A Constructive Introduction to Quantum Field Theory

Lecture 4: Forward–backward equations

M. GUBINELLI
Mathematical Institute, University of Oxford

SISSA, November 2026

Abstract

Building on the scale decomposition and the stochastic observables of Lecture 3, we introduce interactions. The interacting measure can be generated in three ways: as a renormalisation group decomposition of the covariance, as an optimal control problem, whose dynamic programming equation is the Polchinski equation, and as a filtering problem: the field at resolution t is a noisy observation of the terminal field, the Polchinski process is this observation written in its own filtration, and backward in resolution its dynamics is linear, with Gaussian transitions, started from the interacting measure. The three views meet in one forward–backward system for a field and a force, and its natural coordinates are the force coordinates: the forward equation loses its drift, and the conditional covariance becomes the propagator. All this can be said in terms of forces, propagators and covariances. In the gradient case there is a parallel story for potentials: the value function generates the canonical transformation to force coordinates, the Legendre transform of the log-partition function of the filter gives the effective average action, and its flow is the Wetterich equation. Lecture 5 continues with the observables of the interacting theory, the Euclidean field equation and Nelson's Euclidean noise, the field equation at every resolution with its sub-Gaussian noise, and infinite volume, and contains the appendices. The treatment is formal: we compute as if all objects were smooth and integrable, and we indicate where the real difficulties lie.

The results presented in Lectures 4–7 were obtained recently, with the use of AI, in collaboration with Sarah-Jean Meyer.

Table of contents

Introduction . . . 2
Structure . . . 2
What is formal . . . 2
Conventions . . . 2
Acknowledgements . . . 2

Main lecture . . . 2

7 Three ways to generate the interacting measure . . . 2
7.1 Renormalisation: decomposing the covariance . . . 3
The field as a process . . . 4
7.2 Optimal control: the Polchinski equation . . . 5
The control problem . . . 6
Euler–Lagrange equation = FBSDE . . . 6
Entropy . . . 6
Generating functionals . . . 6
Pontryagin's maximum principle . . . 7
The costate is a gradient . . . 7
The Hamiltonian of the field . . . 7
With noise . . . 8
7.3 Filtering: measuring the field . . . 8
The terminal field as a hidden signal . . . 8
Generation by denoising . . . 10
Literature: Polchinski, Eldan and diffusion models . . . 11
7.4 Where the three views meet . . . 11
Exercises . . . 12

8 Force coordinates and potentials . . . 13
8.1 The force and a Burgers equation . . . 13
8.2 Removing the predictable drift . . . 14
8.3 The FBSDE in the forecast coordinate . . . 14
A one-dimensional model: Burgers and Lax–Oleinik . . . 15
8.4 Conditional moments . . . 16
Brascamp–Lieb and Cramér–Rao bounds . . . 16
8.5 Potentials: the value function as a generator . . . 17
Phase space and Lagrangian graphs . . . 17
The free action generates the shear . . . 17
The value function generates the graph . . . 18
No convexity is needed . . . 18
8.6 The Legendre transform and Wetterich's equation . . . 19
Probabilistic meaning of Γ . . . 21
8.7 Variational meaning and the tree level . . . 21
Constrained free energies . . . 21
The tree level . . . 22

Exercises . . . . . 22 **References** 23

Introduction

In Lecture 3 we decomposed the Gaussian free field along a resolution parameter and took as primary objects the stochastic observables: martingales in the filtration of finite-resolution measurements [BCG26]. In this lecture we introduce interactions, and we tell one story from three points of view: renormalisation, optimal control and filtering. See also [Pol84, BK87, Sal99, BBD24] for the renormalisation group side and [BG20, BG23] for the variational side.

Structure. Section 7 generates the interacting measure in the three ways, and shows that they produce the same field and the same force. Section 8 finds the natural coordinates, first in terms of forces, propagators and covariances, then in terms of potentials, where the value function generates the canonical transformation and the Legendre transform leads to the Wetterich equation. Lecture 5 continues with observables, the Euclidean field equation and Nelson's Euclidean noise, the field equation at every resolution with its sub-Gaussian noise, the Dyson–Schwinger equations, also in pathwise form, and infinite volume, and contains the appendices: the algebra of the flow, perturbation theory and counterterms, and the functional inequalities that follow from the localisation.

What is formal. Rigorous statements require a finite-dimensional setting: a finite lattice, or a Fourier cutoff on the torus, together with a finite terminal resolution T. There, all the identities below are theorems under mild growth assumptions on the interaction, and their proofs are the computations we give. The hard problems are the limits $T \to \infty$ (ultraviolet) and infinite volume (infrared), together with the locality of the limiting objects. We mark with (formal) the statements that we only derive at this level.

Conventions. As in Lectures 1–3: $\mathcal{M} = \mathbb{T}_L^d$ or $\mathbb{R}^d$, $A = m^2 - \Delta$, $C = A^{-1}$ with kernel G, μ_m the GFF, $\langle f, g\rangle = \int fg$ for real functions. For a functional F on smooth fields, $\mathrm{D}F(\varphi) \in \mathcal{S}'(\mathcal{M})$ and $\mathrm{D}^2F(\varphi)$ are its first and second derivatives, $\mathrm{D}_h F = \langle \mathrm{D}F, h\rangle$. $H_n(a;c)$ are the Hermite polynomials of variance c, and μ^Q is the centred Gaussian measure with covariance Q.

Acknowledgements. The results presented in Lectures 4–7 are joint work with Sarah-Jean Meyer, obtained recently with the use of AI. M. Gubinelli is partly supported by UK Research and Innovation via the grant "StochFields" EP/Z534328/1 and by the Simons Collaboration on Probabilistic Paths to QFT. S.-J. Meyer is supported by the grant "StochFields" EP/Z534328/1.

Main lecture

7 Three ways to generate the interacting measure

Fix a finite volume $\mathcal{M} = \mathbb{T}_L^d$ and a terminal resolution $T < \infty$, which plays the role of an ultraviolet cutoff: the reference field $X_T \sim \mu^{Q_T}$ is smooth. Let V_T be an interaction, a functional of smooth fields bounded below. Our main example is

$$V_T(\varphi) = \int_{\mathbb{T}_L^d} \Big[\frac{\lambda}{12} H_4(\varphi(x); c_T) + \gamma_T \, H_2(\varphi(x); c_T) \Big] \mathrm{d}x, \tag{7.1}$$

with a mass counterterm γ_T to be fixed in Appendix A.2 of Lecture 5 ($\gamma_T = 0$ for $d = 2$). The coupling is normalised by $\lambda/12$ so that the curvature of the quartic term is $\lambda\,\varphi^2$, as in Lectures 6–7. The interacting measure at cutoff T is

$$\nu_T(\mathrm{d}\varphi) = Z_T^{-1} \mathrm{e}^{-V_T(\varphi)} \mu^{Q_T}(\mathrm{d}\varphi). \tag{7.2}$$

How does the Gaussian scale decomposition of Lecture 3 generate ν_T? There are three answers.

- **Renormalisation** (Section 7.1). Split the covariance into shells and integrate them out one at a time. At every resolution the field is again a Gibbs measure, and it is driven by an effective force.
- **Optimal control** (Section 7.2). Steer the Gaussian field by a drift, at the least cost. The effective potential is the value function, and the Polchinski equation is the dynamic programming equation.
- **Filtering** (Section 7.3). Observe the terminal field through noise, with increasing precision. The field at resolution t is the observation, the force is what it reveals about the terminal field, and backward in resolution the process has a linear dynamics with Gaussian transitions, started from the interacting measure.

Section 7.4 shows that the three constructions produce the same process. Throughout,

$$S_t := Q_T - Q_t = Q_{T,t}, \qquad \dot{S}_t = -\dot{Q}_t, \qquad S_T = 0,$$

is the covariance still unresolved at resolution t.

7.1 Renormalisation: decomposing the covariance

Decompose the covariance as in Lecture 3: $Q_T = Q_t + S_t$, and correspondingly

$$X_T = X_t + (X_T - X_t)$$

with independent terms. Integrating out the second one defines the effective potential.

Definition 7.1. *The* effective potential *at resolution $t \in [0,T]$ is*

$$\mathrm{e}^{-V_t(\varphi)} := \int \mathrm{e}^{-V_T(\varphi+\zeta)}\, \mu^{Q_{T,t}}(\mathrm{d}\zeta). \tag{7.3}$$

V_t is the interaction felt by the field at resolution t once the fluctuations between t and T have been integrated out: it is the Wilson effective action. At $t = 0$, $V_0(0) = -\log Z_T$ is the free energy. Here is the first key fact.

Proposition 7.2. (Marginals) *Let $\mathbb{P}$ be defined by*

$$\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{P}_0} = Z_T^{-1} \mathrm{e}^{-V_T(X_T)},$$

where $\mathbb{P}_0$ is the law of the Gaussian process X. Then $\mathrm{Law}_{\mathbb{P}}(X_T) = \nu_T$, and for every $t \leqslant T$

$$\mathrm{Law}_{\mathbb{P}}(X_t) = \nu_t, \qquad \nu_t(\mathrm{d}\varphi) := Z_T^{-1} \mathrm{e}^{-V_t(\varphi)} \mu^{Q_t}(\mathrm{d}\varphi). \tag{7.4}$$

Moreover, the conditional law of X_T given $\mathcal{F}_t$ is

$$\mu_t^{X_t}(\mathrm{d}\varphi), \qquad \mu_t^{\phi}(G) := \mathrm{e}^{V_t(\phi)} \int G(\phi+\zeta)\, \mathrm{e}^{-V_T(\phi+\zeta)}\, \mu^{Q_{T,t}}(\mathrm{d}\zeta). \tag{7.5}$$

Proof. Under $\mathbb{P}_0$, $X_T = X_t + (X_T - X_t)$ with an independent increment of law $\mu^{Q_{T,t}}$. Hence $\mathbb{E}_{\mathbb{P}}[G(X_t)] = Z_T^{-1}\,\mathbb{E}_0[G(X_t)\mathrm{e}^{-V_T(X_T)}] = Z_T^{-1}\,\mathbb{E}_0[G(X_t)\mathrm{e}^{-V_t(X_t)}]$. The conditional law follows in the same way. □

So at every resolution the field is again a Gibbs measure, with reference covariance Q_t and potential V_t. The measure-valued martingale $(\mu_t^{X_t})_t$ goes from ν_T at $t=0$ to δ_{X_T} at $t=T$: it is a *stochastic localisation* of ν_T in the sense of Eldan [Eld23], adapted to the covariance profile [BBD24, BCG26].

The field as a process. The density of the interacting law with respect to the Gaussian one is a martingale in resolution, so by Girsanov the field acquires a drift, and this drift is the effective force. The resulting process is the **Polchinski process**.

Theorem 7.3. (FBSDE for the interacting field) (formal) *Under $\mathbb{P}$ the process $\Phi_t := X_t$, the* interacting field at resolution t, *solves*

$$\mathrm{d}\Phi_t = \dot{Q}_t\,F_t\,\mathrm{d}t + \dot{Q}_t^{1/2}\,\mathrm{d}W_t^{\mathbb{P}}, \qquad F_t = -\mathrm{D}V_t(\Phi_t) = \mathbb{E}_t^{\mathbb{P}}\,[-\mathrm{D}V_T(\Phi_T)], \tag{7.6}$$

with $\Phi_0 = 0$ and a $\mathbb{P}$-cylindrical Brownian motion $W^{\mathbb{P}}$. In particular the effective force $(F_t)_{t\leqslant T}$ *is a $\mathbb{P}$-martingale: an observable field of the interacting theory.*

Proof. The density process is $M_t := \mathbb{E}_0[\,\mathrm{d}\mathbb{P}\,/\,\mathrm{d}\mathbb{P}_0|\mathcal{F}_t] = Z_T^{-1}\mathrm{e}^{-V_t(X_t)}$, by Proposition 7.2. It is a martingale, so the drift in Itô's formula vanishes and $\mathrm{d}M_t = -M_t\langle \mathrm{D}V_t(X_t), \dot{Q}_t^{1/2}\,\mathrm{d}W_t\rangle$. By Girsanov's theorem, $W_t^{\mathbb{P}} = W_t + \int_0^t \dot{Q}_s^{1/2}\mathrm{D}V_s(X_s)\,\mathrm{d}s$ is a $\mathbb{P}$-Brownian motion, which is (7.6). Differentiating (7.3) gives $\mathrm{D}V_t(\phi) = \mu_t^{\phi}(\mathrm{D}V_T)$, which is the second identity. □

Remark 7.4. (A Doob *h*-transform) Under $\mathbb{P}_0$, X is the GFF martingale of Lecture 3, $X_t = \mathbb{E}_0[X_T|\mathcal{F}_t]$, a Gaussian process with independent increments and generator $L_t = \frac{1}{2}\mathrm{Tr}(\dot{Q}_t\,\mathrm{D}^2)$. The function

$$h_t(\phi) := Z_T^{-1}\mathrm{e}^{-V_t(\phi)} = \mathbb{E}_0[Z_T^{-1}\mathrm{e}^{-V_T(X_T)}|X_t = \phi]$$

is space-time harmonic for X, $(\partial_t + L_t)\,h_t = 0$, and this is the Polchinski equation (7.8) for $V_t = -\log h_t - \log Z_T$. Since $h_0(0) = 1$, the density of $\mathbb{P}$ with respect to $\mathbb{P}_0$ on $\mathcal{F}_t$ is $h_t(X_t)$: the Polchinski process is the *Doob h-transform* of the GFF martingale by the terminal weight e^{-V_T}. Its generator is

$$L_t^h G = h_t^{-1}\,(\partial_t + L_t)\,(h_t G) - \partial_t G = L_t G + \langle \dot{Q}_t\,\mathrm{D}\log h_t, \mathrm{D}G\rangle,$$

and the force is the logarithmic gradient of h, $F_t = \mathrm{D}\log h_t(\Phi_t)$. Conditioning a Brownian motion on its endpoint is the simplest example: h_t is then the Gaussian density of the remaining increment, and the drift is that of the bridge (7.15). The interacting field is a GFF martingale conditioned, in the sense of Doob, to end with the law ν_T. The same transform acts on the backward Kolmogorov equation (9.2) of Lecture 5 (Exercise 9.1 there).

