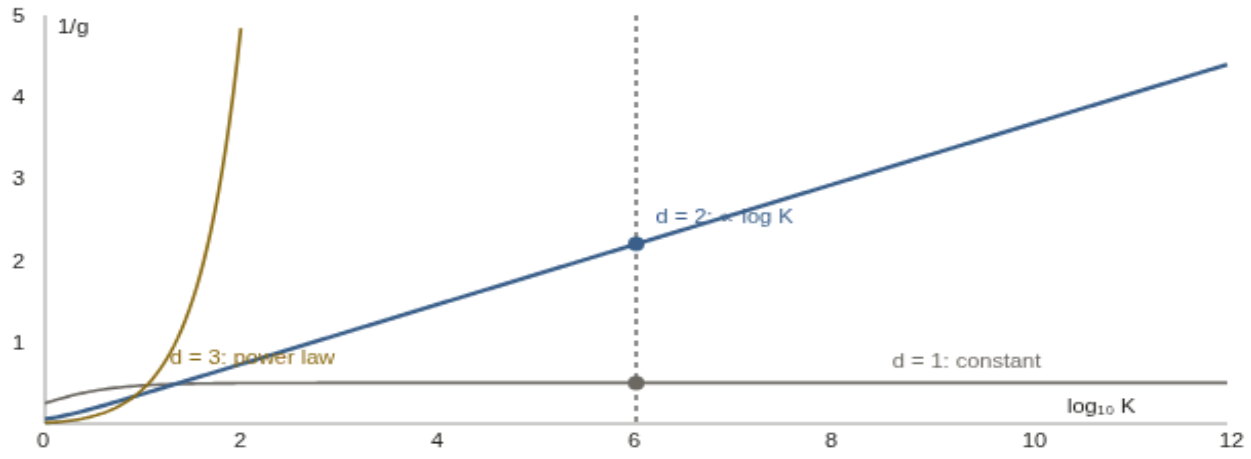
the same form as formula (8). Renormalisation is the operation of fixing E_B and taking $K \to \infty$, which sends the coupling towards 0 as $g(K) = 4\pi/\ln(1 + K^2/E_B)$. For $K^2 \gg E_B$,

$$\beta(g) = K\frac{dg}{dK} = -\frac{g^2}{2\pi} \quad (12)$$

the sign is negative, that is, asymptotically free. The only parameter left in the limit is E_B. Up to here this is an elementary calculation with a momentum cutoff Computed .

On the mathematical side, it is known that in the limit where the cutoff is removed the Hamiltonian converges (in the sense of the resolvent $(H - z)^{-1}$) to a self-adjoint operator with exactly one bound state (energy $-E_B$) (Albeverio et al., Part I, Chapter 5; bibliographic data only) Known . The limit there is taken by shrinking a short-range potential while renormalising the coupling logarithmically, a different route from the momentum cutoff above. As teaching material there is Gosdzinsky–Tarrach 1991 (bibliographic data only).

Figure 2 Moving the slider for the cutoff K shows how the inverse 1/g of the coupling that keeps E_B = 1 moves for d = 1, 2, 3. For d = 2 it is proportional to log K (asymptotically free). "Play" moves K back and forth



K = 10^6.00 : g_1 = 2.0000 g_2 = 0.4548 g_3 = 1.97e-5, g_2 up 1%: E_B ×1.31

Table 5 The coupling $g_d(K)$ that keeps $E_B = 1$ (in units of $\sqrt{E_B}$)

K	$g1$	$g2$	$2\pi/\ln K$	$g3$
10	2.1355	2.7229	2.7288	2.314
10^6	2.0000	0.4548	0.4548	1.974×10^{-5}
10^{12}	2.0000	0.2274	0.2274	1.974×10^{-11}

Computed The numbers in this figure were computed on this machine. The closed form of formula (10) agrees with numerical integration to 11 digits, and $K\,dg/dK$ agrees with $-g^2/2\pi$ for $K \geq 10^4$ (ratio 1.000000). The factor by which E_B changes when g_2 is moved by 1% is 1.05 at $K = 10$ and 1.73 at 10^{12}.

For $d = 1$, $E_B = g^2/4$ is a power of g, and $g(K)$ settles at 2. For $d = 3$, g has to be brought to the critical value $2\pi^2/K$ as a power. These are the cases where the mass dimension of the coupling is positive, zero and negative; four-dimensional Yang–Mills is on the $d = 2$ side, and three-dimensional Yang–Mills on the $d = 1$ side (an analogy).

Detour: Cooper pairs in a metal

A subsection you may skip.

The same logarithm appears for pairs of electrons in a metal. In a metal, the electron states are filled up to the Fermi surface (the surface of the filled sea of electrons). Let ξ be the energy of one electron measured from the Fermi surface, and $N(0)$ the density of states per spin at the Fermi surface (the number of states per unit energy and unit volume). If two electrons on the surface feel an attraction of strength V only within the shell $0 < \xi < \omega_D$ (ω_D is the maximum frequency of lattice vibrations, the cutoff here), the binding condition, by the same steps as formula (10), is

$$1 = N(0)V \int_0^{\omega_D} \frac{d\xi}{2\xi + E_B} = \frac{N(0)V}{2} \ln\left(1 + \frac{2\omega_D}{E_B}\right) \quad (15)$$

Since $N(0)$ is constant, a logarithm appears, just as with $d^2k \propto d(k^2)$ in two dimensions. Solving gives $E_B = 2\omega_D / (e^{2/(N(0)V)} - 1) \approx 2\omega_D\, e^{-2/(N(0)V)}$: even a weak attraction has a bound state, and its size does not show up in the series in powers of V (Cooper 1956; bibliographic data only) Known . The 2 in the exponent comes from the convention of taking $N(0)$ per spin.

Computed python3. The closed form and numerical integration of formula (15) agree within double precision for $N(0)V = 0.05$ to 0.5.

07

Λ changes with the convention, the ratios do not

The value of Λ changes with the definition of the coupling (the scheme). If g of scheme A and $\tilde{g}$ of scheme B are related by $\tilde{g} = g + c\,g^3 + O(g^5)$, then $1/\tilde{g}^2 = 1/g^2 - 2c + O(g^2)$, so from formula (3)

$$\frac{\Lambda_B}{\Lambda_A} = e^{c/b_0} \quad (13)$$

Since b_0, b_1 do not depend on the scheme (4-06), this ratio holds exactly for the Λ of formula (4). The abstract of Celmaster–Gonsalves 1979 writes that the ratio of Λ in two schemes is 'a pure number which we show is exactly deducible from our one-loop results' Known .

Example (Table 6). MS (minimal subtraction: the scheme that subtracts only the poles appearing in dimensional regularisation; 4-04) and $\overline{\text{MS}}$, the PDG's standard (the scheme that also subtracts the constant $\ln 4\pi - \gamma_E$ along with the poles; Bardeen et al. 1978), differ only by the rescaling $\mu^2 \to \mu^2 e^{\gamma_E}/4\pi$, and the ratio is $\sqrt{4\pi e^{-\gamma_E}} = 2.656$. The ratio 28.81 to the lattice Wilson action (SU(3), 4-06) also comes from a shift in $1/g^2$

of only 0.468. The one-loop difference enters the exponent multiplied by $1/b_0$, so Λ changes by an order of magnitude or more.

Table 6 Changing the scheme and the ratio of Λs

Ratio	Group	c	Shift in $1/g^2$	e^{c/b_0}	Ratio from formula (4)
MS / MS	SU(3)	0.0681	−0.136	2.656	–
MS / lattice	SU(3)	0.2341	−0.468	28.81	28.8100
MS / lattice	SU(2)	0.1387	−0.277	19.82	19.8200

Computed python3. Building the β of scheme B from $\tilde{g} = g(1 + cg^2)$ and integrating formula (4) numerically, the ratio is constant to 6 digits for $\mu/\mu_0 = 1$ to 10^{16}. The ratio for the formula with the prefactor power only is slow: 6.76 → 23.08 (SU(3)).

Ratios of physical quantities such as $m/\sqrt{\sigma}$ do not depend on the scheme. $r_0\Lambda$ is a single number only once the scheme is named. To attach MeV, another quantity is borrowed (17-01).

08

How scale symmetry is broken — instantons and the trace anomaly

The classical scale symmetry of § 01 leaves traces in two places in the quantum theory. One is the size of instantons: ρ, free in the classical theory, is distinguished by the running of the coupling (Tong § 2.3.1– § 2.3.2). Using the coupling at the scale $1/\rho$, from formula (3) and $16\pi^2 b_0 = 11N/3$, at one loop

$$e^{-8\pi^2/g^2(1/\rho)} = (\rho\Lambda)^{11N/3} \quad (14)$$

holds identically (checked against Table 4 Computed). Small instantons are suppressed by a power of $\rho\Lambda$, and the suppression stops working at $\rho \sim 1/\Lambda$. The one-loop calculation is 't Hooft 1976 Physics .

The other is the trace of the energy–momentum tensor. The divergence of the conserved current of scale transformations (2-01) is the trace $T^{\mu}{}_{\mu}$, which is 0 classically in four dimensions (3-03). In the quantum theory a term in $F^a_{\mu\nu}F^{a\,\mu\nu}$ proportional to $\beta(g)$ remains (Collins–Duncan–Joglekar 1977; for the formula, 18-04) Physics . § 4 of the problem statement also includes 'the existence of a stress tensor' in the short-distance requirements Known .

09

What can be said and what cannot

On the lattice, a test from formula (7) is whether ratios of correlation lengths $\xi_{\text{lat}} = 1/(am)$ approach constant values (5-05, 5-09). We list the degree of certainty example

by example. The top four are two-dimensional models with a dimensionless coupling, and what has become a theorem is either a model with finitely many degrees of freedom or a form that keeps the cutoff.

Row by row, we separate what is a theorem from what is physics. For the delta well, the binding condition (10) with a cutoff is an elementary calculation, and the operator in the limit with the cutoff removed is in the mathematical literature. For the two-dimensional Gross–Neveu model (a model of interacting fermions), keeping the ultraviolet cutoff and taking the number of components N large (but finite), it is a theorem that the two-point function decays exponentially, that is, that a mass is generated. For the two-dimensional O(N) σ model (a model of fields valued in N-component vectors of length 1), mass generation in a regularised version is also a theorem, but there is still no construction of the theory with the cutoff removed. The $8/e$ of O(3) is a physical value obtained by matching the result of the Bethe ansatz (a method that uses the exact solvability of scattering in two-dimensional integrable models) with perturbation theory. For pure SU(3) in four dimensions, the ratio m/Λ comes only from lattice numbers. The construction of the continuum theory was given by the construction published in 2026-11 (at the level of the committee's assessment; 6-00), but $\Delta > 0$ has not been shown.

Table 7 Examples of dimensional transmutation, and their status in mathematics

Model	Scale	Status in mathematics
Two-dimensional delta well	$K^2 e^{-4\pi/g(K)}$	With a cutoff, an elementary calculation. The limit with the cutoff removed is a self-adjoint operator with exactly one bound state (Albeverio et al.; bibliographic data only)
Two-dimensional Gross–Neveu	Non-analytic mass	With a cutoff and large finite N, exponential decay is a theorem (Kopper–Magnen–Rivasseau 1995)
Two-dimensional O(3) σ model	$m = (8/e)\Lambda_{\overline{\text{MS}}}$	A physical value from matching the Bethe ansatz with perturbation theory (Hasenfratz–Maggiore–Niedermayer 1990)
Two-dimensional O(N) σ model (N large and finite)	Mass m	With an ultraviolet cutoff, mass generation (exponential decay of correlations) is a theorem (Kopper 1999). A construction at the level required did not exist as of the 2004 report (Douglas 2004)
Four-dimensional pure SU(3)	$m(0^{++}) \approx 7\Lambda_{\overline{\text{MS}}}$ (the rough guide of Table 3)	Lattice numbers. The construction of the continuum theory: the construction published in 2026-11 (at the level of the committee's assessment). $\Delta > 0$ is not proved

The problem statement does not mention Λ; it asks for $\Delta > 0$ and for agreement with the predictions of asymptotic freedom at short distances (0-03). If the theory exists and agrees with those predictions, in physics one considers that Λ can be read off from them Physics . The mass gap is then the claim that C, which shows up neither in perturbation theory nor in the classical theory, is positive. How this single scale is used is treated in 17-01 and 17-12.

What this chapter can say

	Content
What is said	The μ-invariance of formulas (2) to (4), the solution (7), that the Taylor coefficients of $e^{-1/x}$ are 0, formulas (10) to (15). These are Known calculations, and the numbers in the tables and figures are Computed
Conditional	"Masses are constant multiples of Λ" assumes that the theory exists and that the renormalisation group describes short distances correctly Physics . Table 3 is an extrapolation from the lattice. The limiting operator of the delta well is taken from the literature (text not checked). The correspondence for each d is an analogy
What is not said	That Λ is defined in four dimensions, $C > 0$, that the values of Table 3 exist as quantities of the continuum theory (they are extrapolations from the lattice, not theorems). This site shows neither the mass gap nor the construction of the continuum theory

Sources and reproduction

Item	Kind	Source / tools
The problem statement, the sentence on Λ, formula (4)	Known	A. Jaffe, E. Witten, "Quantum Yang–Mills Theory" (Clay); PDG, PRD 110 (2024) 030001 § 9.1.1; S. Capitani et al., NPB 544 (1999) 669, hep-lat/9810063 (text of all the above). W. A. Bardeen et al., PRD 18 (1978) 3998 (abstract)
Origin, instantons	Known	D. Tong, *Gauge Theory* § 2.3.1– § 2.3.2, § 2.9 (text); E. J. Weinberg, thesis (1973), hep-th/0507214 Chapter IV (text); S. Coleman, E. Weinberg, PRD 7 (1973) 1888 (abstract only; the first appearance of the term dimensional transmutation not checked)
The classical theorems of § 01	Known	S. Deser, PLB 64 (1976) 463; S. Coleman, CMP 55 (1977) 113 'There are no classical glueballs' (abstracts of both)
Literature for § 04 to § 08	Known	D. J. Gross, A. Neveu, PRD 10 (1974) 3235; W. Celmaster, R. J. Gonsalves, PRD 20 (1979) 1420; G. 't Hooft, PRD 14 (1976) 3432; J. C. Collins et al., PRD 16 (1977) 438 (abstracts of all the above). S. Albeverio et al., *Solvable Models in Quantum Mechanics* (Springer, 1988) Part I, Chapter 5, P. Gosdzinsky, R. Tarrach, Am. J. Phys. 59 (1991) 70, and L. N. Cooper, Phys. Rev. 104 (1956) 1189 are bibliographic data only
Numbers and examples in Tables 3 and 7	Physics	C. Morningstar, M. Peardon, PRD 60 (1999) 034509; A. Athenodorou, M. Teper, JHEP 11 (2020) 172; P. Hasenfratz, M. Maggiore, F. Niedermayer, PLB 245 (1990) 522 (abstract); C. Kopper et al., CMP 169 (1995) 121; C. Kopper, CMP 202 (1999) 89 (the O(N) σ model with large finite N; abstract); M. R. Douglas's report to Clay (2004). Beyond the bibliographic data, based on the material of the checked survey
Tables 2 to 6, the figures, formula (14), formula (15)	Computed	python3 (RK4, scipy). The figures evaluate the same formulas in JavaScript. 28.81 and 19.82 are the values of 4-06

Next chapter to read: 4-08 Gauge fixing and Faddeev–Popov.

Gauge fixing and Faddeev–Popov — correct locally, and globally the signs cancel

This chapter organises established knowledge (sources carry the “Known” label in the text). This book's own investigation is mainly in Part 15 and Volume 3 (Parts 16–28 and the appendices).

What this chapter shows — Perturbation theory needs gauge fixing, and its procedure (Faddeev–Popov) produces a determinant and ghosts. After fixing, BRST symmetry remains and formally guarantees independence from the choice of gauge. When an orbit meets the condition surface many times (Gribov copies) the procedure breaks down, and on the lattice the cancellation of signs gives 0/0.

Prerequisite chapters — 4-03 (perturbation theory, propagators) and 3-04 (orbit space, the geometry of Gribov copies, Singer's theorem). Grassmann numbers (§ 03), BRST cohomology (§ 04) and the Euler characteristic (§ 08) are defined on the spot, as far as needed.

Said first — Faddeev–Popov and BRST are tools established at each order of perturbation theory, not a non-perturbative definition. The figures and tables are calculations in a planar model and for constant fields in a box. This chapter shows neither the mass gap nor the construction of the theory.

Order of this chapter

1. Why fix the gauge — the action is flat along the orbits, and no propagator can be made
2. A finite-dimensional model — "inserting 1" with rotations of the plane
3. In field theory — writing the determinant as an integral over anticommuting fields (ghosts) (Detour: in the abelian case the ghosts disappear)
4. BRST symmetry — the symmetry that remains after fixing, and how physical states are chosen
5. What happens when there are copies — comparing four ways of counting (Figure 1, Table 1; Detour: counting intersections with signs)
6. The Gribov region and the fundamental modular region — convex and bounded in every direction

7. The horizon in a box — the larger the box, the closer the horizon is to the origin (Figure 2, Table 2)
8. The 0/0 on the lattice — counting with signs gives the Euler characteristic 0
9. Gauge fixing seen from the mass gap problem — its place in the axioms and in attempts at construction

---

01

Why fix the gauge — no propagator can be made

Conventions: Euclidean $\mathbb{R}^4$, $\hbar = c = 1$, and we write the Hermitian $A_\mu = A_\mu^a T^a$ and the coupling g separately (the style of 3-02; the connection of the geometry chapters is $\mathcal{A} = -igA$).

$$(D_\mu \omega)^a = \partial_\mu \omega^a + g f^{abc} A_\mu^b \, \omega^c, \qquad \delta A_\mu^a = \frac{1}{g} (D_\mu \lambda)^a, \qquad S = \frac{1}{4} \int F_{\mu\nu}^a F_{\mu\nu}^a \, d^4x \quad (1)$$

Perturbation theory (4-03) uses the inverse of the quadratic part of the action as the propagator. S does not change under gauge transformations, so the quadratic part is flat in the direction $A = \partial\lambda$ and has no inverse.

$$S_2 = \frac{1}{2} \int \frac{d^4k}{(2\pi)^4} \, \tilde{A}_\mu^a(-k) \, K_{\mu\nu}(k) \, \tilde{A}_\nu^a(k), \qquad K_{\mu\nu} = k^2 \delta_{\mu\nu} - k_\mu k_\nu, \qquad Kk = 0 \quad (2)$$

Example. If $k = (0, 0, 0, \kappa)$, then $K = \kappa^2 \, \mathrm{diag}(1, 1, 1, 0)$, and the fourth component does not appear in the action Computed . On Wilson's lattice, where the volume of the orbits (4-02 § 07) is finite, no fixing is needed (5-01 § 06) Known . It is needed when one expands around $A = 0$ while staying in the continuum.

Let us put into words why there is no inverse. The path integral $\int \mathcal{D}A \, e^{-S}$ (4-02) adds up all fields related by gauge transformations (one orbit). Along the orbit e^{-S} does not change, so the Gaussian integral in that direction does not decay and diverges. Speaking in metaphor, the bottom of the valley is a single trench, flat without end along the trench. Precisely, the kernel K of the quadratic form (2) has eigenvalue 0 in the direction of k. So we would like to take the integral along the orbit outside, choose one representative from each orbit and add up the rest. We first see how to take it out in a finite-dimensional model.

02

A finite-dimensional model — Faddeev–Popov with rotations of the plane

The weight $e^{-S(|x|)}$ on the plane does not change under the rotation R_θ, and the orbits are circles centred at the origin. Here the angle of rotation θ plays the role of "the direction of the orbit". The aim is to bring the integral over θ to the surface without changing the value of Z, and for that we put a rewritten 1 into Z. We use the rule for the one-variable δ function.

$$\int d\theta\, \delta\big(G(\theta)\big)\, \big|G'(\theta_0)\big| = 1 \qquad (G \text{ has a single point } \theta_0 \text{ where it equals 0 and no other})$$

If the condition $G(x) = 0$ meets each circle exactly once, $\theta \mapsto G(R_\theta x)$ vanishes at a single point only, so the following "insertion of 1" holds.

$$1 = \Delta_G(x) \int_0^{2\pi} d\theta\, \delta\big(G(R_\theta x)\big), \qquad \Delta_G(x) = \Big|\tfrac{\partial}{\partial\theta} G(R_\theta x)\Big|_{G(R_\theta x)=0} \quad (3)$$

Putting this into $Z = \int d^2x\, e^{-S}$ and changing variables $x \to R_\theta^{-1} x$, since d^2x, S and Δ_G do not change under rotation, the integral over θ leaves the volume of the group 2π.

$$Z = 2\pi \int d^2x\, \delta\big(G(x)\big)\, \Delta_G(x)\, e^{-S(|x|)} \quad (4)$$

This is the whole of the Faddeev–Popov procedure, and the weight Δ_G on the surface is the **Faddeev–Popov determinant**.

Example. Taking the condition to be the positive x_1 axis ($G = x_2$, $x_1 > 0$), $\Delta_G = r$, and (4) is $Z = 2\pi \int_0^\infty r\, e^{-S} dr$: the Jacobian of polar coordinates. The determinant measures the size of the orbit (the same role as the Weyl factor in 3-04 § 04).

