

**On Conditional Wiener Integrals and a Novel Approach
to the Fermion Sign Problem**

A thesis submitted for the degree of

Doctor of Philosophy
in Applied Mathematics

at the University of Leicester

by

Warwick Michael Dumas
Department of Mathematics,
University of Leicester,
United Kingdom.

January 2010

Abstract

On Conditional Wiener Integrals and a Novel Approach to the Fermion Sign Problem

by

Warwick Michael Dumas

Doctor of Philosophy in Applied Mathematics

The path-integral formulation of nonrelativistic quantum mechanics was introduced by Feynman in 1948. The use of Path Integral Monte Carlo can be put on a rigorous footing using conditional Wiener integrals. This thesis addresses the topics both of numerical error and of Monte Carlo error.

A piecewise constant numerical method which is of second order of accuracy for computing conditional Wiener integrals for a rather general class of sufficiently smooth functional is proposed. The method is based on simulation of Brownian bridges via the corresponding stochastic differential equations (SDEs) and on ideas of the weak-sense numerical integration of SDEs. A convergence theorem is proved. Special attention is paid to integral-type functionals. Results of some numerical experiments are presented.

In a further part of the research, the goal is to develop Monte Carlo methods for fermion simulations that are resistant to the explosion of variance which happens due to the fermion sign problem. A novel approach is developed which represents a radical departure from the current approaches. This is based on the principle of using a geometrical interpretation of the problem in order to find ways to maximize the negative covariance between the countersigned functional contributions. The fundamental connection between quantum exchange and the fermion sign problem is exploited. It is shown that this leads to a mathematical proof of the well-known exact solution to the sign problem for 1-dimensional fermion systems, and also to a novel exact solution in the case of a pair of 2-dimensional fermions.

Contents

List of Figures	iv
1 Introduction	5
1.1 Principles and aims of this thesis	5
1.2 Overview of chapters	10
2 Preliminaries	12
2.1 Conditioning and the conditional Wiener measure	13
2.1.1 The conditional Wiener measure	13
2.1.2 The Markov property and conditioning	17
2.1.3 Properties of the conditional Wiener measure	24
2.1.4 Conditioning the Wiener measure on a finite set of terminal points . .	26
2.1.5 Definition of functional integration	28
2.2 The relationship between path integrals and quantum statistical mechanics .	30
2.2.1 Using the Kolmogorov equations for a conditional Wiener integral . .	31
2.2.2 Physical interpretation of the conditional Wiener integral	34
2.2.3 Indistinguishability in terms of measure conditioning	37
2.2.4 Finding expectations of observables	38
2.3 Using noninteracting copies to approximate states of definite energy	41
2.3.1 Results on how to obtain sums of eigenstates	42
3 On numerical integration methods for conditional Wiener integrals	49
3.1 Fréchet derivatives and Taylor's Theorem	50
3.1.1 The class of functionals to be considered	51
3.1.2 Examples of functionals	53
3.2 The piecewise constant method	54
3.2.1 Background: orders of convergence	54
3.2.2 Definition of the piecewise constant method	56
3.3 Proof of the convergence theorem	59
3.4 Integral-type functionals	71
4 Simulations of Boltzmann and bosonic statistics	74
4.1 The probabilistic approach to finding expectations of observables	74
4.1.1 Integrating in the space of loops	75
4.1.2 Making interval estimates for a ratio of bivariate Gaussian variables .	80
4.2 Importance sampling and Markov Chain Monte Carlo	82
4.2.1 Importance sampling and the action measure	82
4.2.2 Sampling according to the action measure via Markov Chain Monte Carlo	85

4.2.3	Including permutations in the random sampling	87
4.2.4	Numerical results still apply under importance sampling	90
4.3	Numerical results for boltzmannons and bosons	91
4.3.1	Experiment results: Correlation function	92
4.3.2	Experiment results: Kinetic energy of 4 1D bosons	94
4.3.3	Experiment results: Potential energy of 64 boltzmannons	95
4.3.4	Experiment results: 1D harmonic oscillator with 1 particle	97
5	Fermionic simulations and the fermion sign problem	104
5.1	Simulating systems of fermions	104
5.1.1	Functional integrals for fermion systems	104
5.1.2	Introduction to the fermion sign problem	105
5.1.3	The fermion sign problem is insoluble	109
5.2	Extant practical approaches to fermion simulations	110
5.2.1	Restricted Path Integral Monte Carlo	110
5.2.2	Multilevel Blocking	112
5.3	Two simple algorithms: linear translation and expansion	113
5.3.1	Linear translation	113
5.3.2	Rearranging to use expanded paths, for noninteracting particles	115
5.4	Numerical results from linear translation and expanding paths algorithms, for 2 1D fermions	119
6	Solution to the fermion sign problem in 1 dimension	125
6.1	The Partitioning Theorem for two fermions in 1D	126
6.2	The Partitioning Theorem for n fermions in 1D	134
6.3	Simulation methods for 1-dimensional non-crossing bridges	135
6.3.1	The pdf for a point on a non-crossing Brownian bridge	137
6.3.2	Drawing a point on a non-crossing Brownian bridge	147
6.3.3	The initial point distribution	158
6.4	Simulation methods for n 1D non-crossing bridges	161
6.5	Numerical results using non-crossing algorithm for 2 fermions in a 1D Harmonic Oscillator	162
7	Multidimensional pairs of fermions	166
7.1	A novel approach to the fermion sign problem	166
7.1.1	Independence and dimensional crossing events	167
7.2	Subdiamond theory for the case of two fermions in a 2-dimensional space	169
7.2.1	Diamonds	169
7.2.2	The iterated subdiamond approach	191
7.3	The Last-to-first Subdiamond algorithm and its implementation	199
7.3.1	Notes on implementation	199
7.3.2	Further developments that give rise to the full algorithm	201
7.3.3	The Last-to-first algorithm in pseudocode	205
7.3.4	Drawing from the distributions of first and last crossing times	208
7.4	Numerical results using the Last-to-first Subdiamond algorithm for 2 fermions in a multidimensional space	212
7.4.1	Additive potentials and the Correlated Oscillator	213
7.4.2	Last-to-first experiments: a pair of 2D fermions	217
7.4.3	Last-to-first experiments: a pair of 3D fermions	219

7.5	Extending the Subdiamond approach to n particles	221
7.5.1	Towards a general method	221
7.5.2	Illustrative examples of crossing sequences	223
8	Conclusions and outlook	232
8.1	Conclusions	232
8.2	Outlook for research	236
	Bibliography	240

List of Figures

2.1	Contributions to $I_{T,2}$ for 2 1D fermions	48
3.1	Illustration of piecewise constant method X^h	57
4.1	<i>Correlation function.</i> The dependence of the correlation function $\Gamma(\theta)$ from (4.27) on θ simulated with $h = 0.2$ and $M = 10^8$ for $T = 10$. The left figure corresponds to the harmonic potential (4.29) and the right figure – to the anharmonic potential (4.30), both with $\omega = 1$	93
4.2	Simulated density, for 1 particle in H.O., $T = 2, h = 0.05, M = 10^6$, against solution	99
4.3	Simulated density for 1 particle in H.O. with $T = 6, h = 0.1, M = 10^6$, against solution	100
4.4	Simulated density for 1 particle in H.O., $T = 10, h = 0.2, M = 10^8$, against solution	101
5.1	Particle density using linear translation algorithm, $T = 2, h = 0.1, M = 10^7$	119
5.2	Particle density using expanding algorithm, $T = 2, h = 0.1, M = 10^6$	120
5.3	Particle density using expanding algorithm, $T = 2, h = 0.1, M = 10^7$	121
5.4	Particle density using linear translation algorithm, $T = 4, h = 0.1, M = 10^8$	122
5.5	Particle density using expanding algorithm, $T = 4, h = 0.05, M = 10^9$	122
5.6	Particle density using linear translation algorithm, $T = 6, h = 0.1, M = 10^{10}$	123
5.7	Particle density from expanding algorithm, $T = 6, h = 0.05, M = 10^{10}$	123
6.1	Schematic of how the pdf of $X(t_k)$ is determined when conditioning on E_0	139
6.2	$f(x,y)$, unnormalized.	140
6.3	unnormalized pdf for $X(t)$ when $X \sim w_{t_{k-1}, x_{k-1}}^{T, (a,b)} \Big _{E_2}$	141
6.4	Decomposing the volume of probability under the pdf curve	149
6.5	RN derivative of candidate measure compared with pdf of target measure. $\Delta_0 = 0.001, \Delta_T = 0.001, h = 0.01, T = 0.03$	156
6.6	RN derivative of candidate measure compared with pdf of target measure. $\Delta_0 = 0.001, \Delta_T = 0.05, h = 0.1, T = 2.0$	157
6.7	RN derivative of candidate measure compared with pdf of target measure. $\Delta_0 = 2.0, \Delta_T = 2.0, h = 0.1, T = 2.0$	158
6.8	RN derivative of candidate measure compared with pdf of target measure. $\Delta_0 = 0.1, \Delta_T = 0.0001, h = 0.01, T = 0.02$	159
6.9	Particle density for 2 1D fermions using Simpson's Rule for S and non-crossing algorithm. $T = 2, h = 0.1, M = 10^7$	163
6.10	Particle density for 2 1D fermions using Simpson's Rule for S and non-crossing algorithm. $T = 4, h = 0.1, M = 10^8$	164

6.11	Particle density for 2 1D fermions using Simpson's Rule for S and non-crossing algorithm. $T = 6, h = 0.1, M = 10^8$	164
6.12	Particle density for 2 1D fermions using Simpson's Rule for S and non-crossing algorithm. $T = 10, h = 0.1, M = 10^9$	165
7.1	Schematic of the contributing path segments for a path in E_x , according to the First Diamond Proposition for finding $\mathcal{G}_T((a_x, 0), (b_x, 0), (a_x, 0), (b_x, 0))$. Red is positive and blue is negative.	174
7.2	Schematic of the contributing path segments for a path in E_x , according to the Generalised Diamond Proposition.	182
7.3	A schematic of the decomposition of $\mathcal{G}_T((a, b), (a, b))$ made available by the Second Diamond Proposition. Blue segments are negative and red segments are positive.	186
7.4	The growth of the expected first crossing time with T , taking the initial and terminal distances to be 1, compared with $T^{1/2}$	188
7.5	How the sign problem may reoccur after 1 iteration of the subdiamond method.	189
7.6	Schematic of path segments in Generalised Second Diamond Proposition.	190
7.7	Schematic of two iterations of the subdiamond method.	194
7.8	How the fermion sign problem may recur if two iterations are performed.	195
7.9	Schematic of three iterations of the subdiamond method.	196
7.10	Diagram of how diamonds occur in the last-to-first method, using the native coordinate basis.	202
7.11	Illustration of some ideas for placing 3-way diamonds given a sequence of crossing dimensions	203
7.12	Diagram of how diamonds occur in the Last-to-first method when time is a circle.	204
7.13	A schematic of how to sample whether a path lies in E_3 , in a case where 3 iterations are required.	212
7.14	Particle density for 2 fermions in noninteracting correlated oscillator, $T=6$	216
7.15	Average sign using the Last-to-first algorithm for a pair of 2D fermions, non-interacting correlated oscillator.	218
7.16	Average sign using the Last-to-first algorithm for a pair of 3D fermions, non-interacting correlated oscillator	220
7.17	Schematic showing a crossing sequence $AB_x BC_x AB_y BC_y$, with permutations (id), (AB), (BC), (AB)(BC).	224
7.18	$AB_x BC_x BC_y AB_y$ can be regarded as $AB_y AB_x BC_x BC_y$ by moving the 0 label when 0 and T are identified to make $[0, T]$ a circle.	227

Acknowledgments

I would like to thank all of those who have supported me in making this work a success.

The foremost amongst these has been my supervisor, Professor Michael Tretyakov. It has been an immense privilege to work with the supervision of someone of the very highest intellectual calibre, who the first time we met, told me that there should be more to a PhD than aiming for a piece of paper. His suggestions and encouragement have been invaluable, and having the benefit of his practical advice is probably the only reason that this thesis was ever completed at all.

I would like to thank Dr Ruslan Davidchack for his friendship throughout and for many helpful discussions, and I am also grateful to Professor Ian Davies for his comments.

When my study of the fermion sign problem began, we had no idea that at the same time, it was being proven insoluble. There have been many times that it seemed that the prospects of success were dismal, and many when something temporarily seemed to offer an avenue of progress, only for it to turn out to be a dead end. I would like to thank my fiancée, Rachel, for loving and supporting me and believing in me throughout the last 3 years. If persistence has been my main virtue in this endeavour, it has been entirely second-hand.

I would also like to thank my parents Andrew and Susan, and my sister Francesca, for their support and encouragement.

I would like to thank Simon d'Angelo, my mathematics teacher in secondary school, who was the first person to recommend me to think about doing a PhD in mathematics.

A part of the computations were performed on the University of Leicester Mathematical Modelling Centre's cluster, which was purchased through the HEFCE Scientific Research Investment Fund.