Read as an equation for an unknown pair (Φ, F), (7.6) is a *forward–backward* system [MY99]. The field is driven forward in resolution by a force that is determined backward, as the conditional expectation of a functional of the terminal field:

$$\Phi_t = \int_0^t \dot{Q}_s F_s\,\mathrm{d}s + \int_0^t \dot{Q}_s^{1/2}\mathrm{d}W_s, \qquad F_t = \mathbb{E}_t\,[-\mathrm{D}V_T(\Phi_T)], \qquad t\in[0,T]. \tag{7.7}$$

Here and below we drop the superscript and write W for the Brownian motion driving the interacting field. No equation for the potential is needed to state (7.7), and $\nu_T = \mathrm{Law}(\Phi_T)$ is recovered from the solution. This is the starting point of FBSDE constructions of Euclidean fields (Figure 7.1): one solves (7.7) with renormalised V_T, with estimates uniform in T and in the volume, and takes the limit. The field coordinate Φ is *not* an observable, only a semi-martingale. Its martingale part is the Gaussian noise, and the observables are built from it and from F.

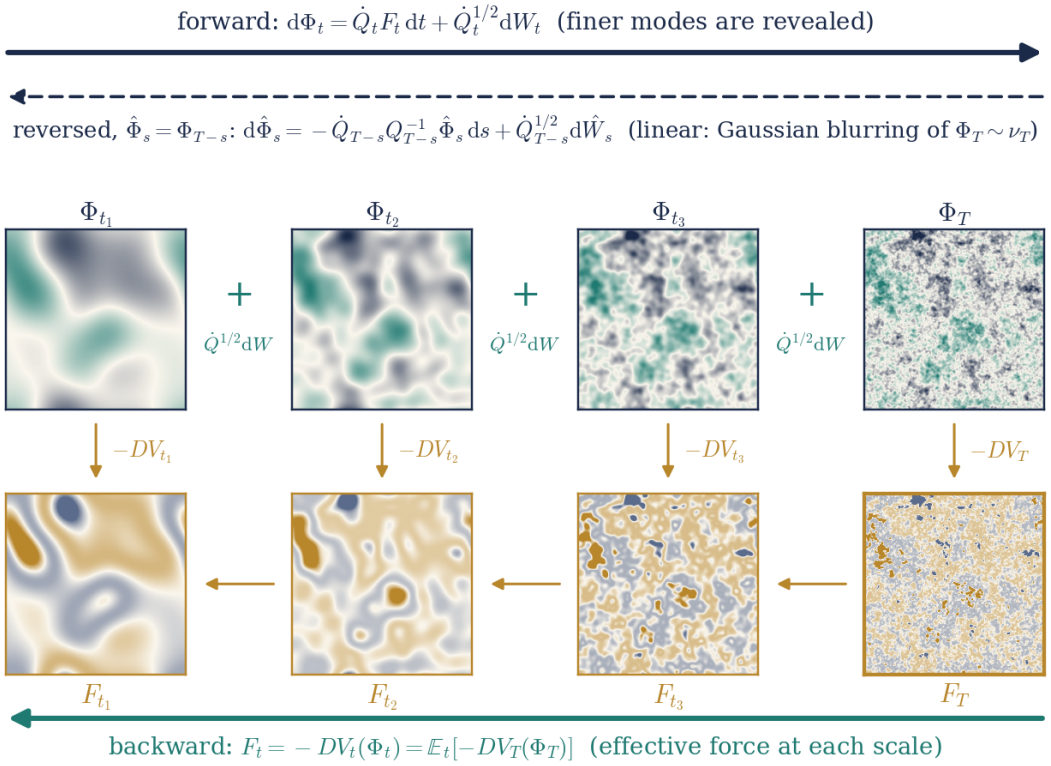


Figure 7.1. The forward–backward structure of (7.7) on $\mathbb{T}^2$, heat profile, at resolutions $t_1 < t_2 < t_3 < T$, all built from the same noise. Top: the field gains the fluctuation of each shell, $\dot{Q}_t^{1/2}\mathrm{d}W_t$, and reveals finer modes. Read backward, the same process solves the linear equation (7.17): a Gaussian blurring of $\Phi_T \sim \nu_T$, which does not see the interaction (dashed arrow). Vertical arrows: at each resolution the force is the function $-\mathrm{D}V_t$ of the field. Bottom: the effective force $F_t = \mathbb{E}_t[-\mathrm{D}V_T(\Phi_T)]$ for the Φ^4 interaction, determined backward from the terminal interaction. It is shown to first order in λ, where it is the Wick cube $-\frac{\lambda}{3}H_3(\Phi_t; c_t)$, and each panel is in its own units.

7.2 Optimal control: the Polchinski equation

The measure ν_T also solves a minimisation problem. Perturb the Gaussian field by an adapted drift, and pay for the drift and for the interaction at the endpoint. The effective potential is the value function of this problem. Its equation in t, which we first derive directly from (7.3), is the dynamic programming equation.

Theorem 7.5. (Polchinski equation [Pol84]) (formal) *V_t solves the backward Hamilton–Jacobi equation*

$$\partial_t V_t = \tfrac{1}{2}\langle \mathrm{D}V_t, \dot{Q}_t\,\mathrm{D}V_t\rangle - \tfrac{1}{2}\mathrm{Tr}(\dot{Q}_t\,\mathrm{D}^2 V_t), \qquad t \in [0,T], \tag{7.8}$$

with terminal condition V_T.

Proof. $h_t = Z_T^{-1} e^{-V_t} = e^{\frac{1}{2}\Delta_{Q_T,t}} h_T$ (Remark 7.4) is a Gaussian convolution, where $\Delta_Q := \mathrm{Tr}(Q\,\mathrm{D}^2)$. Since $\partial_t Q_{T,t} = -\dot{Q}_t$, it solves the backward heat equation $\partial_t h_t = -\frac{1}{2}\mathrm{Tr}(\dot{Q}_t\,\mathrm{D}^2 h_t)$. Insert $\mathrm{D}^2 h = (\mathrm{D}V \otimes \mathrm{D}V - \mathrm{D}^2 V)\,h$ and divide by $-h$. □

The nonlinear term generates new, generally nonlocal, interactions (trees), and the trace term contracts pairs of fields with the covariance of the shell $[t, t+\mathrm{d}t]$ (loops). The substitution $h \propto e^{-V}$, which linearises the equation, is a Cole–Hopf transform: we come back to this in Section 8.1, and in Appendix A of Lecture 5.

The control problem. Let $\mathbb{H}_a$ be the set of progressively measurable drifts $u\colon [0,T] \to L^2(M)$ with $\mathbb{E}\int_0^T \|u_s\|^2 \mathrm{d}s < \infty$, and set

$$I_t(u) := \int_0^t \dot{Q}_s^{1/2} u_s \,\mathrm{d}s, \qquad X^u := X + I(u),$$

the field driven by the control u under the Gaussian reference $\mathbb{P}_0$.

The following result is standard in the stochastic optimal control setting, see e.g. Fleming and Soner's book [FS06]. A useful non-Markovian generalisation is given by Boué–Dupuis [BD98].

Theorem 7.6. (Polchinski is HJB) (formal)

$$-\log \mathbb{E}_0[e^{-V_T(X_T)}] = V_0(0) = \inf_{u\in\mathbb{H}_a} \mathbb{E}_0\Big[V_T(X_T^u) + \frac{1}{2}\int_0^T \|u_s\|^2 \mathrm{d}s\Big]. \tag{7.9}$$

The infimum is attained at the feedback control $u_s^* = -\dot{Q}_s^{1/2}\,\mathrm{D}V_s(X_s^{u^*})$, *and* $\mathrm{Law}(X_T^{u^*}) = \nu_T$. *The effective potential is the value function, and* (7.8) *is the Hamilton–Jacobi–Bellman equation of this control problem.*

Proof. (Verification.) For any u, by Itô's formula and (7.8),

$$\begin{aligned}\mathrm{d}V_t(X_t^u) &= \big[\tfrac{1}{2}\|\dot{Q}_t^{1/2}\mathrm{D}V_t\|^2 + \langle \dot{Q}_t^{1/2}\mathrm{D}V_t, u_t\rangle\big]\mathrm{d}t + \mathrm{d}M_t \\ &= \big[\tfrac{1}{2}\|u_t + \dot{Q}_t^{1/2}\mathrm{D}V_t\|^2 - \tfrac{1}{2}\|u_t\|^2\big]\mathrm{d}t + \mathrm{d}M_t,\end{aligned}$$

with $\mathrm{D}V_t = \mathrm{D}V_t(X_t^u)$ and a martingale M. Integrating and taking expectations,

$$\mathbb{E}_0\Big[V_T(X_T^u) + \frac{1}{2}\int_0^T \|u_t\|^2\mathrm{d}t\Big] = V_0(0) + \frac{1}{2}\,\mathbb{E}_0\int_0^T \|u_t + \dot{Q}_t^{1/2}\mathrm{D}V_t(X_t^u)\|^2\mathrm{d}t. \tag{7.10}$$

So the cost is $\geqslant V_0(0)$, with equality iff $u = u^*$. With this control X^{u^*} solves the same SDE as Φ in (7.6). □

Euler–Lagrange equation = FBSDE. Perturbing $u \to u + \varepsilon v$ in (7.9), the first-order condition is $\mathbb{E}\int_0^T \langle u_s + \dot{Q}_s^{1/2}\,\mathbb{E}_s[\mathrm{D}V_T(X_T^u)], v_s\rangle \mathrm{d}s = 0$ for all adapted v, i.e.

$$u_s = -\dot{Q}_s^{1/2}\;\mathbb{E}_s[\mathrm{D}V_T(X_T^u)] = \dot{Q}_s^{1/2} F_s. \tag{7.11}$$

This is (7.7). The FBSDE is the Euler–Lagrange equation of the variational problem, and the Polchinski equation is its Hamilton–Jacobi equation.

Entropy. By Girsanov, $\frac{1}{2}\,\mathbb{E}_0\int\|u\|^2 \geqslant \mathrm{Ent}(\mathrm{Law}(X^u)\,|\,\mathrm{Law}(X))$, which is at least the entropy of the terminal marginal. So (7.9) is a pathwise lift of the Gibbs variational principle $-\log Z_T = \inf_\nu\{\int V_T\,\mathrm{d}\nu + \mathrm{Ent}(\nu|\mu^{Q_T})\}$. The lift is useful because adapted drifts can be *constructed scale by scale*: one looks for u as a functional of the Gaussian process X, not as an unknown measure.

Generating functionals. Adding a test functional f,

$$\mathcal{W}_T(f) := -\log\int e^{-f}\mathrm{d}\nu_T = \inf_u \mathbb{E}_0\Big[f(X_T^u) + V_T(X_T^u) + \frac{1}{2}\int_0^T \|u_s\|^2\mathrm{d}s\Big] - V_0(0). \tag{7.12}$$

Bounds on $\mathcal{W}_T(f)$ uniform in T, for f in a rich class, give tightness, moment bounds and the identification of limits. For Φ^4_3 the naive problem does not have uniform bounds: the leading-order optimal drift $u_s \approx -\frac{\lambda}{3}\dot{Q}_s^{1/2}{:}X_s^3{:}$ has a divergent cost. One renormalises by the change of variables $u = -\frac{\lambda}{3}\dot{Q}^{1/2}{:}X^3{:}+r$. After it, and after further paracontrolled changes of variables, the singular terms are explicit polynomials in the Wick powers of X, cancelled by *the counterterms*, and the residual functional is coercive. Its control in Besov and paracontrolled norms, uniformly in T, is the construction of Φ^4_3 in [BG20]; see also [BG21].

Pontryagin's maximum principle. Consider first a deterministic problem: minimise the cost $J(u) = g(x_T) + \int_t^T \ell(x_s, u_s)\,\mathrm{d}s$ over controls u on $[t,T]$, where the state solves $\dot{x}_s = b(x_s,u_s)$, $x_t = x$. With the *Hamiltonian*

$$\mathcal{H}(x,p,u) := \langle p, b(x,u)\rangle - \ell(x,u),$$

Pontryagin's principle [PBGM62] says that along an optimal pair (x,u) there is a *costate* p such that

$$\dot{x}_s = \partial_p \mathcal{H}, \qquad \dot{p}_s = -\partial_x \mathcal{H}, \qquad p_T = -\mathrm{D}g(x_T), \qquad u_s \in \arg\max_v \mathcal{H}(x_s,p_s,v).$$

The state runs forward from its initial value, the costate backward from its terminal value: a two-point boundary value problem. In the smooth case the principle is the first-order condition. If $\xi = \partial_\varepsilon x$ is the variation of the state under $u \to u + \varepsilon v$, then $\dot{\xi} = \partial_x b\,\xi + \partial_u b\, v$, $\xi_t = 0$, and the costate equation is exactly what makes $\frac{\mathrm{d}}{\mathrm{d}s}\langle p,\xi\rangle = \langle \partial_x \ell, \xi\rangle + \langle p, \partial_u b\, v\rangle$. Integrating over $[t, T]$ and using the terminal condition,

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon} J(u+\varepsilon v)|_{\varepsilon=0} = -\int_t^T \langle \partial_u \mathcal{H}(x_s,p_s,u_s), v_s\rangle\,\mathrm{d}s.$$

So the costate is the Lagrange multiplier of the constraint $\dot{x} = b$, and an optimal control is a critical point of the Hamiltonian.