03

In field theory — the determinant and ghosts

We repeat (3)(4) with the group taken to be all pointwise gauge transformations and the condition $G^a = \partial_\mu A^a_\mu - \omega^a$. Since $\delta G^a = \frac{1}{g}\partial_\mu (D_\mu \lambda)^a$, the content of the determinant is the following $\mathcal{M}$. The constant factor $1/g$ and the sign become an overall constant of the determinant, absorbed into the overall constant of Z (which cancels in ratios of expectation values), so we drop them.

$$\mathcal{M} = -\partial_\mu D_\mu, \qquad Z = \int \mathcal{D}A\, \det \mathcal{M}\, \delta(\partial_\mu A_\mu - \omega)\, e^{-S} \quad (5)$$

Averaging ω with the weight $e^{-\int \omega^2/2\alpha}$ (many references write the parameter as ξ, but this book uses ξ for the correlation length, so we write α; the same letter as in 4-

03), the δ function turns into a term of the action, and the quadratic form acquires an inverse.

$$S + \frac{1}{2\alpha}\int (\partial_\mu A^a_\mu)^2, \qquad D^\alpha_{\mu\nu}(k) = \frac{1}{k^2}\left(\delta_{\mu\nu} - (1-\alpha)\frac{k_\mu k_\nu}{k^2}\right) \quad (6)$$

$\alpha = 1$ is the Feynman gauge and $\alpha \to 0$ the Landau gauge (the inverse checked numerically Computed). To make $\det \mathcal{M}$ a term of the action as well, we use anticommuting numbers (Grassmann numbers: $\theta_i\theta_j = -\theta_j\theta_i$, $\theta_i^2 = 0$). The rules are $\int d\theta\, 1 = 0$ and $\int d\theta\, \theta = 1$. Since the expansion stops at $e^{-a\bar\theta\theta} = 1 + a\theta\bar\theta$,

$$\int d\bar\theta\, d\theta\, e^{-a\bar\theta\theta} = a, \qquad \int \prod_i d\bar\theta_i d\theta_i\, e^{-\bar\theta M\theta} = \det M \quad (7)$$

Known . The determinant comes out **with its sign** (the source of the 0/0 of § 08). In this way the determinant becomes an integral over anticommuting fields $c^a, \bar c^a$ (**ghosts**; a different meaning from the "ghost" in 4-10, which refers to the negative-weight states of Pauli–Villars).

$$\det \mathcal{M} = \int \mathcal{D}\bar c\, \mathcal{D}c\, e^{-\int \bar c^u (\mathcal{M}c)^u}, \qquad \bar c^a (\mathcal{M}c)^a \simeq \partial_\mu \bar c^a \partial_\mu c^a + g f^{abc} (\partial_\mu \bar c^a) A^b_\mu c^c \quad (8)$$

($\simeq$ is integration by parts.) A ghost propagator δ^{ab}/k^2 and a vertex proportional to $g f^{abc} p_\mu$ come out Known . Ghosts are not particles that come out, and appear only inside loops (they also enter the β function: 4-06 § 03). **Example.** A real Gaussian integral in n components gives $(\det A)^{-1/2}$ (2-07 § 02). In the Feynman gauge ($\alpha = 1$) with $g = 0$ the kernel is $k^2\delta_{\mu\nu}$, so at each momentum the 4 components of the gauge field give $(k^2)^{-4/2}$ and the ghosts give $(k^2)^{+1}$ by (7), together $(k^2)^{-2/2}$: only the two transverse polarisations remain ($4 - 2 = 2$).

Detour: in the abelian case the ghosts disappear

A subsection you may skip.

For $U(1)$, $f^{abc} = 0$, so $\mathcal{M} = -\partial^2$ does not depend on A, and $\det \mathcal{M}$ cancels in ratios as a constant. No ghosts are needed. In the non-abelian case $\mathcal{M}$ depends on A, so the determinant couples to the field and ghost loops join the diagrams. The abstract of Faddeev–Popov (1967) writes that Feynman and DeWitt showed that in gauge-invariant field theories the rules for closed loops must be changed Known . The change of rules is needed when this $\mathcal{M}$ depends on A.

04

BRST symmetry — the symmetry that remains after fixing

The fixed action still has an anticommuting symmetry s that mixes the gauge field and the ghosts (with an auxiliary field b^a added).

$$sA^a_\mu = (D_\mu c)^a, \quad sc^a = -\tfrac{g}{2} f^{abc} c^b c^c, \quad s\bar{c}^a = ib^a, \quad sb^a = 0; \qquad S + \int s\Psi, \quad \Psi = \bar{c}^a (\partial_\mu A^{c}_{\mu}$$

$s\Psi = ib\,\partial A + \frac{\alpha}{2} b^2 + \bar{c}\,\mathcal{M} c$, and integrating out b returns to (6)(8). From $sS = 0$ and nilpotency $s^2 = 0$ (the Jacobi identity), two things follow formally Known .

- **Independent of the choice of gauge.** Changing Ψ changes the action only by $s(\delta\Psi)$, and for O annihilated by s, $\langle O\, s(\delta\Psi)\rangle = \langle s(O\delta\Psi)\rangle = 0$. The abstract of Becchi–Rouet–Stora (1976) writes that they showed that, for semisimple groups, in linear gauges in which all fields are massive, the Slavnov identities (the s symmetry written as relations between correlations) hold to all orders of renormalised perturbation theory, and that in the SU(2) Higgs–Kibble model examined as an example they showed, in the perturbative sense, that the physical S operator is unitary and independent of the gauge parameters (the 1975 paper is on the abelian Higgs–Kibble model). It is not a conclusion about massless pure Yang–Mills.
- **Physical states.** s becomes an operator Q_B acting on states (the BRST charge; $Q_B^2 = 0$). In covariant gauges the state space has an indefinite metric (the inner product $\langle \psi, \psi \rangle$ can be negative or 0), and as it stands it has no probabilistic interpretation. So we take the physical states to be $\mathrm{Ker}\, Q_B / \mathrm{Im}\, Q_B$ (BRST cohomology). This is "the states annihilated by Q_B, identified modulo the image of Q_B". Since $Q_B^2 = 0$, the image $\mathrm{Im}\, Q_B$ is contained in the kernel $\mathrm{Ker}\, Q_B$, and the quotient can be formed. The aim is that states in the image have inner product 0 with physical states, so dropping them by the identification does not change measurable quantities. States of negative or zero norm (the time and longitudinal polarisations, the ghosts) are dropped in this way, and the inner product on the remaining space is expected to be positive. The S operator is the operator that connects the states before and after scattering, and unitarity means preserving the total probability (the same property as e^{-itH} in 2-05 § 02). The abstract of Kugo–Ojima (1979) writes that, within a locally covariant operator formalism (a framework that assumes the theory and its asymptotic fields exist), they showed the unitarity of the physical S matrix from the algebra of the BRS charge Physics . The existence of the theory itself is assumed, and the identification in the continuum theory has not been shown (§ 09).

"BRST" is said to be the initials of Becchi, Rouet, Stora and Tyutin. **Example.** Computing s^2 symbolically on the part with constant fields, $s^2 A^a$ and $s^2 c^a$ had 0 terms for both SU(2) and SU(3). Changing the coefficient $-\frac{1}{2}$ of sc to -1 leaves 2 terms and 25 terms in $s^2 A^a$ Computed .

05

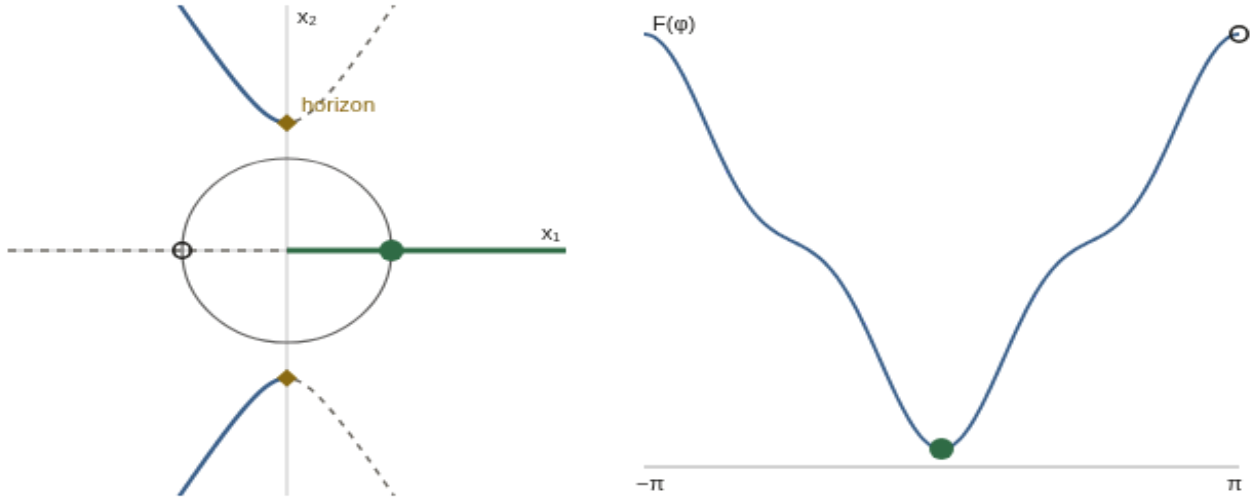
What happens when there are copies — four ways of counting

We build a case in which the "exactly once" of (3) breaks down. Following the Landau gauge (3-04 § 05), we take critical points of F on the circle as representatives, $M = \partial_\varphi^2 F$ as the determinant, and the set of local minima as the analogue of the Gribov region.

$F = -r\cos\varphi - c\,r^2\cos 3\varphi$: $\quad r < r_* = \frac{1}{3c}$ gives 2 copies (1 minimum, 1 maximum), $\quad r >$

For large fields pairs of a local minimum and a local maximum are born, and at those points $(0, \pm r_*)$, $M = 0$ (the analogue of the horizon). With the weight $e^{-r^2/2}$ we count r^2 (true expectation value 2) in four ways (for the difference between $|M|$ and M, see the detour of this section): (i) all copies with $|\det|$, (ii) only the Gribov region (local minima), (iii) only the global minimum of F (the fundamental modular region), (iv) with signs (the BRST way of counting).

Figure 1 Moving c and r changes the intersections (copies) of the condition surface with the orbit in the plane (left or top) and F(φ) on the circle (right or bottom). ● local minimum (Gribov region), ○ local maximum, green ● global minimum (fundamental modular region), ◇ horizon



c = 0.20, r = 1.20: copies 2 (min 1, max 1), sum of signs 0 r* = 1.667

⟨r²⟩: true 2.000, |det| 2.924, Gribov region 2.924, global min 2.000, signed 0/0

Computed The numbers in this figure were computed on this machine (the closed formulas of formula (10) and Table 1 evaluated on the spot; Table 1 agrees with a brute-force count of critical points to within 0.004).

Table 1 Four ways of counting ⟨r²⟩ (true value 2). Closed formula $\frac{4+8e^{-u}(1+u)}{2+4e^{-u}}$, $u = 1/(18c^2)$

c	r_*	(i) $\lvert\det\rvert$	(ii) Gribov region	(iii) Global minimum	(iv) With signs
0.08	4.167	2.005896	2.005896	2	0/0
0.20	1.667	2.924325	2.924325	2	0/0
0.50	0.667	2.273601	2.273601	2	0/0

When the number of copies varies from orbit to orbit, (i) is off, and (ii), which leaves copies inside the region, is not fixed either. (iii) is correct, but choosing the global minimum is a global comparison and cannot be written as a local condition. (iv) is 0/0, since local minima and local maxima come in equal numbers.

Example. The error of (i), $4ue^{-u}/(1+2e^{-u})$ (2×10^{-8} at $c=0.05$), is 0 at every order in powers of c, the same type as $e^{-1/x}$ in 4-07 § 05 Computed .

Detour: counting intersections with signs

A subsection you may skip. We look at the difference between $|M|$ and M in (10) with one variable on a circle.

Take the group to be the circle $\theta\in[0,2\pi)$, the condition a periodic function $f(\theta)=0$, and its zeros θ_i (with $f'\neq0$ there). From the rule for the δ function, $\int\delta(f)\,u\,d\theta=\sum_i u(\theta_i)/|f'(\theta_i)|$,

$$\int_0^{2\pi}d\theta\,\delta\big(f(\theta)\big)\,|f'(\theta)|=\#\{\theta_i\},\qquad\int_0^{2\pi}d\theta\,\delta\big(f(\theta)\big)\,f'(\theta)=\sum_i\operatorname{sign}f'(\theta_i)=0 \tag{10a}$$

The right-hand side is 0 because a periodic function crosses 0 upwards as many times as it crosses downwards. Example: for $f(\theta)=\sin(\theta-\theta_A)$ there are two intersections; the left-hand side is 2 and the right-hand side $1-1=0$. With absolute values the number of intersections remains; without them, the cancellation of signs. The relation to the obstruction to BRS-invariant gauge fixing done non-perturbatively on the lattice (Neuberger 1987; the abstract writes that it is formally present in the continuum as well) is treated in § 08.

06

The Gribov region and the fundamental modular region — convex and bounded in every direction

The abstract of Gribov (1978) writes that in the non-abelian case fixing $\partial_\mu A_\mu$ does not fix the gauge, and that restricting to the range where the FP determinant is positive seems to be enough Known . Reading the Landau gauge as the stationary points of the norm $\|A^g\|^2$,

$$\Omega=\{\partial_\mu A_\mu=0,\ \mathcal{M}(A)>0\}\ \text{(local minima)},\qquad\Lambda=\{A:\ \|A\|\le\|A^g\|\ \forall g\}\ \text{(global minim}$$

Ω is the **Gribov region**, its boundary the **horizon**, and Λ the **fundamental modular region** Known . $\bar\Omega$ is the closure of Ω (including its boundary). Ω is an open set given by $\mathcal{M}>0$, and Λ a closed set given by $\le$; the boundary of Λ has points touching the horizon (van Baal, § 07), so we write it with the closure.

- **Convex.** $\mathcal{M}(A)$ is linear in A, so "$\langle\omega, \mathcal{M}(A)\omega\rangle > 0$ for every ω" is an intersection of half-spaces. The abstract of Zwanziger (1982) also writes 'convex region Ω which is bounded in every direction'. Λ is also convex (the section "Complete gauge fixing" of van Baal 1997).
- **Bounded in every direction** (the abstract of Zwanziger 1982). The outline in finite dimensions: $\mathcal{M}_1$ in $\mathcal{M}(tA) = \mathcal{M}_0 + t\mathcal{M}_1$ is a symmetric matrix with trace 0 ($f^{aba} = 0$), so it has a negative eigenvalue, and for large t the operator stops being positive.
- **Every orbit passes inside the horizon** (the abstract of Dell'Antonio–Zwanziger 1991: the norm attains its minimum on each orbit, and there the orbit meets the inside of the horizon).
- **Copies remain even inside Ω.** Zwanziger (2004) argues that they do not affect expectation values Physics .

The solid line in Figure 1 is the analogue of Ω, and the positive x_1 axis the analogue of Λ. For the effect on the propagator, see 17-02 § 07.

07

The horizon in a box — eigenvalues in a constant background field

Taking a constant SU(2) field $A^a_\mu = a\,\delta^{a3}\delta_{\mu 1}$ in a periodic box of side L, $\mathcal{M}$ is diagonalised by plane waves $p = 2\pi n/L$, and splits into colour component 3 (neutral) and $\omega^1 \pm i\omega^2$ (charged).

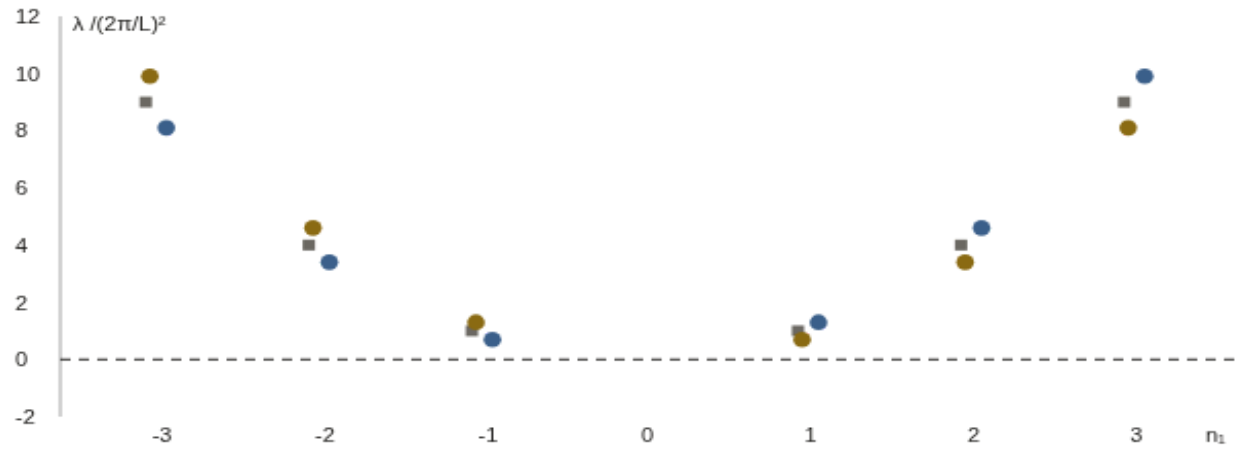
$$\lambda_{\text{neutral}} = \left(\tfrac{2\pi}{L}\right)^2 |n|^2, \qquad \lambda_\pm = \left(\tfrac{2\pi}{L}\right)^2 \left(|n|^2 \pm s\, n_1\right), \qquad s = \frac{gaL}{2\pi} \quad (12)$$

The smallest non-trivial eigenvalue (excluding $n = 0$) is $1 - s$ at $n = (\pm 1, 0, 0, 0)$, and the horizon is at

$$g\, a_{\text{h}} = 2\pi/L \;\longrightarrow\; 0 \quad (L \to \infty) \quad (13)$$

van Baal (the section "Gauge fields on the three-torus" of 1997) does the same calculation in the Coulomb gauge on the three-torus with $C = gaL$, and shows that the Gribov region is the cube $|C_k| \le 2\pi$ and that copies of $A = 0$ under twisted (antiperiodic) gauge transformations lie on the horizon Known . The fundamental modular region is $|C_k| \le \pi$ ($s \le \frac{1}{2}$) when antiperiodic transformations are also identified (pure Yang–Mills). When there are fields in the fundamental representation (quarks) and only periodic transformations are allowed, it is $|C_k| \le 2\pi$ ($s \le 1$), and its boundary touches the horizon Known .

Figure 2 Moving s (the size of the background) moves the eigenvalues of the FP operator (in units of $(2\pi/L)^2$, with transverse n equal to 0). Grey is neutral, blue and orange charged, red negative. The "boundary of Λ (s = 0.5)" in the readout is the value when antiperiodic transformations are also identified (s = 1 with periodic transformations only)




```
s = 0.30: smallest non-trivial eigenvalue 0.70 ×(2π/L)²  → inside Λ and Ω
```

Computed The numbers in this figure were computed on this machine (formula (12) evaluated on the spot). At $s = 0$, 0.5, 1 and 2 the smallest non-trivial eigenvalue is 1, 0.5, 0 and −1 (in order: $A = 0$, the boundary of the fundamental modular region [when antiperiodic transformations are also identified], the horizon, and the copy of $A = 0$ under a periodic transformation).

We also checked on the lattice. On an L^4 lattice, setting the links in direction 1 to $e^{i\theta\sigma 3/2}$ and diagonalising the second derivative of the Landau gauge functional $\sum(1 - \frac{1}{2}\mathrm{Re}\,\mathrm{tr}\, g_x U_{x,\mu} g^\dagger_{x+\mu})$ momentum by momentum, the charged eigenvalues agreed with $\frac{1}{2}[\cos\frac{\theta}{2} - \cos(p_1 \pm \frac{\theta}{2})] + \frac{1}{2}\sum_{\nu\neq 1}(1 - \cos p_\nu)$ (with lattice spacing 1, $\theta = ga$).

Table 2 The horizon of the lattice FP matrix (bisection)

L	4	8	16	32
$\theta_{\mathrm{h}} L/2\pi$	1.000000	1.000000	1.000000	1.000000

Computed python3, NumPy. The difference from the closed formula is at most 3×10^{-16}.

The larger the box, the closer to the origin the horizon is in long-wavelength directions, and the abstract of Gribov writes that this restriction cancels the infrared singularity and leads to a linear interaction at long distances Physics . It has not become a theorem about the four-dimensional measure.

08

The 0/0 on the lattice — it vanishes when counted with signs

The ghosts gave $\det \mathcal{M}$ without absolute value (formula (7)). So the integral of BRST-symmetric gauge fixing counts the copies on an orbit **with** $\mathrm{sign}\det \mathcal{M}$.

$$Z_{\mathrm{GF}}[U] = \int \prod_x dg_x\, \delta\big(G(U^g)\big)\, \det \mathcal{M}(U^g) = \sum_{\text{copies}} \mathrm{sign}\det \mathcal{M} = \chi(G)^N = 0 \quad (14)$$

We define the words used here. The **Euler characteristic** χ: the sum over the critical points (those whose matrix of second derivatives is non-degenerate) of a smooth function on a compact manifold without boundary (1-05), each counted with the sign of the determinant of that matrix ($+1$ if the number of negative eigenvalues is even, -1 if odd). This number does not depend on the choice of function (the Poincaré–Hopf theorem). On a circle a minimum counts $+1$ and a maximum -1, giving $\chi(S^1) = 0$; on the sphere $\chi(S^2) = 2$. For a compact connected Lie group G of positive dimension $\chi(G) = 0$, and for product spaces $\chi(G^N) = \chi(G)^N$ Known . **Morse theory** is the name of the theory that reads the shape of a manifold by counting critical points in this way.

The last two equalities read the sum of signs as the Euler characteristic of G^N (N is the number of lattice points) (Schaden 1999's interpretation via Morse theory, formula (2) of Mehta et al. 2014; on orbits whose critical points are non-degenerate), and $\chi(G) = 0$ Known (7-11 § 05). Expectation values of gauge-invariant quantities are 0/0. Mehta et al. call this "Neuberger's 0/0 problem" and cite Neuberger's papers of 1986 and 1987, and the abstract of the 1987 paper says that this obstruction met non-perturbatively is formally present in the continuum formulation as well Known .

Example. The "sum of signs 0" in Figure 1 is its smallest version ($\chi(S^1) = 0$). The same abstract also argues that at every order of the perturbative expansion the BRS invariance of the lattice does the job of local gauge invariance. This is because the cancellation of signs does not show up at any order of perturbation theory, the same type as $e^{-1/(18c^2)}$ in the example of § 05. The ways around it change the way of counting Known .

- **Do not fix** (Wilson's lattice; Elitzur's theorem: 18-02 § 09), or the maximal tree gauge, which has no Gribov problem (3-04 § 08).
- **Modified Landau gauge by stereographic projection**: map the link variables to flat coordinates by stereographic projection before building the functional (von Smekal et al. 2007; abstract: for compact $U(1)$ it solves the Gribov problem, and used on the maximal abelian subgroup it may avoid the 0/0 of $SU(N)$; the continuum limit of the Landau gauge is unchanged).
- **Orbifolding**: replace the group manifold used for gauge fixing by a generalisation of manifolds that allows singular points (an orbifold), and count with Morse theory on it (Mehta et al. 2014; abstract: the partition function becomes independent of the orbit).
- **Regularising with a mass**: count with an action in which mass terms (the Curci–Ferrari mass) are added to the gauge field and the ghosts, regularising the 0/0 (Ghiotti–von Smekal–Williams 2006, abstract). **Coset space**: BRST counting not on the whole group but on $SU(2)/U(1)$ (the sphere S^2, with $\chi = 2$) (Schaden).

For a comparison with the algebra of divergences, see 7-12.