A selection of the notation used

General

Notation	Formula	Description
w.r.t.		with respect to
s.t.		such that
pdf		probability density function
cdf		cumulative distribution function
SDE		stochastic differential equation
i.i.d.		independent and identically distributed
fdd		finite-dimensional distribution
χ_A		indicator function for A
$\varphi_{\sigma^2} : \mathbb{R}^{2r} \rightarrow \mathbb{R}$	$\frac{1}{(2\pi\sigma^2)^{r/2}} \exp\left(-\frac{\ x-y\ ^2}{2\sigma^2}\right)$	Gaussian pdf, for any dimension $r \in \mathbb{N}$
$\Phi(z)$		standard Gaussian cdf
π		a permutation
π		the ratio between the circumference of a circle and its diameter
$\text{sgn}(\pi)$	-1 for π odd; 1 for π even	signature of the permutation
Π_n		group of permutations of order n
$X \sim \mu$		the law of the r.v. X induces the measure μ .
δ_y	Dirac delta	probability measure with $\delta_y(\{y\}) = 1$.
$\delta(x, y)$	Kronecker delta	a real-valued function: $\delta(x, y) = 1$ if $x = y$; 0 otherwise.
Λ		Lebesgue measure on $\mathcal{B}(\mathbb{R}^{nd})$ or $\mathcal{B}(\mathbb{R})$.
$\biguplus_{i=1}^{\infty} B_i$		a countable union of disjoint sets
H	$\Delta^2 + V$	Hamiltonian operator

Parameters

Notation	Formula	Description
n		number of particles in the system
d		dimension of the space which the particles inhabit
N	T/h	number of time intervals in discretisation
T		Inverse temperature: ie, temperature is $1/k_B T$
h		discretisation timestep
M		number of Monte Carlo draws

Space and measure etc

Notation	Formula	Description
\mathcal{C}_0^T		$\{X : [0, T] \rightarrow \mathbb{R}^{nd}, X \text{ continuous}\}$
\mathcal{C}_{0,x_0}^T		$\{X \in \mathcal{C}_0^T : X(0) = x_0\}$
$\mathcal{C}_{0,x_0}^{T,x_T}$		$\{X \in \mathcal{C}_0^T : X(T) = x_T\}$
$\mathcal{C}_{0,x_0}^{T,\Pi_n x_T}$		$\{X \in \mathcal{C}_0^T : X(0) = x_0\}$
$\mathcal{C}_{0,\Pi_n x_0}^{T,\Pi_n x_T}$		$\{X : [0, T] \rightarrow \mathbb{R}^{nd} / \Pi_n, X(0) = \Pi_n x_0, X(T) = \Pi_n x_T\}$
\mathcal{C}_O		$\{X \in \mathcal{C}_0^T : X(0) = X(T)\}$
\mathcal{C}_Π		$\{X \in \mathcal{C}_0^T : X(0) \in \Pi_n X(T)\}$
w_{0,x_0}^T		Wiener measure on \mathcal{C}_{0,x_0}^T
w_{0,x_0}^{T,x_T}	$w_{0,x_0}^T \big _{X(T)=x_T}$	Conditional Wiener measure on $\mathcal{C}_{0,x_0}^{T,x_T}$
$w_{0,x_0}^{T,\Pi_n x_T}$	$w_{0,x_0}^T \big _{X(T) \in \Pi_n x_T}$	Wiener measure conditioned on $X(T) \in \Pi_n x_T$
$w_{0,\Pi_n x_0}^{T,\Pi_n x_T}$		"Conditional Wiener measure" corresponding to process with state space \mathbb{R}^{nd} / Π_n
w^*		$\Lambda \times w_{0,y}^{T,y}$
$w^{*\Pi}$		$\Lambda \times \alpha(y) w_{0,y}^{T,\Pi_n y}$ where $\alpha(y) = \sum_{\pi \in \Pi_n} \varphi_T(y, \pi y)$
W		Brownian motion
$X_{[a,b]}, X_{(a,b)}$ etc		for $0 \leq a < b \leq T$, the restriction of X to $[a, b]$
E_x		the event of a crossing in the x-coordinate
τ		a first (or last) crossing time
η		measure corresponding to first crossing time and point

Functionals and integrals

Notation	Formula	Description
$V : \mathbb{R}^{nd} \rightarrow \mathbb{R}$		potential (scalar field)
$F : \mathcal{C}_{0,x_0}^T \rightarrow \mathbb{R}$		a functional
\mathcal{U}	$\int F(X) dw_{0,y}^{T,y}$	a conditional Wiener integral
$c(\pi)$	$\text{sgn}(\pi)$ [for fermions] 1 [for bosons]	
$S : \mathcal{C}_{0,x_0}^T \rightarrow \mathbb{R}$	$\int_0^T V(x(t)) dt$	action functional for a path x
$S_{t,x}$		$dS = V(X(t)), S(0) = 0$
$Y : \mathcal{C}_{0,x_0}^T \rightarrow \mathbb{R}$	$\exp(-S(x))$	exponentiated action functional
$u_T(t, x, y)$	$\int_{C_{t,x}^{T,y}} \exp\left(\int_t^T -V(s) ds\right) dw_{t,x}^{T,y}$	action integral
$\mathcal{J}_T(x_0, x_T)$	$\int_{C_{0,x_0}^{T,x_T}} Y(X) dw_{0,x_0}^{T,x_T}(X)$	action integral
$\mathcal{I}_T(x_0, x_T)$	$\int_{C_{0,x_0}^{T,\Pi_n x_T}} c\left(\frac{X(T)}{x_T}\right) Y(X) dw_{0,x_0}^{T,\Pi_n x_T}(X)$	action integral
$\mathcal{G}_T(x_0, x_T)$	$\left(\sum_{\pi \in \Pi_n} \varphi_T(x_0, \pi x_T)\right) \mathcal{I}_T(x_0, x_T)$	action integral
\mathcal{A} or \mathcal{A}_D	$\int_{\mathbb{R}^{nd}} \int_{C_{0,y}^{T,y}} F(x) dw_{0,y}^{T,y} dy$	configurational integral
\mathcal{A}	$\int_{\mathbb{R}^{nd}} \int_{C_{0,y}^{T,\Pi_n y}} F(x) dw_{0,y}^{T,\Pi_n y} dy$	configurational integral
\mathcal{Z} or \mathcal{Z}_D	$\frac{1}{n!} \int_{\mathbb{R}^{nd}} \varphi_T(y, y) \mathcal{J}_T(y, y) dy$	partition function (distinguishable particles)
\mathcal{Z}	$\frac{1}{n!} \int_{\mathbb{R}^{nd}} \left(\sum_{\pi \in \Pi_n} \varphi_T(y, \pi y)\right) \mathcal{I}_T(y, y) dy$	partition function (indistinguishable particles)

Subdiamond notations

Notation	Description
θ_i	for i odd, x ; for i even, y .
τ_i, τ'_i	for $i > 0$, first and last crossing times of θ_i within $(\tau_{i-1}, \tau'_{i-1})$; $(\tau_0, \tau'_0) = (0, T)$.
χ	maximum crossing index: within (τ_χ, τ'_χ) there is no crossing of θ_{i+1} .
ς_i	for $0 \leq i < \chi$, last crossing of θ_i before τ_{i+1} .
ς'_i	for $0 \leq i < \chi$, first crossing of θ_i after τ'_{i+1} .
f_i, f'_i	a measure-preserving bijection (of some description given in the text)
\diamond_i	[in Lemma 7.2.7] $Y(X_{[\tau_{i-1}, \tau_i]}) - Y(f_i(X_{[\tau_{i-1}, \tau_i]}))$
\diamond_i	[from p.197 onwards] $Y(X_{[\varsigma_{i-1}, \tau_i]}) - Y(f_i(X_{[\varsigma_{i-1}, \tau_i]}))$

Simulation-related

Notation	Formula	Description
X^h		Piecewise constant approximation to X
$\overline{F}(X) = F(X^h)$		Piecewise constant approximation to $F(X)$
$\overline{\mathcal{U}}$	$\int_{\mathcal{C}_{0,y}^{T,y}} \overline{F}(X) dw_{0,y}^{T,y}$	
$\overline{\mathcal{A}}$	$\int_{\mathbb{R}^{nd}} \int_{\mathcal{C}_{0,y}^{T,y}} \overline{F}(x) dw_{0,y}^{T,y} dy$	
$\widetilde{F}(X)$		<i>either</i> unspecified approximation to F
		<i>or</i> approximation to F based on using Simpson's Rule for S
$\widetilde{\mathcal{U}}$	$\int_{\mathcal{C}_{0,y}^{T,y}} \widetilde{F}(X) dw_{0,y}^{T,y}$	
$\widetilde{\mathcal{A}}$	$\int_{\mathbb{R}^{nd}} \int_{\mathcal{C}_{0,y}^{T,y}} \widetilde{F}(x) dw_{0,y}^{T,y} dy$	
$\widetilde{\mathcal{A}}^{\text{MC}}$		Monte carlo estimator for $\widetilde{\mathcal{A}}$
$\widetilde{\mathcal{Z}}^{\text{MC}}$		Monte carlo estimator for $\widetilde{\mathcal{Z}}$
$\nu : \mathcal{B}(\mathcal{C}_O) \rightarrow R$	see (4.14)	action measure
$\nu : \mathcal{B}(\mathcal{C}_{\Pi}) \rightarrow R$	see (4.24)	action measure
$\nu : \mathcal{B}(\mathbb{R}^{Nnd}) \rightarrow R$		(discretised) action measure

System description

Notation	Formula	Description
$\langle A \rangle$		expectation of observable A
$\widetilde{\langle A \rangle}$		an approximation to $\langle A \rangle$
$\widehat{\langle A \rangle}$		Monte Carlo estimator for $\widetilde{\langle A \rangle}$
q or q_D		unnormalized thermal density matrix (distinguishable particles)
q		unnormalized thermal density matrix
$\rho : \mathbb{R}^{nd} \rightarrow \mathbb{R}$		system position density
$\varpi : \mathbb{R}^d \rightarrow \mathbb{R}$		particle density
E_0, E_1, \dots		energy level
$r_n(x)$		real eigenfunction of H
ψ_n		wavefunction (for distinguishable)
ϕ_n		wavefunction (for indistinguishable)

Chapter 1

Introduction

1.1 Principles and aims of this thesis

Since R.P. Feynman introduced his path integral formulation of quantum physics [Fey48], inspired by some earlier work of P.A.M. Dirac [Dir33], there has been interest in using functional integrals to calculate statistics of quantum systems. The aim of this thesis is to address certain issues that arise in performing Monte Carlo evaluations of these functional integrals.

We are going to be concerned with statistics of finite-temperature quantum systems with n identical particles inhabiting a d -dimensional space, in a potential $V : \mathbb{R}^{nd} \rightarrow \mathbb{R}$, at inverse temperature T . Using the insights of M. Kac, described in works such as [Kac56, Kac57, Fre85], it is possible to rigorously provide expressions for the expectations of observables in terms of conditional Wiener integrals [MT04] (this is done in Chapter 2). For example, in order to find the thermal average of the potential energy for a system of distinguishable particles ("boltzmannons"), we may take

$$\langle V \rangle = \frac{\int_{\mathbb{R}^{nd}} \int_{\mathcal{C}_{0,x_0}^{T,x_0}} V(x) Y(x) dw_{0,x_0}^{T,x_0}(x) d\Lambda(x_0)}{\int_{\mathbb{R}^{nd}} \int_{\mathcal{C}_{0,x_0}^{T,x_0}} Y(x) dw_{0,x_0}^{T,x_0}(x) d\Lambda(x_0)}, \quad (1.1)$$

where $\mathcal{C}_{0,x_0}^{T,x_0} \subset (\mathbb{R}^{nd})^{[0,T]}$ is the space of continuous functions $x : [0, T] \rightarrow \mathbb{R}^{nd}$ with $x(0) = x_0$ and $x(T) = x_0$, where w_{0,x_0}^{T,x_0} denotes the conditional Wiener measure (defined on p.16), and

where the exponentiated action functional $Y : \mathcal{C} \rightarrow \mathbb{R}$ is given by

$$Y(x) = \exp \left(- \int_0^T V(x(s)) ds \right). \quad (1.2)$$

(Note that our definition of the action shall differ from that which applies in Lagrangian mechanics, especially in that the kinetic energy term is for us, in effect, supplied by the conditional Wiener measure itself and does not need to be included.) In order to estimate the conditional Wiener integrals in (1.1), it is most expedient to use Monte Carlo methods. Any approach to calculating quantum statistics based on these ideas is referred to as Path Integral Monte Carlo; some applied examples are [CM00, HR05, BMNR01, LN04, She05].

The probabilistic approach to numerically evaluating integrals consists in regarding the desired integral as the expectation of a random variable. We shall see that it is possible to construct an evaluable weak approximation \bar{F} to any functional F from a broad class FA (see p.51). We are able to use Monte Carlo to estimate the expectation of such an \bar{F} , since we write

$$\mathcal{U} = \mathbb{E}_{X \sim w_{0,x_0}^{T,x_T}} F(X) \approx \mathbb{E}_{X \sim w_{0,x_0}^{T,x_T}} \bar{F}(X) = \bar{\mathcal{U}} = \mathbb{E} \bar{\mathcal{U}}^{\text{MC}}$$

where

$$\bar{\mathcal{U}}^{\text{MC}} = \frac{1}{M} \sum_{i=1}^M \bar{F}(X_i); \quad X_i \text{ i.i.d., } X_i \sim w_{0,x_0}^{T,x_T}$$

so that in fact, due to the standard Central Limit Theorem (e.g. [Wil01]),

$$\bar{\mathcal{U}}^{\text{MC}} \sim \text{Gaussian} \left(\mathbb{E} \bar{F}(X), \text{Var} \left(\bar{F}(X) \right) M^{-1} \right).$$

Inference about functional integrals

We shall adopt a Bayesian perspective, which, in order to provide clarity for the unfamiliar reader, we set out in detail. We shall model $\mathbb{E} \bar{F}$ and $\bar{\mathcal{U}}^{\text{MC}}$ via continuous random variables θ, ϑ defined over a probability space $(\Omega^0, \mathcal{F}^0, P)$ with a joint pdf over \mathbb{R}^2 . The marginal distribution of θ shall be called the **prior** distribution for $\mathbb{E} \bar{F}$. The conditional distribution of ϑ , given a value of θ , shall naturally be chosen to be exactly the conditional distribution of $\bar{\mathcal{U}}^{\text{MC}}$ given a hypothetical value of $\mathbb{E} \bar{F}$. When a realization of $\bar{\mathcal{U}}^{\text{MC}} = x$ is obtained, the

posterior distribution for $E\bar{F}$ is the name given to the conditional distribution of θ , given that ϑ is equal to x . Intuitively, it is the posterior for $E\bar{F}$ that describes the information about $E\bar{F}$ that is known *a posteriori*, ie after performing the experiment. Meanwhile, the prior for $E\bar{F}$ describes our information about it beforehand. Although $E\bar{F}$ is a deterministic quantity, because it is unknown we describe our information about it, both before and after the experiment, in the form of the distribution of a random variable. Given the prior and the conditional distribution of \bar{U}^{MC} , the posterior may be found directly using Bayes' Rule (see e.g. [HF04]). It is proportional to the product of the prior pdf and the **likelihood**, the latter being defined as the function of y formed by the conditional pdf of ϑ given that $\theta = y$ being evaluated at the realized value x . Note that in our framework, the prior pdf and the posterior pdf are both Radon-Nikodym derivatives of probability measures, whereas if the likelihood is regarded as the Radon-Nikodym derivative of a measure then in general, this *likelihood measure* does not have full measure 1.