The costate is a gradient. Let $V(t,x) = \min_u J(u)$ be the value function. Comparing the cost from (t,x) with the cost from $(t+\mathrm{d}t, x + b\,\mathrm{d}t)$ gives the Hamilton–Jacobi–Bellman equation

$$\partial_t V = \mathcal{H}_*(x, -\mathrm{D}V), \qquad \mathcal{H}_*(x,p) := \max_u \mathcal{H}(x,p,u), \qquad V(T,\cdot) = g.$$

Its characteristics are Hamilton's equations for $\mathcal{H}_*$, and along them $p_s = -\mathrm{D}V(s,x_s)$: the costate is the sensitivity of the optimal cost to a displacement of the state, i.e. the force that the future exerts on the present.

The Hamiltonian of the field. In the problem of Theorem 7.6 the state is the field, $b(x,u) = \dot{Q}_s^{1/2}u$, $\ell = \frac{1}{2}\|u\|^2$ and $g = V_T$. Then $\mathcal{H}(p,u) = \langle p, \dot{Q}_s^{1/2}u\rangle - \frac{1}{2}\|u\|^2$ is maximal at $u = \dot{Q}_s^{1/2}p$, and

$$\mathcal{H}_*(p) = \frac{1}{2}\langle p, \dot{Q}_s p\rangle$$

is the Hamiltonian of a free particle with mass matrix $\dot{Q}_s^{-1}$: it does not depend on the state. Hence $\dot{p} = 0$ and $\dot{x} = \dot{Q}_s p$: the costate is conserved, and from t to T the state moves along the linear shear

$$\Sigma_t{:}\,(x,p) \longmapsto (x + S_t p,\ p). \tag{7.13}$$

Its trajectories are the characteristics of the transport part of (8.1). All the interaction sits in the terminal condition $p = -\mathrm{D}V_T(x + S_t p)$, an implicit equation for the momentum, solved by a sum of tree graphs (Exercise 8.1). The corresponding Hamilton–Jacobi equation $\partial_t V = \frac{1}{2}\langle \mathrm{D}V, \dot{Q}_t \mathrm{D}V\rangle$ is the Polchinski equation (7.8) without its trace term.

With noise. For a controlled diffusion $\mathrm{d}X = b(X,u)\,\mathrm{d}s + \sigma\,\mathrm{d}W$, with σ independent of the control, the stochastic maximum principle [Bis73, Pen90, YZ99] keeps the Hamiltonian and the maximisation, but the costate becomes an adapted process which solves a *backward* stochastic differential equation,

$$\mathrm{d}p_s = -\partial_x \mathcal{H}\,\mathrm{d}s + \Upsilon_s\,\mathrm{d}W_s, \qquad p_T = -\mathrm{D}g(X_T).$$

The martingale term $\Upsilon\,\mathrm{d}W$ is the price of adaptedness: the costate may not anticipate the noise. For the field $\partial_x \mathcal{H} = 0$, so the costate is a martingale, $p_t = \mathbb{E}_t[-\mathrm{D}V_T(X_T)]$, and since the value function is the effective potential (Theorem 7.6), $p_t = -\mathrm{D}V_t(X_t)$. The noise adds the Itô term to the Hamilton–Jacobi equation, which becomes the Polchinski equation (7.8).

In the notation of Theorem 7.6 the controlled field is $\mathrm{d}X = \dot{Q}_t^{1/2}(u\,\mathrm{d}t + \mathrm{d}W)$ and the optimal control is $u_t = \dot{Q}_t^{1/2} p_t$ (Exercise 7.3). Since p_t is the value F_t of the force, the FBSDE (7.7) is the stochastic Pontryagin system of this problem:

the field Φ is the state, and the force F is the costate.

7.3 Filtering: measuring the field

Stochastic localisation was introduced by Eldan [Eld13] to study measures in high dimension (see [Eld23] for a survey). The idea is to write a measure as a mixture of simpler ones, through a measure-valued martingale driven by noise. Proposition 7.2 says that the Polchinski process does exactly this for ν_T. What Eldan's construction adds to the picture of Section 7.1 is a statistical reading:

the Polchinski process is a noisy observation of its own terminal value, and its drift is what an optimal filter learns from this observation.

The terminal field as a hidden signal. Assume that all the deterministic covariance operators are functions of A, hence commute, and let $K = Q_T^{-1}$, $J_t = KQ_t$, so that $J_0 = 0$, $J_T = 1$. For the heat profile, $K = Ae^{A/T^2}$ and $J_t = \mathrm{e}^{-A(t^{-2}-T^{-2})}$ is a heat semigroup, which tends to e^{-A/t^2} as $T \to \infty$. Given its terminal value, the interacting field is still the Gaussian bridge, as the density of $\mathbb{P}$ with respect to $\mathbb{P}_0$ depends on X_T only:

$$\Phi_t = J_t\Phi_T + \zeta_t, \quad \zeta_t \sim N(0, Q_t(1-J_t)), \quad \text{independent of } \Phi_T. \tag{7.14}$$

Read (7.14) as a statistical model. The *signal* Φ_T is hidden, with *prior* ν_T, and the path $(\Phi_s)_{s\leqslant t}$ is an *observation* of it: a copy attenuated by J_t and corrupted by a Gaussian noise independent of the signal. Infinitesimally, the bridge is the Doob transform of X by the Gaussian density of the remaining increment $\Phi_T - \Phi_t \sim \mu^{S_t}$, so that

$$\mathrm{d}\Phi_t = \dot{Q}_t S_t^{-1}(\Phi_T - \Phi_t)\,\mathrm{d}t + \dot{Q}_t^{1/2}\,\mathrm{d}B_t, \tag{7.15}$$

with B a Brownian motion independent of Φ_T, in the filtration enlarged by $\sigma(\Phi_T)$: at each resolution the field is pulled towards the hidden terminal field, and the pull is seen through noise. The *filtering problem* is to estimate Φ_T from $\mathcal{F}_t = \sigma(\Phi_s : s \leqslant t)$, the filtration of the Polchinski process itself. Let $\psi_t = \mathbb{E}_t[\Phi_T]$, $\mathrm{Cov}_t = \mathrm{Cov}(\Phi_T|\mathcal{F}_t)$, and let

$$\mathcal{R}_t := S_t^{-1} - Q_T^{-1}, \qquad \mathcal{R}_0 = 0, \qquad \dot{\mathcal{R}}_t = S_t^{-1}\dot{Q}_t S_t^{-1} \geqslant 0,$$

be the **regulator**.

Proposition 7.7. (Filtering, formal; a theorem at finite cutoff)

a) *The conditional law of* Φ_T *given* $\mathcal{F}_t$ *is the Bayesian posterior*

$$\mu_t^{\Phi_t}(\mathrm{d}y) \propto \mathrm{e}^{\langle S_t^{-1}\Phi_t, y\rangle - \frac{1}{2}\langle y, \mathcal{R}_t y\rangle}\nu_T(\mathrm{d}y). \tag{7.16}$$

It depends on the path only through Φ_t*, a sufficient statistic, and* $\mathcal{R}_t$ *is the precision gained on* Φ_T*.*

b) *The Brownian motion* W *of the FBSDE* (7.6) *is the* ***innovation*** *of the filter,*

$$W_t = B_t + \int_0^t \dot{Q}_s^{1/2} S_s^{-1}(\Phi_T - \psi_s)\,\mathrm{d}s,$$

and the force is the posterior mean of the pull in (7.15)*, per unit* $\dot{Q}_t$*:*

$$F_t = \mathbb{E}_t[S_t^{-1}(\Phi_T - \Phi_t)], \qquad \textit{i.e.} \quad \psi_t = \Phi_t + S_t F_t.$$

So (7.6) *is the observation equation* (7.15) *written in its own filtration.*

c) *Every observable* $O_t = \mathbb{E}_t[G(\Phi_T)]$ *solves the filter equation*

$$\mathrm{d}O_t = \langle \mathrm{Cov}_t(G(\Phi_T), \Phi_T), S_t^{-1}\dot{Q}_t^{1/2}\,\mathrm{d}W_t\rangle,$$

in particular

$$\mathrm{d}\psi_t = \mathrm{Cov}_t S_t^{-1}\dot{Q}_t^{1/2}\,\mathrm{d}W_t.$$

Proof. (a) By Proposition 7.2 the conditional law is $\mu_t^{\Phi_t}$, the tilt of $\mu^{S_t}(\cdot - \Phi_t)$ by e^{-V_T}. The ratio of the Gaussian densities of $\mu^{S_t}(\cdot - x)$ and μ^{Q_T} at y is $\mathrm{e}^{\langle S_t^{-1}x, y\rangle - \frac{1}{2}\langle y, \mathcal{R}_t y\rangle}$ up to a function of x; equivalently, apply Bayes' formula to the likelihood of (7.14). (b) Integrate (7.6) from t to T and condition on $\mathcal{F}_t$: since F is a martingale, $\mathbb{E}_t\int_t^T \dot{Q}_s F_s\,\mathrm{d}s = S_t F_t$, so $\psi_t = \Phi_t + S_t F_t$, and the posterior mean of the drift of (7.15) is $\dot{Q}_t S_t^{-1}(\psi_t - \Phi_t) = \dot{Q}_t F_t$. Comparing the two expressions (7.6) and (7.15) of $\mathrm{d}\Phi_t$ gives the formula for W. (c) $O_t = G_t(\Phi_t)$ with $G_t(x) = \mu_t^x(G)$ is a martingale, so the drift in Itô's formula vanishes, and differentiating (7.16) in x gives $\mathrm{D}G_t(x) = S_t^{-1}\mathrm{Cov}_t(\Phi_T, G(\Phi_T))$. □

So the Polchinski process is a filter. The interacting measure is the prior on a hidden field; at resolution t one has observed the coarse field Φ_t, and $\mu_t^{\Phi_t}$ is what is known about the fine one. The forecast ψ_t is the posterior mean of the hidden field, and $\dot{Q}_t F_t$ is the posterior mean of its pull on the observed one. For the Gaussian prior μ^{Q_T} the increments are independent, so the posterior mean is Φ_t itself and $F = 0$: Φ_t is the estimate that ignores the interaction and ψ_t the one that includes it. The effective potential is a likelihood ratio: by the proof of Theorem 7.3, $Z_T^{-1}\mathrm{e}^{-V_t(\Phi_t)}$ is the likelihood ratio of the observation up to resolution t under the interacting and under the Gaussian prior, and the Polchinski equation is the equation of its logarithm.

Proposition 7.8. (Backward in resolution) (formal) *Under* $\mathbb{P}$ *the reversed process* $\hat{\Phi}_s := \Phi_{T-s}$*,* $s \in [0, T]$*, solves the linear equation*

$$\mathrm{d}\hat{\Phi}_s = -\dot{Q}_{T-s}Q_{T-s}^{-1}\hat{\Phi}_s\,\mathrm{d}s + \dot{Q}_{T-s}^{1/2}\,\mathrm{d}\hat{W}_s, \qquad \hat{\Phi}_0 = \Phi_T \sim \nu_T, \tag{7.17}$$

with $\hat{W}$ a Brownian motion independent of Φ_T, and its solution is (7.14). Backward in resolution, the Polchinski process is a linear Markov process, with Gaussian transition kernels which do not depend on the interaction, started from the interacting measure. It is not a Gaussian process: its marginals ν_t are not Gaussian.

Proof. Under $\mathbb{P}_0$, X is a centred Gaussian process with independent increments, so its reversal is a Gaussian Markov process: for $t<u$, the law of X_t given $(X_r)_{r\geqslant u}$ is $N(Q_tQ_u^{-1}X_u, Q_t - Q_tQ_u^{-1}Q_t)$, and letting $u\downarrow t$ gives the drift and the diffusion of (7.17). Since $\mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{P}_0$ is a function of X_T, the initial value of the reversed process, these transition kernels are the same under $\mathbb{P}$, and only the initial law changes, from μ^{Q_T} to ν_T. □

Generation by denoising. So two dynamics have the same marginals ν_t. Backward in resolution, the interacting field is progressively blurred by Gaussian noise, $\Phi_t = J_t\Phi_T + \zeta_t$, until nothing is left at $t=0$; this dynamics does not know the interaction. Forward, the Polchinski process undoes the blurring, and all the interaction sits in its drift. For $0<t\leqslant T$, at finite cutoff, the time-reversal formula for diffusions says that the forward drift is minus the backward one plus $\dot{Q}_t$ times the score of the marginal: if p_t is the density of ν_t,

$$\dot{Q}_tF_t = \dot{Q}_tQ_t^{-1}\Phi_t + \dot{Q}_t\mathrm{Dlog}\,p_t(\Phi_t), \qquad \text{i.e.} \quad F_t = \mathrm{Dlog}\,p_t(\Phi_t) + Q_t^{-1}\Phi_t:$$

the force is the score of ν_t relative to that of the Gaussian reference μ^{Q_t}, as $\nu_t \propto \mathrm{e}^{-V_t}\mu^{Q_t}$ requires. The forecast is the optimal denoiser, and Tweedie's formula

$$\psi_t = J_t^{-1}(\Phi_t + Q_t(1-J_t)\mathrm{Dlog}\,p_t(\Phi_t))$$

is again $\psi_t = \Phi_t + S_tF_t$. This is the structure of score-based generative models (diffusion models), with ν_T as the data distribution, the blurring as the noising process and the Polchinski process as the generating dynamics.

