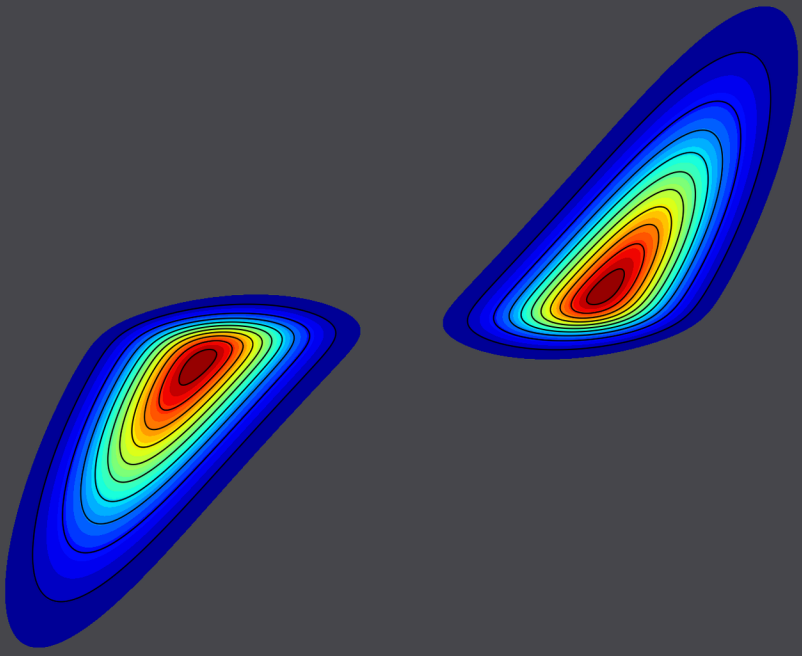


The Schrödinger coupling of two laws.

Conditioned Diffusions and Relative Entropy

Importance Sampling, Large Deviations and Schrödinger Bridges

Noa Ben-Dor

ODIN PRESS

Conditioned Diffusions and Relative Entropy

Copyright © 2026 by **Noa Ben-Dor**.
All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopying, recording, or otherwise—without the prior written permission of the publisher, except in the case of brief quotations embodied in critical reviews and certain other noncommercial uses permitted by copyright law.

Paperback version, First edition: 2026.
ISBN: forthcoming (Bowker allocation pending)
Trim Size: 6×9 inch. Paper color: White.

Algorithm copyright. All algorithms reproduced in this volume, in the chapter bodies and their pseudocode boxes, are the copyrighted intellectual property of **Noa Ben-Dor**. No reproduction, redistribution, adaptation or incorporation into any derivative work is permitted without the prior written permission of the copyright holder.

Published by Odin Press.
`https://odinpress.org`
This book was typeset in LaTeX by Odin Press from a manuscript submitted by the author.

9 8 7 6 5 4 3 2 1 0

Preface

This book began with three computations that turned out to be one. Over a few years I needed, for three unrelated reasons, a drift: the drift that turns a Brownian motion into a Brownian bridge, the drift that makes a rare event cheap to simulate, and the drift that carries one probability law to another in the sense Schrödinger gave the question in 1931. Each was found by its own method and written in its own notation. When I finally set them side by side, each was the logarithmic gradient of a positive solution of the backward heat equation. The three problems had different literatures. They had the same function.

The explanation is a single variational formula. Add a drift to a Brownian path; the price of the new law, measured in relative entropy against Wiener measure, is one half of the expected integral of the squared drift. The Boué–Dupuis representation says that the logarithm of an exponential moment of a functional of the path is an optimum over drifts of the expected functional plus that price. The optimal drift is the drift of a Doob h-transform, and this book is a slow reading of that one sentence in four of its forms.