09

Gauge fixing seen from the mass gap problem

What the problem statement asks for are correlations of gauge-invariant local fields (such as $\operatorname{tr} F_{\mu\nu}F_{\rho\sigma}$) and the gap of the reconstructed Hamiltonian (0-03). Gauge fixing is an indispensable tool in perturbation theory (§ 04). Elsewhere it has the following place Known .

- **Relation to the axioms.** The positivity needed for a probabilistic interpretation (reflection positivity, defined in 4-10) cannot be expected for fields fixed in a covariant gauge. That covariant gauges use a state space with an indefinite metric (§ 04) is a known framework, and Magnen–Rivasseau–Sénéor 1993 write that covariant gauges such as the Feynman gauge lack positivity and so cannot be used from the start of a construction (4-10 § 09). So the reconstruction is done from gauge-invariant quantities. The identification of BRST cohomology with that positive-definite space has not been shown in the continuum (the entry "physical states" of 18-02).
- **Attempts at construction** use it locally. On Bałaban, the exposition of the problem statement writes 'The choices vary locally in space-time, as well as on different length scales.' (Jaffe–Witten § 6.5, 7-05, 7-07). Chevyrev–Shen in two dimensions also use lattice gauge fixing (abstract, 7-09).

What Singer's theorem forbids is only continuous global slices, and the 0/0 too is a phenomenon of fixing a covariant gauge BRST-symmetrically over the whole lattice. The four-dimensional continuum limit of gauge-invariant correlations was built at the level of correlation functions by the construction published in 2026-11 (at the level of the committee's assessment; 6-00). The remaining questions are whether its measure exists and whether the spectrum has a gap (6-04).

What this chapter can say

	Content
What is said	The formal derivation of (3) to (9) Known . Table 1, Table 2, the numbers in the figures, the symbolic calculation of s^2 Computed . (10a) in the detour Paper (elementary)

	Content
Conditional	The Slavnov identities to all orders of perturbation theory in massive linear gauges; unitarity and gauge independence in the perturbative sense in the Higgs–Kibble model. The unitarity of Kugo–Ojima within a framework that assumes the theory exists Physics . The properties of the Gribov region checked at the level of abstracts. (14) on non-degenerate orbits. The effect on the infrared Physics
What is not said	The arrangement of copies in the four-dimensional continuum theory, the measure on the fundamental modular region, the identification of the BRST and OS spaces. Neuberger's argument only as far as the abstract. This site shows neither the mass gap nor the construction of the continuum theory

Sources and reproduction

Item	Kind	Source / tools
The procedure, Gribov copies	Known	L. D. Faddeev, V. N. Popov, Phys. Lett. B 25 (1967) 29; V. N. Gribov, Nucl. Phys. B 139 (1978) 1 (abstracts)
BRST, unitarity	Known	C. Becchi, A. Rouet, R. Stora, Commun. Math. Phys. 42 (1975) 127 (the abelian Higgs–Kibble model; bibliographic data only); Ann. Phys. 98 (1976) 287 (abstract); T. Kugo, I. Ojima, Prog. Theor. Phys. Suppl. 66 (1979) 1 (abstract). I. V. Tyutin, arXiv:0812.0580 (bibliographic data only)
The Gribov region	Known	D. Zwanziger, Nucl. Phys. B 209 (1982) 336; G. Dell'Antonio, D. Zwanziger, Commun. Math. Phys. 138 (1991) 291; D. Zwanziger, Phys. Rev. D 69 (2004) 016002 (abstracts); P. van Baal, hep-th/9711070 (text of the sections "Complete gauge fixing" and "Gauge fields on the three-torus")
The 0/0 on the lattice and ways around it	Known	H. Neuberger, Phys. Lett. B 175 (1986) 69 (bibliographic data only); Phys. Lett. B 183 (1987) 337 (abstract only); D. Mehta, N. S. Daleo, J. D. Hauenstein, C. Seaton, Phys. Rev. D 90 (2014) 054504, arXiv:1406.6678 (abstract, text § I, § II); M. Schaden, Phys. Rev. D 59 (1999) 014508 (bibliographic data only); L. von Smekal et al. arXiv:0710.2410 (abstract); M. Ghiotti et al. hep-th/0611058 (abstract)
Quotations in § 09	Known	A. Jaffe, E. Witten, "Quantum Yang–Mills Theory" (Clay) § 6.5; I. Chevyrev, H. Shen, arXiv:2302.12160 (abstract); J. Magnen, V. Rivasseau, R. Sénéor, Commun. Math. Phys. 155 (1993) 325 (via the quotation in 4-10 § 09)
Tables, figures, the numbers in § 01, § 03, § 04	Computed	python3, NumPy (toy_fp, box_horizon, lattice_formula_check, brst_nilpotent, propagator_check). The figures evaluate formulas (10)(12) in JavaScript

Next chapter to read: 4-09 The axioms of quantum field theory.

The axioms of quantum field theory — "the theory exists" said in seven conditions, and infinitely many degrees of freedom force a choice of representation

This chapter organises established knowledge (sources carry the “Known” label in the text). This book's own investigation is mainly in Part 15 and Volume 3 (Parts 16–28 and the appendices).

What this chapter shows — The seven conditions of the Wightman axioms, which the problem statement names as the content of "existence", and what each of them excludes. That the state space and the fields can be rebuilt from the totality of correlation functions. That with infinitely many degrees of freedom the vacuum may fail to lie in the Fock space, and that this, as Haag's theorem, excludes the interaction picture.

Prerequisite chapters — 1-10 Hilbert spaces and self-adjoint operators, 2-06 The harmonic oscillator and creation and annihilation. We also use test functions and distributions (1-11) and light cones and Lorentz transformations (2-04), but define each in a line where it is used. $\hbar = c = 1$. We write the squared length of a spacetime point $x = (t, \underline{x})$ as $x^2 = -t^2 + |\underline{x}|^2$; the metric signature $(-,+,+,+)$ refers to this arrangement of signs.

Said first — This is an introduction to a known framework, and the proofs are given only in outline. The problem statement of 2000, § 6, writes that no non-trivial theory satisfying the axioms in four dimensions is known. For four-dimensional Yang–Mills, the construction published in 2026-11 built one (at the level of the committee's assessment; section 09). This chapter shows nothing about Yang–Mills. The numbers are for free fields.

Order of this chapter

1. Why axioms — conditions for judging whether a theory has been made
2. Fields have no values at points — average first, then make an operator (Table 1)
3. The Wightman axioms — seven conditions (Table 2)
4. Wightman functions and reconstruction — building the theory from correlations

5. Locality — commuting is not the same as uncorrelated (Figure 1, Table 3)
6. What follows from the axioms — PCT, spin and statistics, clustering. Detour: the vacuum reaches everywhere
7. Representations with infinitely many degrees of freedom — with infinitely many particles, they do not fit in the same space (Figure 2, Table 4)
8. Haag's theorem — an interacting field cannot be placed in the space of the free field
9. What happens in gauge theories — observables, positivity, the form of the axioms

---

01

Why axioms — conditions for judging whether a theory has been made

Physics textbooks calculate on the assumption that an interacting field theory exists (4-03). The mathematical question comes before that. Unless we first decide what must be made for us to say it "exists", we cannot judge whether it has been made. The axioms are the list of conditions for that judgement.

§ 3 of the problem statement defines a quantum field as 'an operator-valued generalized function on space-time obeying certain axioms', and continues 'Gårding and Wightman gave mathematically precise axioms for quantum field theories on R4 with a Minkowski signature, see [45]'. § 4 takes the content of existence to be properties at least as strong as the axioms of [45] (the book of Streater–Wightman) and [35] (the two papers of Osterwalder–Schrader) (0-03) Known . What must be made is the following quadruple.

$$\mathcal{T} = (\mathcal{H},\ U,\ \Omega,\ \phi) \qquad \text{state space, representation of the symmetry, vacuum, field} \ (1)$$

Example. In the quantum mechanics of one particle (2-05), $L^2(\mathbb{R})$, e^{-itH}, the ground state and the position operator are these four, and they are at hand from the start. What is new in field theory is that the symmetry is the Poincaré group (translations and Lorentz transformations; 2-04), that fields have no value at a point (section 02), and that with infinitely many degrees of freedom $\mathcal{H}$ is not determined uniquely (section 07). The axioms on the imaginary-time side are treated in 4-10.

02

Fields have no values at points — average first, then make an operator

Placing the free field of 2-06 § 08 over the whole of space, the vacuum expectation value of the square of the value at a point $\phi(\underline{x})$ becomes the sum of the zero-point fluctuations of all modes, and diverges. Averaged with a function f, $\phi(f) = \int \phi(\underline{x}) f(\underline{x})\, d^D x$ (D is the dimension of space) is finite Known .

$$\langle \Omega, \phi(f)^2\, \Omega \rangle = \int \frac{d^D p}{(2\pi)^D} \frac{|\hat{f}(\underline{p})|^2}{2\omega_p}, \qquad \omega_p = \sqrt{\underline{p}^2 + m^2} \quad (2)$$

$\hat{f}(\underline{p}) = \int f(\underline{x})\, e^{-i\underline{p}\cdot\underline{x}}\, d^D x$ is the Fourier transform: the amplitude of f split into waves of wave number $\underline{p}$ (1-11). An f of width ε removes the modes with $|\underline{p}| \gtrsim 1/\varepsilon$, and as $\varepsilon \to 0$ (2) diverges. Example. For a normalised Gaussian ($\hat{f} = e^{-\varepsilon^2 p^2/2}$), (2) grows logarithmically for $D = 1$, as $e^{\varepsilon^2 m^2/2} K_0(\varepsilon^2 m^2/2)/(4\pi)$ (K_0 is from section 05), and like $1/(8\pi^2\varepsilon^2)$ for $D = 3$.

Computed Table 1 $\langle \phi(f)^2 \rangle$ of the free field averaged with a Gaussian of width ε ($m = 1$, numerical integration of (2)). In brackets, $\frac{1}{2\pi} \ln \frac{1}{m\varepsilon}$ and $\frac{1}{8\pi^2\varepsilon^2}$

$m\varepsilon$	$D = 1$	$D = 3$
0.1	0.43302 (0.36647)	1.2383 (1.2665)
0.01	0.79736 (0.73294)	126.59 (126.65)
0.001	1.16379 (1.09940)	12665 (12665)

The difference for $D = 1$ approaches $(\ln 2 - \gamma/2)/(2\pi) = 0.06439$ (γ is Euler's constant) Computed . It is finite for every $\varepsilon > 0$, and only $\varepsilon = 0$ has no meaning. So we define a field as a linear map from test functions $f \in \mathcal{S}(\mathbb{R}^4)$ to operators $\phi(f)$. This is an operator-valued distribution. $\mathcal{S}(\mathbb{R}^4)$ is the space of functions that are smooth and that, together with their derivatives of every order, decay faster than any power of $|x|$ (1-11). We average over spacetime and not only over space because, for interacting fields, a field at fixed time is thought not necessarily to make sense. In theories that need renormalisation of the field strength (4-04), fixing the time and averaging only over space does not make the quantity corresponding to (2) finite Known .

03

The Wightman axioms — seven conditions, and what they exclude

The axioms are conditions imposed on each part of (1). The numbering differs from book to book, so in this book we call them by the names of the conditions rather than by numbers. Table 2 is for a single real scalar field Known .

Table 2 The Wightman axioms (a neutral scalar field; a summary in the form of Streater–Wightman)

No.	Condition	Content	What it excludes
1	States and symmetry	A strongly continuous unitary representation $U(a, \Lambda)$ of the Poincaré group on a separable Hilbert space $\mathcal{H}$	Indefinite inner products (section 09)
2	Spectral condition	The joint spectrum of $P^\mu = (H, \boldsymbol{P})$ lies in V_+ (formula (3))	Negative energy
3	Uniqueness of the vacuum	A unit vector Ω invariant under U, unique up to a phase	Mixtures of vacua
4	Field axiom	For each $f \in \mathcal{S}(\mathbb{R}^4)$, $\phi(f)$ is defined on a dense $\mathcal{D} \ni \Omega$ with $\phi(f)\mathcal{D} \subset \mathcal{D}$; $\langle \psi, \phi(f)\chi \rangle$ is a tempered distribution; $\phi(f)^* \supset \phi(\bar{f})$	Values at points (section 02)
5	Covariance	$U(a,\Lambda)\phi(f)U(a,\Lambda)^{-1} = \phi(f_{a,\Lambda})$, $f_{a,\Lambda}(x) = f(\Lambda^{-1}(x-a))$	Special coordinate systems
6	Local commutativity	If the supports of f, g are spacelike separated, $[\phi(f), \phi(g)] = 0$	Influences faster than light (section 05)
7	Cyclicity	Linear combinations of $\phi(f_1)\cdots\phi(f_n)\Omega$ are dense in $\mathcal{H}$	States unrelated to the field

We fix the words of the table one line at a time. **Dense**: every element of $\mathcal{H}$ can be approached arbitrarily closely by elements of the set. **Separable**: the linear combinations of countably many vectors are dense. $U(a, \Lambda)$ is the unitary operator assigned to the pair of a spacetime translation a and a Lorentz transformation Λ (a linear transformation preserving x^2; 2-04 § 02), and **strongly continuous** means that as (a, Λ) approaches the identity, $U(a, \Lambda)\psi \to \psi$ for every ψ. A **tempered distribution** is a continuous linear functional on $\mathcal{S}(\mathbb{R}^4)$. Two points x, y are **spacelike separated** if $(x-y)^2 > 0$, that is, if they cannot be connected by a light signal, and the surface $x^2 = 0$ is the **light cone** of the origin (2-04 § 03).

$$\sigma(P) \subset V_+ = \{\, p = (p^0, \boldsymbol{p}) \;:\; p^0 \geq |\boldsymbol{p}| \,\} \qquad (3)$$

From (3) follows $H \geq 0$, and the mass operator $M = \sqrt{H^2 - \boldsymbol{P}^2}$ makes sense (4-11 § 05). The summary in § 3 of the problem statement names the states-and-symmetry axiom, the spectral condition ('positive energy, 0 ≤ H'), the uniqueness of the vacuum ('unique up to a phase'), covariance and local commutativity Known . For spin-1/2 fields one uses the double cover of the Poincaré group (a group in which two elements correspond to each Lorentz transformation; the four-dimensional version of $SU(2) \to SO(3)$ in 1-02), and local commutativity becomes anticommutation between fermions, $\phi(f)\phi(g) + \phi(g)\phi(f) = 0$. Example. The free field of mass m satisfies all seven, and its spectrum is the union of $\{0\}$, the hyperboloid $p^0 = \sqrt{\boldsymbol{p}^2 + m^2}$ and $p^0 \geq \sqrt{\boldsymbol{p}^2 + 4m^2}$ (Figure 4 of 2-06 § 08).

04

Wightman functions and reconstruction — building the theory from correlations

The vacuum expectation values of products of fields are called Wightman functions. By the field axiom (Table 2) they are tempered distributions.

$$W_n(x_1,\dots,x_n) = \langle \Omega,\ \phi(x_1)\cdots\phi(x_n)\,\Omega\rangle \quad (4)$$

The seven conditions translate into properties of W_n: Poincaré invariance, the spectral condition, locality (spacelike-separated neighbours may be swapped) and positivity. Positivity says that the length of $\Psi = \sum_n \phi(f_n)\Omega$ ($\phi(f_n) = \int f_n\,\phi(x_1)\cdots\phi(x_n)\,dx$) is not negative, and from $(\phi(x_1)\cdots\phi(x_n))^* = \phi(x_n)\cdots\phi(x_1)$

$$\|\Psi\|^2 = \sum_{n,k} \int f_n(x_1,\dots,x_n)\ W_{n+k}(x_n,\dots,x_1,y_1,\dots,y_k)\ f_k(y_1,\dots,y_k)\,dx\,dy\ \ge 0 \quad (5)$$

The reconstruction theorem Known . If $\{W_n\}$ satisfying these exist, a theory (1) with them as its Wightman functions can be made, unique up to unitary equivalence. The abstract of Wightman 1956 writes 'The problem of determining a theory of a neutral scalar field given its vacuum expectation values is posed and solved.' The construction is of the same type as in 4-10 § 05: put (5) as an inner product on finite sequences of test functions $(f_0, f_1, \dots)$, divide out the vectors of length 0 and complete, and the result is $\mathcal{H}$. $\Omega = (1,0,0,\dots)$, and $\phi(g)$ is the operation that replaces each component by the function with g attached in front ($(g\otimes f_1)(x_0,x_1) = g(x_0)f_1(x_1)$).

$$\phi(g):\ (f_0,f_1,f_2,\dots)\ \longmapsto\ (0,\ g\,f_0,\ g\otimes f_1,\ g\otimes f_2,\dots)$$

Analyticity follows from the spectral condition. Put $\psi = \phi(f)\Omega$ with a spacetime test function f, and let ρ be the spectral measure of H for ψ (1-10 § 05). Since Ω does not change under time evolution, the left-hand side is the two-point Wightman function with one of the fields shifted by time t and averaged.

$$\langle \psi,\ e^{-itH}\,\psi\rangle = \int_{[0,\infty)} e^{-iEt}\,d\rho(E) \ \xrightarrow{\ t=-i\tau\ }\ \int_{[0,\infty)} e^{-E\tau}\,d\rho(E) \quad (6)$$

Since $E \ge 0$, it extends analytically to $\operatorname{Im} t < 0$, and at $t = -i\tau$ oscillation turns into decay. This is the imaginary-time correlation, and the abstract writes 'The vacuum expectation values are shown to be boundary values of analytic functions.' This direction is the rigorous form of the Euclideanisation of 4-02, and the reverse direction is 4-10 Known . Example. For the free field, W_n is the sum of products of W_2 over all pairings, and reconstruction yields the Fock space of 2-06.

05

Locality — commuting is not the same as having no correlation

Local commutativity is the requirement that fields at two points that cannot be connected by light do not disturb each other. The time order of such two points is reversed for some observers (2-04 § 03), so if the order of the operators mattered there would be a contradiction. Local commutativity does not say that the correlation is 0. We compare with the free field in one space dimension.

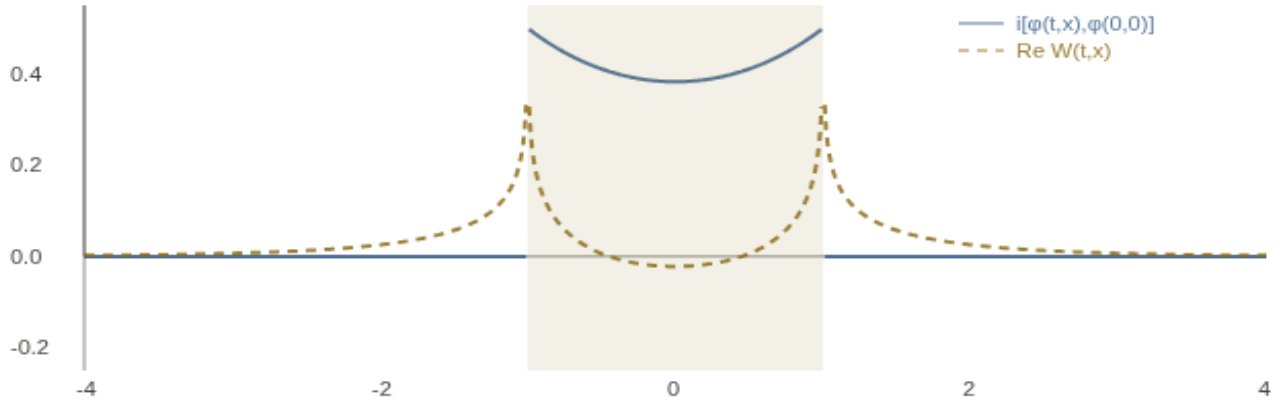
$$W(t,x) = \langle \Omega, \phi(t,x)\phi(0,0)\Omega \rangle = \int_{-\infty}^{\infty} \frac{dp}{2\pi \cdot 2\omega_p} e^{-i(\omega_p t - px)} \quad (7)$$

$$i\,[\phi(t,x),\phi(0,0)] = i\big(W(t,x) - W(-t,-x)\big) = \tfrac{1}{2}\,\mathrm{sgn}(t)\,\theta(t^2 - x^2)\,J_0\big(m\sqrt{t^2-x^2}\big) \quad (8)$$

$$W(0,x) = \frac{1}{2\pi} K_0(m|x|), \qquad K_0(z) \approx \sqrt{\frac{\pi}{2z}}\, e^{-z} \quad (z \to \infty) \quad (9)$$

θ is the step function, and $J_0(z) = \frac{1}{\pi}\int_0^\pi \cos(z\sin\vartheta)\,d\vartheta$ and $K_0(z) = \int_0^\infty e^{-z\cosh u}\,du$ are Bessel functions. The θ in (8) is local commutativity: outside the light cone the commutator is 0. But the correlation (9) is not 0 even outside, and decreases with distance as e^{-mr}. This decay is the spatial face of the mass gap (4-11 § 06). Passing to Euclidean space, the spatial direction and the imaginary-time direction are related by rotation, so it decays at the same rate as the imaginary-time decay of (6) (4-11 § 02). In four dimensions $W(0,x) = mK_1(mr)/(4\pi^2 r)$ Computed .

Figure 1 Changing the time t and the mass m changes the shape of the commutator $i[\phi(t,x),\phi(0,0)]$ (solid line) and the correlation $\mathrm{Re}\,W(t,x)$ (dashed line) of the free field in 1+1 dimensions. The shading is the inside of the light cone. "Play" advances t



t = 1.00, m = 1.00: x = 1.50 (outside): commutator 0, Re W = 0.0567 / at x = 0: commutator 0.3826, Re W = −0.0221

Computed The numbers in this figure were computed on this machine ((8)(9) and the series and integral for Y_0 evaluated in JavaScript; agreeing with SciPy to within 10^{-13}).

Computed Table 3 The cross-section at $m = 1$, $t = 1$. Inside, $\operatorname{Re} W$ is $-\frac{1}{4} Y_0(m\sqrt{t^2 - x^2})$ (a Bessel function of the second kind). Agrees with numerical integration of (7) to 5 digits

x	Position	$i[\phi(t,x), \phi(0,0)]$	$\operatorname{Re} W(t,x)$
0	Inside	0.38260	−0.02206
0.9	Inside	0.47653	0.13611
1.1	Outside	0	0.15875
2	Outside	0	0.02529
4	Outside	0	0.00205

On the light cone $\operatorname{Re} W$ diverges logarithmically. The divergence at $t = x = 0$ is the same as the logarithm for $D = 1$ in section 02, a sign that fields have no value at a point.