Remark 7.9. (Time reversal compensates the drift up to the score) Forward, in the direction of increasing resolution, Φ has drift $b_t = \dot{Q}_tF_t$, which contains the whole interaction. Backward, by (7.17), the reversed process runs in the direction of decreasing resolution and has at $s=T-t$ the drift $\beta_t(x) = -\dot{Q}_tQ_t^{-1}x$, which does not see the interaction at all. Written as a velocity in the direction of increasing resolution, the backward drift is $b_* = -\beta$. The two velocities b and b_* do not agree, and the time-reversal formula [And82, HP86] says what separates them:

$$b_t(x) - b_*(x) = b_t(x) + \beta_t(x) = \dot{Q}_t\mathrm{Dlog}\,p_t(x),$$

$\dot{Q}_t$ times the score of the density p_t of the field at that scale. Reversing the direction of resolution compensates the drift exactly up to the score. For the free field $b_t=0$ and $\mathrm{Dlog}\,p_t(x) = -Q_t^{-1}x$, so the whole backward drift is the Gaussian score; with an interaction, the score carries everything that the backward dynamics ignores. In Nelson's stochastic mechanics [Nel67] b_* is the backward velocity, and the half difference $\frac{1}{2}(b-b_*) = \frac{1}{2}\dot{Q}_t\mathrm{Dlog}\,p_t$ is the *osmotic velocity*. The same score is the noise of the field equation of Lecture 5, the observable $\omega_t = K\Phi_t - J_tF_t$ of Section 9.3 there: by (9.14) of Lecture 5,

$$\omega_t = -J_t\mathrm{Dlog}\,p_t(\Phi_t) = -J_t\dot{Q}_t^{-1}(b_t+\beta_t)(\Phi_t),$$

so the noise at resolution t is, up to the smoothing J_t, the part of the drift that time reversal does not compensate.

Remark 7.10. (The measurement coordinate and the literature) In the literature on stochastic localisation [Eld13, Eld23, BBD24] the observation is written in another coordinate. Let $j_t := S_t^{-1}\Phi_t$. Since $S_t^{-1}J_t = \mathcal{R}_t$, (7.14) becomes a linear measurement of the terminal field with signal-to-noise ratio $\mathcal{R}_t$:

$$j_t = \mathcal{R}_t\Phi_T + N_t, \quad N_t = S_t^{-1}\zeta_t, \quad \mathbb{E}[N_s \otimes N_t] = \mathcal{R}_{s\wedge t}, \quad N \text{ independent of } \Phi_T. \tag{7.18}$$

In this coordinate (7.16) is an exponential family with natural parameter j_t and log-partition function

$$\mathcal{K}_t(j) = \log \int \mathrm{e}^{\langle j,y\rangle - \frac{1}{2}\langle y, \mathcal{R}_t y\rangle} \nu_T(\mathrm{d}y),$$

and

$$\mathcal{K}_t(S_t^{-1}x) = \frac{1}{2}\langle x, S_t^{-1}x\rangle - V_t(x)$$

up to a function of t. Then $\psi_t = \mathrm{D}\mathcal{K}_t(j_t)$, $\mathrm{Cov}_t = \mathrm{D}^2\mathcal{K}_t(j_t)$, $\mathrm{d}j_t = \dot{\mathcal{R}}_t\psi_t\,\mathrm{d}t + \dot{\mathcal{R}}_t^{1/2}\,\mathrm{d}W_t$ with the same innovation W, and

$$\partial_t\mathcal{K}_t + \frac{1}{2}\mathrm{Tr}(\dot{\mathcal{R}}_t\mathrm{D}^2\mathcal{K}_t) + \frac{1}{2}\langle \mathrm{D}\mathcal{K}_t, \dot{\mathcal{R}}_t\mathrm{D}\mathcal{K}_t\rangle = 0, \tag{7.19}$$

the Polchinski equation in the variable j [BBD24] ($\partial_t\mathcal{K} = -\frac{1}{2}\mathbb{E}\langle y, \dot{\mathcal{R}}y\rangle$ and $\mathrm{Tr}(\dot{\mathcal{R}}\mathrm{D}^2\mathcal{K}) = \mathbb{E}\langle y, \dot{\mathcal{R}}y\rangle - \langle\psi, \dot{\mathcal{R}}\psi\rangle$, expectations under the posterior). The basic form of Eldan's scheme is the profile $S_t = (Q_T^{-1} + t)^{-1}$, $t \in [0,\infty)$ (the terminal resolution is then $t = \infty$, and Q_T is the cutoff covariance), where $\mathcal{R}_t = t$ and $j_t = t\Phi_T + N_t$ with N a Brownian motion; Eldan's original construction [Eld13] uses an adapted control matrix. After the change of variable $q = q_0 + t^2$, and with Q_T replaced by the full covariance C, this is the mass-regulator profile of Exercise 8.8, used in Lectures 6–7.

Literature: Polchinski, Eldan and diffusion models. The objects of this section have been studied in three communities. Stochastic localisation was introduced by Eldan [Eld13] for the KLS conjecture and is surveyed in [Eld23]; Chen and Eldan [CE22] made it a general framework, the localisation schemes, for mixing bounds of Markov chains. That Eldan's process and the Polchinski flow of the renormalisation group are the same object in different coordinates is explained in the survey of Bauerschmidt, Bodineau and Dagallier [BBD24], who use the Polchinski flow to prove log-Sobolev inequalities. On the generative side, diffusion models were introduced in [SWMG15, HJA20] and given their SDE form, a noising diffusion reversed in time by a drift built from the score, in [SSK21]; the time-reversal formula goes back to Anderson [And82] and Haussmann and Pardoux [HP86], and Tweedie's formula for the posterior mean is discussed in [Efr11]. The two sides were connected by El Alaoui, Montanari and Sellke [EMS22], who run stochastic localisation as a sampling algorithm, by El Alaoui and Montanari [AM21], who read it as the observation of a signal in Gaussian noise (the filtering point of view of Proposition 7.7), and by Montanari [Mon23], who presents denoising diffusions as instances of stochastic localisation. On the physics side, Cotler and Rezchikov showed that Polchinski's exact renormalisation group is a gradient flow of the relative entropy in the sense of optimal transport [CR23], and proposed to read diffusion models as inverse renormalisation group flows [CR23b].

7.4 Where the three views meet

The three constructions give one process. Renormalisation produces it as the field at resolution t with its effective drift (Theorem 7.3), control as the optimal state (7.11), and filtering as the observation written in its own filtration (Proposition 7.7). Each of them has a name for each object:

	Renormalisation	**Control**	**Filtering**
t	resolved covariance Q_t	time of the problem	information gained $\mathcal{R}_t$
Φ_t	field at resolution t	state	observation; posterior mean for the Gaussian prior
F_t	effective force	costate	posterior mean of the terminal force
W	fresh fluctuation of the shell	noise of the controlled field	innovation
S_t	unresolved covariance	free motion from t to T	posterior covariance for the Gaussian prior
$\psi_t = \Phi_t + S_t F_t$	mean of the conditional law $\mu_t^{\Phi_t}$	terminal state predicted by free motion	posterior mean
Cov_t	covariance of $\mu_t^{\Phi_t}$	$S_t - S_t \mathrm{D}^2 V_t S_t$	posterior covariance
V_t	effective potential	value function	minus the log-likelihood ratio of the observation, up to $\log Z_T$

In the last column, Φ_t is the estimate of the terminal field that ignores the interaction, and ψ_t is the estimate that includes it. Their difference is $S_t F_t$: the force is what the interaction adds to the estimate, in units of the free propagator S_t. The conditional covariance Cov_t replaces S_t as the propagator. Forces, propagators and covariances are the common language of the three views, and the next section writes the system in it. Only the last row of the table, and the formula for Cov_t in the control column, involve a potential.

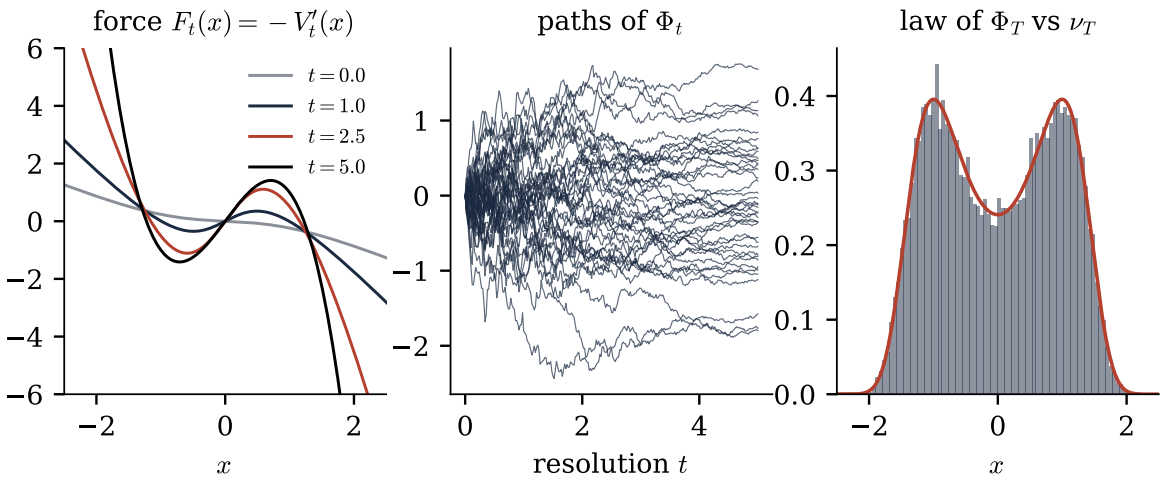


Figure 7.2. A zero-dimensional Polchinski process (Exercise 7.2): $Q_t = 1 - \mathrm{e}^{-t}$, $T = 5$, $V_T(x) = \frac{1}{2}x^4 - \frac{3}{2}x^2$. Left: the effective force $F_t = -V_t'$ is almost linear at low resolution and develops the double-well structure as $t \to T$. Centre: paths of the FBSDE (7.6), which split between the two wells (stochastic localisation). Right: histogram of Φ_T ($2 \cdot 10^4$ paths) against the density of ν_T.

Exercises

Exercise 7.1. (Quadratic interactions) Let $V_T(\varphi) = \frac{\kappa}{2}\langle\varphi,\varphi\rangle$, $\kappa > 0$. (a) Show that $V_t(\varphi) = \frac{1}{2}\langle\varphi, \kappa(1+\kappa Q_{T,t})^{-1}\varphi\rangle +$ const solves (7.8). (b) Show that ν_T is Gaussian with covariance $(Q_T^{-1} + \kappa)^{-1}$: for $Q_T \to (m^2 - \Delta)^{-1}$ this is the GFF of mass $(m^2 + \kappa)^{1/2}$. (c) Write (7.6) explicitly. It is a linear FBSDE. Solve it, and check that F_t is a martingale. *Moral:* the mass can be put either into the covariance or into the force.

Exercise 7.2. (Zero dimensions) Take M a point, $Q_t = q(t)$ scalar, increasing, $V_T(x) = \lambda x^4 - a x^2$ with $\lambda > 0$, $a \geqslant 0$ (a double well if $a > 0$; Figure 7.2 has $\lambda = \frac{1}{2}$, $a = \frac{3}{2}$). (a) Write (7.8) and solve it numerically for $q(t) = 1 - \mathrm{e}^{-t}$. (b) Simulate (7.6) by first computing V_t' on a grid, and compare the histogram of Φ_T with $\mathrm{e}^{-\lambda x^4 + a x^2 - x^2/2q(T)}$. (c) Compute V_t to second order in λ.

Exercise 7.3. (Stochastic maximum principle) (a) For an optimal control u of (7.9), put $p_t := \mathbb{E}_t[-\mathrm{D}V_T(X_T^u)]$, so that $u_t = \dot{Q}_t^{1/2} p_t$ by (7.11). Use the martingale representation theorem to show that p solves the backward equation $\mathrm{d}p_t = \mathrm{Y}_t \mathrm{d}W_t$, $p_T = -\mathrm{D}V_T(X_T^u)$, for an adapted Y, coupled to the forward equation of X^u, and identify (X^u, p) with (Φ, F) in (7.6). (b) Show that $\mathrm{Y}_t = \mathrm{D}F_t(\Phi_t)\,\dot{Q}_t^{1/2}$.

Exercise 7.4. **(Optimal drift)** (a) Check from (7.11) that the optimal u is a martingale up to the factor $\dot{Q}^{1/2}$. (b) For the quadratic interaction of Exercise 7.1, compute the optimal drift and the value (7.9) explicitly.

Exercise 7.5. **(Cost of the leading drift)** For $V_T = \frac{\lambda}{12}\int :\varphi^4:_T$ on $\mathbb{T}^3_L$, show that

$$\frac{1}{2}\mathbb{E}\int_0^T \|\frac{\lambda}{3}\dot{Q}_s^{1/2}:X_s^3:\|^2 \mathrm{d}s = \frac{\lambda^2}{12}L^3\int Q_T(z)^4\,\mathrm{d}z,$$

which diverges linearly in T. Relate this to the $k=4$ term of (A.6) of Lecture 5.

8 Force coordinates and potentials

The three views agree on the forward–backward system (7.7), which from now on is the primary object, with a martingale force $F_t = F_t(\Phi_t)$, $\mathrm{d}F_t = \Upsilon_t\,\mathrm{d}W_t$. We look for the coordinates in which it is simplest. They are suggested by the table above: the posterior mean ψ, which we call the **forecast**, and the force. They remove the predictable part of the forward dynamics, and they integrate exactly the bilinear (tree) part of the force flow.

Sections 8.1–8.4 tell this story in terms of forces, propagators and covariances. The potential enters only through the terminal force and, for a gradient force, through the identification of the response with the conditional covariance; the system in force coordinates makes sense for a force that is not a gradient. Sections 8.5–8.7 tell it again in terms of potentials: the value function generates the canonical transformation to force coordinates, the Legendre transform connects it to the filter, and the equation for the force is the Wetterich equation for the effective average action [Wet93, BTW02].