Chapter 1 builds the instrument: path space and its densities, the stochastic exponential, Girsanov's theorem with the conditions that make it honest, and the identity between the relative entropy of a change of drift and its quadratic cost. It ends with weights that lie, likelihood ratios whose estimates look calm while their variance is infinite. Chapter 2 conditions a Brownian path on its endpoint, on staying positive and on a rare event, and shows that each conditioning is a drift whose price is minus the logarithm of the probability of the event. Chapter 3 proves the variational formula, identifies its optimal drift as an h-transform, and reads importance sampling as its equality case: under the optimal drift the estimator has zero variance. Chapter 4 lets the noise go to zero, and the formula becomes geometry: Schilder's theorem, the Freidlin–Wentzell action, the most likely

transition between the wells of a double-well potential, and Kramers' law for the time it takes. Chapter 5 asks for the cheapest path law with two prescribed marginals, Schrödinger's problem, finds an h-transform with two ends, computes it with Sinkhorn's iteration, and watches it tend to optimal transport as the noise vanishes.

Two habits run through the book. The first is that every theorem is proved, with its hypotheses stated where they bite: a representation proved for bounded functionals is not quoted for unbounded ones, and conditioning on an event of probability zero is treated as the limit it is. The second is that every formula is computed twice. The price of conditioning is computed as a probability and again as an integrated squared drift; the variational formula is evaluated from both of its sides; Kramers' law is checked against an exact quadrature and against simulation; Sinkhorn's iteration is compared with a closed form. The algorithms are written as pseudocode in the text, short enough to carry into any language.

The reader I have in mind knows Brownian motion, Itô's formula, martingales and stopping times, and the measure theory of conditional expectation. Large deviations and optimal transport are not assumed; what the book needs of them is built where it is used, and the results too long to prove here are stated precisely and cited.

The book is short on purpose. It follows one formula through four of its readings and stops where the formula stops.

Noa Ben-Dor

Contents

Preface vii

List of Figures xi

List of Tables xiii

1 Changing the Measure 1
- 1.1 Paths, Laws and Densities . . . 2
- 1.2 The Stochastic Exponential . . . 6
- 1.3 Girsanov's Theorem . . . 11
- 1.4 When the Density Is Honest . . . 18
- 1.5 The Price of a Drift . . . 26
- 1.6 Weights That Lie . . . 32

2 Conditioning Is a Drift 41
- 2.1 Space–Time Harmonic Functions . . . 42
- 2.2 Doob's h-Transform . . . 46
- 2.3 Bridges . . . 52
- 2.4 Conditioning on a Rare Event . . . 58
- 2.5 The Price of Conditioning . . . 66
- 2.6 Computing the Conditioned Path . . . 72

3 The Variational Formula 79
- 3.1 Gibbs' Variational Principle . . . 80
- 3.2 The Boué–Dupuis Representation . . . 85
- 3.3 The Optimal Drift Is an h-Transform . . . 89
- 3.4 Importance Sampling as the Equality Case . . . 98
- 3.5 Near-Optimal Drifts . . . 105
- 3.6 Computing the Formula . . . 110

4 Small Noise 117

4.1 The Laplace Principle 118
4.2 Schilder's Theorem 124
4.3 Freidlin–Wentzell 129
4.4 The Most Likely Transition 134
4.5 The Quasi-Potential and Kramers' Law 141
4.6 Computing the Most Likely Path 148

5 Bridges Between Laws **157**
5.1 The Schrödinger Problem 158
5.2 Two Potentials 164
5.3 Entropic Transport 171
5.4 Sinkhorn's Iteration 175
5.5 Entropic Interpolation 181
5.6 One Formula, Four Readings 188

References **195**

Index **207**

About the Author **209**

Chapter 1

Changing the Measure

What does it cost to make a random path do something it would rarely do on its own? This book answers that question in four settings and finds the same answer each time: a drift. Conditioning a Brownian path on an unlikely event bends it by a drift. The best importance sampler for that event bends it by the same drift. The most likely way for a weakly noisy system to climb out of a well, and the cheapest way to carry one cloud of diffusing particles into another, are drifts again. Every one of them is priced in a single currency, relative entropy, and every one of them solves one variational formula. The chapters that follow read that formula four ways.

This first chapter builds the instrument that makes a drift and a price the same object. It starts on path space, where the law of a process is compared with Wiener measure through a single density, and it shows that such a density, read along the path, is a martingale and is a stochastic exponential $\mathcal{E}(u)$. Girsanov's theorem then says what the density does to the path: it adds the drift u, and nothing else changes. Next come the hypotheses, stated plainly, under which a stochastic exponential is an honest density rather than a positive local martingale that has quietly lost mass.