The usual situation, loosely speaking, is that it is desirable to be able to claim *a posteriori* that the information contained in the prior did not exert much influence on the posterior obtained, giving the posterior the quality of objective information. When reasonable changes to the prior do not affect the conclusions of an analysis as described by the posterior, this is called a robust analysis (see [HF04]). For us, this issue shall be substantially circumvented by the fact that it is reasonable to assume that a Monte Carlo represents a large amount of information, as we shall now explain. (For a general discussion of the issues involved in objective inference, see e.g. [Ber06].) We make several assumptions about the prior for $E\bar{F}$. We have already assumed that a joint pdf for θ, ϑ exists. We further assume that the marginal pdf for θ (that is, the prior pdf for $E\bar{F}$) has a bounded derivative. Moreover, we assume that the prior pdf is nonzero except outside the range of F .

As mentioned, when M is sufficiently large, we may regard the distribution of \bar{U}^{MC} as Gaussian about (the hypothetical value of) $E\bar{F}$, with variance $\text{Var}(\bar{F}(X))M^{-1}$. This means that we think of the likelihood as near-zero outside of a ball about the observed \bar{U}^{MC} , of

diameter, say, $8\text{Var}(\overline{F}(X))^{1/2}M^{-1/2}$. If M is sufficiently large, the diameter of this region will be small compared to the maximum gradient of the prior pdf, and consequently, we may think of the prior pdf as constant over this region. (See also the point of view in [Wil01].) It follows that approximately, the posterior for $\text{E}\overline{F}$ is then Gaussian about the observed realization of \overline{U}^{MC} with variance $\text{Var}(\overline{F}(X))M^{-1}$.

The goals of this thesis

In such a schema there are then two separate errors which may be involved in making inferences about the true value of a functional integral, and hence about a system statistic: there is the numerical error, or bias, $|\text{E}\overline{F}(X) - \text{E}F(X)|$, and the Monte Carlo error, $|\overline{U}^{\text{MC}} - \text{E}\overline{U}^{\text{MC}}|$. It is our goal to make contributions to the study of both of these errors and how to control them.

One achievement of the research in this thesis shall be to prove that for a certain piecewise constant numerical method, the bias is of second order in the discretisation timestep, when F is in FA. This was already known only for exponential-type functionals [MT04ii, MT04, Suz91]; ie, those of the form $F(x(\cdot)) = \exp\left(-\int_0^T f(t, x(t))dt\right)$. The numerical method is based on simulation of the Brownian bridge via stochastic differential equations. Our result is important in applications, because in many cases, the functionals that are relevant to finding statistics are not of exponential type.

In the case of distinguishable particles, the Monte Carlo error is not usually problematic. However, in practice all quantum particles are either bosons or fermions, indistinguishable particles. In this thesis we conceptually approach indistinguishability through the idea of conditioning the Wiener measure on a finite set of terminal points, calling this measure $w_{0,x}^{T,\{x_1,\dots,x_k\}}$. Let us call to mind the following:

Definition 1.1.1 *A **permutation** of a finite set $\{1, \dots, n\}$ is a bijection $\pi : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$. We may use π as an operator over any n -fold product space: where $x = \{x_1, \dots, x_n\}$, $\pi x = \{x_{\pi(1)}, \dots, x_{\pi(n)}\}$.*

The permutations of $\{1, \dots, n\}$ are said to be permutations of order n , and these form a group. We shall use the notation Π_n to denote the group of permutations of order n . Note that for any $i_1, i_2 \in \{1, \dots, n\}$, Π_n may be partitioned into pairs which are closed under premultiplication by the pair-exchange of i_1, i_2 .

Definition 1.1.2 *When the irreducible representation of a permutation π takes the form of an odd number of pair exchanges, we say that π is **odd** and otherwise we say that π is **even**. We define the **signature** of π , $\text{sgn}(\pi)$, to be 1 when π is even and -1 when π is odd.*

In order to find statistics for systems of bosons, rather than performing integrals with respect to w_{0,x_0}^{T,x_T} , we need to perform essentially the same integrals with respect to $w_{0,x_0}^{T,\Pi_n x_T}$. Again, in this case, ways of managing the Monte Carlo error are already well-understood. If instead the particles are fermions, then we must still integrate with respect to $w_{0,x_0}^{T,\Pi_n x_T}$, but must also introduce a factor in the integrand that is equal to $\text{sgn}(\pi)$, where the terminal point $X(T) = \pi x_T$. The phenomenon of particles exchanging places between the initial and terminal points is called *quantum exchange* [Cep92]. For fermions, a greater propensity for quantum exchange means that the magnitude of negative contributions to the integral is becoming similar to the magnitude of the positive contributions, and this gives rise to the so-called *fermion sign problem* (see e.g. [Cep96]), which is the subject of the second half of this thesis.

In this further part of the research, the goal is to develop Monte Carlo methods for fermions that are resistant to the explosion of variance which happens due to the fermion sign problem. We explore the idea of generating a negative covariance between functional samples (ie, a positive covariance of the magnitudes of samples with opposite signs) by choosing sample paths which are close to each other in space. Focusing on exponential-type functionals, we consider a novel approach, the chief virtue of which is that it is based on taking account of the event that fermion coordinates coincide, which is in some sense the root cause of the problem, since it is what enables quantum exchange to take place. Examining first the 1-dimensional case, we see that it is possible to reexpress the functional integrals as

sums of integrals over disjoint events described in terms of particle coincidences. This gives rise to a mathematical proof of the well-known 1-dimensional exact solution that involves concentrating the integration measure on those paths where particles never meet. The same logic is then extended to the case of multidimensional fermion systems, and an exact solution is found for case of a 2-dimensional fermion pair, in the sense that as T increases the efficiency of the simulation does not tend to zero. It seems very doubtful that it is possible to avoid the sign problem as n grows (cf [TW05]); that is, it must be assumed that the complexity of the simulation grows exponentially with n .

1.2 Overview of chapters

Let us now provide a synopsis of the thesis chapters that are to follow.

Chapter 2 introduces the conditional Wiener measure and its properties, and develops the idea that indistinguishability results in conditioning on a set of terminal points. We then explain how conditional Wiener integrals may be related to functional statistics of quantum systems at finite temperature, using the Kolmogorov equations for a conditional Wiener integral (see also [DT82, DT82ii]). We prove a result which expresses the density matrix in a state of definite energy in terms of the zero-temperature limit of a functional integral expression.

In *Chapter 3*, we propose a probabilistic numerical method of second order of accuracy for computing conditional Wiener integrals of sufficiently smooth functionals. This method exploits a Markovian representation of the Brownian bridge. We begin by recalling the Fréchet derivative of a functional, and the Taylor Theorem for functionals, which is a key building block for the proof of the main result. This chapter is based on our paper [DT10].

In *Chapter 4*, we then proceed to offer a discussion of simulation procedures, concentrating on the cases of boltzmannons and bosons. It is seen that the conclusions of Chapter 3 also apply in the case of the Markov Chain Monte Carlo methods (sampling according to the action measure) which are popular in this area. Some simulation results are then presented,

illustrating the result of Chapter 3 and some further points.

Chapter 5 explains the fermion sign problem which is the main difficulty involved in performing path integral simulations of fermion systems. The way in which the problem emerges is discussed, and a brief overview is given of two extant methods. To illustrate the difficulty of the problem, results from some simple algorithms are reported.

Chapter 6 explores the solution to the fermion sign problem in 1 dimension, which is a very important case to understand before embarking upon the more general case. It is seen that when two fermions first meet on a path, this gives rise to a measure-preserving bijection between positive and negative contributing paths from that point, such that the value of the functional is also preserved, leading to complete cancellation. This leads to the well-known conclusion that in 1 dimension the solution is to prevent fermion paths from crossing. We then explain how this is to be achieved. Some numerical results are presented which illustrate the solution.

In *Chapter 7*, an approach, based upon the same principles that yielded the 1-dimensional solution, is developed for the case of 2 fermions in a multidimensional space. This allows the sign problem to be avoided, at least in the 2-dimensional case, in the sense that if T , the path length, is increased, then the average sign of contributions apparently does not tend to 0. This is in contrast to some existing approaches which, rather, are more effective for a large number of particles at sufficiently high temperatures. The efficacy of the approach is tested empirically.

In *Chapter 8*, we summarize the conclusions of the thesis and provide some remarks about possible directions of further research.

Chapter 2

Preliminaries

In this Chapter we shall begin by giving a treatment of some relatively well-known material based on sources such as [KS98, Shi89, Fre85, RW94], introducing essential concepts such as measure conditioning, the Wiener measure and the Markov property. The reader is advised that we have intentionally avoided directly repeating the "standard" presentation given in other works, and concentrated mainly on facts that are required for the endeavours of this thesis. Our treatment does, however, bear some commonalities with that of [Fre85]. Following works such as [Kac57, Kac56, Fre85], we then explain the relationship between conditional Wiener integrals and quantum statistical mechanics, via the probabilistic representation of the solutions to a certain parabolic partial differential equation. The connection with finding expectations of observables is briefly explained.

In Section 2.3, we then prove a novel result, which establishes a functional integral expression whose zero-temperature limit yields a density corresponding to the sum of the first k eigenstates. A closely related result in the special case of a system of noninteracting particles was demonstrated in Section 5.6 of [Iva05]; see also [GIV98]; but here the more general case is considered.

2.1 Conditioning and the conditional Wiener measure

For convenience, let us recall some well-known facts about measure conditioning. Let (Ω, \mathcal{F}, P) be a probability space. Suppose θ to be a random variable on (Ω, \mathcal{F}, P) with codomain \mathbb{R}^d . It has been demonstrated (see e.g. [Shi89] p.229) that it is possible to define the regular conditional probability with respect to the σ -algebra generated by θ , with uniqueness up to different versions. This has the implication that we may define conditional measures (see [Shi89] p.226). That is, given $A \in \sigma(\theta)$, we may define the conditional measure $P|_A$ to be given by $P|_A(B) = P(B|A)$ for $B \in \mathcal{F}$. Naturally this means that in order for $P|_A$ to be well-defined when $P(A) = 0$, a canonical version of conditional probability $P(B|\sigma(\theta))$ must first be assumed for every event B . (Fortunately, for us this shall generally be straightforward to do, at least in the cases of interest in this work.)

2.1.1 The conditional Wiener measure

The Wiener measure and Brownian motion

For any dimension $r \in \mathbb{N}$ and $s > 0$, we shall use the notation, $\varphi_s : \mathbb{R}^{2r} \rightarrow \mathbb{R}$ to indicate the independent Gaussian pdf with variance s in all dimensions, viz

$$\varphi_s(x, y) = (2\pi s)^{-r/2} \exp\left(-\frac{\|x - y\|^2}{2s}\right).$$

We shall use \mathcal{C}_{0,x_0}^T to mean $\left\{x(\cdot) \in (\mathbb{R}^d)^{[0,T]} : x(\cdot) \text{ continuous and } x(0) = x_0\right\}$, the set of continuous paths over $[0, T]$ with initial point x_0 . The reader should make themselves familiar with the Borel σ -algebra on \mathcal{C}_{0,x_0}^T , which is known to be generated by the collection of simple sets (also known as cylinders) - that is, those of the form

$$B = \{x \in \mathcal{C}_{0,x_0}^T : x(t_i) \in G_i, i = 1, \dots, \kappa\} \tag{2.1}$$

for some set of t_i with $0 < t_1 < \dots < t_\kappa \leq T$ and some products of intervals $G_i \subseteq \mathbb{R}^d$, referred to as gates. (This is described thoroughly in [Fre85] for the 1-dimensional case, which is essentially the same.) It is clear that measures on $\mathcal{B}(\mathcal{C}_{0,x_0}^T)$ are specified fully by the measure

which they assign to simple sets. Moreover, it follows from the Caratheodory Extension Theorem that given a measure over the algebra of simple sets, there exists an extension to $\mathcal{B}(\mathcal{C}_{0,x_0}^T)$. For us, the measure ascribed to simple sets shall be called the finite-dimensional distribution; note that this is different from the standard usage, but more convenient for our purposes¹. Since, as mentioned, we are able to consider a measure on $\mathcal{B}(\mathcal{C}_{0,x_0}^T)$ to be well-defined through the measure ascribed to simple sets, we are able to make the following definition.

Definition 2.1.1 *The **Wiener measure**, denoted w_{0,x_0}^T , is that measure on $\mathcal{B}(\mathcal{C}_{0,x_0}^T)$ such that for any simple set B given by (2.1),*

$$w_{0,x_0}^T(B) = \int \cdots \int_{G_1 \times G_2 \times \cdots \times G_\kappa} \prod_{i=0}^{\kappa-1} \varphi_{t_{i+1}-t_i}(x_i, x_{i+1}) dx_1 \cdots dx_\kappa \quad . \quad (2.2)$$

The Wiener measure, introduced in [Wie24], is discussed thoroughly in works such as [KS98, Fre85, IM74, RW94].

Definition 2.1.2 *A **random process** X is a family of random variables parametrized by a time variable t ; that is, either $X = \{X(t)\}_{t \in [0,T]}$ for some $T > 0$, or $X = \{X(t)\}_{t \geq 0}$.² The codomain of the random variables, together with the co- σ -algebra with respect to which they are measurable, is called the **state space** of the process.*

Definition 2.1.3 *The function of t , for a fixed $\omega \in \Omega$, given by $X(\omega; t) = X(t)$ shall be called a **trajectory** of the process X .*

In general, we may therefore regard a random process with state space \mathbb{R}^d as a random variable with codomain $(\mathbb{R}^d)^{[0,T]}$, the space of \mathbb{R}^d -valued functions. However, if the trajectories of the process are all continuous then we may regard it as a random variable with codomain \mathcal{C}_0^T . Further, if $X(0) = x_0 \in \mathbb{R}^d$ for every ω then we may consider the codomain of X to be \mathcal{C}_{0,x_0}^T ; and so on.

¹If we used the standard definition of finite-dimensional distribution then we would need to require certain consistency conditions in order for the corresponding measure on simple sets to be well-defined; in this case the Daniell-Kolmogorov Extension Theorem would provide the extension from finite-dimensional distributions to measures on $\mathcal{B}(\mathcal{C}_{0,x_0}^T)$; see e.g. [KS98, p.50].

²We shall usually be interested only in the case that $t \in [0, T]$.

Definition 2.1.4 (cf [IM74] p.16) A random process $W : \Omega \rightarrow \mathcal{C}_{0,x_0}^T$, defined on any probability space (Ω, \mathcal{F}, P) , shall be called *Brownian motion*, or a **Wiener process**, when its law induces the Wiener measure w_{0,x_0}^T on $\mathcal{B}(\mathcal{C}_{0,x_0}^T)$.