8.1 The force and a Burgers equation

Differentiating (7.8), the effective force $F_t = -\mathrm{D}V_t$, now as a function of the field, solves

$$\partial_t F_t = -\mathrm{D}F_t\,[\dot{Q}_t F_t] - \tfrac{1}{2}\mathrm{Tr}(\dot{Q}_t \mathrm{D}^2 F_t), \qquad F_T = -\mathrm{D}V_T, \tag{8.1}$$

an infinite-dimensional, backward, viscous *Burgers equation*. The Cole–Hopf transform $F = \mathrm{D}\log h$ linearises it, which is the statement that $h \propto \mathrm{e}^{-V}$ solves a heat equation. Two remarks on (8.1):

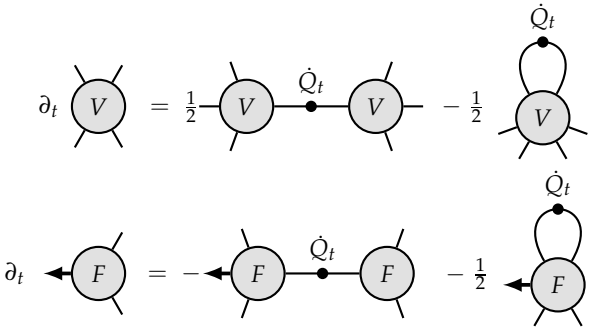


Figure 8.1. The Polchinski equation (7.8) and the Burgers equation (8.1) as graphs. A blob with n legs is the n-th derivative of V_t, or of the force F_t, whose root is the arrow; a line with a dot is the kernel $\dot{Q}_t$. The first term joins two blobs and makes a tree; the second closes a loop on one blob. Integrating from V_T and iterating generates the connected Feynman graphs, with lines $Q_{T,t}$.

i. it uses only the force, which is a local quantity (*intensive*), while V_t is *extensive*, of order of the volume. This matters in infinite volume (Section 9.5 of Lecture 5);

ii. the Burgers flows for different covariances commute when $\mathrm{D}F_t$ is symmetric, i.e. when F_t is a gradient [BCG26]: then $F_t = -\mathrm{D}V_t$, the flows are gradients of Polchinski flows, i.e. of Gaussian convolutions after the Cole–Hopf transform, and Gaussian convolutions commute. For a non-gradient force, as for the general Wilson–Itô diffusions of [BCG26], the flows do not commute.

The function $F_t(\cdot)$ with $F_t = F_t(\Phi_t)$ is the *decoupling field* of the FBSDE, and the Burgers equation (8.1) is its quasilinear PDE; $\mathrm{Y}_t = \mathrm{D}F_t(\Phi_t)\,\dot{Q}_t^{1/2}$ is the martingale representation integrand of the costate. The two terms of (8.1) have distinct origins. The transport term $-\mathrm{D}F\,[\dot{Q}F]$ comes from the drift of the state, and generates the tree diagrams. The diffusion term $-\frac{1}{2}\mathrm{Tr}(\dot{Q}\,\mathrm{D}^2F)$ is the Itô correction, and generates the loops.

8.2 Removing the predictable drift

The drift $\dot{Q}_t F_t$ of the state is predictable, and the costate is a martingale. Hence the drift that remains to be accumulated on $[t,T]$ is predicted, at resolution t, by transporting the current state with the current costate along the free flow (7.13).

Proposition 8.1. (The forecast) (formal) *The conditional mean of the terminal field is*

$$\psi_t := \mathbb{E}_t[\Phi_T] = \Phi_t + S_t F_t, \tag{8.2}$$

i.e. $(\psi_t, F_t) = \Sigma_t(\Phi_t, F_t)$. *It is a martingale, with*

$$\mathrm{d}\psi_t = \dot{Q}_t^{1/2}\mathrm{d}W_t + S_t\mathrm{Y}_t\,\mathrm{d}W_t = (1 + S_t\mathrm{D}F_t(\Phi_t))\,\dot{Q}_t^{1/2}\mathrm{d}W_t. \tag{8.3}$$

Proof. The identity is in the proof of Proposition 7.7(b), and uses no commutativity. For the differential, the drift $\dot{Q}_t F_t$ of Φ cancels against $\dot{S}_t F_t$, and $S_t\,\mathrm{d}F_t = S_t\mathrm{Y}_t\,\mathrm{d}W_t$. □

So ψ is a closed observable, the *field observable* of [BCG26]. The field splits as

$$\Phi_t = \psi_t - S_t F_t\text{:} \tag{8.4}$$

a martingale, minus the drift still expected on the unresolved scales, which vanishes at $t = T$. The noise of ψ in (8.3) has two sources: the fresh fluctuation of the shell, $\dot{Q}_t^{1/2}\mathrm{d}W_t$, and the revision of the forecast of the remaining drift, $S_t\,\mathrm{d}F_t$. The coupling between the forward and the backward equations, which was in the drift of Φ, now sits in the volatility of ψ. The coordinate ψ is the *interaction picture* with respect to the noiseless (classical) part of the dynamics.

8.3 The FBSDE in the forecast coordinate

Assume that the map

$$x \mapsto \psi = x + S_t F_t(x)$$

is invertible, and write its inverse as

$$x = \psi + S_t\,\mathcal{F}_t(\psi), \qquad \mathcal{F}_t(\psi) := -F_t(x), \tag{8.5}$$

so that $F_t = -\mathcal{F}_t(\psi_t)$: the costate is now a function of the forecast. Introduce the *response*, and recall the *regulator* of the filter:

$$H_t(\psi) := \big(S_t^{-1} + \mathrm{D}\mathcal{F}_t(\psi)\big)^{-1}, \qquad \mathcal{R}_t := S_t^{-1} - Q_T^{-1}, \qquad \dot{\mathcal{R}}_t = S_t^{-1}\,\dot{Q}_t\,S_t^{-1} \geqslant 0. \tag{8.6}$$

Differentiating the two maps $x \leftrightarrow \psi$ gives

$$1 + S_t\mathrm{D}F_t(x) = (1 + S_t\mathrm{D}\mathcal{F}_t(\psi))^{-1} = H_t(\psi)\,S_t^{-1}.$$

Theorem 8.2. (FBSDE in force coordinates) (formal) *In the coordinates* $(\psi, \mathcal{F})$*, the system* (7.7) *becomes*

$$\mathrm{d}\psi_t = H_t(\psi_t)\; S_t^{-1}\,\dot{Q}_t^{1/2}\,\mathrm{d}W_t, \qquad \mathcal{F}_t(\psi_t) = \mathbb{E}_t[\mathrm{D}V_T(\psi_T)], \tag{8.7}$$

with $\psi_0 = S_0 F_0 = \mathbb{E}[\Phi_T]$ *and* $\psi_T = \Phi_T$*. The forward equation has no drift, and its covariance is*

$$C_t := H_t\,\dot{\mathcal{R}}_t\,H_t^*.$$

The decoupling field $\mathcal{F}$ *solves the pure diffusion equation*

$$\partial_t \mathcal{F}_t + \tfrac{1}{2}\mathrm{Tr}\big[C_t\,\mathrm{D}^2\mathcal{F}_t\big] = 0, \qquad \mathcal{F}_T = \mathrm{D}V_T. \tag{8.8}$$

Proof. The forward equation is (8.3) with $1 + S_t\mathrm{D}F_t = H_t S_t^{-1}$, and its covariance is

$$H_t S_t^{-1}\,\dot{Q}_t\,S_t^{-1}H_t^* = H_t\,\dot{\mathcal{R}}_t H_t^*.$$

The backward equation is $F_t = \mathbb{E}_t[F_T]$ with $\psi_T = \Phi_T$. Since $\mathcal{F}_t(\psi_t)$ is a martingale and ψ has no drift, the drift in Itô's formula for $\mathcal{F}_t(\psi_t)$ is the left-hand side of (8.8). □

Compare (8.8) with the Burgers equation (8.1): the transport term has been absorbed exactly by the change of coordinates, and only the Itô term is left. If it is dropped, $\mathcal{F}_t = \mathrm{D}V_T$ does not move: all the tree diagrams are contained in the coordinate change (8.5). The loops are generated only by the quadratic variation C_t of the forecast, i.e. by the randomness of the noise and of the force revisions, and the nonlinearity has moved into the field dependence of C_t. None of this uses a gradient structure or a Gaussian convolution: Theorem 8.2 holds for any FBSDE of the form (7.7) with a Markov martingale force, gradient or not, for which the coordinate change is invertible. For a general terminal force $F_T = \mathfrak{f}(\Phi_T)$, replace $\mathrm{D}V_T$ by $-\mathfrak{f}$ in (8.7) and (8.8).

A one-dimensional model: Burgers and Lax–Oleinik. The passage to force coordinates is the passage from Eulerian to Lagrangian coordinates for Burgers' equation. Consider first the inviscid Burgers equation on $\mathbb{R}$,

$$\partial_s u + u\,\partial_x u = 0, \qquad u(0,x) = g'(x).$$

Along the characteristics $x = y + s\,g'(y)$ the velocity is constant, $u(s, y + s\,g'(y)) = g'(y)$: in the Lagrangian label y the transport term disappears, and nothing moves. As long as the characteristics do not cross, i.e. $1 + s\,g'' > 0$, the map $y \mapsto x$ is invertible, and $u = \partial_x G$ with the potential given by the *Lax–Oleinik formula* (or Hopf–Lax formula)

$$G(s,x) = \min_y \left\{ g(y) + \frac{(x-y)^2}{2s} \right\}, \qquad u(s,x) = \frac{x - y(s,x)}{s} = g'(y(s,x)),$$

whose minimiser $y(s,x)$ is the Lagrangian label of the point x. After the characteristics cross, the formula still selects one label, the entropy solution, and u has shocks. The viscous equation $\partial_s u + u\,\partial_x u = \varepsilon\,\partial_x^2 u$ is linearised by the Cole–Hopf transform,

$$G^\varepsilon(s,x) = -2\,\varepsilon \log \int \mathrm{e}^{-(g(y) + (x-y)^2/2s)/2\varepsilon}\,\mathrm{d}y,$$

and $G^\varepsilon \to G$ as $\varepsilon \to 0$ by Laplace's method.

The dictionary with this section is: the unresolved covariance S_t is the time s, the field x is the Eulerian point, the forecast ψ is the Lagrangian label y, $-F_t = \mathrm{D}V_t$ is the velocity u (and V_t is G), and V_T is the datum g. The shear (7.13), i.e. (8.5) at tree level, $x = \psi + S_t\mathrm{D}V_T(\psi)$, is the characteristic map; the noiseless Polchinski equation is the Hamilton–Jacobi equation $\partial_s G + \frac{1}{2}(\partial_x G)^2 = 0$, whose gradient solves the inviscid Burgers equation; the tree-level potential (8.17) is the Lax–Oleinik formula; and the condition for the characteristics not to cross, $S_t^{-1} + \mathrm{D}^2 V_T > 0$, is $1 + s\,g'' > 0$. The Polchinski equation is the viscous Hamilton–Jacobi equation, with $\varepsilon = \frac{1}{2}$ in the metric $\dot{Q}_t$, and (7.3) is the Cole–Hopf formula.

Theorem 8.2 is the Lagrangian description of the *viscous* problem. With noise the label is no longer a function of the Eulerian point: it is its posterior mean, the mean ψ of y under the Cole–Hopf measure $\propto \mathrm{e}^{-g(y)-(x-y)^2/2s}$, which is μ_t^x. In this label the velocity is no longer constant but a martingale, and its evolution (8.8) contains only the loops. The Lax–Oleinik form survives exactly: the Moreau formula (8.12) of Section 8.5 is the Lax–Oleinik formula with the datum V_T replaced by Γ_t, and at finite cutoff no shocks form: the noise keeps the label map invertible, locally by Proposition 8.3 and globally by the convexity argument of Section 8.5. Figure 8.3 shows the construction for a double well.

8.4 Conditional moments

For the Polchinski process the response has a direct meaning.

Proposition 8.3. (formal) *If $F_t = -\mathrm{D}V_t$ then*

$$H_t(\psi_t) = \mathrm{Cov}_t = \mathrm{Cov}^{\mathbb{P}}(\Phi_T \,|\, \Phi_t) = S_t - S_t\,\mathrm{D}^2 V_t(\Phi_t)\,S_t.$$

In particular $H_t > 0$ is symmetric, $1 + S_t\,\mathrm{D}F_t = \mathrm{Cov}_t S_t^{-1}$ is always invertible, and $H_t \leqslant S_t$ when V_T is convex.

Proof. By (7.5), the law of Φ_T given $\mathcal{F}_t$ depends only on Φ_t, so it is also the law given Φ_t. The conditional law μ_t^ϕ is the tilt of $\mu^{S_t}(\cdot - \phi)$ by e^{-V_T}; differentiating twice in ϕ gives $\mathrm{Cov}(\mu_t^\phi) = S_t - S_t\,\mathrm{D}^2 V_t(\phi)\,S_t$. On the other hand $H_t S_t^{-1} = 1 + S_t\,\mathrm{D}F_t = 1 - S_t\,\mathrm{D}^2 V_t$. The last claim is the Brascamp–Lieb inequality. □

So the forecast and the response are the first two conditional moments of the stochastic localisation $\mu_t^{\Phi_t}$ of Proposition 7.2: $(\psi_t, \mathrm{Cov}_t) = (\mathbb{E}[\Phi_T \,|\, \Phi_t], \mathrm{Cov}(\Phi_T \,|\, \Phi_t))$ (Figure 8.2). They are tied by Itô's formula: since $\mathbb{E}_t[\Phi_T \otimes \Phi_T] = \psi_t \otimes \psi_t + \mathrm{Cov}_t$ is a martingale,

$$\mathrm{d}\langle \psi, \psi \rangle_t = C_t\,\mathrm{d}t = \mathrm{Cov}_t \dot{\mathcal{R}}_t \mathrm{Cov}_t\,\mathrm{d}t, \qquad \mathrm{dCov}_t = -\mathrm{Cov}_t \dot{\mathcal{R}}_t \mathrm{Cov}_t\,\mathrm{d}t + \mathrm{d}(\text{martingale}). \tag{8.9}$$

Conditional variance is resolved at the rate $\mathrm{Cov}_t \dot{\mathcal{R}}_t \mathrm{Cov}_t$, which for the Gaussian reference is $S_t \dot{\mathcal{R}}_t S_t = \dot{Q}_t$.