The chapter delivers three things the whole book leans on: Girsanov's theorem together with the conditions that make it valid; the identity that prices a drift u at half its expected energy, $H(\mathbb{Q} \mid \mathbb{P}) = \frac{1}{2}\mathbb{E}_{\mathbb{Q}} \int_0^T u_s^2 \, ds$; and a measured account, by quadrature and by simulation, of the heavy tails that the likelihood-ratio weights of a strong drift grow. The six sections follow that order: paths and densities, the stochastic exponential, Girsanov's the-

orem, the conditions that make a density honest, the price of a drift, and the weights that lie.

1.1 Paths, Laws and Densities

Let $\Omega = C([0,T])$ be the space of continuous paths on a deterministic horizon $T > 0$: every law in this book lives on it, priced against Wiener measure.

Definition 1.1 (Path space and Wiener measure)**.** Path space is $\Omega = C([0,T])$, the continuous functions $\omega\colon [0,T] \to \mathbb{R}$ with the supremum norm. The canonical process is $W_t(\omega) = \omega(t)$, its natural filtration is $\mathcal{F}_t^0 = \sigma(W_s : s \le t)$, and $\mathcal{F}_T^0$ is the Borel σ-algebra of Ω. Wiener measure $\mathbb{P}$ is the unique probability on $(\Omega, \mathcal{F}_T^0)$ under which W is a standard Brownian motion started at 0: $W_0 = 0$, and for $0 \le s < t \le T$ the increment $W_t - W_s$ is independent of $\mathcal{F}_s^0$ with law $N(0, t-s)$.

A point of path space is a whole trajectory, and the canonical process only reads it; the randomness lives entirely in the measure. Under Wiener measure the paths are Brownian, and under another measure the same coordinate maps describe a different process. Existence and uniqueness of $\mathbb{P}$ are classical [1, 2]; uniqueness holds because the cylinder sets $\{W_{t_1} \in B_1, \dots, W_{t_n} \in B_n\}$ generate $\mathcal{F}_T^0$, are closed under intersection, and have probabilities fixed by Brownian motion.

For stochastic calculus under $\mathbb{P}$ we use the augmentation $(\mathcal{F}_t)_{t\in[0,T]}$ of $(\mathcal{F}_t^0)$ by the $\mathbb{P}$-null sets of $\mathcal{F}_T^0$. It satisfies the usual conditions, complete and right-continuous [3], and $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ with $\mathcal{F} = \mathcal{F}_T$ is the filtered probability space of the whole book. We write $\mathbb{E}$ for expectation under $\mathbb{P}$ and $\mathbb{E}_{\mathbb{Q}}$ under any other probability $\mathbb{Q}$. When we compare two laws that may be singular we compare them on $\mathcal{F}_T^0$, where both are defined; a law that is absolutely continuous with respect to $\mathbb{P}$ extends uniquely to $\mathcal{F}_T$, because every $\mathbb{P}$-null set is contained in a set of $\mathcal{F}_T^0$ that the law also gives mass zero.

Any process with continuous paths, on any probability space $(\Omega', \mathcal{F}', \mathbb{P}')$, is a random point X of Ω with law $\mathbb{P}' \circ X^{-1}$, and two such processes have the same law exactly when their finite-dimensional distributions agree. Every question about a process is therefore a question about a probability on one space, and the natural comparison with $\mathbb{P}$ is a density.

Definition 1.2 (Absolute continuity and the density process)**.** A probability $\mathbb{Q}$ on $(\Omega, \mathcal{F}_T^0)$ is absolutely continuous with respect to $\mathbb{P}$, written $\mathbb{Q} \ll \mathbb{P}$, if $\mathbb{P}(A) = 0$ implies $\mathbb{Q}(A) = 0$. Its Radon–Nikodym density is the nonnegative $\mathcal{F}_T$-measurable $Z = d\mathbb{Q}/d\mathbb{P}$ with $\mathbb{Q}(A) = \mathbb{E}[Z\mathbf{1}_A]$ for every $A \in \mathcal{F}_T$, and its density process is $Z_t = \mathbb{E}[Z \mid \mathcal{F}_t]$ for $t \in [0, T]$. The two laws are equivalent, $\mathbb{Q} \sim \mathbb{P}$, if each is absolutely continuous with respect to the other, and mutually singular, $\mathbb{Q} \perp \mathbb{P}$, if some $A \in \mathcal{F}_T^0$ has $\mathbb{P}(A) = 1$ and $\mathbb{Q}(A) = 0$.