It is explained in [KS98] (p.71) that for W to be a Wiener process, one possible choice of (Ω, \mathcal{F}, P) and W is given by $\Omega = \mathcal{C}_{0,x_0}^T$, $\mathcal{F} = \mathcal{B}(\mathcal{C}_{0,x_0}^T)$, $P = w_{0,x_0}^T$ with $W(t; \omega) = \omega(t)$ for all $t \in [0, T]$, $\omega \in \Omega$. The authors of that work consider this to be the "canonical" probability space on which to define the Wiener process. For an heuristic discussion of the Wiener process and its properties, see e.g. [Ein26] or [RW94]. We make three further definitions relevant to random processes (see e.g. [RY99] p.41-2):

Definition 2.1.5 A **filtration** on a measurable space (Ω, \mathcal{F}) is an increasing family (\mathcal{F}_t) of sub- σ -algebras of \mathcal{F} . That is, for each $t \in [0, T]$, we have a sub- σ -algebra $\mathcal{F}_t \subseteq \mathcal{F}$ and when $s < t$, $\mathcal{F}_s \subseteq \mathcal{F}_t$. A random process X on (Ω, \mathcal{F}) such that $X(t)$ is measurable w.r.t. \mathcal{F}_t for all $t \in [0, T]$ is said to be **adapted** to the filtration (\mathcal{F}_t) .

Definition 2.1.6 A measurable space endowed with a filtration is said to be a **filtered space**, and to say that a random process is defined on a particular filtered space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ implies that it is \mathcal{F}_t -adapted.

Definition 2.1.7 The **natural filtration** of a random process X is the minimal filtration to which X is adapted, ie it is given by $\mathcal{F}_t^* = \sigma(X_s, s \in [0, t])$.

We therefore may, for instance, consider the process W to be defined on the filtered probability space $(\mathcal{C}_{0,x_0}^T, \mathcal{B}(\mathcal{C}_{0,x_0}^T), \mathcal{F}_t, w_{0,x_0}^T)$ where (\mathcal{F}_t) is simply the natural filtration of W .

Remark 2.1.8 The existence of the Wiener process implies that for any measure $w_{0,x_0}^T|_A$ obtained from w_{0,x_0}^T through conditioning, there exists a process defined on the same Ω , and adapted to the same filtration, whose distribution induces the conditional measure $w_{0,x_0}^T|_A$. For we may take the same X and take P to be this conditional measure. This also applies iteratively when further conditioning is applied.

Definition of the conditional Wiener measure

We shall be interested in conditioning the Wiener measure with respect to the event that $W(T)$ takes a particular value. Our definition must be in agreement with the required definition of conditioning on the event that $W(T)$ lies in a certain interval. As indicated at the beginning of this section, defining conditional measures is equivalent to choosing a version of conditional probability, for every event in $\mathcal{B}(\mathcal{C}_{0,x_0}^T)$; ie, to define $P|_A$ we must define $P(B|A)$ for each $B \in \mathcal{B}(\mathcal{C}_{0,x_0}^T)$. However, it is clear that it is sufficient to consider only simple events B , given by (2.1); in fact we temporarily make the further restriction that $t_\kappa < T$. A natural choice of canonical version is then given if we consider the probability density function (pdf) corresponding to the finite dimensional distribution of the Wiener process and apply the usual elementary interpretation of conditional probability. Viz,

$$w_{0,x_0}^T(B|W(T) = x) = \int \cdots \int_{G_1 \times G_2 \times \dots \times G_\kappa} \prod_{i=0}^{\kappa-1} \varphi_{t_{i+1}-t_i}(x_i, x_{i+1}) \varphi_{T-t_\kappa}(x_\kappa, x) \frac{1}{\varphi_T(x_0, x)} dx_1 \cdots dx_\kappa \quad (2.3)$$

This then allows the definition of the conditional measure $w_{0,x_0}^T|_A$ for any $A = \{\omega \in \Omega : W(T) \in \Gamma\}$, both for a product of intervals Γ , and hence also for Γ the countable union of some products of intervals. The following proposition allows us to be flexible with the domain of definition of such conditional measures. We shall use the notation w_{0,x_0}^{T,x_T} for the so-called conditional Wiener measure $w_{0,x_0}^T|_{\{X(T)=x_T\}}$.

Proposition 2.1.9 *Let \mathcal{F} be a σ -algebra over a non-empty set \mathbb{A} and let \mathcal{G} be a σ -algebra over a non-empty set $\mathbb{B} \in \mathcal{F}$ s.t. $\mathcal{G} \subseteq \mathcal{F}$, and assume that \mathcal{G} is sufficiently rich that every $A \in \mathcal{F} \setminus \mathcal{G}$ contains at least one element of $\mathbb{A} \setminus \mathbb{B}$. Then if μ is a probability measure defined on $(\mathbb{B}, \mathcal{G})$ then $\mu' : \mathcal{F} \rightarrow \mathbb{R}$ defined by $\mu'(A) = \mu(A \cap \mathbb{B})$ for $A \in \mathcal{F}$ is a probability measure on $(\mathbb{A}, \mathcal{F})$. Moreover, if μ' is any probability measure on $(\mathbb{A}, \mathcal{F})$ s.t. $\mu'(\mathbb{A} \setminus \mathbb{B}) = 0$ then $\mu : \mathcal{G} \rightarrow \mathbb{R}$ defined by $\mu(B) = \mu'(B)$ for $B \in \mathcal{G}$ is a probability measure on $(\mathbb{B}, \mathcal{G})$.*

Proof. To prove the first direction amounts to noting firstly that $A \cap \mathbb{B} \in \mathcal{G}$ due to our assumptions, and then that $\mu'(\emptyset) = 0$, that countable additivity is inherited from μ , and

that $\mu'(\mathbb{A}) = \mu(\mathbb{B}) = 1$. To prove the second direction, note that $\mu(\emptyset) = 0$, that countable additivity is inherited from μ' , and that $\mu(\mathbb{B}) = \mu'(\mathbb{B}) = \mu'(\mathbb{A}) = 1$. ■

Thus it is obvious that the conditional Wiener measure may equally well be regarded as a probability measure on $\mathcal{B}(\mathcal{C}_{0,x_0}^{T,x_T})$, since $w_{0,x_0}^{T,x_T}(\mathcal{C}_{0,x_0}^{T,x_T}) = 1$, or as a probability measure on \mathcal{C}_{0,x_0}^T which happens to assign full measure to the w_{0,x_0}^T -nullset $\mathcal{C}_{0,x_0}^{T,x_T}$. We shall usually have in mind the latter situation, although the former is perhaps more standard in the literature.

It is notable that we could define w_{0,x_0}^{T,x_T} by simply ascribing to it the finite-dimensional distributions inducing, for B given by (2.1) with $t_\kappa < T$, (cf (2.3))

$$w_{0,x_0}^{T,x_T}(B) = \int \int_{G_1 \times G_2 \times \dots \times G_\kappa} \prod_{i=0}^{\kappa-1} \varphi_{t_{i+1}-t_i}(x_i, x_{i+1}) \varphi_{T-t_\kappa}(x_\kappa, x_T) \frac{1}{\varphi_T(x_0, x_T)} dx_1 \dots dx_\kappa \quad . \quad (2.4)$$

Since this is a consistent family of finite-dimensional distributions with the same boundedness properties as w_{0,x_0}^T (cf [Fre85] p.27), we could proceed to use the same techniques to prove the existence of a corresponding process as are demonstrated in [KS98, Fre85, IM74] in the case of Wiener measure. However, it shall become clear that the concept of measure conditioning is rather important in what follows.

2.1.2 The Markov property and conditioning

Definition 2.1.10 (see e.g. [Shi89] p.248) *We shall consider a random process X , or the corresponding measure on a path space, to have the **Markov property** when it is such that if we define the family of transition measures $Q(s, x; t)$ by*

$$\forall G \in \mathcal{B}(\mathbb{R}^d) : \forall s, t \in [0, T] \text{ with } s < t : Q(s, x_1; t)(G) = P(X(t) \in G | X(s) = x_1)$$

then the Kolmogorov-Chapman equation is satisfied:

$$\forall r, s, t \in [0, T] \text{ with } r < s < t : Q(r, x_1; t)(G) = \int_{\mathbb{R}^d} Q(s, x; t)(G) dQ(r_1, x_1; s)(x) \quad . \quad (2.5)$$

Equivalently, by applying the above equality repeatedly, we could say that for a simple set B given by (2.1),

$$P(X \in B) = \int_{G_1} \int_{G_2} \dots \int_{G_\kappa} dQ(t_{\kappa-1}, x; t_\kappa)(x_\kappa) \dots dQ(t_1, x_1; t_2)(x_2) dQ(0, x_0; t_1)(x_1). \quad (2.6)$$

This implies that when the transition measures Q are absolutely continuous and thus there is a corresponding family of probability densities $p(s, x_1; t, x_2)$, we could write more straightforwardly:

$$P(X \in B) = \int \cdots \int_{G_1 \times G_2 \times \cdots \times G_\kappa} \prod_{i=0}^{\kappa-1} p(t_i, x_i; t_{i+1}, x_{i+1}) dx_1 \cdots dx_\kappa \quad . \quad (2.7)$$

It is not difficult to prove that our definition here is equivalent to the definition used in [KS98] (p.74), which defines a Markov process as s.t. almost surely, for $s, t \in [0, T]$ and $\Gamma \in \mathcal{B}(\mathbb{R}^d)$, where (\mathcal{F}_s) is the natural filtration of X ,

$$P(X(t+s) \in \Gamma | \mathcal{F}_s) = P(X(t+s) \in \Gamma | \sigma(X(s))) \quad . \quad (2.8)$$

Moreover, this specifically means that for $X(s)$ -almost all $y \in \mathbb{R}^d$, whenever some $B \in \mathcal{F}_s$ with $P(B) > 0$ and $\{X(s) = y\} \subseteq B$,

$$P(X(t+s) \in \Gamma | X \in B, X(s) = y) = P(X(t+s) \in \Gamma | X(s) = y) \quad . \quad (2.9)$$

This may in turn be recognised to be equivalent to: for $0 < r < s < t < T$,

$$P(X(t) \in \Gamma | X(r) \in G, X(s) = y) = P(X(t) \in \Gamma | X(s) = y) \quad . \quad (2.10)$$

Definition 2.1.11 *We shall say that a Markov process (or measure) is **time-homogeneous** in the case that for $s_1, s_2 \in [0, T]$, for all $x \in \mathbb{R}^d$, and for all $t > 0$ such that both measures are defined:*

$$Q(s_1, x; s_1 + t) = Q(s_2, x; s_2 + t) \quad (2.11)$$

It follows that there is then truly only one transition measure $Q(t-s; x)$ and if this is absolutely continuous then we have one transition density for the process, $p(t-s; x_1, x_2)$.

Thus for any simple set B we shall have (see e.g. [RY99] p.36.):

$$P(X \in B) = \int \int_{G_1 \times G_2 \times \cdots \times G_\kappa} \prod_{i=0}^{\kappa-1} p(t_{i+1} - t_i; x_i, x_{i+1}) dx_1 \cdots dx_\kappa \quad . \quad (2.12)$$

Considering (2.2), it is immediately evident that the Wiener process is Markov and also is time-homogeneous. Meanwhile, we can clearly see that (2.7) is satisfied by (2.4) under $t_\kappa < T$, and hence it follows that (2.6) is satisfied for any simple set B . On the other hand, (2.12) is not satisfied by (2.4).

The Strong Markov property

For us, a **random time** shall be a random variable with values in $[0, T]$.

Definition 2.1.12 *A random time τ defined on a filtered probability space $(\mathcal{C}_{0,x_0}^T, \mathcal{B}(\mathcal{C}_{0,x_0}^T), \mathcal{F}_t, w_{0,x_0}^T)$ shall be called a **Markov time** when it has the property that for $t \in [0, T]$, $\{\tau < t\} \in \mathcal{F}_t$.*

Definition 2.1.13 *(cf [KS98] p.81, [Shi89] p.127) We shall say that a random process X , or the corresponding measure on a path space, has the **strong Markov property** when it is such that for any Markov time τ , almost surely, whenever some $B \in \mathcal{F}_\tau$ with $P(B) > 0$ and $\{X(\tau) = y\} \cap B \neq \emptyset$,*

$$P(X(t + \tau) \in \Gamma | X \in B, X(\tau) = y) = P(X(t + \tau) \in \Gamma | X(\tau) = y). \quad (2.13)$$

In other words, when a process is strong Markov, conditioning on even the value at a Markov time still makes the subsequent process independent from what has historically occurred up to that point. It is proven in works such as [KS98] that the Wiener process is strong Markov.

Conditioning on sets of non-zero measure

Let P be a Markov measure on $\mathcal{B}(\mathcal{C}_{0,x_0}^T)$. Recall that when $P(A) > 0$, for any simple set B we must have

$$P|_A(B) = \frac{P(A \cap B)}{P(A)}. \quad (2.14)$$

We now develop the idea of defining measures conditional on nullsets, in specific cases.