Brascamp–Lieb and Cramér–Rao bounds. The conditional law μ_t^ϕ has density proportional to e^{-U}, with $U(y) = V_T(y) + \frac{1}{2}\langle y - \phi, S_t^{-1}(y - \phi)\rangle$ and $\mathrm{D}^2 U = S_t^{-1} + \mathrm{D}^2 V_T$. Two classical inequalities bound its covariance by the Hessian of U. If $\mathrm{D}^2 U > 0$, the Brascamp–Lieb inequality [BL76], $\mathrm{Var}_\mu(f) \leqslant \mathbb{E}_\mu \langle \mathrm{D}f, (\mathrm{D}^2 U)^{-1}\,\mathrm{D}f\rangle$, applied to linear functions $f = \langle h, \cdot \rangle$, and the Cramér–Rao inequality, which needs no convexity, give respectively the upper and the lower bound in

$$(\mathbb{E}_t[S_t^{-1} + \mathrm{D}^2 V_T(\Phi_T)])^{-1} \leqslant \mathrm{Cov}_t \leqslant \mathbb{E}_t[(S_t^{-1} + \mathrm{D}^2 V_T(\Phi_T))^{-1}].$$

The lower bound follows from $\mathbb{E}_\mu[(y - m) \otimes \mathrm{D}U(y)] = 1$ (integration by parts) and the Cauchy–Schwarz inequality, with the Fisher information $\mathbb{E}_\mu[\mathrm{D}U \otimes \mathrm{D}U] = \mathbb{E}_\mu \mathrm{D}^2 U$ (Exercise 8.9). The two bounds coincide when V_T is quadratic, and their gap is a Jensen gap for the inverse. Through Proposition 8.3 they are two-sided bounds on the Hessian of the effective potential, i.e. on the response H_t, which is the propagator of the Wetterich equation below. Some consequences:

- If V_T is convex, then $\mathrm{Cov}_t \leqslant S_t$: the interaction can only reduce the conditional variance. Equivalently $\mathrm{D}^2 V_t \geqslant 0$: convexity is preserved along the Polchinski flow (as also follows from the Prékopa–Leindler inequality).

- If $\mathrm{D}^2 V_T \geqslant \kappa$ with $\kappa > 0$, then $\mathrm{Cov}_t \leqslant (S_t^{-1} + \kappa)^{-1}$ uniformly in the conditioning: a uniform mass in the interaction is a uniform mass in every conditional law.
- For Φ_3^4 convexity is already lost through the Wick ordering: $H_4(\varphi; c_T) = \varphi^4 - 6\,c_T\,\varphi^2 + 3\,c_T^2$, so (7.1) has curvature $-\lambda\,c_T + 2\gamma_T$ at $\varphi = 0$, which diverges as $T \to \infty$ (this happens already for Φ_2^4). The upper bound is lost, and at low resolution the conditional law may spread over several wells (Figure 8.2). This is the large-field problem. In Lecture 6 it is handled by a clipped approximate force, whose global coercivity (Lemma 12.1 there) is a bound of Brascamp–Lieb type on the response.

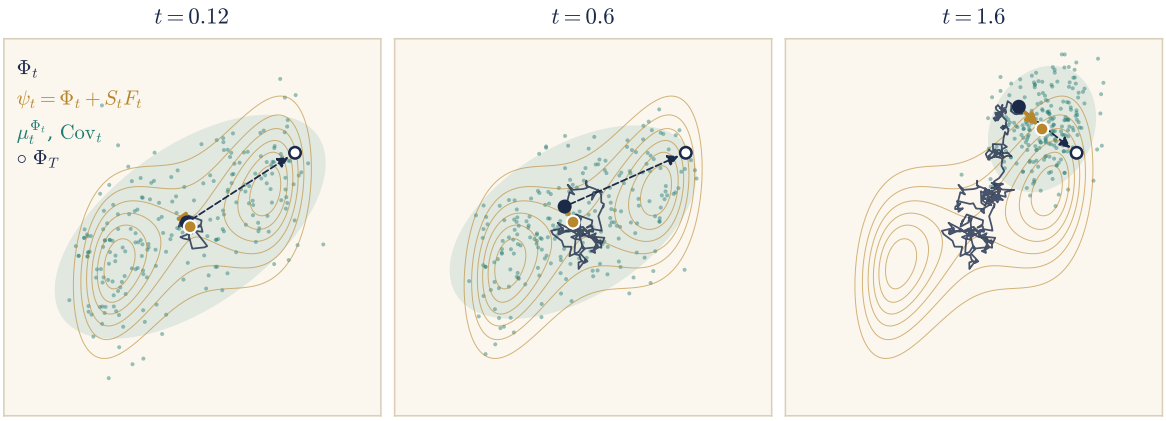


Figure 8.2. Stochastic localisation for a two-component field: $Q_t = (1 - \mathrm{e}^{-t})\,1$, $T = 4$, V_T a tilted double well, and the contours show the density of ν_T. One path of (7.6) up to resolution t (navy), with its terminal value Φ_T (open circle). The forecast $\psi_t = \Phi_t + S_t F_t$ (gold) is the mean of the conditional law $\mu_t^{\Phi_t}$: teal samples, and the ellipse is two standard deviations of Cov_t. At low resolution the conditional law covers both wells; as t grows it localises on one of them.

8.5 Potentials: the value function as a generator

We now use that the force is a gradient, $F_t = -\mathrm{D}V_t$, whose graph is a Lagrangian submanifold, and that V_t is the value function of Theorem 7.6. In Hamilton–Jacobi theory a value function is more than a cost: it generates a canonical transformation. A complete family of solutions of the Hamilton–Jacobi equation generates the transformation that trivialises the dynamics. Force coordinates are this transformation.

Phase space and Lagrangian graphs. Pairs (x, p) of a field and a value of the force form the phase space $\mathbb{R}^n \times \mathbb{R}^n$, with the symplectic form $\Omega = \sum_i \mathrm{d}x_i \wedge \mathrm{d}p_i$. A submanifold L of dimension n is *Lagrangian* if Ω vanishes on it. The graph $L = \{(x, F(x))\}$ of a force field is Lagrangian if and only if $\sum_{i,j} \partial_j F_i \,\mathrm{d}x_i \wedge \mathrm{d}x_j = 0$, i.e. $\mathrm{D}F$ is symmetric, i.e. F is (locally) a gradient: the gradient sector is the sector of Lagrangian graphs. A map that preserves Ω (a *symplectic* map) sends Lagrangian submanifolds to Lagrangian submanifolds. The shear (7.13) is symplectic, since $\Sigma_t^* \Omega = \Omega + \sum_{i,j} (S_t)_{ij}\,\mathrm{d}p_j \wedge \mathrm{d}p_i = \Omega$ for S_t symmetric. So it maps the graph of a gradient to a Lagrangian submanifold; if the image is again a graph over the new base variable, it is the graph of a gradient.

The free action generates the shear. The cheapest way to reach the field ψ at resolution T from the field x at resolution t, without interaction, is to minimise $\frac{1}{2}\int_t^T \|u_s\|^2\,\mathrm{d}s$ under the constraint $x + \int_t^T \dot{Q}_s^{1/2} u_s\,\mathrm{d}s = \psi$. The minimiser is $u_s = \dot{Q}_s^{1/2} p$, with the constant costate $p = S_t^{-1}(\psi - x)$, and the minimal cost is the **free action**

$$\mathcal{A}_t(x, \psi) = \frac{1}{2}\langle \psi - x, S_t^{-1}(\psi - x)\rangle. \tag{8.10}$$

The relation $\psi = x + S_t p$ is the shear (7.13). Since $\mathrm{D}_x \mathcal{A}_t = -p$ and $\mathrm{D}_\psi \mathcal{A}_t = p$, at fixed t

$$\mathrm{d}\mathcal{A}_t = \langle p,\, \mathrm{d}\psi\rangle - \langle p,\, \mathrm{d}x\rangle,$$

so $\mathcal{A}_t$ is the generating function of the canonical transformation Σ_t. Moreover

$$\partial_t \mathcal{A}_t = \frac{1}{2}\langle \mathrm{D}_x \mathcal{A}_t, \dot{Q}_t \mathrm{D}_x \mathcal{A}_t\rangle:$$

for every ψ it solves the noiseless Polchinski equation, so it is a complete integral of that equation, with parameter ψ. Jacobi's theorem then says that in the new coordinates the Hamiltonian $\mathcal{H}_*$ vanishes: ψ and p are constants of motion.

This is what force coordinates are:

(ψ_t, F_t) are the Jacobi coordinates of the free Hamiltonian of the field.

The noise does not destroy this: the constants of motion become martingales, ψ_t by Proposition 8.1 and F_t by Theorem 7.3. The only dynamics left is their quadratic variation.

The value function generates the graph. The graph of the force is Lagrangian, hence so is its image under the shear. The image is a graph over ψ: the map $x \mapsto x + S_t F_t(x)$ is S_t times the gradient of $\frac{1}{2}\langle x, S_t^{-1} x\rangle - V_t(x)$, which is strictly convex since its Hessian is $S_t^{-1}\mathrm{Cov}_t S_t^{-1} > 0$ (Proposition 8.3), so the map is injective. Hence the image is the graph of a gradient:

$$\mathcal{F}_t = \mathrm{D}\Gamma_t, \qquad H_t = (S_t^{-1} + \mathrm{D}^2\Gamma_t)^{-1}, \qquad \Gamma_T = V_T. \tag{8.11}$$

These two generating functions are related by the Moreau (infimal convolution) transform

$$V_t(x) = \inf_\psi \left\{\Gamma_t(\psi) + \tfrac{1}{2}\langle x - \psi, S_t^{-1}(x-\psi)\rangle\right\}, \tag{8.12}$$

whose optimality conditions $x = \psi + S_t \mathrm{D}\Gamma_t(\psi)$, $\mathrm{D}V_t(x) = \mathrm{D}\Gamma_t(\psi)$ are (8.5). It is well defined by the same convexity: as a function of $j = S_t^{-1} x$, the function $\frac{1}{2}\langle x, S_t^{-1} x\rangle - V_t(x)$ is the Legendre transform of $\Gamma_t + \frac{1}{2}\langle \cdot, S_t^{-1}\cdot\rangle$. The value function V_t itself need not be convex. We use (8.12) to fix the additive constant in Γ_t. Figure 8.3 shows the construction in zero dimensions.

At the minimiser, $V_t(x) = \Gamma_t(\psi) + \mathcal{A}_t(x, \psi)$: the value function is the cost Γ_t attached to the forecast, plus the cost of reaching the forecast by free motion. This is the rule for transporting a generating function along a canonical transformation:

$$\mathrm{d}V_t = \langle \mathrm{D}\Gamma_t,\, \mathrm{d}\psi\rangle + \langle p,\, \mathrm{d}\psi\rangle - \langle p,\, \mathrm{d}x\rangle = -\langle F_t,\, \mathrm{d}x\rangle,$$

since $\mathrm{D}\Gamma_t(\psi) = -F_t = -p$.

No convexity is needed. The construction uses no convexity of V_T, V_t or Γ_t. The convex function is $g_t(x) = \frac{1}{2}\langle x, S_t^{-1} x\rangle - V_t(x)$: with $j = S_t^{-1} x$ it is, up to a constant, the log-Laplace transform of the finite measure $\mathrm{e}^{-V_T(y) - \frac{1}{2}\langle y, S_t^{-1} y\rangle}\,\mathrm{d}y$, for every V_T bounded below. Its gradient in j is the conditional mean ψ, and it is strictly convex because $\mathrm{Cov}_t > 0$. In finite dimensions, when V_T is finite, this exponential family is regular with full support, so $j \mapsto \psi$ is a bijection of $\mathbb{R}^n$: the image of the shear is a graph over the whole ψ space, also when V_t is a double well.

For Γ_t one only gets semiconvexity, $S_t^{-1} + \mathrm{D}^2\Gamma_t = \mathrm{Cov}_t^{-1} > 0$, i.e. $\Gamma_t + \frac{1}{2}\langle \cdot, S_t^{-1}\cdot\rangle$ is strictly convex, which is what (8.12) requires. At $t = 0$, where $S_0 = Q_T$, this semiconvexity is the convexity of the effective average action of Section 8.6, Γ_0^{EAA}. Γ_t itself is convex if and only if $\mathrm{Cov}_t \leqslant S_t$, which holds for convex V_T but may fail for a double well at low resolution. Convexity matters only without noise: the tree-level potential (8.17) is merely semiconcave, and unless $S_t^{-1} + \mathrm{D}^2 V_T > 0$ the characteristics cross, producing caustics and shocks of the force. The noise removes them.

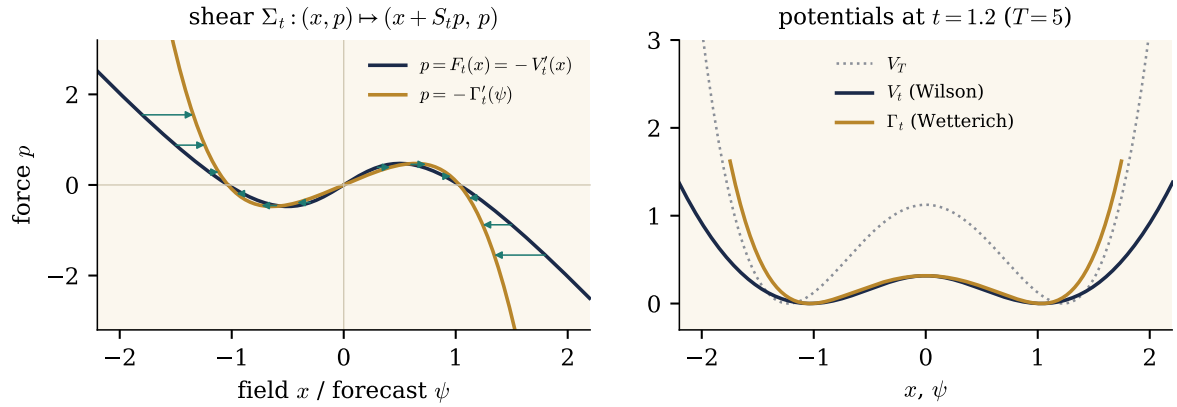


Figure 8.3. The zero-dimensional model of Figure 7.2 at $t = 1.2$. Left: the shear Σ_t (teal arrows) maps the graph of the force $F_t = -V_t'$ (navy) to a graph over the forecast ψ (gold), which is the graph of $-\Gamma_t'$. Right: the bare potential V_T, the effective potential V_t and Γ_t, each shifted to have minimum 0.