Absolute continuity says that $\mathbb{Q}$ never charges what $\mathbb{P}$ ignores, and the Radon–Nikodym theorem turns it into a function: a density $Z \geq 0$ with $\mathbb{E}Z = 1$, unique up to null sets, with $\mathbb{E}_{\mathbb{Q}}[G] = \mathbb{E}[ZG]$ for every bounded or nonnegative $\mathcal{F}_T$-measurable G [1]. The density is a weight on paths: $\mathbb{Q}$ favours the paths where $Z > 1$, disfavours those where $Z < 1$, and abandons $\{Z = 0\}$. Reading Z only through what is known at time t gives the density process, and the first fact of the book is that this reading is a martingale.

Proposition 1.3 (The density process is a martingale)**.** Let $\mathbb{Q} \ll \mathbb{P}$ on $\mathcal{F}_T$ with density Z. The density process $(Z_t)_{t\in[0,T]}$ is a nonnegative $\mathbb{P}$-martingale with $\mathbb{E}Z_t = 1$ for every t, and for every $t \in [0, T]$ the restriction of $\mathbb{Q}$ to $\mathcal{F}_t$ is absolutely continuous with respect to the restriction of $\mathbb{P}$, with density

$$\frac{d\mathbb{Q}|_{\mathcal{F}_t}}{d\mathbb{P}|_{\mathcal{F}_t}} = Z_t.$$

Proof. Since $Z \geq 0$ and $\mathbb{E}Z = 1$, the conditional expectation $Z_t = \mathbb{E}[Z \mid \mathcal{F}_t]$ is well defined, nonnegative and integrable, and $\mathbb{E}Z_t = \mathbb{E}Z = 1$ by the tower property. For $s \leq t$ the tower property again gives $\mathbb{E}[Z_t \mid \mathcal{F}_s] = \mathbb{E}[\mathbb{E}[Z \mid \mathcal{F}_t] \mid \mathcal{F}_s] = \mathbb{E}[Z \mid \mathcal{F}_s] = Z_s$, and Z_t is $\mathcal{F}_t$-measurable by construction; hence (Z_t) is a martingale. For $A \in \mathcal{F}_t$ the indicator $\mathbf{1}_A$ is $\mathcal{F}_t$-measurable, so that

$$\mathbb{Q}(A) = \mathbb{E}[Z\mathbf{1}_A] = \mathbb{E}\big[\mathbb{E}[Z \mid \mathcal{F}_t]\mathbf{1}_A\big] = \mathbb{E}[Z_t\mathbf{1}_A].$$

If $\mathbb{P}(A) = 0$ the right side vanishes, so the restriction of $\mathbb{Q}$ to $\mathcal{F}_t$ is absolutely continuous with respect to that of $\mathbb{P}$, and since Z_t is $\mathcal{F}_t$-measurable the identity says that Z_t is its Radon–Nikodym density. Uniqueness of the density up to null sets of $\mathcal{F}_t$ completes the proof. □

Chapter 3

The Variational Formula

Chapter 1 put a price on every drift: a change of measure on path space that adds a drift costs half the drift's expected energy in relative entropy. Chapter 2 spent that currency on conditioning. The drift $\partial_x \log h$ of a positive space–time harmonic function h makes a Brownian path end above a level, stay away from zero or return to a point; the price of conditioning on an event A is $-\log \mathbb{P}(A)$, and no other law that makes A certain is cheaper. That last fact was proved by hand, for one kind of constraint. This chapter asks why, and finds the answer in a single formula.

The formula is the Boué–Dupuis representation [30]. It writes the logarithm of an exponential moment, $-\log \mathbb{E}\, e^{-F(W)}$, as the least expected cost of steering a Brownian path with a drift, the cost being the functional F of the steered path plus half the energy spent on steering. Read one way it is a control problem; read another, a statement about relative entropy; read a third way, a recipe for importance sampling. The three readings meet at one drift. For a functional of the endpoint the minimizing drift is the drift of an h-transform, the optimally steered path has the Gibbs law, and sampling under that drift estimates the exponential moment with zero variance. The conditioning drifts of Chapter 2 return here twice, once as the cheapest control and once as the perfect sampler.