Conditioning on simple sets

Firstly let us suppose that A consists of the event that X passes through a gate $A = \{x \in \mathcal{C}_{0,x_0}^T : x(s) \in \Gamma\}$, and $P(A) > 0$, and for simplicity let us also assume B also consists of just one gate, $B = \{x \in \mathcal{C}_{0,x_0}^T : x(t) \in G\}$. (From this point onwards, " $x \in \mathcal{C}_{0,x_0}^T$ " or

equivalent shall sometimes be taken as read, and we shall simply write $\{x(s) \in \Gamma\}$.) Then clearly,

$$\begin{aligned} \text{case } s < t : P|_A(B) &= \frac{\int_{\Gamma \times G} dQ(0, x_0; s)(y) dQ(s, x_1; t)(x)}{\int_{\Gamma} dQ(0, x_0; s)(y)} \\ \text{case } s > t : P|_A(B) &= \frac{\int_{G \times \Gamma} dQ(0, x_0; t)(x) dQ(t, x_1; s)(y)}{\int_{\Gamma} dQ(0, x_0; s)(y)} \\ \text{case } s = t : P|_A(B) &= \frac{\int_{G \cap \Gamma} dQ(0, x_0; s)(y)}{\int_{\Gamma} dQ(0, x_0; s)(y)} \end{aligned}$$

In general, if $B = \{x \in \mathcal{C}_{0, x_0}^T : x(t_i) \in G_i, i = 1, \dots, \kappa\}$ then, for instance, in the case that $s \in (t_i, t_{i+1})$,

$$P|_A(B) = \frac{\int_{G_1 \times \dots \times G_i \times \Gamma \times G_{i+1} \times \dots \times G_\kappa} dQ(0, x_0; t_1)(x_1) \cdots dQ(t_i, x_i; s)(y) dQ(s, y; t_{i+1}) \cdots dQ(t_{\kappa-1}, x_{\kappa-1}; t)(x_\kappa)}{\int_{\Gamma} dQ(0, x_0; s)(x)}$$

For events such as $A = \{x(s) = y\}$ we shall use the notation $P|_{(s,y)}$ to indicate $P|_A$. We now consider the particular case that the transition measures $Q(\cdot, \cdot; s)$ are absolutely continuous at $y \in \mathbb{R}^d$, and we consider transition densities $p(\cdot, \cdot; s, y)$ (that is, continuous versions of the Radon-Nikodym derivatives with respect to Lebesgue measure). Then it is clear that we may choose a canonical version of conditional probability such that

$$\begin{aligned} P|_{(s,y)}(B) &= \frac{1}{p(0, x_0, s)(y)} \int_{G_1 \times \dots \times G_\kappa} p(t_i, x_i; s, y) dQ(0, x_0, t_1)(x_1) \cdots \\ &\cdots dQ(t_{i-1}, x_{i-1}; t_i)(x_i) dQ(t_i, x_i; s)(y) dQ(s, y; t_{i+1})(x_{i+1}) \cdots dQ(t_{\kappa-1}, x_{\kappa-1}; t_\kappa)(x_\kappa) \quad (2.15) \end{aligned}$$

since it is not difficult to verify that this leads to, for $A = \{x(s) \in \Gamma\}$,

$$P|_A(B) = \int_{\Gamma} P|_{(s,y)}(B) dQ(0, x_0; s)(y)$$

as required. Moreover, this is the version familiar from elementary probability, in the case that X has absolutely continuous transition measures. For, suppose we consider the case $s > t$. We may say that the joint density for $X(t), X(s)$ is $p(0, x_0; t, x)p(t, x; s, y)$ and then consider that the conditional density for $X(t)$ given $X(s)$ is the ratio of this joint density to the marginal density for $X(s)$.

We may use (2.15) in a more convenient form: for a suitable choice of (s, y) ,

$$P_{0,x_0}^T|_{(s,y)} = P_{0,x_0}^{s,y} \times P_{s,y}^T. \quad (2.16)$$

In the case that P is strong Markov, then for any Markov time τ , (2.16) also holds conditionally given that $\tau = s$.

Moreover, we may generalise to the case that A is some simple set

$$A = \{x \in \mathcal{C}_{0,x_0}^T : x(s_i) \in \Gamma_i, i = 1, \dots, \varkappa\}.$$

Note that since for any $(s_1, s_2, \dots, s_\varkappa)$, clearly $X(s_1), X(s_2), \dots, X(s_\varkappa)$ jointly form a Euclidean-valued random variable, we are entitled to construct conditional measures based on $X(s_1), X(s_2), \dots, X(s_\varkappa)$. We consider the conditional measure of a simple set $B = \{x \in \mathcal{C}_{0,x_0}^T : x(t_i) \in G_i, i = 1, \dots, \kappa\}$, whose gates all occur at different times to those of A . Suppose that all s_i and t_i (other than $s_0 = t_0 = 0$) are relabelled as t'_i so that $0 < t'_1 < \dots < t'_{\kappa+\varkappa}$ and correspondingly, the G_i and Γ_i are renamed as C_i . Moreover, suppose (for simplicity) that X has absolutely continuous transition measures. Then it is clearly consistent to further define:

$$P|_A(B) = \frac{\int \int_{C_1 \times C_2 \times \dots \times C_{\kappa+\varkappa}} \prod_{i=0}^{\kappa+\varkappa-1} p(t'_i, x_i; t'_{i+1}, x_{i+1}) d\mu_1(x_1) \cdots d\mu_{\kappa+\varkappa}(x_{\kappa+\varkappa})}{\int \int_{\Gamma_1 \times \Gamma_2 \times \dots \times \Gamma_\varkappa} \prod_{i=0}^{\varkappa-1} p(s_i, x_i; s_{i+1}, x_{i+1}) d\mu_1(x_1) \cdots d\mu_\varkappa(x_\varkappa)}$$

where for any gate of B , μ_i is Lebesgue measure and for any gate of A it is given as follows. Suppose this gate Γ has a decomposition over dimensions as $\Gamma = \Gamma^{(1)} \times \Gamma^{(2)} \times \dots \times \Gamma^{(d)}$; then $\mu_i = \mu_i^{(1)} \times \dots \times \mu_i^{(d)}$ s.t. whenever $\Gamma^{(k)} = \{y\}$, $y \in \mathbb{R}$, the measure $\mu_i^{(k)}$ is the Dirac delta δ_y , and otherwise $\mu_i^{(k)}$ is Lebesgue.

Bayes' Rule

Now that $P(A|X(t_1) = x_1, \dots, X(t_\kappa) = x_\kappa)$ has been defined (except in pathological cases), we may deduce from (2.14) a Bayes' Rule (cf [HF04]), which applies when $P(A) > 0$. Let Q be the measure on $\mathbb{R}^{\kappa d}$ induced by $X(t_i), i = 1, \dots, \kappa$ under P . Then when $B =$

$$\{x \in \mathcal{C}_{0,x_0}^T : x(t_i) \in G_i, i = 1, \dots, \kappa\},$$

$$P|_A(B) = \frac{\int \cdots \int_{G_1 \times \cdots \times G_\kappa} P(A|X(t_i) = x_i; i = 1, \dots, \kappa) dQ(x)}{\int \cdots \int_{\mathbb{R}^{\kappa d}} P(A|X(t_i) = x_i; i = 1, \dots, \kappa) dQ(x)}.$$

With this in mind, we make a further choice of definition. Let Y be a random variable with codomain \mathbb{R}^{d_0} , so that the remarks at the beginning of Section 2.1 apply. Let P^* be the measure induced by the distribution of Y , and let $P^*|_x$ be the measure induced by the distribution of Y conditional on $X(t_1) = x_1, \dots, X(t_\kappa) = x_\kappa$, and suppose that $P^*|_x$ is absolutely continuous in a neighbourhood of $y \in \mathbb{R}^{d_0}$; let $p^*(\cdot|X(t_1) = x_1, \dots, X(t_\kappa) = x_\kappa)$ be the corresponding density. Then we shall write

$$P|_{\{Y=y\}}(B) = \frac{\int \cdots \int_{G_1 \times \cdots \times G_\kappa} p^*(y|X(t_i) = x_i; i = 1, \dots, \kappa) dQ(x)}{\int \cdots \int_{\mathbb{R}^{\kappa d}} p^*(y|X(t_i) = x_i; i = 1, \dots, \kappa) dQ(x)} \quad (2.17)$$

and this is clearly a valid definition in the sense that it satisfies

$$P|_A(B) = \int_{\Gamma} P|_{\{Y=y\}}(B) dP^*(y).$$

Using finite sets in place of gates

Now let us consider the situation that $A = \{x \in \mathcal{C}_{0,x_0}^T : x(s) \in \Gamma\}$ s.t. $\Gamma = \{y_1, y_2, \dots, y_{|\Gamma|}\}$, and define the measure of a simple set $B = \{x \in \mathcal{C}_{0,x_0}^T : x(t_i) \in G_i, i = 1, \dots, \kappa\}$, again for simplicity in the case that the transition measures are absolutely continuous and $s \in (t_i, t_{i+1})$.

It is clearly consistent with our other definitions to write:

$$P|_A(B) = \frac{\sum_{j=1}^{|\Gamma|} \int_{G_1 \times \cdots \times G_\kappa} p(0, x_0; t_1, x_1) \cdots p(t_i, x_i; s, y_j) p(s, y_j; t_{i+1}, x_{i+1}) \cdots p(t_{\kappa-1}, x_{\kappa-1}; t_\kappa, x_\kappa) dx}{\sum_{j=1}^{|\Gamma|} p(0, x_0; s, y_j)} \quad (2.18)$$

In the light of (2.17), we may envisage a similar definition in the case of conditioning on $\{Y \in \{y_1, y_2, \dots, y_\Theta\}\}$ for a Euclidean-valued random variable Y .

Joint conditioning is repeated conditioning

Let $A = A_1 \cap A_2$ be an event with $P(A) > 0$. Then we may immediately write, for any $B \in \mathcal{B}(\mathcal{C}_{0,x_0}^T)$,

$$\begin{aligned} P(B|A) &= \frac{P(A \cap B)}{P(A)} = \frac{P(A_1 \cap A_2 \cap B)}{P(A_1 \cap A_2)}; \\ P|_{A_1}|_{A_2}(B) &= \frac{P|_{A_1}(A_2 \cap B)}{P|_{A_1}(A_2)} = \frac{P(A_1 \cap A_2 \cap B)}{P(A_1)} \bigg/ \frac{P(A_1 \cap A_2)}{P(A_1)} \end{aligned}$$

and so clearly $P|_A = P|_{A_1}|_{A_2}$ in the case that $P(A) > 0$. It is relevant to explore to what extent this property holds in the cases of conditioning on nullsets as defined above.

When X is a random variable with values in \mathcal{C}_0^T , and I is an interval subset of $[0, T]$, we shall define the notation X_I to indicate the restriction of X to I ; that is, X_I is a random variable whose values are functions over I , and which agrees with X , ie $X_I(\omega; s) = X(\omega; s)$ for $s \in I$.

Consider the case that $A = A_1 \cap A_2$ with A_1 measurable with respect to $X_{[0,t]}$ (or in other words, $A_1 \in \mathcal{F}_t$) and with A_2 measurable w.r.t. $X_{(t,T]}$; it is clear that every Borel set has a decomposition of this kind. Then if $P|_{(t,x)}(A) > 0$, we may already note that

$$\begin{aligned} P|_{(t,x)}|_A &= P|_{(t,x)}|_{A_1}|_{A_2} = \left(P_{0,x_0}^{t,x}|_{A_1} \times P_{t,x}^T \right)|_{A_2} = P_{0,x_0}^{t,x}|_{A_1} \times P_{t,x}^T|_{A_2} \\ P|_A|_{(t,x)} &= P|_{A_1}|_{A_2}|_{(t,x)} = \left(P|_{A_1}|_{A_2} \right)_{0,x_0}^{t,x} \times \left(P|_{A_1}|_{A_2} \right)_{t,x}^T = \left(P|_{A_1} \right)_{0,x_0}^{t,x} \times \left(P|_{A_2} \right)_{t,x}^T \\ &= P_{0,x_0}^{t,x}|_{A_1} \times P_{t,x}^T|_{A_2} \quad . \end{aligned} \tag{2.19}$$

Now let us consider some random variable $Y = (Y_1, Y_2)$ where Y_1 is measurable with respect to $\sigma(X_{[0,t]})$ and Y_2 is measurable w.r.t. $\sigma(X_{(t,T]})$. Note then that $P|_{(t,x)}(Y_1 \in \Gamma_1, Y_2 \in \Gamma_2) = P|_{(t,x)}(Y_1 \in \Gamma_1) P|_{(t,x)}(Y_2 \in \Gamma_2)$. Hence if the induced measure for Y under $P|_{(t,x)}$ has a density at a particular point $y = (y_1, y_2)$ then this density is the product of those for Y_1 and for Y_2 . Let us call these p_1^* and p_2^* . Suppose B is a simple set and let $B = B_1 \cap B_2$ where B_1 is $X_{[0,t]}$ -measurable and B_2 is $X_{(t,T]}$ -measurable, and let B_1 consist of gates $G_1, \dots, G_{\kappa'}$ and let B_2 consist of gates $G_{\kappa'+1}, \dots, G_{\kappa}$. Then let Q_1 be the measure on $X(t_1), \dots, X(t_{\kappa'})$ induced under $P|_{(t,x)}$, and let Q_2 be the measure on $X(t_{\kappa'+1}), \dots, X(t_{\kappa})$ induced under $P|_{(t,x)}$. Then we may use (2.17) to see that:

$$\begin{aligned}
P|_{(t,x)}|_{\{Y=y\}}(B) &= \frac{\int \cdots \int_{G_1 \times \cdots \times G_\kappa} p_1^*(y_1|X(t_i) = x_i \forall i) p_2^*(y_2|X(t_i) = x_i \forall i) d(Q_1 \times Q_2)(x)}{\int \cdots \int_{\mathbb{R}^{\kappa d}} p_1^*(y_1|X(t_i) = x_i \forall i) p_2^*(y_2|X(t_i) = x_i \forall i) d(Q_1 \times Q_2)(x)} \\
&= \frac{\int \cdots \int_{G_1 \times \cdots \times G_{\kappa'}} p_1^*(y_1|X(t_i) = x_i \forall i) dQ_1(x) \int \cdots \int_{G_{\kappa'+1} \times \cdots \times G_\kappa} p_2^*(y_2|X(t_i) = x_i \forall i) dQ_2(x)}{\int \cdots \int_{\mathbb{R}^{\kappa' d}} p_1^*(y_1|X(t_i) = x_i \forall i) dQ_1(x) \int \cdots \int_{\mathbb{R}^{(\kappa-\kappa')d}} p_2^*(y_2|X(t_i) = x_i \forall i) dQ_2(x)} \\
&= P_{0,x_0}^{t,x}|_{\{Y_1=y_1\}} \times P_{t,x}^T|_{\{Y_2=y_2\}} \quad . \quad (2.20)
\end{aligned}$$

2.1.3 Properties of the conditional Wiener measure

As we mentioned in Chapter 1, we shall be concerned with systems of n particles, each with d coordinates, so that the "dimension of the system" is nd . The discussion given thus far, in Subsections 2.1.1 and 2.1.2, applies wholly when considering the full nd -dimensional system however, since our d in the above may be taken to be nd when we wish to describe the whole system. When X has state space \mathbb{R}^{nd} , we shall let X^i indicate the i th coordinate of X ; if the trajectories of X represent system trajectories then we shall use $X^{(j)}$ to represent the d -dimensional coordinate vector for the j th particle, and we shall let $X^{(j),i}$ represent the i th coordinate of the j th particle.