06

What follows from the axioms — PCT, spin and statistics, clustering

From the seven conditions alone, properties known from experience come out as theorems. The title of [45], 'PCT, Spin and Statistics and all That', refers to two of them. The PCT theorem says that the combination of spatial reflection, time reversal and the exchange of particles and antiparticles is a symmetry of every theory; the spin–statistics theorem says that, at spacelike-separated points, fields of integer spin commute and fields of half-integer spin anticommute Known . The third is clustering: under the uniqueness of the vacuum, when groups of fields are moved apart in a spatial direction, correlations split into products.

$$W_{n+k}(x_1, \dots, x_n,\ y_1 + \lambda a, \dots, y_k + \lambda a) \xrightarrow[\lambda \to \infty]{} W_n(x_1, \dots, x_n)\, W_k(y_1, \dots, y_k) \quad (10)$$

With a mass gap the convergence is exponentially fast. § 5 of the problem statement writes 'An important consequence of the existence of a mass gap is clustering' (4-11 § 06) Known . Example. For the free field, the difference in (10) is a sum of terms e^{-mr} from (9).

Detour: the vacuum reaches everywhere

A subsection you may skip.

The Reeh–Schlieder theorem says that the density in cyclicity (Table 2) holds even with only test functions supported in an arbitrarily small open set of spacetime (attributed to Reeh–Schlieder 1961) Known . Operating only on the fields inside a room, one can make distant states as closely as one likes. Strictly it is an approximation, and the operators used are in general not unitary. This can happen because, as the dashed line in Figure 1 shows, the vacuum has

correlations reaching outside the light cone. On the other hand, the commutator (solid line) is 0 outside, so signals cannot be sent faster than light.

07

Representations with infinitely many degrees of freedom — with infinitely many particles, they do not fit in the same space

With finitely many oscillators, the irreducible representation of the commutation relations is unique (Stone–von Neumann; 2-06 § 09). With infinitely many it is not. We look at two free fields with different masses. For a single oscillator, the number n of quanta in the ground state of angular frequency ω', counted with the annihilation operator a_ω of ω, and the overlap between the ground states are

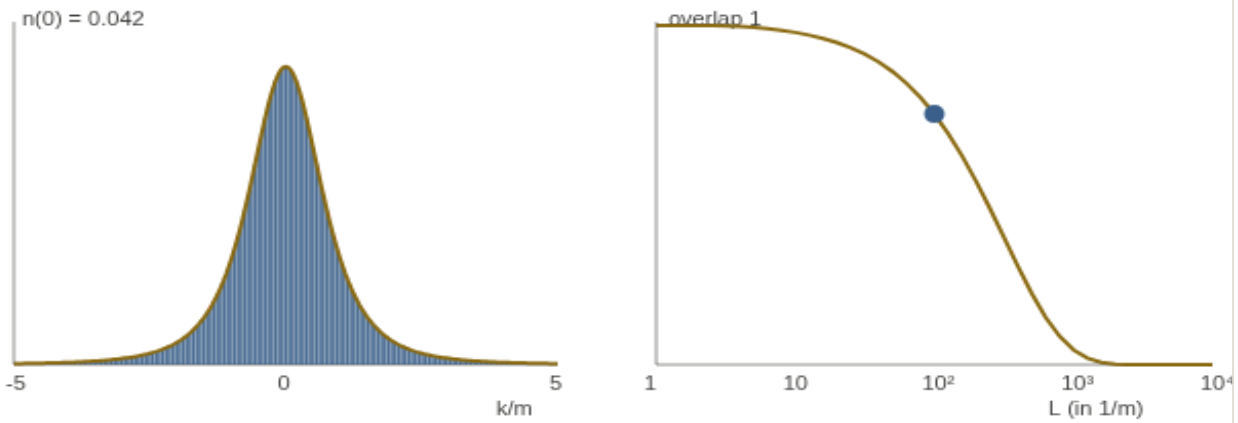
$$n = \langle 0\omega' | a_\omega^\dagger a_\omega | 0\omega' \rangle = \frac{(\omega - \omega')^2}{4\omega\omega'}, \qquad \langle 0\omega | 0\omega' \rangle = \left(\frac{2\sqrt{\omega\omega'}}{\omega + \omega'} \right)^{1/2} = (1+n)^{-1/4} \tag{11}$$

Known Computed . For a field on a ring of length L (2-06 (14)), each mode $k = 2\pi j / L$ has $\omega_k = \sqrt{k^2 + m^2}$ and $\omega'_k = \sqrt{k^2 + m'^2}$, and the vacuum is a product over modes, so

$$N(L) = \sum_k n_k, \qquad \langle \Omega_m, \Omega_{m'} \rangle = \prod_k (1+n_k)^{-1/4} \tag{12}$$

n_k is the n of (11) taken for mode k, and $N(L)$ is the expected number of quanta of mass m contained in the vacuum of mass m'. Since $n_k \approx (m'^2 - m^2)^2 / (16 k^4)$ decreases, for finite L the sum is finite and the overlap is positive. Enlarging L, both the number of modes and $N(L)$ are proportional to L, and the overlap goes to 0 as e^{-sL}.

Figure 2 Changing L and m'/m changes, on the left, n_k for each mode (vertical bars are the modes of the ring, the curve is $n(k)$), and on the right, the relation between the vacuum overlap and L (logarithmic scale). "Play" enlarges L



L = 100/m, m'/m = 1.50: number of quanta N = 1.269 (per 1/m: 0.0127), overlap 0.731

Computed The numbers in this figure were computed on this machine ((12) summed over $|k| \le 60m$, with the rest filled in by the k^{-4} tail; agrees with Table 4 to 6 digits).

Computed Table 4 $N(L)$ and the overlap on a ring of length L (in units of $1/m$) (formula (12))

L	N $(m' = 1.5m)$	Overlap	N $(m' = 2m)$	Overlap
10	0.12695	0.96915	0.43539	0.90040
100	1.2692	0.73104	4.3534	0.35025
1000	12.692	0.043592	43.534	2.778e−5
10000	126.92	2.478e−14	435.34	2.740e−46

Enlarging L, the overlap goes to 0. Precisely, in infinite volume there is no vector in the Fock space of m corresponding to the vacuum of m', and the two give unitarily inequivalent representations of the same commutation relations of the time-zero fields Known . From (12) one can read that the overlap is positive if and only if $\sum_k n_k < \infty$. A criterion of the same form is known for general Bogoliubov transformations (transformations that replace a_k by a linear combination of a_k and $a^\dagger_{-k}$; the relation between a_ω and $a_{\omega'}$ in (11) is an example) (attributed to Shale 1962). The infinite-volume density $\int \frac{d^D k}{(2\pi)^D} n_k$ $(m' = 1.5m)$ is finite: $0.012692,\ 0.0049736,\ 0.0030929$ (in units of m^D) for space dimension $D = 1, 2, 3$. Even for $D = 3$, which corresponds to real spacetime (four-dimensional spacetime), the density is finite, and the breakdown comes from the volume alone. For $D = 4$, where space is four-dimensional (spacetime five-dimensional), the density diverges logarithmically, and the overlap is 0 even in a finite box Computed . A figure for the three-dimensional lattice is in 7-11 § 07.

08

Haag's theorem — an interacting field cannot be placed in the space of the free field

Perturbation theory in physics takes the Fock space of the free field as its stage, splits $H = H_0 + V$ (H_0 free, V the interaction), and uses the picture of passing from the free vacuum to the interacting vacuum (the interaction picture). In section 07, even a mild change that only alters the mass gave a new vacuum that did not lie in the original Fock space.

Haag's theorem says this in general form Known . If the time-zero fields and conjugate momenta are unitarily related to those of the free field, and both theories have a unique vacuum invariant under spatial translations and rotations, then the first four Wightman functions agree with those of the free field, and a theory whose two-point function is the same as the free field's is the free field of the same mass (Haag 1955, Hall–Wightman 1957; the consolidated form is in Streater–Wightman, and a reading in Earman–Fraser 2006). The central lemma is elementary: in the Fock space of mass $m >$

0, the only vectors invariant under all spatial translations are constant multiples of the vacuum.

$$e^{-i\vec{a}\cdot(\vec{p}_1+\cdots+\vec{p}_n)}\,\psi_n(p_1,\ldots,p_n)=\psi_n(p_1,\ldots,p_n)\ \ (\forall \vec{a})\implies \psi_n=0\quad (n\geq 1)\ \text{(13)}$$

The n-particle component ψ_n vanishes off the surface where the total momentum is 0, and that surface has volume 0, so as an element of L^2 it is 0. Example. The vacuum of m' in (12) is invariant under translations and differs from the free vacuum, so it has no place in the Fock space.

In practice there are two ways of handling this. Physics calculates with a finite box and an ultraviolet cutoff, and takes as answers only the quantities with the cutoff removed by renormalisation (4-04). Constructive field theory builds the Hamiltonian in a space with a cutoff and obtains a new representation in the limit (7-03). Particles are also made from the spectrum. § 6 of the problem statement writes 'If m is an isolated eigenvalue of M, then one infers from the Wightman axioms and Haag–Ruelle scattering theory that asymptotic scattering states of an arbitrary number of particles exist' Known .

09

What happens in gauge theories — observables, positivity, the form of the axioms

It is known that the gauge field A_μ itself does not become a field satisfying the Wightman axioms (attributed to Strocchi 1967, 1970) Known . There are three reasons. Observables: A_μ changes under gauge transformations (3-01). § 4 of the problem statement asks that local operators correspond to 'the gauge-invariant local polynomials in the curvature F and its covariant derivatives, such as Tr Fij Fkl (x)', and in the footnote on local fields (footnote 1 of Jaffe–Witten) warns that, because of renormalisation, there is no natural one-to-one correspondence between classical polynomials and quantum operators Known .

Positivity: treating A_μ as a field in a Lorentz-covariant gauge needs a space with an indefinite inner product, and the states-and-symmetry axiom is broken. In the Gupta–Bleuler method (for the electromagnetic field), one takes from the large space the subspace of states satisfying a condition, and recovers positivity by dividing out the states of length 0 in it Known . For the non-abelian BRST method (4-08 § 04), a construction of the same form (take a subspace and divide; BRST cohomology) is known at the level of perturbation theory (attributed to Kugo–Ojima 1979). At the non-perturbative level this construction is not established, because of Gribov copies (4-08 § 05) Physics . The form of observables: Wilson loops (3-02) are quantities on curves and do not fit into the field axiom, which averages fields at points (4-10 § 09). § 3 of the

problem statement also mentions the Haag–Kastler framework, which places an algebra of operators on each region, and writes 'One hopes that the continued mathematical exploration of quantum field theory will lead to refinements of the axiom sets that have been in use up to now'.

Where things stand. § 6 of the problem statement writes 'at present we do not know any non-trivial relativistic field theory that satisfies the Wightman (or any other reasonable) axioms in four dimensions' (an assessment from 2000). In two and three dimensions, interacting scalar fields satisfying the axioms (ϕ^4_2, ϕ^4_3 and others) have been constructed (7-03) Known . Three-dimensional Yang–Mills has not been constructed (7-06). For four-dimensional Yang–Mills, the construction published in 2026-11 built, as the scaling limit of Wilson's lattice gauge theory, a theory satisfying the construction's own version of the Wightman axioms (the committee's assessment is that it is proved in Lean; 6-00, 12-08). The local fields are renormalised insertions corresponding to gauge-invariant polynomials in the field strength (8-04), and Wilson loops are not treated (6-04). The road that makes fields from the Euclidean side is the next chapter.

What this chapter can say

Content	What is said	What is not said
The axioms (section 03) and their correspondence with the problem statement	Known . § 3 to § 6 of the problem statement checked against the text	Table 2 is a summary of the standard form (the text of Streater–Wightman not obtained)
Reconstruction, analyticity (section 04)	Known (the abstract of Wightman 1956) and the outline of the construction	The details of the proofs
Tables 1 and 3, Figure 1, formulas (2)(7) to (9)	Computed : agreement in double precision for the free field	The form of correlations of interacting fields
Table 4, Figure 2, formulas (11)(12)	Computed : the overlap between vacua of free fields goes to 0 with the volume	The proof of Shale's criterion (text not read)
PCT, spin and statistics, Reeh–Schlieder, Haag's theorem	Presentation of the Known claims, and the lemma (13)	The proofs of the theorems (original papers bibliographic data only)

Content	What is said	What is not said
The axioms for gauge theories (section 09)	Organising the requirements of the problem statement and the known obstructions	Recovery of positivity by BRST at the non-perturbative level is not established. The form of axioms including Wilson loops is not settled. A theory satisfying the axioms for four-dimensional Yang–Mills was built by the construction published in 2026-11 (at the level of the committee's assessment; 6-00). This chapter shows nothing

Sources and reproduction

Item	Kind	Source / tools
Quotations from the problem statement	Known	A. Jaffe, E. Witten, "Quantum Yang–Mills Theory" (Clay) § 3 to § 6, footnote 1. Checked word for word against the text
Axioms, PCT, spin and statistics, fields at fixed time (section 02)	Known	R. Streater, A. Wightman, *PCT, Spin and Statistics and all That*, Benjamin, 1964. Bibliographic data only
Reconstruction, analyticity	Known	A. S. Wightman, Phys. Rev. 101 (1956) 860. Checked in the abstract
Haag's theorem	Known	R. Haag, Mat.-fys. Medd. Dan. Vid. Selsk. 29 no. 12 (1955): bibliographic data only (secondary sources). D. Hall, A. S. Wightman, 'A theorem on invariant analytic functions with applications to relativistic quantum field theory', same journal 31 no. 5 (1957): bibliographic data only (consistent across secondary sources). J. Earman, D. Fraser, Erkenntnis 64 (2006) 305–344: bibliographic data only
The criterion for Bogoliubov transformations	Known	D. Shale, Trans. Amer. Math. Soc. 103 (1962) 149–167. Bibliographic data only
Detour	Known	H. Reeh, S. Schlieder, Nuovo Cim. 22 (1961) 1051–1068. Bibliographic data only
Haag–Kastler, Haag–Ruelle	Known	R. Haag, D. Kastler, J. Math. Phys. 5 (1964) 848–861: bibliographic data only (bibliographic data checked with Crossref). D. Ruelle, Helv. Phys. Acta 35 (1962) 147: bibliographic data only, not checked. R. Haag, *Local Quantum Physics*, Springer, 1992, and Phys. Rev. 112 (1958) 669. Bibliographic data only
Gauge fields and positivity (section 09)	Known, Physics	F. Strocchi, Phys. Rev. 162 (1967) 1429–1438, Phys. Rev. D 2 (1970) 2334–2340. T. Kugo, I. Ojima, Prog. Theor. Phys. Suppl. 66 (1979) 1–130. Bibliographic data only (bibliographic data checked with Crossref)
Tables 1, 3 and 4, Figures 1 and 2, (8)(9)(11)(12) and the formula with K_1	Computed	python3, SciPy (`smear_variance.py`, `wightman_1p1.py`, `free_3p1.py`, `bogoliubov.py`). The figures evaluate the same formulas in JavaScript

Next chapter to read: 4-10 The Osterwalder–Schrader axioms and reconstruction.

The Osterwalder–Schrader axioms and reconstruction — reflection positivity turns probability back into quantum mechanics

> This chapter organises established knowledge (sources carry the “Known” label in the text). This book's own investigation is mainly in Part 15 and Volume 3 (Parts 16–28 and the appendices).

What this chapter shows — The five conditions imposed on Schwinger functions (regularity, Euclidean covariance, reflection positivity, permutation symmetry and clustering; the Osterwalder–Schrader axioms E0 to E4), and the procedure that builds from them a Hilbert space, $H \geq 0$ and a vacuum. That the key, reflection positivity, is for one-dimensional Gaussian processes the same as "the correlation is a positive superposition of decaying exponentials". Regularisations that break reflection positivity, and the lattice, which keeps it.

Prerequisite chapters — 4-02 The path integral and Euclideanisation, 4-09 The axioms of quantum field theory. Tools from 1-08, 1-10, 1-11, 2-06 and 2-07. $\hbar = c = 1$.

Said first — This is an introduction to known theorems, and the proofs are given only in one-dimensional outline. For four-dimensional Yang–Mills, the construction published in 2026-11 built a Hilbert space from the Schwinger functions of the limit by the reconstruction of this chapter (the committee's assessment is that it is proved in Lean; 6-00, 12-08). A probability measure satisfying the axioms has not been constructed (6-04). This chapter shows nothing. Section 06 is a calculation in a one-particle model.

Order of this chapter

1. Making the way back a theorem
2. Schwinger functions
3. Five conditions (Table 1)
4. Reflection positivity (Figure 1, Table 2)
5. Reconstruction
6. Running reconstruction on the smallest model (Figure 2, Table 3)

7. Detour: noisy data
8. The free field
9. Regularisations that break reflection positivity, and the lattice that keeps it
10. What gauge theories add

01

Making the way back a theorem — from imaginary time to Hilbert space

What can be made on the Euclidean side is a probability measure μ and its moments (4-02, One sign is different). What the problem statement asks for is a Hilbert space $\mathcal{H}$, $H \geq 0$ and a vacuum Ω (4-09), and § 4 writes 'Existence includes establishing axiomatic properties at least as strong as those cited in [45, 35].' [35] is the two papers of Osterwalder–Schrader (OS) Known .

The way there (from spacetime to imaginary time) is analytic continuation. On the way back, Schwinger said in the discussion at the 1958 CERN conference 'The question of to what extent you can go backwards, remains unanswered' (according to Jaffe 2018). Nelson built the way back under the Markov property (the property that, given the present, past and future are independent), and OS replaced it by reflection positivity (problem statement § 3) Known .

$$\mu \;\longrightarrow\; \{S_n\} \xrightarrow[\text{reconstruction}]{(E0)\sim(E4)} (\mathcal{H},\ H \geq 0,\ \Omega) \xrightarrow[\text{analytic continuation}]{(E0')} \{W_n\} \quad (1)$$

Example. For one particle, $\mathbb{E}[x(\tau)x(0)] = \langle x\psi_0, e^{-\tau(H-E_0)}x\psi_0\rangle$ is the way there (formula (8) of 2-06). Section 06 is the reverse.

02

Schwinger functions — energy is written into the correlations

We write points as $x = (\tau, \underline{x}) \in \mathbb{R}^d$ and call τ the Euclidean time. The n-point **Schwinger function** is a moment of μ; precisely, the object is $S_n(f)$ smoothed with test functions (1-11). It is related to the Wightman function formally by the substitution $x^0 \to -i\tau$, $S_n = W_n(x_1^E, \dots, x_n^E)$ (Summers § 3) Known .

$$S_n(x_1, \dots, x_n) = \int \phi(x_1)\cdots\phi(x_n)\, d\mu(\phi) \quad (2)$$

Example: the free field of mass m is the Gaussian measure with covariance $(-\Delta + m^2)^{-1}$ (Δ is the Laplacian $\sum_\mu \partial_\mu^2$, not the Δ of the mass gap) (problem statement § 6.2). Integrating over p_0 alone (formula (11) of 2-04) gives

$$S_2(x,y) = \int \frac{d^d p}{(2\pi)^d} \frac{e^{ip\cdot(x-y)}}{p^2+m^2}, \qquad \int \frac{dp_0}{2\pi} \frac{e^{ip_0\tau}}{p_0^2+\omega^2} = \frac{e^{-\omega|\tau|}}{2\omega} \quad (3)$$

where $\omega = \sqrt{\underline{p}^2 + m^2}$ is the energy of a particle of momentum $\underline{p}$. The energy is written into the rate at which the correlation decreases: this is the seed of reconstruction. For $d = 1$ the right-hand side is itself $S_2(\tau)$, which becomes the example of section 04. The higher S_n are sums of products of S_2 over pairings (1-08).

03 Five conditions — from regularity to clustering, and linear growth

Let $\Theta(\tau, \underline{x}) = (-\tau, \underline{x})$ be time reversal.

Known Table 1 The OS conditions (scalar field; the numbering is said to be that of OS 1975)

Condition	Content	On the spacetime side
Regularity (E0)	Tempered away from coinciding points	Operator-valued distributions
Euclidean covariance (E1)	Invariant under rotations and translations	Poincaré covariance
Reflection positivity (E2)	For f_n supported at positive times, $\sum_{n,m} S_{n+m}(\Theta f_n \otimes f_m) \geq 0$	Positive inner product, $H \geq 0$
Permutation symmetry (E3)	Invariant under permutations of the points	Separated fields commute
Clustering (E4)	$S_{n+m} \to S_n S_m$ when moved apart in a spatial direction	A unique vacuum
Linear-growth condition (E0′)	$\lvert S_n(f)\rvert \leq \sigma_n \lvert f\rvert_{C_n}$, $\sigma_n \leq A(n!)^B$ (the order of smoothness used in the estimate is proportional to n: the origin of the name "linear growth")	Continuation to W_n

Notation of the table. The **support** of a function f is (the closure of) the set of points where $f \neq 0$, and "supported at positive times" means $f = 0$ for $\tau \leq 0$. $\otimes$ is the function with separated variables multiplied, $(f \otimes g)(x,y) = f(x)g(y)$, and Θf is f with the time of each variable reversed. $\lvert f\rvert_k$ is the supremum of the derivatives of f up to order k multiplied by polynomial weights, measuring the smoothness and decay of rapidly decreasing functions (1-11).

That regularity excludes coinciding points agrees with the fact that S_2 of the free field diverges at $x = y$ (for $d \geq 2$). The proof in the 1973 paper had an error concerning the continuity of tensor products of distributions (Jaffe 2018), and the second paper of 1975 is said to have given a correct proof with the linear-growth condition added as an assumption (according to secondary sources; OS 1975 bibliographic data only). It has been shown that the S_n made from a Wightman theory satisfy the five conditions, but they do not necessarily satisfy the linear-growth condition. So it is not known whether the OS axioms including the linear-growth condition are equivalent to the Wightman axioms themselves (Summers § 3, footnote 15). What OS 1975 showed is equivalence with a modified form of Wightman theory (Jaffe 2018; the modification is said to add a growth condition) Known . Example. The free field satisfies all of them, and the linear-growth condition follows from S_{2k} being a sum over $(2k-1)!! \leq (2k)!$ pairings. The number of pairings $(2k-1)!!$ goes into σ_n, and each factor is bounded by a seminorm of fixed order.