The chapter delivers the formula with its optimal drift identified, importance sampling proved to be its equality case, and both sides checked on a computer. Section 3.1 proves Gibbs' variational principle on path space; Section 3.2 turns its minimum over laws into a minimum over drifts; Section 3.3 identifies the optimal drift; Section 3.4 proves zero variance; Section 3.5

measures what a near-optimal drift loses; Section 3.6 computes both sides of the formula.

3.1 Gibbs' Variational Principle

Chapter 2 found conditioning to be the cheapest way to make an event certain; that minimality is one instance of a principle valid for every path functional F.

Throughout, $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ is the path space of Chapter 1: $\Omega = C([0,T])$ with the canonical process W, its filtration $(\mathcal{F}_t)$ under the usual conditions, and Wiener measure $\mathbb{P}$, under which W is a standard Brownian motion started at 0. Let $F : \Omega \to \mathbb{R}$ be bounded and $\mathcal{F}_T$-measurable, a cost charged to each path. In the exponential moment $\mathbb{E}_{\mathbb{P}}\, e^{-F}$ a path counts with the weight e^{-F}: cheap paths count for nearly their Wiener probability, expensive ones for almost nothing. The number

$$-\log \mathbb{E}_{\mathbb{P}}\, e^{-F}$$

is the free energy of F, in the language of statistical mechanics. It lies between $\inf F$ and $\sup F$, and by Jensen's inequality it is at most $\mathbb{E}_{\mathbb{P}} F$: averaging the weights favours the cheap paths. The principle of this section says by exactly how much.

Definition 3.1 (Gibbs measure)**.** For a measurable $F : \Omega \to \mathbb{R}$ with $\mathbb{E}_{\mathbb{P}}\, e^{-F} < \infty$, the Gibbs measure of F is the probability $\mathbb{Q}^*$ on $(\Omega, \mathcal{F}_T)$ with density

$$\frac{d\mathbb{Q}^*}{d\mathbb{P}} = \frac{e^{-F}}{\mathbb{E}_{\mathbb{P}}\, e^{-F}}.$$

When F is bounded the density lies between two positive constants, and $\mathbb{Q}^*$ and $\mathbb{P}$ are equivalent.

The Gibbs measure leans the cloud of Brownian paths toward the region where F is small without moving any single path: a change of measure in the sense of Chapter 1, whose density is written down from the start. What singles it out is a minimization.

Theorem 3.2 (Gibbs' variational principle)**.** Let F be bounded and measurable. Then

$$-\log \mathbb{E}_{\mathbb{P}}\, e^{-F} = \inf_{\mathbb{Q}} \{\mathbb{E}_{\mathbb{Q}} F + H(\mathbb{Q} \mid \mathbb{P})\},$$

the infimum taken over all probability measures $\mathbb{Q}$ on $(\Omega, \mathcal{F}_T)$, with $H(\mathbb{Q} \mid \mathbb{P}) = +\infty$ when $\mathbb{Q}$ is not absolutely continuous with respect to $\mathbb{P}$. The infimum is attained at the Gibbs measure $\mathbb{Q}^*$ and at no other law.

Proof. Write $c = \mathbb{E}_{\mathbb{P}}\, e^{-F}$. Laws with $H(\mathbb{Q} \mid \mathbb{P}) = +\infty$ give the value $+\infty$, so let $H(\mathbb{Q} \mid \mathbb{P}) < \infty$. Then $\mathbb{Q} \ll \mathbb{P}$ with density Z, the set $\{Z > 0\}$ has $\mathbb{Q}$-probability one, and $\log Z$ is $\mathbb{Q}$-integrable: its positive part because the entropy is finite, its negative part because $\mathbb{E}_{\mathbb{Q}}(\log Z)^- = \mathbb{E}_{\mathbb{P}}[Z(\log Z)^-] \le e^{-1}$. On $\{Z > 0\}$ let $Y = e^{-F}/Z$. Its logarithm $-F - \log Z$ is $\mathbb{Q}$-integrable with mean $-\mathbb{E}_{\mathbb{Q}} F - H(\mathbb{Q} \mid \mathbb{P})$, and $\mathbb{E}_{\mathbb{Q}} Y = \mathbb{E}_{\mathbb{P}}[e^{-F}\mathbf{1}_{\{Z>0\}}] \le c$. Jensen's inequality for the strictly concave logarithm gives