Brownian bridge SDE and incremental simulation procedure Let $a, b \in \mathbb{R}^{nd}$. Let us consider the \mathcal{F}_t -adapted stochastic process X given by the following stochastic differential equation (SDE from here onwards) and initial and terminal conditions: for $0 \leq t < T$ and for $i = 1, \dots, nd$,

$$dX^i = \frac{b^i - X^i}{T-t} dt + dW^i(t); \quad X^i(0) = a^i; \quad X^i(T) = b^i, \quad (2.21)$$

where by $W^i(t)$ we understand an independent Wiener process for each coordinate. By a standard result (see [IW81, RW94, KS98]), for any measurable $A \subseteq \mathcal{C}_{0,a}^{T,b}$, the limiting proportion of trajectories within A is equal to $w_{0,a}^{T,b}(A)$. The solution of (2.21) can be written as

$$X^i = a^i(T-t) + b^i t + (T-t) \int_0^t \frac{dW^i(s)}{T-s} \quad . \quad (2.22)$$

Hence, it is clear that for any $0 \leq \Delta < T - t$,

$$X(t + \Delta) = X(t) + \Delta \frac{b - X(t)}{T - t} + (T - t - \Delta) \int_t^{t+\Delta} \frac{dW(s)}{T - s}, \quad (2.23)$$

where $\int_t^{t+\Delta} \frac{dW(s)}{T-s}$ is interpreted as a vector of stochastic integrals. Therefore we have

$$\begin{aligned} \mathbb{E} \left[(T - t - \Delta) \int_t^{t+\Delta} \frac{dW(s)}{T - s} \middle| X(t) \right] &= 0; \\ \mathbb{E} \left[\left((T - t - \Delta) \int_t^{t+\Delta} \frac{dW(s)}{T - s} \right)^2 \middle| X(t) \right] &= \left(1 - \frac{\Delta}{T - t} \right) \Delta. \end{aligned} \quad (2.24)$$

It follows that we can exactly simulate the solution of (2.21) by a simple recurrent procedure based on the formula

$$X(t + \Delta) = X(t) + \Delta \frac{b - X(t)}{T - t} + \Delta^{1/2} \sqrt{\frac{T - t - \Delta}{T - t}} \xi, \quad t < T, \quad (2.25)$$

where ξ is a random vector of which the components are independent Gaussian random variables with zero mean and unit variance, and which are independent of $X(t)$.

Dimensional independence Where $y_0, y'_0 \in \mathbb{R}^r$ and $x_0, x'_0 \in \mathbb{R}$, $w_{0,(x_0,y_0)}^{T,(x'_0,y'_0)} = w_{0,(x_0)}^{T,(x'_0)} \times w_{0,(y_0)}^{T,(y'_0)}$.

Moreover, a similar fact holds true when coordinates are rotated, because of the rotational symmetry of the Gaussian transition density. Where $(x_1, x_2, \dots, x_r) \in \mathbb{R}^r$ is represented as (z_1, z_2, \dots, z_r) in a particular coordinate basis, and where (x'_1, \dots, x'_r) is represented as (z'_1, \dots, z'_r) , then $w_{0,x}^{T,x'}$ is equal to $w_{0,x_1}^{T,x'_1} \times w_{0,x_2}^{T,x'_2} \times \dots \times w_{0,x_r}^{T,x'_r}$ and is also equal to $w_{0,z_1}^{T,z'_1} \times w_{0,z_2}^{T,z'_2} \times \dots \times w_{0,z_r}^{T,z'_r}$ where components correspond to the new basis vectors. In particular, note that for a 2-dimensional Brownian bridge X representing two 1-dimensional particles, we have that $\frac{X^{(1)}(t) - X^{(2)}(t)}{\sqrt{2}}$ and $\frac{X^{(1)}(t) + X^{(2)}(t)}{\sqrt{2}}$ are independent Brownian bridges. This is evident from considering the finite-dimensional distributions.

Conditioning on passing through a point It follows from the discussion above that the conditional Wiener measure has conditional independence: where $x \in \mathbb{R}^{nd}$ and $t \in (0, T)$,

$$w_{0,x_0}^{T,x_T} \bigg|_{(t,x)} = w_{0,x_0}^{t,x} \times w_{t,x}^{T,x_T}.$$

Joint conditioning Suppose for simplicity that $d = 1$, suppose $x_0^{(1)} > \dots > x_0^{(n)}$ and $x_T^{(1)} > \dots > x_T^{(n)}$, and consider conditioning on a set such as $E_0 = \{X \in \mathcal{C}_{0,x_0}^{T,x_T} : X^{(1)} > X^{(2)} > \dots > X^{(n)}\}$. Then where $E_0 = E_1 \cap E_2$, s.t. E_1 is $\sigma(X_{[0,t]})$ -measurable and E_2 is $\sigma(X_{(t,T]})$ -measurable,

$$w_{0,x_0}^{T,x_T} \Big|_{E_0} \Big|_{(t,x)} = w_{0,x_0}^{t,x} \Big|_{E_1} \times w_{t,x}^{T,x_T} \Big|_{E_2} .$$

Definition 2.1.14 The *first crossing time* is defined as the first exit time of X from $D = \{x^{(1)} > x^{(2)} > \dots > x^{(n)}\}$, where if necessary the particles are relabelled so that $X^{(1)}(0) \geq X^{(2)}(0) \geq \dots \geq X^{(n)}(0)$. In other words, since our attention is restricted to continuous paths X , it is the minimum t s.t. $X^{(i)}(t) = X^{(j)}(t)$ for some $i \neq j$.

The first crossing time is a Markov time and we denote it by τ . Let $E(t, x) = \{\tau = t, X(\tau) = x\}$ be the event that the first crossing time is at t and at system position x . Then it follows from (2.13) that

$$w_{0,x_0}^{T,x_T} \Big|_{E(t,x)} = w_{0,x_0}^{t,x} \Big|_{E(t,x)} \times w_{t,x}^{T,x_T} \Big|_{E(t,x)} = w_{0,x_0}^{t,x} \Big|_{E(t,x)} \times w_{t,x}^{T,x_T} . \quad (2.26)$$

Linear translation Where $l_{a,b} : [0, T] \rightarrow \mathbb{R}^d$ is given by $l_{a,b}(t) = \frac{t}{T}b + \frac{T-t}{T}a$, $w_{0,a}^{T,b}(l_{a,b} + B) = w_{0,0}^{T,0}(B)$ where for any event B , $l_{a,b} + B$ denotes the event $\{X + l_{a,b} : X \in B\}$.

We can view this in another way. If we consider two processes $X^{(1)}, X^{(2)}$ such that $X^{(1)}$ satisfies (2.21) with endpoints $a^{(1)}, b^{(1)}$ whereas $X^{(2)}$ satisfies (2.21) with endpoints $a^{(2)}, b^{(2)}$, but with the same Wiener process W , then $X^{(2)} = X^{(1)} + l_{a^{(2)}-a^{(1)}, b^{(2)}-b^{(1)}}$.

2.1.4 Conditioning the Wiener measure on a finite set of terminal points

We have noted that the choice of a canonical version of conditional probability, with respect to the σ -algebra generated by a real-valued random vector, such as $X(T)$, allows us to choose a suitable definition for measures such as $w_{0,x_0}^T|_{X(T) \in \Gamma}$ when $\Gamma \in \sigma(X(T))$ is a finite

set, $\Gamma = \{y_1, \dots, y_k\}$. Let $B = \{x \in \mathcal{C}_{0,x_0}^T : x(t_i) \in G_i, i = 1, \dots, \kappa\}$. According to (2.18),

$$w_{0,x_0}^T \big|_{X(T) \in \Gamma} (B) = \frac{\sum_{j=1}^k \int_{G_1 \times \dots \times G_\kappa} \prod_{i=0}^{\kappa-1} \varphi_{t_{i+1}-t_i}(x_i, x_{i+1}) \varphi_{T-t_\kappa}(x_\kappa, y_j) dx}{\sum_{j=1}^k \varphi_T(x_0, y_j)}$$

From this definition it is evident that the finite-dimensional distributions of $w_{0,x_0}^T \big|_{X(T) \in \Gamma}$ induce a mixing measure between the conditional Wiener measures corresponding to the k termini, with coefficients representing the relative probabilities of the termini under w_{0,x_0}^T . That is,

$$w_{0,x_0}^T \big|_{X(T) \in \Gamma} (B) = \sum_{j=1}^k \frac{\varphi_T(x_0, y_j)}{\sum_{j'=1}^k \varphi_T(x_0, y_{j'})} w_{0,x_0}^{T,y_j}(B) .$$

When the coordinates of X represent sets of coordinates describing positions of separate particles, we may be especially interested in allowing that the terminal value is any permutation of the point positions in some system position $x_T \in \mathbb{R}^d$. We shall use the notation $w_{0,x_0}^{T,\{x_1,\dots,x_k\}}$ and in the case that $\{x_1, \dots, x_k\} = \{\pi_1 x_T, \pi_2 x_T, \dots, \pi_k x_T\}$ for a set of permutations π_i which form a group Π , then we shall write $w_{0,x_0}^{T,\Pi x_T}$ for $w_{0,x_0}^{T,\{x_1,\dots,x_k\}}$.

We can see that $w_{0,x_0}^{T,\Pi_n x_T}$ is Markov; especially, it is evident that for $x \in \mathbb{R}^{nd}, t \in (0, T)$,

$$w_{0,x_0}^{T,\Pi_n x_T} \big|_{(t,x)} = w_{0,x_0}^{t,x} \times w_{t,x}^{T,\Pi_n x_T} .$$

Moreover, we may consider what happens if we take the state space to be the identification space \mathbb{R}^{nd} / Π_n , ie the quotient space given by applying to \mathbb{R}^{nd} the equivalence relation $x \equiv \pi x, \pi \in \Pi_n$ (see e.g. [BM99]). We shall now see that we can create a measure on continuous \mathbb{R}^{nd} / Π_n -valued functions with endpoints $\mathfrak{X}(0) = \Pi_n x_0, \mathfrak{X}(T) = \Pi_n x_T$, and we shall call this measure $w_{0,\Pi_n x_0}^{T,\Pi_n x_T}$. In order to show that our definition (2.27) makes this well-defined, we shall first need to prove a preliminary fact.

It is notable that for $\gamma \in \Pi_n$, if we fix $t \in (0, T)$ then the law for $X(t)$ under $X \sim w_{0,x_0}^{T,\Pi_n x_T}$ and that under $X \sim w_{0,\gamma x_0}^{T,\Pi_n x_T}$ are conjugated by γ . That is, for $B \in \mathcal{B}(\mathbb{R}^{nd})$, where we define

$$\gamma B := \{y \in \mathbb{R}^{nd} : \gamma x = y \text{ some } x \in B\},$$

$$\begin{aligned} w_{0,x_0}^{T,\Pi_n x_T} \{X(t) \in B\} &= \sum_{\pi \in \Pi_n} \frac{\varphi(x_0, \pi x_T)}{\sum_{\pi' \in \Pi_n} \varphi(x_0, \pi' x_T)} w_{0,x_0}^{T,\pi x_T} \{X(t) \in B\} \\ &= \sum_{\pi \in \Pi_n} \frac{\varphi(\gamma x_0, \pi \gamma x_T)}{\sum_{\pi' \in \Pi_n} \varphi(\gamma x_0, \pi' \gamma x_T)} w_{0,\gamma x_0}^{T,\pi \gamma x_T} \{X(t) \in \gamma B\} \\ &= \sum_{\pi \gamma^{-1} \in \Pi_n} \frac{\varphi(\gamma x_0, \pi x_T)}{\sum_{\pi' \gamma^{-1} \in \Pi_n} \varphi(\gamma x_0, \pi' x_T)} w_{0,\gamma x_0}^{T,\pi x_T} \{X(t) \in \gamma B\} \\ &= \sum_{\pi \in \Pi_n} \frac{\varphi(\gamma x_0, \pi x_T)}{\sum_{\pi' \in \Pi_n} \varphi(\gamma x_0, \pi' x_T)} w_{0,\gamma x_0}^{T,\pi x_T} (\gamma B) = w_{0,\gamma x_0}^{T,\Pi_n x_T} \{X(t) \in \gamma B\}, \end{aligned}$$

where here we have used the self-evident fact that postmultiplication by γ^{-1} maps Π_n to itself. But then if we fix $t \in (0, T)$ and consider the law of $\mathfrak{X}(t)$, a random variable with codomain \mathbb{R}^{nd}/Π_n , under $\mathfrak{X} \sim w_{0,\Pi_n x_0}^{T,\Pi_n x_T}$, clearly it makes sense when $B \in \mathcal{B}(\mathbb{R}^{nd})$, to define

$$w_{0,\Pi_n x_0}^{T,\Pi_n x_T} \{\mathfrak{X}(t) \in \Pi_n B\} := w_{0,x_0}^{T,\Pi_n x_T} \{X(t) \in \Pi_n B\} = w_{0,\gamma x_0}^{T,\Pi_n x_T} \{X(t) \in \Pi_n B\} \quad (2.27)$$

where here, obviously $\Pi_n B$ takes on two different meanings according to whether it signifies a set in $\mathcal{B}(\mathbb{R}^{nd})$ or its equivalent in $\mathcal{B}(\mathbb{R}^{nd}/\Pi_n)$. Since we therefore have defined the finite-dimensional distribution, and hence the transition measure, corresponding to a Markov process \mathfrak{X} with state space \mathbb{R}^{nd}/Π_n , it follows that $w_{0,\Pi_n x_0}^{T,\Pi_n x_T}$ is well-defined as a (Markov) measure on the Borel σ -algebra generated by cylinders over \mathbb{R}^{nd}/Π_n .