04

Reflection positivity — measuring lengths with the mirror of time

Let F be a quantity determined by the fields at positive times only (for example $F = \sum_j c_j \phi(\tau_j, \underset{\sim}{x}_j)$, $\tau_j > 0$), and ΘF the same with time reversed. Reflection positivity has the following form.

$$\langle F, F \rangle := \mathbb{E}\big[\, \Theta F \; F \,\big] \;\geq\; 0 \qquad (4)$$

Speaking in metaphor, it is a condition that measures the length of a future quantity by its overlap with its mirror image. Precisely, it is the condition that the quadratic form of correlations between past and future is not negative, and unlike $\mathbb{E}[|F|^2] \geq 0$ it is not trivial. For a one-dimensional stationary Gaussian process (covariance $C(\tau - \tau')$), taking $F = \sum_j c_j \phi(\tau_j)$,

$$\sum_{i,j} c_i c_j C(\tau_i + \tau_j) \geq 0 \;\; (\tau_j > 0), \qquad \text{in particular} \;\; C(2\tau) \geq 0, \;\; C(\tau_1 + \tau_2)^2 \leq C(2\tau_1) C(2\tau_2$$

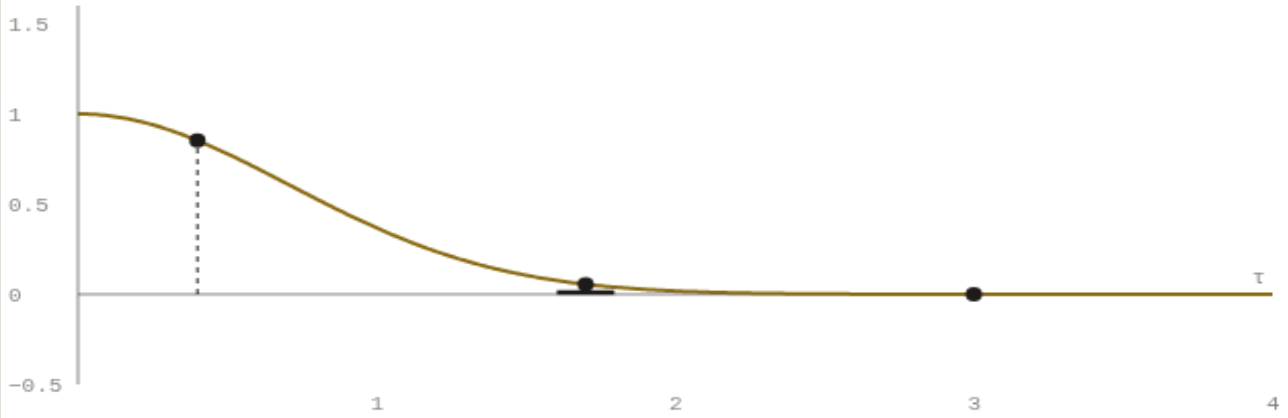
The right-hand side is the diagonal and determinant of the 2×2 matrix of two points, and requires the convexity of $\log C$. A sufficient condition: if $C(\tau) = \int_{[0,\infty)} e^{-E\tau}\, d\rho(E)$ ($\rho \geq 0$), the left-hand side becomes $\int |\sum_j c_j e^{-E\tau_j}|^2 d\rho \geq 0$.

Example: $e^{-\tau^2}$ is not reflection positive

$e^{-\tau^2}$ is a legitimate covariance (its Fourier transform is positive; **Bochner's theorem**: a continuous function C is positive definite, that is, $\sum_{i,j} c_i c_j C(\tau_i - \tau_j) \geq 0$ for any τ_i

and complex c_i, if and only if its Fourier transform is a positive measure), but $\log C$ is concave, and at $\tau 1 = 0.2$, $\tau 2 = 1.5$, from $C(0.4) = 0.852144$, $C(1.7) = 0.055576$, $C(3.0) = 1.234 \times 10^{-4}$ the determinant is -2.98×10^{-3} Computed . (Why it cannot be written in this form: for positive ρ, $C(\tau) \geq \rho([0, E])e^{-E\tau}$, so it cannot decrease faster than an exponential. That this is the cause of the violation becomes clear once necessity is shown in section 05.)

Figure 1 Choosing a covariance and moving τ_1, τ_2 moves the three points $C(2\tau_1), C(\tau_1 + \tau_2), C(2\tau_2)$ (circles) and the upper bound of (5) (horizontal bar). It is violated if an end circle is below 0, or if the absolute height of the middle circle exceeds the bar



e^−τ², τ₁=0.20, τ₂=1.50: C(2τ₁)=0.8521, C(τ₁+τ₂)=0.0556, C(2τ₂)=0.00012, smallest eigenvalue −0.00349 (violated)

Computed The numbers in this figure were computed on this machine (the matrix of (5) computed on the spot).

Computed Table 2 "Matrix" is the smallest eigenvalue of $[C(\tau_i + \tau_j)]$ ($\tau_i = 0.1i$, $i \leq 40$), and "two points" the minimum of the determinant of (5) for $\tau_1, \tau_2 \in (0, 2]$.

$C(\tau)$	Fourier transform (positive)	Matrix	Two points
$e^{-\lvert\tau\rvert}$	$\frac{2}{1+\omega^2}$	≈ 0	≈ 0
$e^{-\lvert\tau\rvert} + \frac{1}{2}e^{-3\lvert\tau\rvert}$	$\frac{2}{1+\omega^2} + \frac{3}{9+\omega^2}$	≈ 0	≈ 0
$e^{-\tau^2}$	$\sqrt{\pi}\, e^{-\omega^2/4}$	−0.462	−0.238
$(1 - \lvert\tau\rvert)_+$	$\left(\frac{\sin(\omega/2)}{\omega/2}\right)^2$	−0.366	−0.239
$e^{-\lvert\tau\rvert}\cos 3\tau$	$\frac{1}{1+(\omega-3)^2} + \frac{1}{1+(\omega+3)^2}$	−2.13	−0.376
$\frac{1}{2}e^{-\lvert\tau\rvert} - \frac{1}{6}e^{-3\lvert\tau\rvert}$	$\frac{8}{(1+\omega^2)(9+\omega^2)}$	−0.054	−0.011

All six are legitimate as covariances, and the bottom four violate reflection positivity ($(a)_+ = \max(a, 0)$). The damped oscillation is the form of a propagator with complex poles $\omega = \pm 3 \pm i$, and the Gribov–Zwanziger form of 17-02 also has complex poles. By (5), a field whose correlation in the time direction becomes negative cannot be read as

a field of a reflection-positive theory. In lattice calculations in the Landau gauge, the correlation of gluons in the time direction has been observed to become negative, and physics reads this as a sign that gluons are not observed particles Physics .

05

Reconstruction — inner product, contraction, Hamiltonian

Given (4), the rest proceeds by linear algebra Known . **(1) Inner product.** On the space $\mathcal{E}_+$ of all bounded functions of the fields at positive times, put $\langle F, G\rangle := \mathbb{E}[\Theta F\, G]$. By (4), $\langle F, F\rangle \geq 0$, but $\langle F, F\rangle = 0$ is possible even with $F \neq 0$ (positive semidefinite). Even so, Cauchy–Schwarz $|\langle F, G\rangle|^2 \leq \langle F, F\rangle\langle G, G\rangle$ holds (from the discriminant of the quadratic $\langle F + tG, F + tG\rangle \geq 0$ in t). From this the set $\mathcal{N}$ of elements of length 0 is a subspace, and we complete $\mathcal{E}_+/\mathcal{N}$, which identifies elements that differ only by something of length 0, to get $\mathcal{H}$ (for completion, 1-10). **(2) Shifting.** The shift U_s by $s \geq 0$ in the time direction takes quantities at positive times to quantities at positive times, and from the translation invariance of μ and $\Theta U_s \Theta = U_{-s}$,

$$\langle F, U_s G\rangle = \langle U_{s/2} F, U_{s/2} G\rangle = \langle U_s F, G\rangle \quad (6)$$

(3) Contraction. Repeating (6) and Cauchy–Schwarz n times gives

$$\|U_s F\|^2 = \langle F, U_{2s} F\rangle \leq \|F\|\,\|U_{2s} F\| \implies \|U_s F\| \leq \|F\|^{1-2^{-n}} \|U_{2^n s} F\|^{2^{-n}} \quad (7)$$

and since $\|U_u F\| \leq \sup |F|$, as $n \to \infty$ we get $\|U_s F\| \leq \|F\|$. Length 0 goes to length 0, and U_s defines $T(s)$ on $\mathcal{H}$. **(4) Hamiltonian.** $T(s)$ is symmetric with $0 \leq T(s) = T(s/2)^2 \leq 1$ and $T(s)T(u) = T(s+u)$, so if it is continuous, the spectral theorem (1-10) gives

$$T(s) = e^{-sH}, \qquad H = H^* \geq 0, \qquad \Omega := [1], \qquad H\Omega = 0 \quad (8)$$

This is the Hamiltonian.

Necessity in one dimension, and real time

In one dimension $C(\tau + 2\varepsilon) = \langle \psi_\varepsilon, e^{-\tau H} \psi_\varepsilon\rangle = \int e^{-E\tau}\, d\rho_\varepsilon(E)$ ($\psi_\varepsilon = [\phi(\varepsilon)]$, and ρ_ε is the spectral measure; ϕ is not bounded, but for Gaussians $\mathcal{E}_+$ may be extended to square-integrable functions), so the sufficient condition of section 04 is also necessary. Therefore a C that decreases faster than an exponential, like $e^{-\tau^2}$, cannot be reflection positive. The form $\int e^{-E\tau}\, d\rho(E)$ is called the **Laplace transform** of ρ, and on this side it corresponds to the characterisation of Widder 1934. The connection with reflection positivity is said to have been pointed out by Klein–Landau (Jaffe 2018). Since $H \geq 0$, e^{-zH} makes sense for $\operatorname{Re} z \geq 0$, and at $z = it$ it is time evolution e^{-itH}.

Boosts are obtained in the OS reconstruction from the analytic continuation of the Wightman functions. Fröhlich–Osterwalder–Seiler 1983 and Klein–Landau abstracted this into the form of extending the rotation angle to imaginary values (0-04 § 02) (Jaffe 2018 § 3.3.1). The linear-growth condition is used at the stage of obtaining W_n Known . Example. If $C = e^{-m|\tau|}/(2m)$ and $F = \phi(\tau_0)$, then $\|U_s F\| = e^{-ms}\|F\|$, and on this vector H acts as multiplication by m.

06

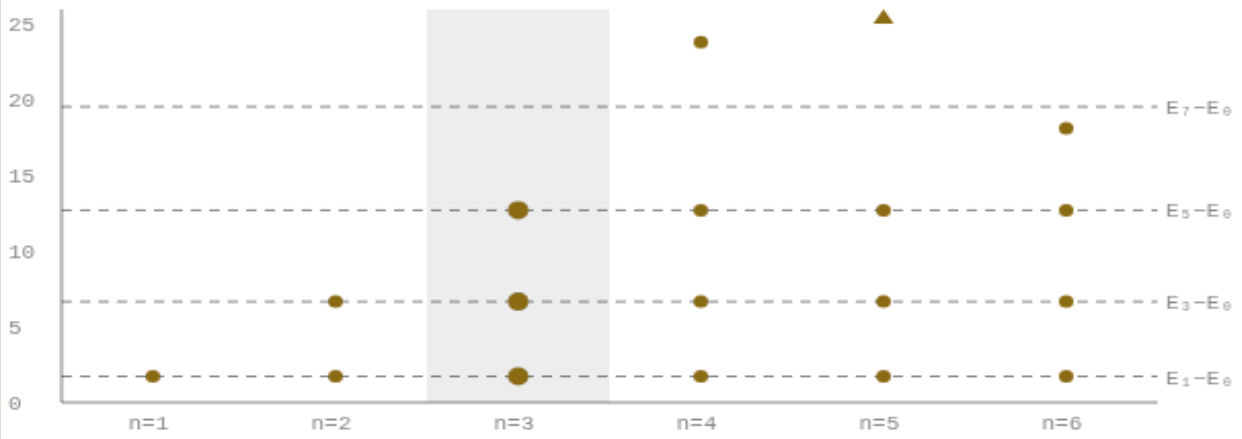
Running reconstruction on the smallest model — reading levels from correlations alone

With $H = \frac{1}{2}p^2 + x^4$ of Table 4 of 2-07 we make $C(\tau) = \mathbb{E}[x(\tau)x(0)]$ (diagonalising in a basis of 240 functions), then forget H and carry out section 05 with C alone. With $s = 0.25$ and $F_i = x((i+\frac{1}{2})s)$ $(i < n)$,

$$G_{ij} = \langle F_i, F_j\rangle = C\big((i+j+1)s\big), \qquad G^{(1)}_{ij} = \langle F_i, T(s) F_j\rangle = C\big((i+j+2)s\big) \quad (9)$$

from the generalised eigenvalue problem $G^{(1)}v = \lambda G v$ (the same as the eigenvalues of $G^{-1/2}G^{(1)}G^{-1/2}$ if G is positive definite) we read $E = -\frac{1}{s}\log\lambda$. Since these are eigenvalues of $T(s)$ restricted to the subspace spanned by F_i, by the variational principle (1-10) they are **upper bounds** on the true levels Known . That G is positive definite (reflection positivity) is presupposed. x is odd, so what can be read are $E_1 - E_0$, $E_3 - E_0$, …, and the weights $|\langle \psi_0, x\psi_n\rangle|^2$ fall sharply, from 0.287 at $n = 1$ to 3.9×10^{-9} at $n = 7$ Computed .

Figure 2 Changing the number n of correlations and the noise ε changes the levels read (circles; one n per column). The dashed lines are the true levels, the triangles above 26



ε=0 (rounding only), n=3: smallest eigenvalue of G 3.6e−10. Levels read: 1.72566、6.66774、12.71344

Computed The numbers in this figure were computed on this machine (results of the same calculation as Table 3).

Computed Table 3 Levels read from the table of correlations alone (no noise), and the smallest eigenvalue of G.

n	E_1-E_0	E_3-E_0	E_5-E_0	E_7-E_0	min G
1	1.72807				1.9e-1
2	1.72566	6.67039			2.4e-5
3	1.72566	6.66774	12.71344		3.6e-10
4	1.72566	6.66774	12.71165	23.80937	2.5e-16
True	1.725658	6.667744	12.711350	19.552863	

At $n = 3$ three levels come out to 4 or 5 digits, all above the true values. At $n = 4$ the smallest eigenvalue of G reaches the size of double-precision rounding, and beyond that nothing can be distinguished.

Detour: noisy data — the way back is unstable

A subsection you may skip.

Correlations from Monte Carlo (5-07) have errors. We multiplied $C(ks)$ by normally distributed random errors of relative size ε (with a fixed seed) and ran the same procedure Computed . At $n = 4$ with $\varepsilon = 10^{-12}$, the third level is 12.70331, below the true 12.71135, and the upper-bound property is lost. At $\varepsilon = 10^{-8}$ the smallest eigenvalue of G becomes negative, -6.4×10^{-10}, and the data violate (4) (the matrix form on the left of (5)) ($E3 - E0$ is 6.66936). At $\varepsilon = 10^{-4}$ with $n = 6$, a spurious level of 1.52597, below the true gap, appears (Figure 2). Because the Laplace transform is being traced backwards, tiny errors move the upper levels a lot. The reconstruction theorem is a claim about S_n without errors, and reading levels from finite-precision data is a different, ill-conditioned problem. Glueball measurements (5-08) also use matrices of the form (9).

07

The free field — Fock space comes out

For Gaussian measures the reconstruction can be written out completely. With the one-dimensional $C(\tau) = e^{-m|\tau|}/(2m)$ (an oscillator of angular frequency m), using **Wick products**, the powers of ϕ adjusted by the variance $\sigma^2 = C(0)$ ($:\phi^2: = \phi^2 - \sigma^2$, $:\phi^3: = \phi^3 - 3\sigma^2\phi$, and in general $:\phi^k: = \sigma^k He_k(\phi/\sigma)$ with the Hermite polynomials $He_k(x) = (-1)^k e^{x^2/2}\frac{d^k}{dx^k}e^{-x^2/2}$),

$$\mathbb{E}\big[{:}\phi(\tau)^k{:}\,{:}\phi(0)^l{:}\big] = \delta_{kl}\,k!\,C(\tau)^k = \delta_{kl}\,\frac{k!}{(2m)^k}e^{-km|\tau|} \tag{10}$$

and the decay rate km is the level. For this process (the stationary Gaussian process with covariance $e^{-m|\tau|}/(2m)$ is called the **Ornstein–Uhlenbeck process**), $[{:}\phi^k{:}]$ is an eigenvector with eigenvalue km, and $H = m\,a^\dagger a$ Known . The levels $E_k - E_0 = km$ of 2-06 (2-06) come out from the correlations alone. Example. At $m = 1$, $k = 2$ it is $\frac{1}{2}e^{-2\tau}$, and the rate 2 is a two-particle state.

In d dimensions, by (3) positive exponentials line up momentum by momentum, so reflection positivity holds (problem statement § 6.2); reconstruction yields Fock space, and the spectrum is $\{0\} \cup [m, \infty)$ (2-06 § 08). Known The action of Yang–Mills has no mass term (3-02), and the gap must be an output of the interaction (4-11).

08

Regularisations that break reflection positivity, and the lattice that keeps it

Regularisations that suppress the ultraviolet (4-04) can break reflection positivity. Example: the Pauli–Villars type. Subtracting the propagator of a heavy mass M makes it decrease fast, as p^{-4}, at large p.

$$\frac{1}{p^2+m^2} - \frac{1}{p^2+M^2} = \frac{M^2-m^2}{(p^2+m^2)(p^2+M^2)} > 0, \qquad C_{\mathrm{PV}}(\tau) = \frac{e^{-m|\tau|}}{2m} - \frac{e^{-M|\tau|}}{2M} \tag{11}$$

It is legitimate as a covariance, but in the time direction the exponential weight is negative at $E = M$. At $m = 1$, $M = 3$, $CC'' - C'^2 = -\frac{1}{3}e^{-4\tau} < 0$, so $\log C$ is concave everywhere, and the test of (5) is violated at any two distinct points (the last row of Table 2) Computed . A negative weight corresponds to states with negative inner product (ghosts; a different meaning from the "ghost" of 4-08 § 03, which refers to the Faddeev–Popov anticommuting fields) Known . § 6.5 of the problem statement writes that there are few ways of recovering reflection positivity lost through regularisation, and gives dimensional regularisation and Pauli–Villars regularisation as examples (quoted in 0-03 § 03).

The lattice is the opposite. The Wilson action keeps reflection positivity (Osterwalder–Seiler 1978, problem statement § 6.5), and the transfer matrix and the Hamiltonian stand up (for an example, 0-04 § 06; for gauge theories, 5-03). In exchange, rotations are cut down to steps of 90°. Reflection positivity is an inequality, so it passes to the limit (7-01 § 07) Known .

09

What gauge theories add — quantities on curves, gauge fixing, volume

Observables. What has meaning are gauge-invariant quantities, and the local ones are composite quantities such as $\mathrm{Tr}\, F_{\mu\nu}F_{\rho\sigma}(x)$ (problem statement § 4). On Wilson loops (3-02), Chatterjee, citing Seiler, writes 'the key observables in these theories do not take values at points, but at curves'. **Gauge fixing.** Correlations of fixed fields are in general not reflection positive, and covariant gauges use spaces with indefinite inner products (4-08). Magnen–Rivasseau–Sénéor 1993 say that covariant gauges such as the Feynman gauge 'cannot be used directly at the beginning because of their lack of positivity', use a regularised axial gauge, and go no further than expecting that OS positivity 'could be shown to hold with some additional work'. **Volume.** Even on T^4, if a reflection-positive measure can be made, a finite-volume Hilbert space and transfer matrix are obtained (this is so for the periodic box of the lattice; the construction itself on the continuum T^4 is unfinished; problem statement § 6), but a Poincaré-covariant theory on $\mathbb{R}^4$ needs $T^4 \to \mathbb{R}^4$ (0-04 § 09) Known .

Example. From Driver's 1989 measure in two dimensions, the Euclidean invariance of gauge-invariant expectation values follows, but the full set of axioms is not claimed (7-02). § 6.4 of the problem statement calls the two-dimensional abelian Higgs model (Brydges–Fröhlich–Seiler) 'the only complete example of an interacting gauge theory satisfying the axioms' (an assessment from 2000), and § 6.5 continues 'Establishing a quantum mechanical Hilbert space is part of the solution to this Millennium problem.' Known . The construction published in 2026-11 built a Hilbert space from the Schwinger functions of the lattice limit by the reconstruction of this chapter, and checks the construction's own version of the Wightman axioms (the committee's assessment is that it is proved in Lean; 12-08). The volume limit is along a subsequence, and the limit is obtained as a mixture of phases (12-07).

What this chapter can say

Content	Status
Sections 01 to 03, 07 to 09	Known (the text of the OS papers not obtained)
The argument of sections 04 and 05	Proof written in one-dimensional outline Known . Real time, boosts and W_n only stated
The reading on gluons at the end of section 04	Physics (lattice numbers and their interpretation; not a theorem)
Tables 2 and 3, Figures 1 and 2, the test of (11), the detour	In double precision Computed

Content	Status
Axioms, Hilbert space and gap for four-dimensional Yang–Mills	A theory satisfying the axioms and a Hilbert space were built by the construction published in 2026-11 (at the level of the committee's assessment; 6-00). The gap has not been shown (6-04). This chapter shows nothing

Sources and reproduction

Item	Kind	Source / tools
Tables 2 and 3, the figures, the numbers in the text	Computed	python3, NumPy, SciPy (eigenvalues, diagonalisation, the generalised eigenvalue problem, random seed 1)
Quotations from the problem statement	Known	A. Jaffe, E. Witten, *Quantum Yang–Mills Theory* (Clay Mathematics Institute, 2000) § 3, § 4, § 6.2, § 6.4, § 6.5. Checked in the text
The five conditions (E0 to E4) and the linear-growth condition (E0′)	Known	K. Osterwalder, R. Schrader, Comm. Math. Phys. 31 (1973) 83–112; 42 (1975) 281–305. Bibliographic data only. The numbering and the form of the linear-growth condition follow Summers § 3 and a secondary source (Wikipedia "Osterwalder–Schrader theorem")
Schwinger's remark	Known	Discussion at the 1958 CERN conference (proceedings p.140, a reply to a question by Pauli). Via Jaffe 2018; the proceedings not read
History, $S_n = W_n(x^E)$, the role of the linear-growth condition	Known	A. Jaffe, arXiv:1802.07880; S. J. Summers, arXiv:1203.3991 (§ 3, footnote 15). Text checked
Nelson, boosts, the Laplace transform	Known	E. Nelson, J. Funct. Anal. 12 (1973) 97–112, 211–227; J. Fröhlich, K. Osterwalder, E. Seiler, Ann. of Math. 118 (1983) 461–489; A. Klein, L. Landau, J. Funct. Anal. 44 (1981) 121–137; D. V. Widder, Bull. Amer. Math. Soc. 40 (1934) 321–326. Bibliographic data only
The argument of section 05, the Wilson action	Known	J. Glimm, A. Jaffe, *Quantum Physics*, 2nd ed. (Springer, 1987), bibliographic data only (the relevant chapter not checked); K. Osterwalder, E. Seiler, Ann. Phys. 110 (1978) 440–471. Bibliographic data only
The gluon correlations at the end of section 04	Physics	A. Cucchieri, T. Mendes, A. R. Taurines, Phys. Rev. D 71 (2005) 051902; P. O. Bowman et al., Phys. Rev. D 76 (2007) 094505 (lattice QCD with quarks). Bibliographic data only (volume and page checked with Crossref)
Section 09	Known	S. Chatterjee, arXiv:1803.01950 § 1; J. Magnen, V. Rivasseau, R. Sénéor, Comm. Math. Phys. 155 (1993) 325–383 (§ I); B. K. Driver, Comm. Math. Phys. 123 (1989) 575–616 (introduction). The MRS quotations are from § I. Bochner's theorem is defined in the text

Next chapter to read: 4-11 The definition of the mass gap. For reflection positivity on the lattice, 5-03.