8.6 The Legendre transform and Wetterich's equation

Theorem 8.4. (Wetterich equation [Wet93]) (formal) Γ_t *solves*

$$\partial_t \Gamma_t = \tfrac{1}{2}\mathrm{Tr}\Big[\dot{\mathcal{R}}_t \big((S_t^{-1} + \mathrm{D}^2\Gamma_t)^{-1} - S_t\big)\Big], \qquad \Gamma_T = V_T. \tag{8.13}$$

Equivalently, the effective average action $\Gamma_t^{\mathrm{EAA}}(\psi) := \Gamma_t(\psi) + \frac{1}{2}\langle \psi, Q_T^{-1}\psi\rangle$ *solves, up to a field-independent term,*

$$\partial_t \Gamma_t^{\mathrm{EAA}} = \tfrac{1}{2}\mathrm{Tr}\Big[\dot{\mathcal{R}}_t\, (\mathrm{D}^2\Gamma_t^{\mathrm{EAA}} + \mathcal{R}_t)^{-1}\Big]. \tag{8.14}$$

It interpolates between the bare action

$$\Gamma_T^{\mathrm{EAA}} = \frac{1}{2}\langle \psi, Q_T^{-1}\psi\rangle + V_T(\psi)$$

($\mathcal{R}_T = \infty$) and the 1PI effective action Γ_0^{EAA} *of* ν_T *($\mathcal{R}_0 = 0$).*

Proof. Integrate (8.8) in field space. With $\mathcal{F} = \mathrm{D}\Gamma$, the derivative of $H = (S^{-1} + \mathrm{D}^2\Gamma)^{-1}$ in a direction v is $-H\ \mathrm{D}^3\Gamma[v]\ H$, so by the total symmetry of $\mathrm{D}^3\Gamma$,

$$\mathrm{D}_v \mathrm{Tr}(\dot{\mathcal{R}}_t H_t) = -\mathrm{Tr}(H_t\, \dot{\mathcal{R}}_t H_t\, \mathrm{D}^3\Gamma_t[v]) = -\mathrm{Tr}(C_t\, \mathrm{D}^2\mathcal{F}_t)[v].$$

Hence (8.8) reads $\mathrm{D}\big(\partial_t\Gamma_t - \frac{1}{2}\mathrm{Tr}(\dot{\mathcal{R}}_t H_t)\big) = 0$, and (8.13) holds up to a function of t alone. That this function is $-\frac{1}{2}\mathrm{Tr}(\dot{\mathcal{R}}_t S_t)$ with the normalisation (8.12) follows from the Polchinski equation by the envelope theorem (Exercise 8.4). For (8.14), note that $\mathrm{D}^2\Gamma_t^{\mathrm{EAA}} + \mathcal{R}_t = \mathrm{D}^2\Gamma_t + S_t^{-1}$. For the endpoints: $S_T = 0$; and at $t = 0$, $S_0 = Q_T$, so the formula for $\mathcal{K}_0$ in Remark 7.10 (the Cameron–Martin formula) identifies $\frac{1}{2}\langle x, Q_T^{-1}x\rangle - V_0(x) + V_0(0)$ with the cumulant generating function of ν_T at $Q_T^{-1}x$. By (8.12), its Legendre transform is $\Gamma_0^{\mathrm{EAA}} - V_0(0)$ (Exercise 8.6).

Near $t = T$ the two traces $\mathrm{Tr}(\dot{\mathcal{R}}_t(S_t^{-1} + \mathrm{D}^2\Gamma_t)^{-1})$ and $\mathrm{Tr}(\dot{\mathcal{R}}_t S_t) = \mathrm{Tr}(S_t^{-1}\dot{Q}_t)$ in (8.13) diverge separately (like $n/(T-t)$ for $Q_t = t$ in dimension n); only their difference, which is $-\mathrm{Tr}(\dot{Q}_t \mathrm{D}^2 V_t)$ by Exercise 8.4, is integrable. For the same reason the field-independent term dropped in (8.14) is not integrable up to T: Γ_t^{EAA} is defined up to an infinite constant. □

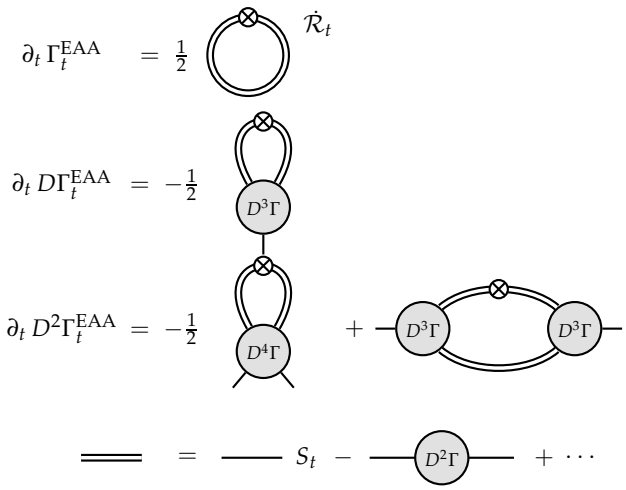


Figure 8.4. The Wetterich equation (8.14) as graphs. The double line is the full response $(\mathrm{D}^2\Gamma_t^{\mathrm{EAA}} + \mathcal{R}_t)^{-1} = (S_t^{-1} + \mathrm{D}^2\Gamma_t)^{-1}$ and the crossed circle is one insertion of $\dot{\mathcal{R}}_t$. There is exactly one loop: the tree term of Figure 8.1 has been absorbed by the Legendre transform. Each field derivative brings down a higher vertex, so the hierarchy does not close at a finite order. The bottom row expands the response around S_t.

Thus the Wetterich equation is the gradient-integrated form of the force equation (8.8), i.e. of the backward half of the FBSDE in forecast coordinates [Mor94, BCG26]. Every object in it is an observable of the Polchinski process:

Wetterich	FBSDE
field argument ψ	forecast $\psi_t = \mathbb{E}_t[\Phi_T]$, the state transported by Σ_t
$\mathrm{D}\Gamma_t(\psi_t)$	minus the costate, $-F_t = \mathbb{E}_t[\mathrm{D}V_T(\Phi_T)]$
propagator $(\mathrm{D}^2\Gamma_t^{\mathrm{EAA}} + \mathcal{R}_t)^{-1}$	conditional covariance Cov_t
regulator $\mathcal{R}_t = S_t^{-1} - Q_T^{-1}$	precision of the unresolved shell, minus that of the prior
$\partial_t \Gamma_t$	$\frac{1}{2}\mathrm{Tr}[\dot{\mathcal{R}}_t(\mathrm{Cov}_t - S_t)]$

The flow of Γ is driven by the difference between the interacting and the free conditional covariances, paired with the shell $\dot{\mathcal{R}}_t$. It is *one-loop exact*, with the full propagator inside the trace. In the Gaussian case $\mathrm{D}^2\Gamma_t$ is constant and only the constant term of Γ flows. For the heat profile with $T = \infty$, $S_t = C(1 - \mathrm{e}^{-A/t^2})$ and

$$\mathcal{R}_t = \frac{A}{\mathrm{e}^{A/t^2} - 1}, \tag{8.15}$$

which behaves like a mass term t^2 on the modes $|k| \ll t$ (for $m \ll t$) and is exponentially small on the modes $|k| \gg t$: the exponential regulator of the functional renormalisation group (Figure 8.5).

Proposition 8.5. (The Legendre transform, formal; a theorem at finite cutoff) *With $\mathcal{K}_t$ as in Remark 7.10, the Legendre transform of the log-partition function of the filter gives the effective action:*

$$\mathcal{K}_t^*(\psi) = \Gamma_t(\psi) + \frac{1}{2}\langle \psi, S_t^{-1}\psi \rangle = \Gamma_t^{\mathrm{EAA}}(\psi) + \frac{1}{2}\langle \psi, \mathcal{R}_t \psi \rangle$$

up to a function of t, and differentiating at fixed ψ gives the Wetterich equation (8.13) again.

Proof. By (8.12) and the formula for $\mathcal{K}$. At fixed ψ, with $j = \mathrm{D}\mathcal{K}^*(\psi)$, the envelope theorem and (7.19) give

$$\partial_t \mathcal{K}^* = \frac{1}{2}\mathrm{Tr}(\dot{\mathcal{R}}\mathrm{Cov}) + \frac{1}{2}\langle \psi, \dot{\mathcal{R}}\psi \rangle,$$

and $\partial_t \frac{1}{2}\langle\psi, S^{-1}\psi\rangle = \frac{1}{2}\langle\psi, \dot{\mathcal{R}}\psi\rangle$. Hence $\partial_t \Gamma_t = \frac{1}{2}\mathrm{Tr}(\dot{\mathcal{R}}\mathrm{Cov})$ up to a function of t, which is $-\frac{1}{2}\mathrm{Tr}(\dot{\mathcal{R}}S)$ with the normalisation (8.12). □

So one Lagrangian graph has three generators: V_t in the field x, $\mathcal{K}_t$ in the coordinate $j = S_t^{-1}x$ of Remark 7.10, and Γ_t in the forecast ψ. The Polchinski flow is the Hamilton–Jacobi equation of the log-partition function of the localisation; the Wetterich flow is its Legendre dual [Mor94], whose argument is the posterior mean, and whose regulator is the information gained. At $t = 0$, where $\mathcal{R}_0 = 0$, $\mathcal{K}_0$ is the generating function Λ of the cumulants of ν_T below.

Probabilistic meaning of Γ. At $t = 0$ the effective average action is the 1PI effective action of ν_T,

$$\Gamma_0^{\mathrm{EAA}}(\psi) = V_0(0) + \sup_j \{\langle j, \psi\rangle - \Lambda(j)\}, \qquad \Lambda(j) = \log \int \mathrm{e}^{\langle j,\varphi\rangle}\, \nu_T(\mathrm{d}\varphi),$$

the generating functional of the vertex functions introduced by Jona-Lasinio [JL64]. It is a probabilistic object, as Jona-Lasinio emphasised in his probabilistic reading of the renormalisation group [JL75, JL01]; for the link with large deviations see Section VIII of [JL01]. By Cramér's theorem the empirical mean of N independent samples of ν_T takes the value ψ with probability

$$\approx \exp\left(-N\left(\Gamma_0^{\mathrm{EAA}}(\psi) - V_0(0)\right)\right),$$

so $\Gamma_0^{\mathrm{EAA}} - V_0(0)$ is the rate function of the large deviations of averages of the field. In particular Γ_0^{EAA} is convex even when V_T is not, it is minimal at the mean of ν_T, and its Hessian there is the inverse covariance. Formally the same holds for the field averaged over a large volume, with the volume as the large parameter and the effective potential of quantum field theory (Γ per unit volume at constant ψ, not V_t) as rate function; in that limit flat parts may appear (the Maxwell construction).

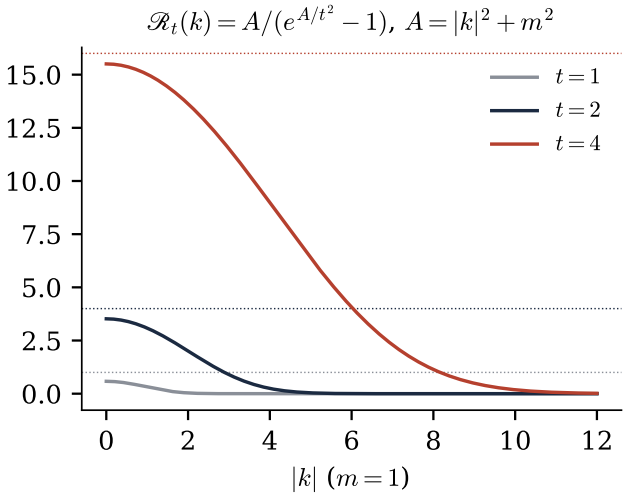


Figure 8.5. The regulator $\mathcal{R}_t$ of the heat profile ($T = \infty$, $m = 1$): close to the mass term t^2 (dotted) on the modes $|k| \ll t$, exponentially small on the modes $|k| \gg t$.