2.1.5 Definition of functional integration

Naturally, when we are able to define a measure P on $\mathcal{B}(\mathcal{C}_{0,x_0}^T)$, it follows immediately that functional integrals with respect to this measure are already defined, via the usual Lebesgue integration formula (see for example [KF57, Coh80]). Let $F : \mathcal{C}_{0,x_0}^T \rightarrow \mathbb{R}$ be measurable with respect to P . Then for instance, if for some constants $y \leq F \leq y'$, then where we shall loosely write y_k for $y + (y' - y)\frac{k}{K}$,

$$\int_{\mathcal{C}_{0,x_0}^T} F(x(\cdot)) dP = \lim_{K \rightarrow \infty} \sum_{k=0}^K y_k P \left(F^{-1} \left(y_k - \frac{1}{2K}, y_k + \frac{1}{2K} \right) \right).$$

However, it is possible to find a sufficient condition to make this equivalent to the popular working definition of $\int_{\mathcal{C}_{0,x_0}^T} F(x(t)) dP$, as used in [FH65, GY56, Cep95], as follows. Let us

write h for T/N . Let P be a measure defined on $\mathcal{B}(\mathcal{C}_{0,x_0}^T)$, and let μ_N be the measure induced on $\mathcal{B}(\mathbb{R}^{Nr})$ by the distribution of $x(h), x(2h), \dots, x(T)$ under P . Given any vector $\underline{x} = \{x_1, \dots, x_N\} \in \mathbb{R}^{Nr}$, let it serve as the parameter for a function $\ell(\{x_1, \dots, x_N\}) : [0, T] \rightarrow \mathbb{R}^r$ to be given by linear interpolation. That is to say, letting $t_k = kh$ for $k = 0, 1, \dots, N$, we have that for $k = 0, 1, \dots, N-1$, for $t \in [t_k, t_{k+1})$:

$$\ell(\{x_1, \dots, x_N\}; t) = \frac{t - t_k}{h} x_k + \frac{t_{k+1} - t}{h} x_{k+1} \quad .$$

It is immediately evident that for each coordinate we have

$$\lim_{N \rightarrow \infty} \int_0^T [\ell(\{x^i(h), x^i(2h), \dots, x^i(T)\}; t) - x^i(t)] dt = 0$$

so clearly, if we assume F to be continuous in the sense that whenever $x(\cdot)$ is the L_1 limit of a sequence $(x^{(j)}(\cdot))$ over \mathcal{C}_{0,x_0}^T , we have that $F(x(\cdot))$ is the limit of $(F(x^{(j)}(\cdot)))$, then it shall follow that for all $x \in \mathcal{C}_{0,x_0}^T$,

$$\lim_{N \rightarrow \infty} F(\ell(\{x(h), x(2h), \dots, x(T)\})) = F(x(\cdot)). \quad (2.28)$$

But then

$$\begin{aligned} \int_{\mathcal{C}_{0,x_0}^T} F(x(t)) dP(x) &= \lim_{N \rightarrow \infty} \int_{\mathcal{C}_{0,x_0}^T} F(\ell(\{x(h), x(2h), \dots, x(T)\})) dP(x) \\ &= \lim_{N \rightarrow \infty} \int_{\mathbb{R}^{Nr}} F(\ell(\underline{x})) d\mu_N(\underline{x}). \end{aligned}$$

This means in particular that the conditional Wiener integral of F is given by (see [GY56]):

$$\int_{\mathcal{C}_{0,a}^{T,b}} F(x(\cdot)) dw_{0,x_0}^{T,x_T} = \lim_{N \rightarrow \infty} \int \cdots \int_{\mathbb{R}^{Nr}} F(\ell(x_1, \dots, x_T)) \frac{\prod_{i=0}^{N-1} \varphi_h(x_i, x_{i+1})}{\varphi_T(x_0, x_T)} dx_1 \cdots dx_{N-1} \quad . \quad (2.29)$$

In the special case that F is the action functional $F(x) = Y(x)$ (cf (1.2)), for an alternative derivation see [Kac57, p.165]; this relates to the Wiener integral but it is clear that it could be easily generalised to apply for any Markov measure P .

2.2 The relationship between path integrals and quantum statistical mechanics

In this section, we prove a theorem based on spectral decomposition of solutions to the backward and forward Kolmogorov equations for the conditional Wiener integral of the action functional (cf [Kry99]). The following exposition is largely based on the work of Kac; see, for example, [Kac51, Kac56] and [Kac57, pp.168-173]. Theorem 2.2.1 is an analogue of the Feynman-Kac Theorem which applies in the case of the unconditional Wiener measure. We shall see in Subsection 2.2.2 that this allows us to introduce the connection between quantum statistical mechanics and functional integrals (see also [DT82ii]).

Notation for action integrals Let $V : \mathbb{R}^{nd} \rightarrow \mathbb{R}$ be a Borel-measurable function, which in physical terms shall represent the potential field for the system. We shall let $u : [0, T] \times \mathbb{R}^{nd} \times \mathbb{R}^{nd} \rightarrow \mathbb{R}$ be given by, for $t \in [0, T)$ and $x_t, x_T \in \mathbb{R}^{nd}$,

$$u(t, x_t, x_T) = \int_{\mathcal{C}_{t, x_t}^{T, x_T}} \exp \left(\int_t^T -V(x(s)) ds \right) dw_{t, x_t}^{T, x_T} \quad (2.30)$$

and for $x_T, x'_T \in \mathbb{R}^{nd}$,

$$u(T, x_T, x'_T) = 1 \quad . \quad (2.31)$$

Moreover, define

$$\mathcal{J}_T(x_0, x_T) = u(0, x_0, x_T) = \int_{\mathcal{C}_{0, x_0}^{T, x_T}} \exp \left(\int_0^T -V(x(s)) ds \right) dw_{0, x_0}^{T, x_T} = \int_{\mathcal{C}_{0, x_0}^{T, x_T}} Y(x) dw_{0, x_0}^{T, x_T} \quad (2.32)$$

Such an integral \mathcal{J}_T is called an action integral. Further, where E is an event in $\mathcal{B} \left(\mathcal{C}_{0, x_0}^{T, x_T} \right)$ such that $w_{0, x_0}^{T, x_T} \Big|_E$ has been defined in Section 2.1, we shall use the notation

$$\mathcal{J}_T^E(x_0, x_T) = \int_{\mathcal{C}_{0, x_0}^{T, x_T}} \exp \left(\int_0^T -V(x(s)) ds \right) dw_{0, x_0}^{T, x_T} \Big|_E \quad (2.33)$$

Time translation invariance

Note that because V is not time-dependent,

$$\int_{\mathcal{C}_{T_1, x_{T_1}}^{T_2, x_{T_2}}} \exp \left(\int_{T_1}^{T_2} -V(x(s)) ds \right) dw_{T_1, x_{T_1}}^{T_2, x_{T_2}} = \mathcal{J}_{T_2-T_1}(x_{T_1}, x_{T_2}) \quad . \quad (2.34)$$

2.2.1 Using the Kolmogorov equations for a conditional Wiener integral

In this subsection, we prove a Theorem which provides an expansion of the functional integral u (and hence \mathcal{J}_T) in terms of the real-valued eigenfunctions of the Hamiltonian operator. This shall then allow us, in what follows, to rigorously explain how functional integration can be used to find information about quantum statistics.

Theorem 2.2.1 *Let $x, y \in \mathbb{R}^{nd}$ and let $V : \mathbb{R}^{nd} \rightarrow \mathbb{R}$ be continuous, bounded from below and have the property that $V(x) \rightarrow \infty$ as $\|x\| \rightarrow \infty$. Then there is a countable set of orthonormal (ie, normalized and orthogonal) solutions to the eigenvalue problem*

$$\frac{1}{2} \sum_{i=1}^{nd} \frac{\partial^2 r}{\partial x_i^2}(x) - V(x)r(x) = -Er(x) \quad . \quad (2.35)$$

Let the solutions be labelled r_m with corresponding eigenvalues E_m , with $E_m \leq E_{m+1}$. Then where u is as defined in (2.30), we shall have

$$u(t, x, y) = \frac{1}{\varphi_{T-t}(x, y)} \sum_{m=0}^{\infty} \exp(-E_m(T-t)) r_m(x) r_m(y) \quad . \quad (2.36)$$

Proof. In works such as [Kac56, Dyn65, Fre85], it has been shown, via the probabilistic representation

$$u(t, x, y) = \mathbb{E} [\exp (S_{t,x}(T))] \quad (2.37)$$

where X is given by (2.21):

$$dX_i = \frac{y_i - X_i}{T-s} ds + dW_i; \quad t \leq s < T; \quad X_i(t) = x_i, \quad (2.38)$$

and $S_{t,x}$ is given by

$$dS = -V(X(t))dt; \quad S_{t,x}(t) = 0 \quad (2.39)$$

that u satisfies the Cauchy problem with the backward Kolmogorov equation:

$$\forall t < T : \forall x \in \mathbb{R}^{nd} : Lu - V(x)u = 0; \quad (2.40)$$

$$\forall x \in \mathbb{R}^{nd} : u(T, x, y) = 1 ,$$

where

$$L = \frac{\partial}{\partial t} + \sum_{i=1}^d \frac{y_i - x_i}{T - t} \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2}. \quad (2.41)$$

However, a transformation of u also gives us the forward Kolmogorov equation, or Fokker-Planck equation: let $v : [0, T] \times \mathbb{R}^{nd} \times \mathbb{R}^{nd} \rightarrow \mathbb{R}$ be given by

$$v(t, x, y) = \varphi_t(x, y)u(T - t, x, y) \quad (2.42)$$

and we will have, instead of (2.40),

$$\forall t > 0 : \frac{\partial v}{\partial t} = \frac{1}{2} \sum_{i=1}^{nd} \frac{\partial^2 v}{\partial x_i^2} - V(x)v , \quad (2.43)$$

$$\lim_{t \rightarrow 0} v(t, x, y) = \delta(\|x - y\|) , \quad (2.44)$$

where in this last, δ is understood to signify the Dirac delta. (We have exchanged the singularity in operator L for singularity in the initial condition.) It is proven in works such as [Fri64] that the Cauchy problem (2.43)-(2.44) has a classical solution (cf Theorem 16 of [Fri64]; the conditions of this theorem are satisfied because we assumed V to be bounded below). We shall attempt to find this solution via separation of variables, writing

$$v(t, x, y) = v(t; y)r(x) \quad (2.45)$$

with $r : \mathbb{R}^{nd} \rightarrow \mathbb{R}$ and $v(\cdot; y) : [0, T] \rightarrow \mathbb{R}$. The solutions are characterised by the eigenvalue problem, for $E \in \mathbb{R}$ (cf (2.35)):

$$v(t; y) = A(y)\exp(-Et) \quad (2.46)$$

$$\frac{1}{2} \sum_{i=1}^{nd} \frac{\partial^2 r}{\partial x_i^2}(x) - V(x)r(x) = -Er(x) \quad (2.47)$$

By a result proven in [Tit58, Fri73], we have that since we assumed $V \rightarrow \infty$ as $\|x\| \rightarrow \infty$, the spectrum of solutions to (2.35) is discrete; in fact, in general when V is greater than some

value β for all sufficiently large $\|x\|$, the spectrum is discrete for $E < \beta$ (cf Theorem 16.5 of [Tit58]). We shall apply an index $m \in \{0, 1, 2, \dots\}$ to the eigenfunctions and corresponding eigenvalues, with $E_m \leq E_{m+1}$. For our purposes in the following we shall choose the r_m to be normalized. Now for $t > 0$, v is given by

$$v(t, x, y) = \sum_{m=0}^{\infty} A_m(y) \exp(-E_m t) r_m(x) \quad (2.48)$$

and we may determine the coefficients $A_m(y)$ from the initial condition as follows. Firstly, rewrite (2.48), for $t \in (0, T]$, as

$$v(t, x, y) = \sum_{m=0}^{\infty} v_m(t; y) r_m(x)$$

and notice that

$$v_m(t; y) = \int_{\mathbb{R}^{nd}} \sum_{l=0}^{\infty} v_l(t; y) r_l(x) r_m(x) dx = \int_{\mathbb{R}^{nd}} v(t, x, y) r_m(x) dx \quad .$$

However, we know from our initial condition that for any measurable function $f : \mathbb{R}^{nd} \rightarrow \mathbb{R}$,

$$\lim_{t \downarrow 0} \int_{\mathbb{R}^{nd}} v(t, x, y) f(x) dx = f(y)$$

and so it is apparent that $\lim_{t \downarrow 0} v_m(t) = r_m(y)$, where we write $t \downarrow 0$ to indicate the limit from above. Therefore, clearly $A_m(y) = r_m(y)$. Hence for $t \in [0, T]$, we may write

$$v(t, x, y) = \sum_{m=0}^{\infty} \exp(-E_m t) r_m(x) r_m(y) \quad . \quad (2.49)$$

Therefore

$$u(t, x, y) = \frac{1}{\varphi_{T-t}(x, y)} \sum_{m=0}^{\infty} \exp(-E_m(T-t)) r_m(x) r_m(y) \quad . \quad (2.50)$$

■

In the case of a Coulombic potential, the spectrum is discrete up to a certain threshold, and in physical terms the essential spectrum at higher energies corresponds to the escape of the electron from the atom. However, to avoid this issue we shall always deal with potentials where the condition $V \rightarrow \infty$ as $\|x\| \rightarrow \infty$ is imposed, and hence where Theorem 2.2.1 applies.

2.2.2 Physical interpretation of the conditional Wiener integral

In this subsection, we seek to explain the ramifications of Theorem 2.2.1, in terms of the physical significance of the action integral. We reflect on the relationship between (2.35), as considered hitherto, and the time-independent Schrödinger equation, in order to conclude that functional integrals can be used to represent the so-called thermal density matrix for a canonical system.

Position density in the case of distinguishable particles

It will be recognised that if in the place of real-valued r , we consider a complex-valued spatial wavefunction ψ , neglecting spin, then (2.35) is the stationary Schrödinger equation (e.g. [AF97, LMR97]), which describes wavefunctions ψ_m corresponding to states of definite energies E_m ; for the left-hand side is $H\psi$ where H is the Hamiltonian operator. However, it is then clear that any solution ψ must be given by some combination $r_1 + ir_2$ where r_1 and r_2 are real-valued functions solving (2.35) which share the same eigenvalue. It is also clear that any linear combination of r_1, r_2 with these conditions is an eigenfunction; we restrict our attention to normalized wavefunctions. In summary, then, the complex-valued solutions to (2.35) with eigenvalue E_l are given by $\sum_{i=1}^{M_l} \alpha_{li} r_{li}$ where r_{li} are the M_l real-valued orthonormal eigenfunctions corresponding to E_l and $\alpha_{li} \in \mathbb{C}$ have $\sum_{i=1}^{M_l} |\alpha_{li}|^2 = 1$.

It is a well-known fact in statistical mechanics (see e.g. [Fey72, p.60]) that for a system of n identical particles which is at a fixed temperature $1/k_B T$, we have the Boltzmann distribution over the states of definite energy, and this means that the normalized density for system position $\rho(x)$ can be found via

$$q_D(x_0, x_T) = \sum_{m=0}^{\infty} \exp(-E_m T) \psi_m(x_0) \psi_m^*(x_T) \quad (2.51)$$

$$\mathcal{Z} = \sum_{m=0}^{\infty} \exp(-E_m T) = \int_{\mathbb{R}^{nd}} q_D(x, x) dx \quad (2.52)$$

$$\rho(x) = \frac{q_D(x, x)}{\mathcal{Z}}. \quad (2.53)$$

Usually q_D is called the unnormalized thermal density matrix and \mathcal{Z} is called the partition

function. Since the physically relevant eigenstates will be orthonormal (see e.g. [AF97, LMR97]), they must have the same multiplicity M_l as the r_{li} . This justifies retaining the same index m in (2.51) as in (2.36).