The definition of the mass gap — the lowest energy above the vacuum, and the uniform decay of correlations

> This chapter organises established knowledge (sources carry the “Known” label in the text). This book's own investigation is mainly in Part 15 and Volume 3 (Parts 16–28 and the appendices).

What this chapter shows — The mass gap is "the infimum of the energy of states orthogonal to the vacuum", and in imaginary time this is equivalent to "all connected correlations decay at least at one rate". A single correlation, or measurements over a finite range, give only upper bounds. How particle states, clustering and the correlation length are like and unlike the gap.

Prerequisite chapters — 4-10 The Osterwalder–Schrader axioms and reconstruction, 2-08 Statistical mechanics and phase transitions. Tools from 1-10 (the spectral theorem, the variational principle). $\hbar = c = 1$, $d = 4$.

Said first — This organises definitions and known correspondences, and shows nothing about the gap of four-dimensional Yang–Mills. The Hilbert space and Hamiltonian that asking about a gap presupposes were built by the construction published in 2026-11 (the committee's assessment is that it is proved in Lean; 6-00). Whether that H has a gap has not been shown (6-04). All the computed examples are free fields.

Order of this chapter

1. Definition — the infimum of the energy of states orthogonal to the vacuum
2. The imaginary-time face — equivalent to uniform decay of all correlations
3. Decay observable by observable (Figure 1, Table 1)
4. A finite window — measurements give only upper bounds (Figure 2, Table 2)
5. How it differs from particle states
6. Clustering
7. The correlation length (Table 3)
8. Finite boxes and uniformity

9. Reading it in gauge theory
10. Appendix table: the uses of the word "gap" — equivalent, upper bound, different (appendix table)

01

Definition — the infimum of the energy of states orthogonal to the vacuum

The definition in § 4 of the problem statement is 'A quantum field theory has a mass gap Δ if H has no spectrum in the interval $(0,\ \Delta)$ for some $\Delta > 0$. The supremum of such Δ is the mass m, and we require m < ∞.' (for a word-by-word reading, 0-03 § 06). This sentence presupposes (a) a Hilbert space $\mathcal{H}$ and a self-adjoint $H \geq 0$ (what the reconstruction of 4-10 § 05 gives) and (b) a vacuum Ω unique up to a phase, with $H\Omega = 0$ (clustering in Table 1 of 4-10), and (c) asks about states orthogonal to the vacuum. In the words of § 1, 'every excitation of the vacuum has energy at least Δ'. Under (b), the definition is the same as the following Known .

$$m = \inf \sigma\big(H|_{\Omega^\perp}\big) = \inf_{\psi \perp \Omega,\ \psi \neq 0} \frac{\langle \psi, H\psi\rangle}{\langle \psi, \psi\rangle}, \qquad \text{there is a gap} \iff m > 0 \quad (1)$$

The second equality is the variational principle (1-10 § 06). That (b) is needed can also be read from (1). If there were a state Ω' of energy 0 orthogonal to the vacuum, (1) would be 0 with $\psi = \Omega'$, but the first sentence of the problem statement alone would miss it, since 0 lies outside $(0, \Delta)$. Example. For the free field of mass m, (1) is also m (we use the same letter), and for mass 0 it is 0 (4-10 § 07).

02

The imaginary-time face — the gap is equivalent to "all correlations decay uniformly"

In the notation of 4-10 § 05, a bounded quantity F at positive times gives a vector $[F]$, and the shift U_τ by τ in the time direction becomes $e^{-\tau H}$. Putting $\psi_F = [F] - \mathbb{E}[F]\,\Omega$, for the **connected correlation** (the correlation with the vacuum component subtracted),

$$\mathbb{E}\big[\Theta G\, U_\tau F\big] - \mathbb{E}[G]\,\mathbb{E}[F] = \langle \psi_G, e^{-\tau H}\psi_F\rangle, \qquad \big|\langle \psi_G, e^{-\tau H}\psi_F\rangle\big| \leq e^{-m\tau}\,\|\psi_G\|\,\|\psi_F\| \quad (2)$$

The inequality follows from the spectral theorem (1-10 § 05) Known . The converse also holds. If $\langle \psi_F, e^{-\tau H}\psi_F\rangle \leq e^{-\Delta\tau}\|\psi_F\|^2$ for every bounded F, then $m \geq \Delta$.

To see this we use the **spectral measure** μ_F of ψ_F: the distribution of weight along the energy axis E when ψ_F is split into eigencomponents of H. If there is an eigenbasis e_k ($He_k = \lambda_k e_k$), then $\mu_F = \sum_k |\langle e_k, \psi_F\rangle|^2\, \delta_{\lambda_k}$ (δ_λ is a weight 1 placed at the point λ), and even for continuous spectrum a distribution with the same meaning is fixed by the spectral theorem. Then $\langle \psi_F, e^{-\tau H}\psi_F\rangle = \int e^{-E\tau} d\mu_F(E)$. The left-hand side satisfies $\int e^{-E\tau} d\mu_F \geq e^{-\Delta'\tau}\mu_F([0,\Delta'))$ ($\Delta' < \Delta$), so as $\tau\to\infty$, $\mu_F([0,\Delta')) = 0$. The set of all ψ_F is **dense** in $\Omega^\perp$ (every vector of $\Omega^\perp$ can be approached arbitrarily closely), so $H \geq \Delta$ on $\Omega^\perp$. Since this includes $E = 0$, the uniqueness of the vacuum comes out as well Known . **The gap is equivalent to all connected correlations decaying at least at one rate that does not depend on the observable.**

A single correlation $C_F(\tau) = \int e^{-E\tau} d\mu_F(E)$ is **log-convex**, that is, $\log C_F$ is convex ($C(\tau_1+\tau_2)^2 \leq C(2\tau_1)\,C(2\tau_2)$ in (5) of 4-10 corresponds to this). So the **effective mass** decreases monotonically.

$$m_{\text{eff}}(\tau) = -\frac{d}{d\tau}\log C_F(\tau) \;\searrow\; E_F := \min \operatorname{supp}\mu_F, \qquad m = \inf_F E_F \quad (3)$$

$\operatorname{supp}\mu_F$ (the **support**) is the smallest closed set carrying the weight of μ_F, and E_F is the lowest energy that ψ_F touches. m_{eff} at any time is an upper bound on E_F: from above it is bounded by a single F, but from below every F is needed (the same shape as the box "The asymmetry of upper and lower bounds" in 1-10). On the lattice U_a is the transfer matrix $T = e^{-aH}$, $m_{\text{lat}} = \log(\lambda_0/\lambda_1)$ and $\Delta = m_{\text{lat}}/a$. λ_0 and λ_1 are the largest and second-largest eigenvalues of T (in gauge theory, on gauge-invariant states; section 08) ((1) of 2-09 is its smallest example).

03

Decay observable by observable — a single correlation sees only the part its observable touches

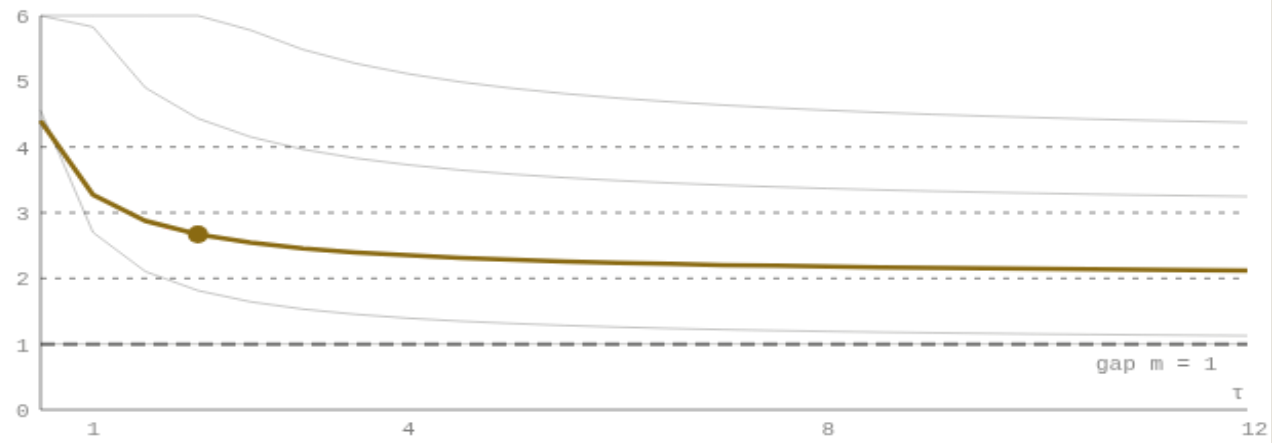
We see that E_F changes with the observable in the four-dimensional free field ($m = 1$). The correlation of O_k, the Wick products of 4-10 § 07 integrated over space to give momentum 0, is

$$O_k(\tau) = \int d^3x :\!\phi^k\!:(\tau,\vec{x}), \qquad \frac{\mathbb{E}[O_k(\tau)O_k(0)]}{\text{volume}} = k!\int d^3x\, S(\tau,\vec{x})^k, \qquad S(x) = \frac{m\,K_1(m|x|)}{4\pi^2|x|}$$

(S is the two-point function of the four-dimensional free field, K_1 a modified Bessel function). O_k creates exactly k particles with total momentum 0, so $EO_k = km$. At large τ the energy of each particle can be approximated as $m + |\vec{p}|^2/(2m)$, and the correlation of k particles with total momentum 0 is $e^{-km\tau}$ multiplied by $\tau^{-3(k-1)/2}$

from the Gaussian integral over the three-dimensional momenta of the remaining $k-1$. So near the lower end $m_{\text{eff}} \approx km + \frac{3(k-1)}{2\tau}$ Computed . Only $k=1$ sees the gap 1; however precisely $:\phi^2:$ is measured, it only approaches 2. This is because the number of particles is conserved for the free field; even with interactions, symmetries split the space into parts (for Yang–Mills, the channels of section 09). For the unprojected point-to-point $S(\tau, 0)$, even with the same ϕ, it decreases slowly, like $1 + \frac{3}{2\tau}$. The energies of moving particles form a continuum, and the integral over one three-dimensional momentum multiplies by $\tau^{-3/2}$ for the same reason; this is why lattice measurements sum over space (5-08).

Figure 1 Choosing an observable and moving τ, the effective mass (thick line and circle) comes down towards the lower end for that observable (thin dashed line). Only φ comes down to the gap m = 1 (thick dashed line)



:φ²:, τ = 2.0: m_eff = 2.671 (lower end 2, gap 1)

Computed The numbers in this figure were computed on this machine (a table of (4) at steps of 0.5 in τ is embedded).

Computed Table 1 $m_{\text{eff}}(\tau)$ of the free field ($m=1$). In brackets, the lower end E_F. The momentum-0 ϕ is 1.00000 at every τ, agreeing with the closed form $e^{-\tau}/2$ within double-precision rounding.

τ	φ point (1)	:φ²: (2)	:φ³: (3)	:φ⁴: (4)
2	1.81431	2.67113	4.43494	6.22414
4	1.39395	2.35013	3.72854	5.11460
8	1.19273	2.18027	3.36864	4.55915
16	1.09513	2.09177	3.18575	4.28029

Every column decreases monotonically (checked at all points at steps of 0.25 for $\tau \le 12$), agrees with the asymptotic form to within 0.13% at $\tau = 16$, and stays above the lower end to the last. What can be read is an upper bound on E_F Computed .

04

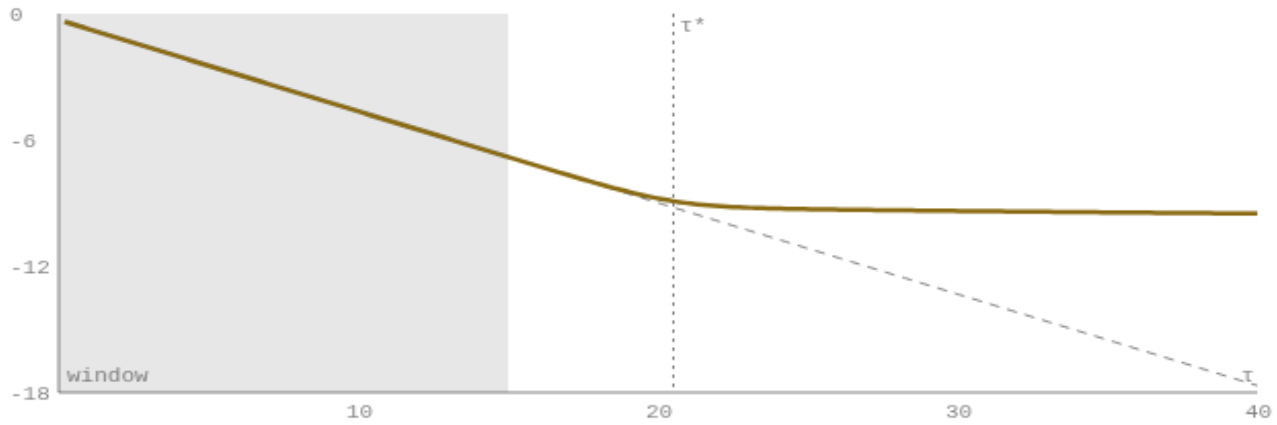
A finite window — measuring the decay gives no lower bound on the gap

We call the range of τ that a measurement can see a **window** (precisely, a finite interval $[\tau_{\min}, \tau_{\max}]$). The infimum (3) involves infinitely many F and $\tau \to \infty$, so a measurement within a window sees only part of it. We look at a theory with no gap, made by placing side by side, independently, a free field ϕ of mass 1 and a free field χ of mass 0. The cross terms vanish, and the correlation of $O = \int d^3x\, [\phi + \varepsilon :\chi^2:]$ is

$$\frac{\mathbb{E}[O(\tau)O(0)]}{\text{volume}} = \frac{e^{-\tau}}{2} + \frac{\varepsilon^2}{8\pi^2\tau}, \qquad m_{\text{eff}}(\tau) = \frac{\frac{1}{2}e^{-\tau} + \frac{\varepsilon^2}{8\pi^2\tau^2}}{\frac{1}{2}e^{-\tau} + \frac{\varepsilon^2}{8\pi^2\tau}} \quad (5)$$

(the coefficient of the second term checked to 12 digits by numerical integration Computed). Up to the τ^* at which the two terms are equal, $m_{\text{eff}} \approx 1$, and it looks like "a theory with gap 1".

Figure 2 Making ε smaller, the point τ* where the correlation (thick line, vertical axis in common logarithms) leaves $e^{-\tau}/2$ (dashed line) runs away to the right. Moving the edge of the window changes m_{eff} there



ε = 1e−3: τ* = 20.51. At the window edge τ = 15.0, m_eff = 0.9949

Computed The numbers in this figure were computed on this machine ((5) computed on the spot; τ* by bisection).

Computed Table 2 τ^* and m_{eff} of (5).

ε	τ*	m_eff(20)
10^{-1}	10.646	0.0502
10^{-2}	15.636	0.0652
10^{-3}	20.512	0.6384
10^{-4}	25.328	0.9942
10^{-6}	34.858	1.0000

Setting the two terms equal gives $\tau^* = \ln(1/\varepsilon^2) + \ln(4\pi^2\tau^*)$, so each time ε becomes an order of magnitude smaller, τ^* moves out by about 5, $2\ln 10 \approx 4.6$ plus a logarithmic correction, and for any finite window there is an ε that looks like "gap 1" (far enough out, $m_{\text{eff}} \approx 1/\tau$). Measurements in a finite window give no lower bound on the gap Known . For how to read lattice numbers, 16-08 § 07.

05

How it differs from particle states — an isolated eigenvalue is a separate requirement from the gap

The gap of the mass operator $M = \sqrt{H^2 - \vec{P}^{\,2}}$ had the same width as the gap of H (0-03 § 06). **A one-particle state of mass** μ is a state belonging to the eigenvalue μ of M (Wigner's 1939 idea of seeing particles as irreducible representations of the Poincaré group), and it is **isolated** if there are no other points of $\sigma(M)$ near μ. § 5 of the problem statement places this outside the requirements, writing 'one would like to prove the existence of an isolated one-particle state (an upper gap, in addition to the mass gap)' Known .

The difference shows in the two-point function. The Euclidean two-point function of a scalar field ϕ ($\langle\Omega, \phi\,\Omega\rangle = 0$) can be written with a positive measure ρ as

$$\widehat{S}_2(p) = \int_{[0,\infty)} \frac{d\rho(s)}{p^2 + s}, \qquad \int d^3x\, S_2(\tau, \vec{x}) = \int \frac{e^{-\sqrt{s}\,\tau}}{2\sqrt{s}}\, d\rho(s) \quad (6)$$

which is called the Källén–Lehmann representation Known (the right-hand side is (3) of 4-10 superposed over s). If there is a gap m, the support of ρ lies in $[m^2, \infty)$, and if ϕ touches a particle of mass μ, ρ has a point weight at $s = \mu^2$.

Example 1: a gap but no particle. Take ρ to be the uniform weight on $[m^2, 4m^2]$, and take the Gaussian measure with covariance $\int d\rho(s)(-\Delta + s)^{-1}$. Reflection positivity is a condition preserved by positive superpositions of covariances, so if it holds for each s (4-10 § 07) it also holds after integrating with the positive measure ρ. On reconstruction, the "one-particle" M fills $[m, 2m]$ continuously, and $\sigma(H) = \sigma(M) = \{0\} \cup [m, \infty)$. The gap is m, and M has no eigenvalue other than 0. Fields of this kind are called generalised free fields (Greenberg 1961) Known . **Example 2: a particle but no gap.** In the theory of section 04, the one-particle state of ϕ has $M = 1$, but two-particle states of χ fill $[0, \infty)$, and the eigenvalue 1 is buried in the continuous spectrum.

What physics expects for Yang–Mills is an isolated glueball at the lower end of the gap Physics . For the difficulty of telling an isolated point from the edge of a continuum using correlations, 16-08 § 05.

06

Clustering — how fast do things become independent when moved apart in a spatial direction

Clustering without regard to rate (two quantities moved apart approach independence) is Osterwalder–Schrader axiom E4, paired with the uniqueness of the vacuum (Table 1 of 4-10) Known . The massless free field also satisfies it, but its correlation decreases only as a power, $1/(4\pi^2|x|^2)$. Formula (2) of § 5 of the problem statement says, for a local field $\mathcal{O}(\underset{\sim}{x}) = e^{-i\underset{\sim}{P}\cdot\underset{\sim}{x}}\mathcal{O}e^{i\underset{\sim}{P}\cdot\underset{\sim}{x}}$ with $\langle\Omega, \mathcal{O}\Omega\rangle = 0$,

$$|\langle\Omega, \mathcal{O}(\underset{\sim}{x})\,\mathcal{O}(\underset{\sim}{y})\,\Omega\rangle| \le e^{-C|\underset{\sim}{x}-\underset{\sim}{y}|} \qquad (0 < C < \Delta,\ \ |\underset{\sim}{x}-\underset{\sim}{y}| \text{ large enough}) \quad (7)$$

The outline on the Euclidean side is (2) and rotation invariance (Euclidean covariance; Osterwalder–Schrader axiom E1). If two quantities localised within radius δ are a distance R apart, rotating the direction joining them onto the τ axis, (2) gives a connected correlation at most (a constant independent of R) $\times e^{-m(R-2\delta)}$. At equal times and spacelike-separated points, the Wightman function (the vacuum expectation value of a product of fields; 4-09 § 04) equals the value of the Schwinger function at time 0 (the analytic continuation of 4-10 § 05), so this decay of the Euclidean correlation becomes (7) directly. The point field $\mathcal{O}(\underset{\sim}{x})$ is not a bounded quantity, so precisely one considers the field slightly smeared over radius δ. The "$C < \Delta$" and "large enough" in (7) account for the constant in front and the width 2δ. The theorem on the Minkowski side is introduced in the introduction of Nachtergaele–Sims 2006: 'in [1] Fredenhagen proves a general theorem applicable to arbitrary strictly local theories' ([1] is Fredenhagen 1985) Known .

The converse does not hold in general; the same introduction writes that there are quantum lattice systems with a unique ground state and exponentially decaying correlations but no gap, and gives Example 2 of Nachtergaele 1996 as an example Known . In a rotation-invariant Euclidean theory, time is one of the directions. If "it decays at a uniform rate in every direction", it decays at the same rate in the time direction too, so the converse argument of section 02 (where the constant in front may depend on F) applies directly, and this is equivalent to the gap. A state mixing two phases (2-09 § 08) breaks clustering itself, and assumption (b) of (1) fails.

07

The correlation length — the inverse of the gap, said as a length

(6) of 2-08 defined a correlation length from a single correlation. The correlation length of the theory is the supremum over observables.

$$\frac{1}{\xi_F} = -\lim_{\tau\to\infty}\frac{1}{\tau}\log C_F(\tau) = E_F, \qquad \xi := \sup_F \xi_F = \frac{1}{m} \tag{8}$$

Both are restatements of (3) Known (restoring units, $\xi = \hbar c/\Delta$; on the lattice $\xi_{\text{phys}} = a\,\xi_{\text{lat}} = 1/\Delta$). The lattice correlation-length problem of Chatterjee 2018 (Chatterjee's Problem 5.1) defines the lattice correlation length by plaquette correlations, and writes 'The number ξ represents the reciprocal of the mass of the lightest glueball' Known .