$$-\mathbb{E}_{\mathbb{Q}} F - H(\mathbb{Q} \mid \mathbb{P}) = \mathbb{E}_{\mathbb{Q}} \log Y \le \log \mathbb{E}_{\mathbb{Q}} Y \le \log c,$$

which is the inequality $\mathbb{E}_{\mathbb{Q}} F + H(\mathbb{Q} \mid \mathbb{P}) \ge -\log c$. At $\mathbb{Q} = \mathbb{Q}^*$, $\log Z = -F - \log c$, and the total cost is $\mathbb{E}_{\mathbb{Q}^*} F - \mathbb{E}_{\mathbb{Q}^*} F - \log c = -\log c$. Conversely, if $\mathbb{Q}$ attains the infimum, both inequalities above are equalities. Equality in Jensen's inequality forces Y to equal a constant $a > 0$, $\mathbb{Q}$-almost surely, so $Z = e^{-F}/a$ on $\{Z > 0\}$; equality in the second gives $\mathbb{E}_{\mathbb{P}}[e^{-F}\mathbf{1}_{\{Z=0\}}] = 0$, hence $\mathbb{P}(Z = 0) = 0$ because $e^{-F} > 0$. Thus $Z = e^{-F}/a$ $\mathbb{P}$-almost surely, normalization gives $a = c$, and $\mathbb{Q} = \mathbb{Q}^*$. □

Nothing in the proof used that $\mathbb{P}$ is Wiener measure: the principle holds for any reference probability, and we use it below with others. It is also called the Donsker–Varadhan formula, and the weak convergence approach to large deviations is built on it [31]. Replacing the inequality by an identity sharpens it.

Corollary 3.3 (The cost of any other law)**.** Let F be bounded and measurable, with Gibbs measure $\mathbb{Q}^*$. For every probability $\mathbb{Q}$ on $(\Omega, \mathcal{F}_T)$,

$$\mathbb{E}_{\mathbb{Q}} F + H(\mathbb{Q} \mid \mathbb{P}) + \log \mathbb{E}_{\mathbb{P}}\, e^{-F} = H(\mathbb{Q} \mid \mathbb{Q}^*),$$

both sides being $+\infty$ together. The excess cost of a law over the free energy is its relative entropy with respect to the Gibbs measure.

Proof. Let $c = \mathbb{E}_{\mathbb{P}}\, e^{-F}$. Since $\mathbb{Q}^*$ and $\mathbb{P}$ are equivalent, $\mathbb{Q} \ll \mathbb{P}$ if and only if $\mathbb{Q} \ll \mathbb{Q}^*$; when this fails both sides are $+\infty$. Let $\mathbb{Q} \ll \mathbb{P}$ with density Z. The chain rule for densities gives, $\mathbb{Q}$-almost surely,

What does it cost to make a random path do something unlikely? This book answers with one formula. On Wiener space every equivalent change of measure is a drift, and its price is relative entropy: one half of the expected integral of the squared drift.

Condition a Brownian path on its endpoint, on staying positive or on a rare event, and the conditioned path is Brownian motion with a drift, the drift of Doob's h-transform; its price is minus the logarithm of the event's probability. The Boué–Dupuis formula makes that drift optimal and reads importance sampling as its equality case: under the cheapest drift, the estimator has zero variance. Let the noise vanish and the formula becomes the geometry of most likely paths and Kramers' law. Prescribe the law at both ends and it becomes Schrödinger's bridge, an h-transform with two ends.

Every theorem is proved in full, and every formula is computed twice, once from each of its sides.

Noa Ben-Dor is a professor of probability who works on change of measure, conditioned diffusions and large deviations.

ISBN 13 978-927-13-2876-1
Paperback
9 789271 328761