8.7 Variational meaning and the tree level

Constrained free energies. (formal) By the Gibbs variational principle applied to the conditional law (7.5), and the identity

$$\mathrm{Ent}(\nu|\mu^S(\cdot - x)) = \mathrm{Ent}(\nu|\mu^S(\cdot - \psi)) + \frac{1}{2}\langle x - \psi, S^{-1}(x - \psi)\rangle$$

for ν of mean ψ,

$$\Gamma_t(\psi) = \inf\Big\{ \int V_T \,\mathrm{d}\nu + \mathrm{Ent}\big(\nu \,|\, \mu^{S_t}(\cdot - \psi)\big) \colon \int \varphi \; \nu(\mathrm{d}\varphi) = \psi \Big\}. \tag{8.16}$$

So Γ_t, the effective average action up to the quadratic term $\frac{1}{2}\langle \psi, Q_T^{-1}\psi\rangle$, is the free energy of the unresolved shell constrained to the forecast ψ, and (8.12) is the minimisation over the forecast. Strictly, the Gibbs principle gives only (8.12) with Γ_t replaced by the right-hand side $\tilde\Gamma_t$ of (8.16). Since the mean is ψ, $\tilde\Gamma_t(\psi) + \frac{1}{2}\langle\psi, S_t^{-1}\psi\rangle = \inf\{\int V_T \mathrm{d}\nu + \mathrm{Ent}(\nu|\mu^{S_t})\colon \text{mean } \psi\}$, an infimum of a convex functional under a linear constraint, hence convex in ψ; so is $\Gamma_t + \frac{1}{2}\langle\psi, S_t^{-1}\psi\rangle$ (Proposition 8.5). Both are recovered from the same V_t by Legendre duality, and therefore $\tilde\Gamma_t = \Gamma_t$. In the language of Theorem 7.6: V_t is the value function of the control problem, and Γ_t is its Moreau transform, i.e. the value function expressed in the forecast coordinate, the interaction picture of the classical part of the dynamics.

The tree level. Dropping the Itô term in (8.8) gives $\Gamma_t = V_T$, and then (8.12) becomes the Hopf–Lax (Lax–Oleinik) formula

$$V_t^{\mathrm{tree}}(x) = \inf_y \Big\{ V_T(y) + \tfrac{1}{2}\langle y - x, S_t^{-1}(y-x)\rangle \Big\}, \tag{8.17}$$

the solution of the noiseless Polchinski equation $\partial_t V = \frac{1}{2}\langle \mathrm{D}V, \dot Q\, \mathrm{D}V\rangle$. It is the saddle point of the Gaussian convolution (7.3). Its minimiser solves the classical field equation $S_t^{-1}(y-x) + \mathrm{D}V_T(y) = 0$ in the shell $[t,T]$ with background x, and expanding it in V_T produces exactly the tree diagrams.

Exercises

Exercise 8.1. (Characteristics) (a) Show that the transport equation $\partial_t F + \mathrm{D}F\,[\dot Q F] = 0$ is solved by $F_t(x) = F_T(y)$, $y = x + S_t F_T(y)$, and that its characteristics are the trajectories of (7.13). (b) Verify that (8.17) solves $\partial_t V = \frac{1}{2}\langle \mathrm{D}V, \dot Q\, \mathrm{D}V\rangle$, and that it is the limit $\hbar \to 0$ of $-\hbar \log \int \mathrm{e}^{-V_T(x+\zeta)/\hbar} \mu^{\hbar S_t}(\mathrm{d}\zeta)$. (c) For $V_T(y) = \lambda y^4$ in zero dimensions, solve the equation for y as a power series in λ, and identify the coefficients as tree graphs.

Exercise 8.2. (Conditional moments) (a) Prove $\mathrm{Cov}(\mu_t^\phi) = S_t - S_t \mathrm{D}^2 V_t(\phi)\, S_t$. (b) Prove (8.9).

Exercise 8.3. (A symplectic shear) (a) Show that Σ_t preserves $\sum_i \mathrm{d}x_i \wedge \mathrm{d}p_i$ iff S_t is symmetric. (b) Show that the image of the graph of $-\mathrm{D}V_t$ under Σ_t is the graph of $-\mathrm{D}\Gamma_t$, with Γ_t given by (8.12). (c) For a non-gradient force, show that Theorem 8.2 still holds, but that $\mathcal{F}_t$ need not be a gradient.

Exercise 8.4. (The constant) Differentiate (8.12) in t at fixed x (envelope theorem), and use (7.8) to show that the quadratic terms cancel and $\partial_t \Gamma_t(\psi) = -\frac{1}{2}\mathrm{Tr}(\dot Q_t \mathrm{D}^2 V_t(x))$, where $x = \psi + S_t \mathrm{D}\Gamma_t(\psi)$ is the field whose forecast is ψ. Deduce (8.13) with the stated constant.

Exercise 8.5. (Gaussian check) For $V_T(\varphi) = \frac{\kappa}{2}\langle\varphi,\varphi\rangle$, show that $\Gamma_t(\psi) = \frac{\kappa}{2}\langle\psi,\psi\rangle + \text{const}$, and solve the FBSDE (8.7) explicitly. Compare with Exercise 7.1.

Exercise 8.6. (1PI effective action) (a) Let $\tilde{\mathcal{K}}(j) = \log \int \mathrm{e}^{\langle j,\varphi\rangle} \mathrm{e}^{-V_T(\varphi)} \mu^{Q_T}(\mathrm{d}\varphi)$, the unnormalised log-Laplace transform, so that $\tilde{\mathcal{K}} = \mathcal{K}_0 + \log Z_T$ with $\mathcal{K}_0$ as in Remark 7.10. Show that $\Gamma_0^{\mathrm{EAA}} = \tilde{\mathcal{K}}^*$, in agreement with $\Gamma_0^{\mathrm{EAA}} = \mathcal{K}_0^* + V_0(0)$ after Proposition 8.5, and that $\mathrm{D}^2\Gamma_0^{\mathrm{EAA}}(\psi)$ is the inverse covariance of the tilted measure with mean ψ. (b) Integrate (8.13) with $\mathrm{D}^2\Gamma_t$ replaced by $\mathrm{D}^2 V_T$, and obtain the one-loop formula $\Gamma_t \approx V_T + \frac{1}{2}\mathrm{Tr}\log(1 + S_t \mathrm{D}^2 V_T)$.

Exercise 8.7. (Trees versus loops) Using (8.12), show that $V_t = \Gamma_t - \frac{1}{2}\langle \mathrm{D}\Gamma_t, S_t \mathrm{D}\Gamma_t\rangle + O(\lambda^3)$ for the Φ^4 interaction (7.1). Deduce from Proposition A.2 of Lecture 5 that, to order λ^2, Γ_t is obtained from V_t by removing exactly the term $k=1$, the one-particle-reducible tree; the bubble ($k=2$), the sunset ($k=3$) and the vacuum term ($k=4$) are 1PI and remain. Where does the sunset come from in (8.13)?

Exercise 8.8. (Regulators) Derive (8.15). Compute $\mathcal{R}_t$ for the sharp Fourier profile, and for $\hat Q_t(k) = (|k|^2 + m^2)^{-1} - (|k|^2 + m^2 + t^2)^{-1}$ (whose kernels are not smooth, but the formulas make sense). In the last case show that $\mathcal{R}_t = t^2$ when $T = \infty$: a mass regulator.

Exercise 8.9. (**Brascamp–Lieb and Cramér–Rao**) Let $\mu(\mathrm{d}y) \propto \mathrm{e}^{-U(y)}\,\mathrm{d}y$ on $\mathbb{R}^n$, with mean m. (a) Show that $\mathbb{E}_\mu[(y-m)\otimes \mathrm{D}U] = 1$ and $\mathbb{E}_\mu[\mathrm{D}U \otimes \mathrm{D}U] = \mathbb{E}_\mu \mathrm{D}^2 U$, and deduce $\mathrm{Cov}_\mu \geqslant (\mathbb{E}_\mu \mathrm{D}^2 U)^{-1}$. (b) For $\mathrm{D}^2 U > 0$, prove the Brascamp–Lieb inequality for linear f by Hörmander's L^2 method: solve $Lg = f - \mathbb{E}_\mu f$ with $L = -\Delta + \langle \mathrm{D}U, D\rangle$ and use $\mathbb{E}_\mu[(Lg)^2] \geqslant \mathbb{E}_\mu\langle \mathrm{D}g, \mathrm{D}^2 U\, \mathrm{D}g\rangle$. (c) In zero dimensions, with $S_t = S$ and $V_T(y) = \lambda y^4$, compute both bounds numerically and compare with the conditional variance; repeat for the double well $V_T(y) = \frac{1}{2}y^4 - \frac{3}{2}y^2$ with $S < 1/3$, so that U stays convex. What goes wrong for $S \geqslant 1/3$?

Exercise 8.10. (**The free action**) (a) Prove (8.10), and that $\partial_t \mathcal{A}_t = \frac{1}{2}\langle \mathrm{D}_x \mathcal{A}_t, \dot{Q}_t \mathrm{D}_x \mathcal{A}_t\rangle$. (b) From the Moreau formula (8.12), recover the relations (8.5) by differentiating $V_t(x) = \Gamma_t(\psi) + \mathcal{A}_t(x,\psi)$.

References

[AM21] A. El Alaoui, A. Montanari. An information-theoretic view of stochastic localization. Preprint, arXiv:2109.00709, 2021.

[And82] B. D. O. Anderson. Reverse-time diffusion equation models. *Stochastic Process. Appl.* 12 (1982), 313–326.

[BBD24] R. Bauerschmidt, T. Bodineau, B. Dagallier. Stochastic dynamics and the Polchinski equation: an introduction. *Probab. Surveys* 21 (2024), 200–290.

[BCG26] I. Bailleul, I. Chevyrev, M. Gubinelli. Stochastic observables and Wilson–Itô diffusions. In preparation.

[BD98] M. Boué, P. Dupuis. A variational representation for certain functionals of Brownian motion. *Ann. Probab.* 26 (1998), 1641–1659.

[BG20] N. Barashkov, M. Gubinelli. A variational method for Φ^4_3. *Duke Math. J.* 169 (2020), 3339–3415.

[BG21] N. Barashkov, M. Gubinelli. The Φ^4_3 measure via Girsanov's theorem. *Electron. J. Probab.* 26 (2021), 1–29.

[BG23] N. Barashkov, M. Gubinelli. On the variational method for Euclidean quantum fields in infinite volume. *Probab. Math. Phys.* 4 (2023), 761–801.

[Bis73] J.-M. Bismut. Conjugate convex functions in optimal stochastic control. *J. Math. Anal. Appl.* 44 (1973), 384–404.

[BK87] D. C. Brydges, T. Kennedy. Mayer expansions and the Hamilton–Jacobi equation. *J. Stat. Phys.* 48 (1987), 19–49.

[BL76] H. J. Brascamp, E. H. Lieb. On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *J. Funct. Anal.* 22 (1976), 366–389.

[BTW02] J. Berges, N. Tetradis, C. Wetterich. Non-perturbative renormalization flow in quantum field theory and statistical physics. *Phys. Rep.* 363 (2002), 223–386.

[CE22] Y. Chen, R. Eldan. Localization schemes: a framework for proving mixing bounds for Markov chains. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, 2022, 110–122.

[CR23] J. Cotler, S. Rezchikov. Renormalization group flow as optimal transport. *Phys. Rev. D* 108 (2023), 025003.

[CR23b] J. Cotler, S. Rezchikov. Renormalizing diffusion models. Preprint, arXiv:2308.12355, 2023.

[Efr11] B. Efron. Tweedie's formula and selection bias. *J. Amer. Statist. Assoc.* 106 (2011), 1602–1614.

[Eld13] R. Eldan. Thin shell implies spectral gap up to polylog via a stochastic localization scheme. *Geom. Funct. Anal.* 23 (2013), 532–569.

[Eld23] R. Eldan. Analysis of high-dimensional distributions using pathwise methods. In *Proceedings of the ICM 2022*, EMS Press, 2023, 4246–4270.

[EMS22] A. El Alaoui, A. Montanari, M. Sellke. Sampling from the Sherrington–Kirkpatrick Gibbs measure via algorithmic stochastic localization. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, 2022.

[**FS06**] W.H. Fleming, H.M. Soner. *Controlled Markov Processes and Viscosity Solutions*. 2nd ed., Stochastic Modelling and Applied Probability 25, Springer, 2006.

[**HJA20**] J. Ho, A. Jain, P. Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems 33* (*NeurIPS 2020*), 2020. arXiv:2006.11239.

[**HP86**] U. G. Haussmann, É. Pardoux. Time reversal of diffusions. *Ann. Probab.* 14 (1986), 1188–1205.

[**JL01**] G. Jona-Lasinio. Renormalization group and probability theory. *Phys. Rep.* 352 (2001), 439–458.

[**JL64**] G. Jona-Lasinio. Relativistic field theories with symmetry-breaking solutions. *Nuovo Cimento* 34 (1964), 1790–1795.

[**JL75**] G. Jona-Lasinio. The renormalization group: a probabilistic view. *Nuovo Cimento B* 26 (1975), 99–119.

[**Mon23**] A. Montanari. Sampling, diffusions, and stochastic localization. Preprint, arXiv:2305.10690, 2023.

[**Mor94**] T. R. Morris. The exact renormalization group and approximate solutions. *Int. J. Mod. Phys. A* 9 (1994), 2411–2449.

[**MY99**] J. Ma, J. Yong. *Forward-Backward Stochastic Differential Equations and their Applications*. Lecture Notes in Math. 1702, Springer, 1999.

[**Nel67**] E. Nelson. *Dynamical Theories of Brownian Motion*. Princeton University Press, Princeton, 1967.

[**PBGM62**] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, E. F. Mishchenko. *The Mathematical Theory of Optimal Processes*. Interscience, 1962.

[**Pen90**] S. Peng. A general stochastic maximum principle for optimal control problems. *SIAM J. Control Optim.* 28 (1990), 966–979.

[**Pol84**] J. Polchinski. Renormalization and effective Lagrangians. *Nuclear Phys. B* 231 (1984), 269–295.

[**Sal99**] M. Salmhofer. *Renormalization: An Introduction*. Springer, 1999.

[**SSK21**] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, B. Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations* (*ICLR*), 2021. arXiv:2011.13456.

[**SWMG15**] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, S. Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning* (*ICML*), 2015. arXiv:1503.03585.

[**Wet93**] C. Wetterich. Exact evolution equation for the effective potential. *Phys. Lett. B* 301 (1993), 90–94.

[**YZ99**] J. Yong, X. Y. Zhou. *Stochastic Controls: Hamiltonian Systems and HJB Equations*. Springer, 1999.