$\xi_{2\text{nd}}^2 = \sum_x |x|^2 G(x) / \big(2d\sum_x G(x)\big)$ from second moments is also often used, and on the lattice the rate of decay also depends on the direction (4-02). What equals the inverse of the gap is the length along the direction of the transfer matrix (an axis). For the free field on the four-dimensional lattice (covariance $(-\Delta_{\text{lat}} + m_0^2)^{-1}$) there are closed forms.

$$\cosh\frac{1}{\xi_{\text{axis}}} = 1 + \frac{m_0^2}{2}, \qquad \xi_{2\text{nd}} = \frac{1}{m_0}, \qquad \text{decay rate along } e = \max\Big\{q\cdot e : \sum_\mu (2\cosh q_\mu -$$

The e in the third is a vector of length 1: on an axis $e = (1,0,0,0)$, on the diagonal $e = (1,1,1,1)/2$.

Computed Table 3 The free field on the four-dimensional lattice. The two columns on the right are ratios to ξ_{axis} (diagonal means the direction $e = (1,1,1,1)/2$). The third formula of (9) was checked by computing the two-point function on a two-dimensional periodic lattice with the fast Fourier transform, agreeing to within 0.11% on the axis and the diagonal.

m_0	ξ_{axis}	$\xi_{\text{diag}}/\xi_{\text{axis}}$	$\xi_{2\text{nd}}/\xi_{\text{axis}}$
2	0.5673	0.91579	0.88137
1	1.0390	0.97228	0.96242
0.2	5.0083	0.99876	0.99834
0.05	20.0021	0.99992	0.99990

When the correlation length is of the order of the lattice spacing ($m_0 = 1, 2$), the three lengths disagree by 3 to 12%; at $\xi_{\text{axis}} = 20$ the differences are of order 10^{-4}, and rotational symmetry returns (the continuum limit; 2-08 § 08). Outside free fields there is no guarantee that the lengths agree, and what is used for the gap is ξ_{axis}.

08

Finite boxes and uniformity — the gap is asked about on the side of the limit

On a finite lattice, a positive gap is almost trivial. In the temporal gauge, the transfer matrix T (5-03 § 05) is an integral operator with a positive continuous kernel acting on the space $L^2(G^n)$ of values of the links at one time (G a compact group, n the number of links). A continuous kernel on a compact space makes T a compact operator (precisely, a norm limit of finite-dimensional matrices), and its eigenvalues form a discrete sequence accumulating only at 0. Further, since the kernel is positive, Jentzsch's theorem (the infinite-dimensional version of the Perron–Frobenius theorem) makes the largest eigenvalue λ_0 simple, and $\lambda_1 < \lambda_0$ Known . If the group is finite, T is a finite matrix with positive entries, and it is literally the Perron–Frobenius theorem. Simplicity of the largest eigenvalue alone leaves open that the other eigenvalues crowd up to λ_0, so discreteness is needed. What is asked is whether $m_{\text{lat}}(a, L)/a$ stays positive in the limits of volume $L \to \infty$ and lattice spacing $a \to 0$. § 6.5 of the problem statement writes 'New ideas are needed to prove the existence of a mass gap that is uniform in the volume of space-time.', and § 5 also says that a uniform gap can play a role in the proof of the infinite-volume limit Known . (2) is an inequality between correlations (the $\|\psi_F\|^2 = \mathbb{E}[\Theta F F] - |\mathbb{E}[F]|^2$ on the right-hand side is also a correlation). So if it holds along a sequence of approximations with the same Δ and the correlations converge, it holds in the limit too. Using the converse of section 02 in the limiting theory gives $m \geq \Delta$ (that the set of all ψ_F is dense in $\Omega^\perp$ comes from the way the reconstruction is built; 7-01).

The ways a finite-box gap closes as $L \to \infty$ are massless particles ($2\pi/L$ in a periodic box; 2-05 § 06) and symmetry breaking (the tunnelling splitting between phases; 2-09 § 08). In the latter each phase in infinite volume has a gap, and § 6.3 of the problem statement writes that each phase of two-dimensional ϕ^4 has 'a non-zero mass gap and a unique, Poincaré-invariant vacuum state' Known . As $a \to 0$, if Δ is held fixed then $m_{\text{lat}} = a\Delta \to 0$, and a gap in lattice units that is uniformly positive means $\Delta = \infty$ in the continuum limit (5-09).

09

Reading it in gauge theory — the gap on gauge-invariant states

For Yang–Mills, the $\mathcal{H}$ of (1) is the space of gauge-invariant states (4-01), and the observables of (2) and (3) are local quantities built from the field strength such as $\operatorname{Tr} F_{ij} F_{kl}(x)$ (problem statement § 4) and Wilson loops (4-10 § 09). States split by the combination J^{PC} of spin, parity and charge conjugation (channels) (J is the spin, and P and C are the signs $\pm$ under spatial reflection and under the operation that replaces the field by its complex conjugate [the antiparticle]), and, as with k in section 03, the minimum over channels of the lowest energy in each is m. In lattice numbers 0^{++} is

the lowest (about 0.7 times 2^{++} for pure $SU(3)$; Chen et al. 2006; 5-08), and physics reads it as the lower end of the gap Physics .

Correlations of gauge-fixed fields are in general not guaranteed to be reflection positive (4-10 § 04, § 09), and for the gluon propagator it is read as violated Physics . The decay of a correlation that is not reflection positive does not give points of $\sigma(H)$. The string tension is fixed by the energy of a pair of static charges, and is a different quantity from excitations above the vacuum (16-08 § 06). In mathematics there are long-standing results on the gap in lattice units at strong lattice coupling (small β) (5-04). Shen–Zhu–Zhu 2023, in 't Hooft's convention (writing the coupling of the action in the form $N\beta$; for conversion to other conventions, Table 4 of 5-01), showed that for $SU(N)$ with $|\beta| < 1/(16(d-1))$ the correlations of a certain family of observables decay exponentially. The abstract calls this 'strictly positive mass gap', and also writes 'Our method improves earlier results' Known . To use it in the equivalence of section 02, that family must be dense. This is a result short of the limits of section 08. For the four-dimensional continuum theory, the $\mathcal{H}$ and H of (1) were built by the construction published in 2026-11 (at the level of the committee's assessment; 12-08), but their gap has not been shown (6-04; for where things stood before that, 7-10).

10

Appendix table: the uses of the word "gap" — the same word, equivalent, upper bound, or different

More than ten quantities called "gap" appear in this book. Some are **equivalent** to the Δ of the problem statement, some are only **upper bounds** on Δ, and some are **different** (the gap of another operator, or an additional requirement); confusing them changes the direction of a claim. The table below lists, one row each, the definition, the main chapters and the relation to Δ (the assignment of relations is this book's organisation; for the direction of upper and lower bounds, the box "The asymmetry of upper and lower bounds" in 1-10).

Appendix table The word "gap". "Equivalent" in the relation column means the same number under the stated assumptions.

#	Name	Definition (one line)	Chapter	Relation to Δ
1	Gap of the Hamiltonian (the mass gap)	The infimum Δ of the spectrum of H on $\Omega^{\perp}$	§ 01, 0-03 § 06	Reference (the Δ of the problem statement)
2	Gap of the mass operator	The next point after 0 for $M = \sqrt{H^2 - P^2}$	§ 05, 0-03 § 06	Equivalent (same width)

#	Name	Definition (one line)	Chapter	Relation to Δ
3	Uniform decay rate	The largest m such that all connected correlations decay at least as fast as $e^{-m\tau}$	§ 02	Equivalent ((2), (3); under reflection positivity)
4	Inverse correlation length	$1/\xi$, $\xi = \sup_F \xi_F$ (along the direction of the transfer matrix)	§ 07	Equivalent ((8)). $\xi_{2\mathrm{nd}}$ and the length along the diagonal are different (close for free fields)
5	Effective mass	$m_{\mathrm{eff}}(\tau) = -\frac{d}{d\tau}\log C_F(\tau)$	§ 02, § 04, 18-01 § 08	Upper bound ($m_{\mathrm{eff}}(\tau) \geq E_F \geq \Delta$)
6	Gap of the transfer matrix (lattice units)	$m_{\mathrm{lat}} = \log(\lambda_0/\lambda_1)$	§ 02, 5-03	Equivalent on the lattice ($T = e^{-aH}$)
7	Gap in physical units	$\Delta = m_{\mathrm{lat}}/a$	§ 08, 5-09	Related to 6 by a. This is what is asked about in the continuum limit
8	Finite-volume gap	$\lambda_1 < \lambda_0$ of T in a box of side L	§ 08	Always positive (Jentzsch's theorem). Says nothing about Δ
9	Uniform gap	$\inf m_{\mathrm{lat}} > 0$ over the volume L (and β in a fixed range)	§ 08, 5-04, 16-05	A lower bound on 6 in infinite volume. A gap in lattice units uniform in β is, in physical units, on the side of $\Delta = \infty$ (a false friend; 18-01)
10	Upper gap	The width by which the one-particle eigenvalue μ is separated from the rest of $\sigma(M)$	§ 05	Different (an additional requirement in § 5 of the problem statement)
11	Langevin gap	The gap $1/C_P$ of the generator L of the path measure (the lattice Wilson measure)	1-10 § 09, 18-03 § 05	Different (the gap of an artificial time). For free fields $1/C_P = m^2$ (the square of the gap). Assuming they line up is 23-01 The line of construction and the diffusive-relaxation-time assumption
12	Poincaré constant of the time-zero measure	C_P of the ground-state distribution $\phi_0^2 dx$	2-05 § 08	Equivalent (ground-state transformation; with the normalisation $C_P = \hbar^2/(2m\Delta)$)

11 and 12 are both Poincaré constants, but 11 concerns the measure on whole paths in imaginary time and 12 the measure at a single time, and only 12 is equivalent (2-05 § 08). 5 and 9 point in opposite directions: 5 is an upper bound obtained from a single observable, 9 a lower bound over all observables and volumes.

What this chapter can say

Content	Status
(1) to (3) and (8), the examples of section 05, the finite-lattice gap of section 08	Known . Outlines written in the text
Tables 1 to 3, Figures 1 and 2, the checks of (5) and (9), the comparison with the asymptotic forms of section 03	For free fields, in double precision Computed
The Euclidean side of (7)	Outline only. The Minkowski side is Fredenhagen 1985 (bibliographic data only)
"Equivalent, upper bound, different" in the appendix table of section 10	This book's organisation. The grounds for each row are the text and the chapters cited (restatements of Known results)
The glueballs of section 09	Physics (lattice numbers and extrapolation)
The gap of four-dimensional Yang–Mills	$\mathcal{H}$ and H were built by the construction published in 2026-11 (at the level of the committee's assessment; 12-08). The gap has not been shown (6-04). This chapter shows nothing

Sources and reproduction

Item	Kind	Source / tools
Tables, figures, numbers in the text	Computed	python3, NumPy, SciPy (numerical integration, bisection, FFT)
Problem statement §1, §4, §5, §6.3, §6.5, the lattice correlation-length problem	Known	A. Jaffe, E. Witten, *Quantum Yang–Mills Theory* (Clay, 2000); S. Chatterjee, arXiv:1803.01950 §5. Checked in the text
Section 06	Known	B. Nachtergaele, R. Sims, Comm. Math. Phys. 265 (2006) 119–130 (arXiv:math-ph/0506030), introduction checked; K. Fredenhagen, 'A Remark on the Cluster Theorem', Comm. Math. Phys. 97 (1985) 461–463, bibliographic data only; B. Nachtergaele, Comm. Math. Phys. 175 (1996) 565–606, Example 2, bibliographic data only
Section 05	Known	H. Lehmann, Nuovo Cimento 11 (1954) 342–357; O. W. Greenberg, Ann. Phys. 16 (1961) 158–176; E. Wigner, Ann. of Math. 40 (1939) 149; G. Källén 1952 'On the Definition of the Renormalization Constants in Quantum Electrodynamics' (journal and volume of first publication not checked). Bibliographic data only
Section 08	Known	K. Osterwalder, E. Seiler, Ann. Phys. 110 (1978) 440–471, bibliographic data only (for the construction of the transfer matrix, 5-03)

Item	Kind	Source / tools
Section 09	Physics, Known	Y. Chen et al., Phys. Rev. D 73 (2006) 014516, Table VIII (text); H. Shen, R. Zhu, X. Zhu, Comm. Math. Phys. 400 (2023) 805–851 (abstract of arXiv:2204.12737)

Next chapter to read: 4-12 Detour: the perturbation series diverges. For what to hold fixed in the limit, 5-09.

Detour: the perturbation series diverges — the factorial in the coefficients is often the shadow of an e^{-c/g^2} that the series does not show

This chapter organises established knowledge (sources carry the “Known” label in the text). This book's own investigation is mainly in Part 15 and Volume 3 (Parts 16–28 and the appendices).

What this chapter shows — Where coefficients growing like $n!$ come from (negative coupling, saddle points, the ends of momentum integrals). Optimal truncation and the Borel sum, and how their ambiguity pairs with terms of size e^{-c/g^2}. The singularities expected in the Borel plane of Yang–Mills.

Prerequisite chapters — 4-04 Divergences and renormalisation (the running coupling). We use the asymptotic series and the zero-dimensional model (2) of 4-03 § 08. From complex analysis, only Cauchy's integral formula is used, stated on the spot.

Said first — The formulas for the models can be derived, and the numbers were checked on this machine. The divergence of the series of four-dimensional Yang–Mills, renormalons, and the relation between the series and the theory are expectations at the level of physics, not theorems. This chapter does not show anything about Yang–Mills.

Order of this chapter

1. Asymptotic series — approaching without converging: the coefficients are fixed by the function, not the other way round
2. Dyson's argument — flipping the sign of the coupling: negative coupling gives an imaginary part $e^{-3/(2s)}$
3. The coefficients are fixed on the flipped side — dispersion relations and saddle points: from the imaginary part, $n!\,(2/3)^n$
4. Optimal truncation — stopping where the terms are smallest: the floor of the error is $e^{-3/(2u)}$
5. The Borel sum — divide by $n!$, restore with an integral: in the model it returns the exact value

6. When there is a singularity on the positive axis — ambiguity and exponentially small terms: the sum splits
7. Renormalons — $n!$ from the ends of momentum integrals: it stops at the order at which the peak reaches Λ
8. The Borel plane of Yang–Mills — a map of the expected singularities: renormalons are the closest
9. Seen from mathematics — what is a theorem and what is an expectation: in four dimensions asymptoticity has not been shown

---

01

Asymptotic series — approaching without converging

The problem of this chapter is that the series of coefficients rendered finite by renormalisation **does not itself converge** (a different matter from the divergences of 4-04). Even so it has uses. If, fixing the order N and letting $u \to 0$, the truncation error vanishes as fast as u^{N+1}, we say that the series is **asymptotic** to f, and write $f(u) \sim \sum a_n u^n$.

$$\left| f(u) - \sum_{n=0}^{N} a_n u^n \right| \le K_{N+1}\, u^{N+1} \qquad (0 < u < u_0,\ N = 0, 1, 2, \dots) \tag{1}$$

The order of the limits is the reverse of convergence. The coefficients are determined uniquely by f Known , but the converse does not hold. $e^{-1/u}$ approaches 0 faster than any u^N, so f and $f + e^{-1/u}$ have the same series (7-12 § 07). The theme of this chapter is that this "part not shown in the coefficients" is recorded in the way the coefficients grow.

Example. That all the coefficients of $e^{-1/u}$ are 0 can be checked in the form of (1) divided by u^N. Putting $v = 1/u$, $e^{-1/u}/u^N = v^N e^{-v}$, and as $u \to 0$ ($v \to \infty$) the exponential beats the power and it approaches 0. This is so for every N, so $a_0 = a_1 = \cdots = 0$. An example in the other direction is the model (2) of this chapter. (14) of 4-03 § 08 shows $\left| Z(u) - \sum_{n<N} a_n u^n \right| \le |a_N|\, u^N$ for $u \ge 0$, and $|a_N|$ can be taken as the K_N of (1). K_N grows factorially with N, but (1) needs only an estimate for each fixed N, so the series is asymptotic to Z.

02

Dyson's argument — flipping the sign of the coupling

The abstract of Dyson (1952) says it presents an argument leading 'tentatively' to the conclusion that the power series of quantum electrodynamics all diverge even after renormalisation Known . Its core is usually presented in the form: "for $e^2 < 0$, like charges attract and the vacuum collapses, so the series cannot continue smoothly to both sides of $e^2 = 0$". It is a heuristic argument. For a small model everything can be computed.

$$Z(u) = \int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi}}\, e^{-S_u(x)}, \qquad S_u(x) = \frac{1}{2}x^2 + \frac{u}{4!}x^4 \tag{2}$$

For $u = -s < 0$ the integral does not exist (the "collapse of the vacuum" in the model). Even so, turning the variable as $x \to e^{-i\theta/4}x$ brings the quartic term of $u = s\,e^{i\theta}$ back to $s\,x^4$, so Z extends to complex u. There are two reasons why the contour may be turned. One is that the integral of a function complex-differentiable on and inside a closed path, taken along that path, is 0 (Cauchy's theorem Known ; we use it in the form of the integral formula in § 03). The integrand is the exponential of a polynomial in x, so it is complex-differentiable everywhere. The other is that on the far arc of the sector between the original real axis and the turned line, the real part of the quartic term stays positive and acts at the rate $|x|^4$, so the contribution of the arc vanishes as $|x| \to \infty$. Approaching the negative axis from above,

$$Z(-s + i0) = e^{-i\pi/4} \int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi}}\, e^{\,\frac{i}{2}x^2 - \frac{s}{24}x^4} \tag{3}$$

takes a non-real value, and from below its complex conjugate. **The values above and below the negative axis disagree: this is the source of the divergence.** The imaginary part is fixed by the height of the wall. $S_{-s}(x) = \frac{1}{2}x^2 - \frac{s}{24}x^4$ takes its maximum $\frac{3}{2s}$ at $x^2 = 6/s$, and the rough probability of leaking from the valley over the wall is $e^{-3/(2s)}$. Strictly, there is no tunnelling in a zero-dimensional integral. What is meant is that the value $\frac{3}{2s}$ of the action at the saddle point $x^2 = 6/s$ of the integrand of (3) fixes the exponent of the imaginary part, and the leading term as $s \to 0$ is $-e^{-3/(2s)}/\sqrt{2}$ (Table 1).

Computed Table 1 Numerical integration of (3). The imaginary part approaches $-e^{-3/(2s)}/\sqrt{2}$ as $s \to 0$

s	$\operatorname{Im} Z$	$e^{-3/(2s)}/\sqrt{2}$	Ratio
0.50	-3.351e-2	3.521e-2	0.952
0.20	-3.825e-4	3.911e-4	0.978
0.10	-2.138e-7	2.163e-7	0.988

03

The coefficients are fixed on the flipped side — dispersion relations and saddle points

The tool we use is **Cauchy's integral formula**: a function complex-differentiable on and inside a closed path satisfies $f(u) = \frac{1}{2\pi i}\oint \frac{f(w)}{w-u}\,dw$ (anticlockwise) at points inside Known . Z is complex-differentiable on the plane minus the negative axis and decreases as $|w|^{-1/4}$ at infinity, so taking the path to be a large circle and the upper and lower sides of the negative axis, only the difference $2i\,\mathrm{Im}\,Z$ between the sides remains.

$$Z(u) = -\frac{1}{\pi}\int_0^\infty \frac{\mathrm{Im}\,Z(-s+i0)}{s+u}\,ds \quad (4)$$

This is the **dispersion relation** (numerically integrating the right-hand side with the imaginary part computed from (3), it agrees with the left-hand side to order 10^{-6} at $u = 0.5, 1, 2$ Computed). Expanding $\frac{1}{s+u} = \sum_n (-u)^n/s^{n+1}$ and integrating term by term,

$$a_n = \frac{(-1)^{n+1}}{\pi}\int_0^\infty \frac{\mathrm{Im}\,Z(-s+i0)}{s^{\,n+1}}\,ds \quad (5)$$

(since it is not complex-differentiable at $u = 0$, what is obtained is an asymptotic series), and for large n only the imaginary part $\approx -e^{-3/(2s)}/\sqrt{2}$ near $s \to 0$ matters. From $\int_0^\infty e^{-c/s} s^{-n-1}\,ds = \Gamma(n)/c^n$,

$$a_n \approx (-1)^n \left(\frac{2}{3}\right)^n \frac{\Gamma(n)}{\pi\sqrt{2}} \quad (6)$$

The factorial comes from $\Gamma(n) = (n-1)!$, and the base $2/3$ is the inverse of the coefficient of the wall height $3/(2s)$; the alternation of signs comes from the singularity lying on the negative axis.

Computed Table 2 The exact coefficients $a_n = \frac{(4n-1)!!}{n!}\left(-\frac{1}{24}\right)^n$ (rationals) and their ratio to (6)

n	a_n	a_n/(6)
1	-0.1250	0.833
10	1.3901e+3	0.981
20	8.1571e+12	0.991
40	4.1327e+38	0.995

Bringing this line of reading large orders from the action at a saddle point into field theory is attributed to Lipatov (1977), and the review of Beneke (1999) credits Lipatov 1977 with factorial divergence due to instantons (classical solutions in imaginary time; defined in § 06) Known .

The large orders of the ground-state energy $\frac{1}{2}+\frac{3}{4}g-\frac{21}{8}g^2+\cdots$ of $H=\frac{1}{2}p^2+\frac{1}{2}x^2+gx^4$ in quantum mechanics were studied by Bender–Wu (1969, 1973) (abstracts) Known . The ratio of the known form $E_n \approx (-1)^{n+1}\frac{\sqrt{6}}{\pi^{3/2}}3^n\Gamma(n+\frac{1}{2})$ (the text of the source not checked) to the coefficients computed with rationals up to order 150 is 0.832 at $n=10$ and 0.991 at 150 Computed . The base 3 is the inverse of the coefficient of the action $\frac{1}{3|g|}$ for passing through the wall of $V=\frac{1}{2}x^2-|g|x^4$ in imaginary time.

04

Optimal truncation — stopping where the terms are smallest

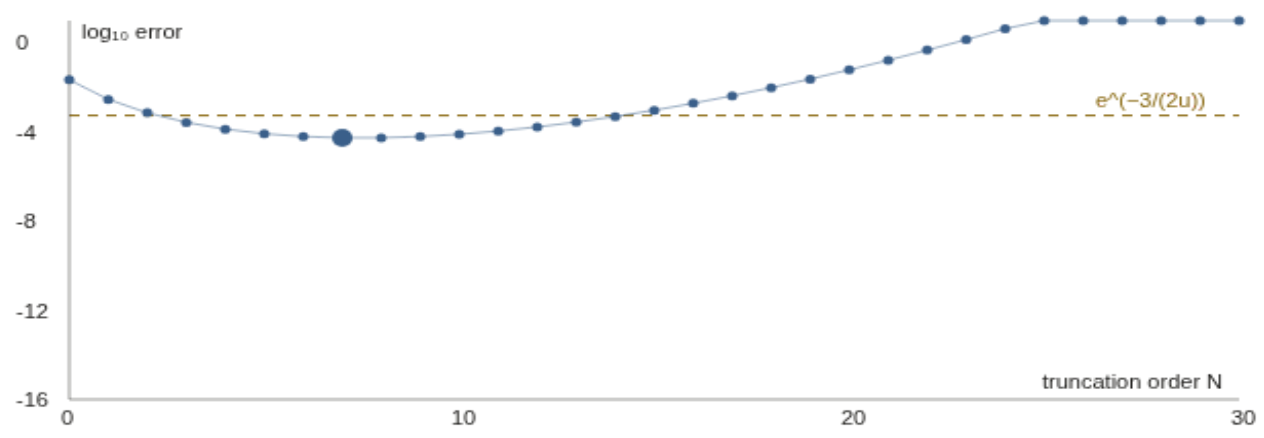
The naive method is to **stop where the terms are smallest**. The terms $\Gamma(n)(2u/3)^n$ of (6) decrease up to $n \approx 3/(2u)$ and grow beyond it. Putting Stirling's formula $\Gamma(n) \approx \sqrt{2\pi/n}\,(n/e)^n$ in there gives

$$\min_n |a_n u^n| \;\approx\; \frac{1}{\pi\sqrt{2}}\sqrt{\frac{2\pi}{N^*}}\,e^{-N^*} \;\propto\; \sqrt{u}\,e^{-3/(2u)}, \qquad N^* \approx \frac{3}{2u} \quad (7)$$

The floor of the error is the size of **an exponentially small quantity fixed by the height of the wall**, and the series can say nothing finer than that. The "part not shown in the coefficients" of section 01 appears as a limit on precision (the ratio on the right moves because of $\sqrt{u}$).

Why the error is about the smallest term. In this model, by (14) of 4-03 § 08, the error of stopping after N terms does not exceed the first discarded term $|a_N|u^N$ Known . Stopping where the terms are smallest, the error is at most that smallest term. Since the signs of the coefficients alternate, the actual error settles at about half the smallest term (0.47 to 0.50 times in the 6 rows of Table 3 Computed). Thus "the best error ≈ the smallest term ≈ (7)". **How to read Table 3.** Each row fixes one u, compares the partial sums from order 0 to order N ($N<60$) with direct numerical integration $Z(u)$, and shows the N with the smallest error and that error. The best N comes around $3/(2u)$. The rightmost column is the error divided by $e^{-3/(2u)}$, and stays within 0.07 to 0.26 as u varies from 0.1 to 2. That is, the error becomes small at the same rate as $e^{-3/(2u)}$.

Figure 1 Changing u moves the bottom of the valley of the truncation error $|S_N(u)-Z(u)|$ (logarithmic scale) parallel to the dashed line $e^{-3/(2u)}$ (about one order of magnitude below it)



u = 0.20, Z(u) = 0.97791537, best N = 7, smallest error 5.52e−5, e^(−3/(2u)) = 5.53e−4

Computed The numbers in this figure were computed on this machine (Simpson's rule and the recurrence for the coefficients; agreeing with Table 3).

Computed Table 3 Optimal truncation ($Z(u)$ is direct numerical integration. The Borel sum (8) of section 05 agrees in all 6 rows to within 5×10^{-16}, within double-precision rounding)

u	Best N	Smallest error	Error $/e^{-3/(2u)}$
0.1	15	2.18e−8	0.071
0.2	7	5.52e−5	0.100
0.3	5	7.96e−4	0.118
0.5	3	7.31e−3	0.147
1.0	1	4.39e−2	0.197
2.0	1	1.22e−1	0.259

05

The Borel sum — divide by $n!$, restore with an integral

The **Borel sum** first divides by $n!$ to make the series converge, then multiplies back with $\int_0^\infty e^{-t} t^n \, dt = n!$.

$$B(t) = \sum_{n \geq 0} \frac{a_n}{n!} t^n, \qquad \tilde{Z}(u) = \frac{1}{u} \int_0^\infty e^{-t/u} \, B(t) \, dt \tag{8}$$

B is called the **Borel transform**, and the complex plane of t the **Borel plane**. Term by term $\frac{1}{u} \int_0^\infty e^{-t/u} \frac{t^n}{n!} dt = u^n$, so the asymptotic series of $\tilde{Z}$ is the original series. For the zero-dimensional model, from $(4n-1)!! = 16^n (\frac{1}{4})_n (\frac{3}{4})_n$ ($(a)_n = a(a+1)\cdots(a+n-1)$),

$$B(t) = \sum_{n \geq 0} \frac{(\frac{1}{4})_n (\frac{3}{4})_n}{(n!)^2} \left(-\frac{2t}{3}\right)^n = {}_2F_1\left(\tfrac{1}{4}, \tfrac{3}{4}; 1; -\tfrac{2t}{3}\right) \tag{9}$$

The hypergeometric function ${}_2F_1(a,b;c;z)$ extends to the plane minus a cut running from $z=1$ Known . **The only singularity in the Borel plane is the single point $t=-3/2$, at the same position as the coefficient of the wall height**. The path of integration $t\geq 0$ does not touch the singularity, and (8) returns to $Z(u)$ as Table 3 shows Computed .

The theorem guaranteeing that the Borel sum agrees with the original function requires analyticity of the original function in a disc tangent at the origin, and a uniform estimate bounding the K_N of (1) by $A\sigma^N N!$ (A,σ constants independent of N). (1) alone is not enough. Beneke cites this as the Watson–Nevanlinna–Sokal theorem (Sokal 1980), and writes 'These requirements are too strong for renormalizable theories' Known .

06

When there is a singularity on the positive axis — ambiguity and exponentially small terms

For a series whose terms have the same sign and grow as $a_n\approx K\,a^n n!$ ($a>0$), the singularity comes at $t=1/a$ on the positive axis, on the path of (8). Going round above or below gives two sums **whose imaginary parts disagree by an amount of size** $e^{-1/(au)}$. We look at the simplest case, $a_n=n!\,a^n$. The Borel transform is $B(t)=\sum_n (at)^n=\frac{1}{1-at}$, with a pole at $t=1/a$. Subtracting the integral going round below from the one going round above gives an integral once clockwise round the pole, and from the residue

$$\tilde{Z}_{\text{above}}(u)-\tilde{Z}_{\text{below}}(u)=\frac{2\pi i}{au}\,e^{-1/(au)}$$

(the imaginary part of each is $\pm\pi e^{-1/(au)}/(au)$). This difference is of a size that shows up in no power of u. Tables and figures are in 7-12 § 08. Beneke shows, in examples where the exact function is known, that the two ambiguities, of the Borel integral and of the exponentially small terms, cancel each other, and sets up the following chain Known .

factorial divergence with equal signs $\Longrightarrow$ ambiguity of the Borel integral $\Longrightarrow$ add exponen

He then warns 'it is not rigorous'. It is not a theorem Physics .

Example. The gap of the double well of 2-05 § 07 has size e^{-S_0} and appears at no order of perturbation theory. A classical solution that moves in imaginary time from one valley to the other is called an **instanton**, and its action is the S_0 of (9) in 2-05. A solution moving the other way is an **anti-instanton** (for the counterpart in Yang–Mills,

see 3-05). The ambiguity of the Borel sum of the series around a valley has size e^{-2S_0} (the weight of an instanton–anti-instanton pair), and is said to cancel against the ambiguity of the pair's contribution (Bogomolny 1980, Zinn-Justin 1981) Physics . The way the series diverges announces the existence of terms the series does not show.

07

Renormalons — $n!$ from the ends of momentum integrals

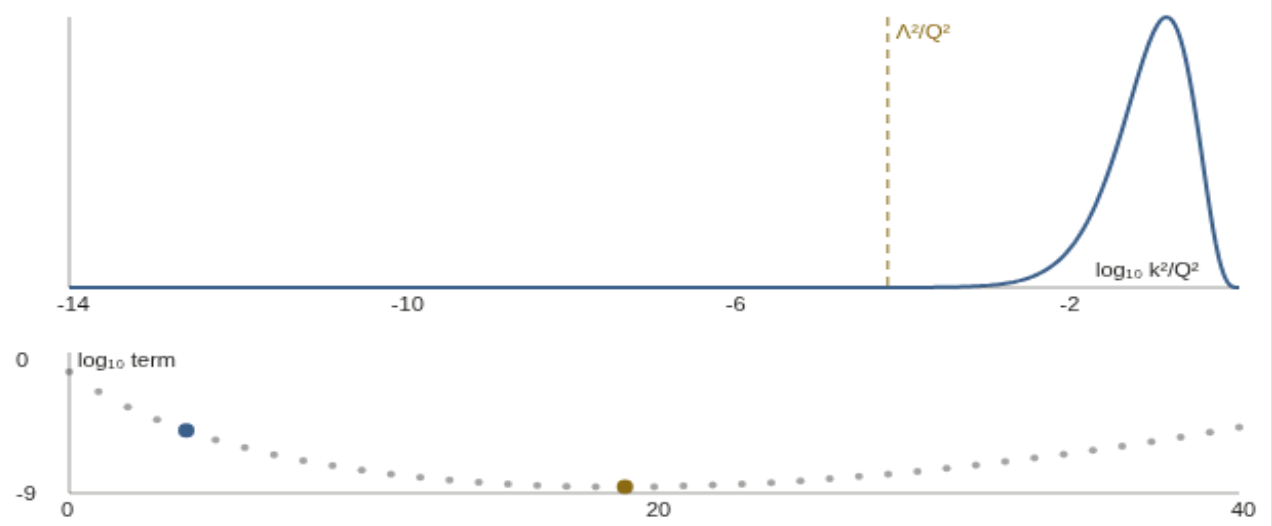
Renormalisable theories have a source of $n!$ other than saddle points: **renormalons** (according to Beneke, found by Gross–Neveu 1974, Lautrup 1977 and 't Hooft 1977, with the name coming from 't Hooft's Erice lectures Known). The source is the **bubble chain** diagram: a propagator with n one-loop bubbles (loops in which a pair of particles is created and annihilated) strung on it; summed over n, it becomes the one-loop running coupling $\alpha(k)$ inserted into a single line (Beneke § 2.2) Known . It is a different thing from the vacuum bubbles of 4-03 § 03.

We use a model that shrinks this to one variable. Suppose there is a quantity measured by injecting a momentum Q from outside, and the integral over the loop momentum k inside it carries the running coupling $\alpha(k) = \alpha/(1 - b\alpha \ln(Q^2/k^2))$ ($\alpha = \alpha(Q)$, $b = \beta_0/4\pi$, $\beta_0 = 11N/3$ for pure $SU(N)$; for conversion between the four ways of writing the coefficient, Table 1 of 4-06; 4-04 § 05, 4-06). Expanding $\alpha(k)$ in α, the order $n+1$ contains the n-th power of $\ln(Q^2/k^2)$. About the rest of the integrand we assume only that, measured with $d\ln k^2$, it vanishes in proportion to k^4 at small k (x^2) and decreases as $1/k^2$ at large k (x^{-1}). With $x = k^2/Q^2$,

$$r_n^{\mathrm{IR}} = \int_0^1 x\,\big(b\ln\tfrac{1}{x}\big)^n\,dx = \frac{b^n\,n!}{2^{n+1}}, \qquad r_n^{\mathrm{UV}} = \int_1^\infty \frac{dx}{x^2}\big(-b\ln x\big)^n = (-b)^n\,n! \tag{11}$$

each term is finite, yet the coefficients grow as $n!$ (the closed form agrees with numerical integration to 6 digits Computed). The reason is the position of the peak. The infrared integrand (measured in $\ln x$, $x^2(\ln\frac{1}{x})^n$) peaks at $x = e^{-n/2}$, and **as the order rises the peak runs away to smaller momenta**. The ultraviolet peak is at $x = e^n$. (2.24) and (2.25) of Beneke give the same positions for the bubble chain diagram Known . On the infrared side the signs are the same (positive axis, with ambiguity), and on the ultraviolet side they alternate. Changing the power at small k to k^a, the infrared renormalon sits at $u = a/2$ on the scale of § 08, and the size of the ambiguity is $(\Lambda/Q)^a$ (Beneke § 2.2) Known . This model has $a = 4$.

Figure 2 Raising n moves the infrared peak of (11) (top panel, horizontal axis $\log_{10} k^2/Q^2$) to the left, and the term $r_n\alpha^{n+1}$ (bottom panel, logarithmic) becomes smallest about where it crosses the dashed line Λ^2/Q^2. Λ moves with α




```
n = 4, peak k²/Q² = 1.35e−1, Λ²/Q² = 6.24e−5 (N = 3), term = 1.01e−5, smallest term at n* = 19
```

Computed The numbers in this figure were computed on this machine (the formula of (11) evaluated inside the figure; agreeing with Table 4).

Where does the peak stop? At one loop $\Lambda^2/Q^2 = e^{-1/(b\alpha)}$ (4-07). The term $r_n^{\mathrm{IR}}\alpha^{n+1}$ is smallest at $n^* \approx 2/(b\alpha)$, and then the peak is at $e^{-n^*/2} \approx \Lambda^2/Q^2$. **The order at which the series stops being useful is the order at which the loop momentum reaches Λ,** and the smallest term has the size of $(\Lambda/Q)^4$ times a factor of the order of a power of α.

Computed Table 4 The order of the smallest term and the position of the peak in the model for pure SU(3) ($b = 11/4\pi$)

$\alpha(Q)$	Λ^2/Q^2	n^*	Peak $/(\Lambda^2/Q^2)$	Smallest term	Smallest term $/(\Lambda/Q)^4$
0.300	2.22e−2	7	1.36	5.09e−4	1.03
0.200	3.31e−3	11	1.24	9.23e−6	0.84
0.118	6.24e−5	19	1.20	2.53e−9	0.65
0.080	6.29e−7	28	1.32	2.11e−13	0.54

Beneke notes that the positions of renormalons do not depend on the higher coefficients of the β function, and are in general logically separate from the Landau pole Known .

08

The Borel plane of Yang–Mills — a map of the expected singularities

Transforming the series in α in the manner of (8) and measuring positions by $u = \beta_0 t/4\pi$, $e^{-t/\alpha} = (\Lambda^2/Q^2)^u$, so **a singularity at position u corresponds to an ambiguity of size $(\Lambda/Q)^{2u}$**. We prepare two terms. The **operator product expansion** is a physical prescription that expands a quantity measured at short distance $1/Q$ as a sum of "coefficient $\times$ $(\Lambda/Q)^d$ $\times$ the vacuum expectation value of a combination of fields of dimension d" Physics . In pure Yang–Mills, the gauge-invariant combination of fields of

lowest dimension is $F^a_{\mu\nu}F^a_{\mu\nu}$, of dimension 4, and its expectation value $\langle F^2\rangle$ is called the **gluon condensate**. There is no gauge-invariant combination of dimension 2 ($A_\mu A_\mu$ has dimension 2 but changes under gauge transformations).

For the two-point correlation functions of QCD that Beneke summarises (quantities measured by injecting a momentum Q from outside; an example is the correlation of quark currents) Physics , the infrared renormalons are at $u = 2, 3, \ldots$ (there is none at $u = 1$ because the operator product expansion has no term of dimension 2, Parisi 1979; $u = 2$ corresponds to the $(\Lambda/Q)^4$ of the gluon condensate, and the weight k^4 of the model of § 07 is this $u = 2$), the ultraviolet renormalons at $u = -1, -2, \ldots$, and the instanton–anti-instanton pairs at $t = 4\pi n$ (Bogomolny–Fateyev 1977). At which u they are present depends on the observable. Converting to $t = 4\pi u/\beta_0$ with the β_0 of pure $SU(N)$, the renormalons move towards the origin in proportion to $1/N$, and the pairs stay at 4π. For $N = 3$, $u = 2$ is at $t = 2.285$ and the pair 5.5 times farther; for $N = 10$, 18 times Computed .

The weight of a pair $e^{-4\pi/\alpha} = e^{-16\pi^2/g^2}$ is $(\Lambda/Q)^{22N/3}$, a far smaller quantity than the $(\Lambda/Q)^4$ of $u = 2$ (for a single one, (14) of 4-07 § 08). Beneke also expects that in QCD the instanton singularities are far away and do not determine the large orders Physics .

Observation on the lattice. Bali–Bauer–Pineda (2014) computed numerically the perturbative coefficients of the average plaquette of pure SU(3) up to order 35 in α, and write 'The on-set of the dominance of the dimension four renormalon associated to the gluon condensate is clearly observed' (abstract) Physics . The PDG review also says that the renormalon divergence is 'not usually visible' at the orders available, but is 'closely connected' with non-perturbative contributions.

09

Seen from mathematics — what is a theorem and what is an expectation

On the side of theorems. (4) to (6) and (9) for the zero-dimensional model are checkable mathematics Known . There are papers with Borel summability in their titles, for the anharmonic oscillator (Graffi–Grecchi–Simon 1970) and for $P(\phi)_2$ and ϕ^4_3 (Eckmann–Magnen–Sénéor 1975, Magnen–Sénéor 1977; 7-03) (bibliographic data only). According to their titles, they show Borel summability for the eigenvalues of the oscillator and for the Schwinger functions (correlation functions) of $P(\phi)_2$ and ϕ^4_3 respectively. In those cases, by the Watson–Nevanlinna–Sokal theorem of § 05, the series determines the value uniquely.

On the side of expectations. For four-dimensional Yang–Mills the coefficients are well defined and finite at each order (4-05), but the relation between the series and the

theory has not been shown. The non-perturbative theory that the question presupposes was built by the construction published in 2026-11 (at the level of the committee's assessment; 6-00), but neither agreement with perturbation theory at all orders nor that the series is an asymptotic expansion of that theory has been shown (6-04). Before the published construction, Beneke (1999) wrote 'Since there is no rigorous non-perturbative definition of R in theories such as QED and QCD, we cannot even ask whether series expansions are asymptotic' Known . The problem statement (Jaffe–Witten § 6.1) also writes that in most known examples perturbation series are 'divergent expansions' and 'even Borel and other resummation methods have limited applicability' Known .

To sum up: (i) the $n!$ in the coefficients is the shadow of quantities of size e^{-c/g^2} that the series does not show Physics . Precisely, in the model, in the sense of (4) to (6), the exponent c of the imaginary part at negative coupling fixes the base $1/c$ of the coefficients. For Yang–Mills it is an expectation at the level of the chain (10), and ultraviolet renormalons (with alternating signs) carry no ambiguity of exponentially small terms. (ii) If the gap is proportional to $\Lambda \propto e^{-8\pi^2/(\beta_0 g^2)}$, it appears at no order of the series (4-07 § 05), and the only clue left is on the side of how the series diverges Physics . (iii) The formalism that assembles the unseen terms from the way the series diverges (transseries) is in 7-12 § 09, and for pure Yang–Mills on $\mathbb{R}^4$ it is not established even at the level of physics.

Outside the scope of this chapter

For obstacles other than divergence, see 7-11; for the planar theory, in which the number of diagrams grows only exponentially, 17-04.

What this chapter can say

	Content
What is said	(4) to (6) and (9) for the zero-dimensional model and the agreement of the Borel sum (Tables 1 to 3). The coefficients of the oscillator to order 150. The closed form of (11), and that the smallest term comes at the order at which the peak reaches Λ (Table 4) Computed . The conversion of positions in the Borel plane
Conditional	The chain (10) Physics . The cancellation in the double well is taken from the literature (text not checked). The positions of renormalons are expectations based on the bubble-chain approximation and the operator product expansion. $u = 2$ for the plaquette is a numerical observation

	Content
What is not said	The divergence, asymptoticity and the reality of renormalons for the series of four-dimensional Yang–Mills are none of them theorems. Dyson's argument is heuristic. This site shows neither the mass gap nor the construction of the continuum theory

Sources and reproduction

Item	Kind	Source / tools
Beneke's statements and quotations in §03 and §05 to §09	Known	M. Beneke, Phys. Rept. 317 (1999) 1, hep-ph/9807443 (text §1, §2)
Quotations from the problem statement and the PDG	Known	Jaffe–Witten (the Clay problem statement) §6.1; J. Huston, K. Rabbertz, G. Zanderighi, '9. Quantum Chromodynamics' (revised August 2023), in S. Navas et al. (Particle Data Group), Phys. Rev. D 110 (2024) 030001, footnote 1 (text of both)
Dyson, the anharmonic oscillator	Known	Dyson, Phys. Rev. 85 (1952) 631; Bender–Wu, Phys. Rev. 184 (1969) 1231, PRD 7 (1973) 1620 (abstracts of all the above)
The lattice plaquette	Physics	Bali–Bauer–Pineda, PRD 89 (2014) 054505, arXiv:1401.7999 (abstract)
Others	Bibliographic data only	Lipatov, JETP 45 (1977) 216; 't Hooft, Erice 1977 lectures (Plenum, 1978); Gross–Neveu, PRD 10 (1974) 3235; Lautrup, PLB 69 (1977) 109; Parisi, NPB 150 (1979) 163; Bogomolny–Fateyev, PLB 71 (1977) 93; Bogomolny, PLB 91 (1980) 431; Zinn-Justin, NPB 192 (1981) 125; Sokal, JMP 21 (1980) 261; Graffi–Grecchi–Simon, PLB 32 (1970) 631; Eckmann–Magnen–Sénéor, CMP 39 (1975) 251; Magnen–Sénéor, CMP 56 (1977) 237
Tables 1 to 3, Figure 1, the check of the dispersion relation, the ratio to the smallest term in §04	Computed	`zero_dim.py` (SciPy, rationals), `min_term.py`. Figure 1 uses Simpson's rule in JavaScript
The coefficients of the anharmonic oscillator to order 150	Computed	`oscillator.py` (recurrence with rationals; the first 4 terms agree with the known values)
(11), Table 4, Figure 2, the positions in §08	Computed	`renormalon.py`. The figure evaluates the same formula in JavaScript

Next chapter to read: 5-01 Wilson's lattice. The detour continues in 7-12.