It is usually assumed in the physics literature (as in, for example, the treatment given in [Fey72]) that in general where ψ_{li} is any choice of orthonormal basis for $\text{span}(\{r_{li}\})$, at any pair of system points $x_0, x_T \in \mathbb{R}^{nd}$,

$$\sum_{i=1}^{M_l} \psi_i(x_0) \psi_i^*(x_T) = \sum_{i=1}^{M_l} r_i(x_0) r_i(x_T). \quad (2.54)$$

This is not difficult to verify in the case of multiplicity 2.

If we adopt this assumption, it implies in particular that any choice of orthonormal basis ψ_{li} for the space spanned by the r_{li} will have the same sum of squared moduli of the basis functions when evaluated at any system point. That is, for any $x \in \mathbb{R}^{nd}$, if ψ_{li} are an orthonormal basis for $\text{span}(r_{li})$,

$$\sum_{i=1}^{M_l} \exp(-E_l T) |\psi_{li}(x)|^2 = \sum_{i=1}^{M_l} \exp(-E_l T) r_{li}^2(x) \quad .$$

(For example, in the case that the eigenfunctions have multiplicity 1, in fact any solution ψ_m with eigenvalue E_m is $r_m e^{i\theta_m}$ for some phase θ_m . The value of θ_m will then in fact be determined by the time-dependent Schrödinger equation, but we shall not discuss this further since the phases have no impact on (2.51).)

Thus in fact, it now follows from (2.36) that

$$\begin{aligned} q_D(x, x) &= \sum_{l=0}^{\infty} \sum_{i=1}^{M_l} \exp(-E_l T) |\psi_{li}(x)|^2 = \sum_{l=0}^{\infty} \sum_{i=1}^{M_l} \exp(-E_l T) r_{li}^2(x) = \sum_{m=0}^{\infty} \exp(-E_m T) (r_m^2(x)) \\ &= \varphi_T(x, x) \mathcal{J}_T(x, x) \end{aligned}$$

and therefore

$$\mathcal{Z} = \int_{\mathbb{R}^{nd}} \varphi_T(x, x) \mathcal{J}_T(x, x) dx = \int_{\mathbb{R}^{nd}} \varphi_T(y, y) \int_{C_{0,y}^{T,y}} \exp \left[- \int_0^T V(x(t)) dt \right] dw_{0,y}^{T,y}(x) dy \quad (2.55)$$

$$\rho(x) = \frac{q_D(x, x)}{\mathcal{Z}} = \frac{\mathcal{J}_T(x, x)}{\int_{\mathbb{R}^{nd}} \mathcal{J}_T(x, x) dx} \quad .$$

In other words, we can find $\rho(x)$ in terms of functional integrals. Moreover, considering (2.54),

$$\begin{aligned} q_D(x_0, x_T) &= \sum_{l=0}^{\infty} \sum_{i=1}^{M_l} \exp(-E_l T) \psi_{li}(x_0) \psi_{li}^*(x_T) = \sum_{m=0}^{\infty} \exp(-E_m T) r_m(x_0) r_m(x_T) \\ &= \varphi_T(x_0, x_T) \mathcal{J}_T(x_0, x_T) \end{aligned} \quad (2.56)$$

which is important for the following discussion.

Position density in the case of indistinguishable particles

We shall let Π_n indicate the permutation group of order n , and write πx for $(x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(n)})$.

We note from p.61 and p.64 of [Fey72] that for indistinguishable particles, the unnormalized density may be found via

$$q(x, x') = \frac{1}{n!} \sum_{\pi \in \Pi_n} c(\pi) q_D(x, \pi x') \quad (2.57)$$

where for bosons, $c(\pi)$ is always 1, and for fermions, $c(\pi)$ represents $\text{sgn}(\pi)$, the signature of the permutation π . Here we consider only a system described by either a symmetric or anti-symmetric spatial wavefunction; the spin wavefunction is not considered. The corresponding partition function is then

$$\begin{aligned} \mathcal{Z} &= \int_{\mathbb{R}^{nd}} q(x, x) dx = \frac{1}{n!} \int_{\mathbb{R}^{nd}} \sum_{\pi \in \Pi_n} c(\pi) q_D(x, \pi x) dx \\ &= \frac{1}{n!} \int_{\mathbb{R}^{nd}} \sum_{\pi \in \Pi_n} c(\pi) \varphi_T(x, \pi x) \mathcal{J}_T(x, \pi x) dx. \end{aligned} \quad (2.58)$$

This is sometimes called the **trace** of the (unnormalized) density matrix q . Therefore we should conclude that for indistinguishable particles,

$$\rho(x_0) = \frac{q(x_0, x_0)}{\int_{\mathbb{R}^{nd}} q(x, x) dx} = \frac{\left(\frac{1}{n!} \sum_{\pi \in \Pi_n} \varphi_T(x_0, \pi x_0) \right) \mathcal{I}_T(x_0, x_0)}{\int_{\mathbb{R}^{nd}} \left(\frac{1}{n!} \sum_{\pi \in \Pi_n} \varphi_T(x, \pi x) \right) \mathcal{I}_T(x, x) dx} \quad (2.59)$$

$$= \left(\frac{1}{n!} \sum_{\pi \in \Pi_n} \varphi_T(x_0, \pi x_0) \right) \frac{\mathcal{I}_T(x_0, x_0)}{\mathcal{Z}} \quad (2.60)$$

$$= \frac{1}{n!} \frac{\mathcal{G}_T(x_0, x_0)}{\mathcal{Z}} = \frac{\mathcal{G}_T(x_0, x_0)}{\int_{\mathbb{R}^{nd}} \mathcal{G}_T(x, x) dx}, \quad (2.61)$$

where

$$\mathcal{I}_T(x_0, x_T) = \frac{1}{\sum_{\pi \in \Pi_n} \varphi_T(x_0, \pi x_T)} \sum_{\pi \in \Pi_n} c(\pi) \varphi_T(x_0, \pi x_T) \mathcal{J}_T(x_0, \pi x_T) \quad (2.62)$$

$$= \frac{n!}{\sum_{\pi \in \Pi_n} \varphi_T(x_0, \pi x_T)} q(x_0, x_T) \quad (2.63)$$

$$= \frac{1}{\sum_{\pi \in \Pi_n} \varphi_T(x_0, \pi x_T)} \sum_{m=0}^{\infty} \exp(-E_m T) \sum_{\pi \in \Pi_n} c(\pi) r_m(x_0) r_m(\pi x_T) \quad , \quad (2.64)$$

$$\mathcal{G}_T(x_0, x_T) = \left(\sum_{\pi \in \Pi_n} \varphi_T(x_0, \pi x_T) \right) \mathcal{I}_T(x_0, x_T) . \quad (2.65)$$

The reader is invited to compare (2.63) with (2.56). As we shall see, (2.59) enables $\rho(x)$ and associated quantities to be calculated quite efficiently for bosons. For fermions, computations are made more challenging by the so-called fermion sign problem, which is the subject of Chapters 5-7. Lastly, note that it follows that where ϕ is used for the wavefunctions that apply in the indistinguishable case,

$$\phi_m(x_0) \phi_m^*(x_T) = \frac{1}{n!} \sum_{\pi \in \Pi_n} c(\pi) r_m(x_0) r_m(\pi x_T) . \quad (2.66)$$

2.2.3 Indistinguishability in terms of measure conditioning

Considering (2.62), it will be apparent that we can write it as one functional integral, using a different conditional measure. When $X(T) = \pi x$, let $c\left(\frac{X(T)}{x}\right)$ mean $c(\pi)$. Then

$$\mathcal{I}_T(x, x) = \int_{\mathcal{C}_{0,x}^{T, \Pi_n x}} c\left(\frac{X(T)}{x}\right) \exp\left(\int_0^T -V(X(s)) ds\right) dw_{0,x}^{T, \Pi_n x} . \quad (2.67)$$

This suggests an alternative heuristic for the finding of the previous subsection, that $\mathcal{I}_T(x, x)$ represents the position density of the system at inverse temperature T . For if it is accepted that for distinguishable particles, $\rho(x) = \mathcal{J}_T(x, x) / \int_{\mathbb{R}^{nd}} \mathcal{J}_T(x, x) dx$, then for indistinguishable particles the only change necessary is to allow all possible particle permutations at the terminal point. (The introduction of $\text{sgn}\left(\frac{X(T)}{x}\right)$ for fermions is unexplained in such an heuristic.)

Note that we may, of course, regard V as a function defined on \mathbb{R}^{nd} / Π_n since it is necessarily independent of any permutation of the particles. Consequently Y may be viewed as a

functional defined on continuous paths with values in \mathbb{R}^{nd}/Π_n . Then, recalling Subsection 2.1.4, since $w_{0,\Pi_n x}^{T,\Pi_n x}(\Pi_n B) = w_{0,x}^{T,\Pi_n x}(\Pi_n B)$ for any $B \in \mathcal{B}(\mathcal{C}_0^T)$,

$$\mathcal{I}_T(x, x) = \int_{\mathcal{C}_{0,\Pi_n x}^{T,\Pi_n x}} c\left(\frac{X(T)}{x}\right) Y(X) dw_{0,\Pi_n x}^{T,\Pi_n x} := \mathcal{I}_T(\Pi_n x, \Pi_n x) . \quad (2.68)$$

The fact that this holds provides another intuition regarding the nature of \mathcal{I}_T , and is often a helpful perspective to recall when thinking about results involving \mathcal{I}_T . It is especially useful for understanding how to implement permutation sampling (cf Subsection 4.2.3). In a similar spirit to that of (2.68), we may note that

$$\begin{aligned} \mathcal{I}_T(x_0, x_T) &= \int_{\mathcal{C}_{0,x_0}^{T,\Pi_n x_T}} c\left(\frac{X(T)}{x_T}\right) \exp\left(\int_0^T -V(X(s))ds\right) dw_{0,x_0}^{T,\Pi_n x_T} \\ &= \int_{\mathcal{C}_{0,\Pi_n x_0}^{T,\Pi_n x_T}} c\left(\frac{X(T)}{x_T}\right) Y(X) dw_{0,\Pi_n x_0}^{T,\Pi_n x_T} := \mathcal{I}_T(\Pi_n x_0, \Pi_n x_T) . \end{aligned}$$

2.2.4 Finding expectations of observables

For the expectation of an observable whose corresponding operator is diagonal in the position representation [Kle95],

$$\langle A \rangle = \int_{\mathbb{R}^{nd}} A(x) \rho(x) dx \quad (2.69)$$

$$= \frac{\int_{\mathbb{R}^{nd}} A(x) \left(\sum_{\pi \in \Pi_n} \varphi_T(x, \pi x) \right) \mathcal{I}_T(x, x) dx}{\int_{\mathbb{R}^{nd}} \left(\sum_{\pi \in \Pi_n} \varphi_T(x, \pi x) \right) \mathcal{I}_T(x, x) dx} . \quad (2.70)$$

More generally, according to Chapter 2 of [Fey72], if an observable corresponds to an operator A then we may write the expectation of this observable as (see also [Kle95] p.108):

$$\langle A \rangle = \int \int_{\mathbb{R}^{2nd}} A(x, x') \frac{q(x, x')}{\mathcal{Z}} dx dx' \quad (2.71)$$

where q is the unnormalized density, defined by (2.57). (Our $A(x, x')$ here is referred to as $\langle x|A|x' \rangle$ in some texts.)

Potential energy

One observable which obeys (2.69) is potential energy, as mentioned in Chapter 1:

$$\langle V \rangle = \frac{\int_{\mathbb{R}^{nd}} V(x) \left(\sum_{\pi \in \Pi_n} \varphi_T(x, \pi x) \right) \mathcal{I}_T(x, x) dx}{\int_{\mathbb{R}^{nd}} \left(\sum_{\pi \in \Pi_n} \varphi_T(x, \pi x) \right) \mathcal{I}_T(x, x) dx}$$

Two-point correlation function

For $d = 1$, a two-point correlation function $\Gamma(\theta)$, $0 \leq \theta \leq T$, has the form (see [Lob96]):

$$\begin{aligned} \Gamma(\theta) &= \langle x(0)x(\theta) \rangle \\ &= \frac{1}{\mathcal{Z}} \int_{-\infty}^{\infty} \int_{\mathcal{C}_{0,y}^{T,y}} x(0)x(\theta) \exp \left(- \int_0^T V(t, x(t)) dt \right) dw_{0,y}^{T,y}(x) dy \\ &= \frac{1}{\mathcal{Z}} \int_{-\infty}^{\infty} \int_{\mathcal{C}_{0,y}^{T,y}} yx(\theta) \exp \left(- \int_0^T V(t, x(t)) dt \right) dw_{0,y}^{T,y}(x) dy, \end{aligned} \tag{2.72}$$

Correlation functions contain important information about quantum-mechanical systems and they are observable in scattering experiments (see, e.g. [Kle95]).

Kinetic energy

In order to discuss kinetic energy we need to take account of mass. Let us consider a more general definition of \mathcal{J}_T :

$$\mathcal{J}_T(x, \pi x) = \mathbb{E} \exp \left(-S \left(X_{0,x}^{T,\pi x} \right) \right)$$

where we let $X_{0,x}^{T,\pi x}(t)$ solve the nd -dimensional system of SDEs

$$dX = \frac{\pi x - X}{T - t} dt + \frac{1}{\sqrt{m}} dw(t), \quad 0 \leq t < T, \quad X(0) = x. \tag{2.73}$$

where m here represents the mass of one particle, taken in our customary definition of \mathcal{J}_T to be 1. Taking account of mass, the partition function for a system of bosons has the form:

$$\mathcal{Z} = \int_{\mathbb{R}^{nd}} \sum_{\pi \in \Pi_n} \varphi_{T/m}(x, \pi x) \mathcal{J}_T(x, \pi x) dx \tag{2.74}$$

$$= \left(\sum_{\pi \in \Pi_n} \varphi_{T/m}(x, \pi x) \right) \int_{\mathbb{R}^{nd}} \mathcal{I}(x, x) dx. \tag{2.75}$$

It can be shown to follow from (2.71) that the kinetic energy is given by (see also [TI84]):

$$\langle K \rangle = \frac{m}{T\mathcal{Z}} \frac{\partial \mathcal{Z}}{\partial m}.$$

We have, in the case of bosons,

$$\begin{aligned} \frac{\partial \mathcal{Z}}{\partial m} &= \int_{\mathbb{R}^{nd}} \sum_{\pi \in \Pi_N} \varphi_{T/m}(x, \pi x) \\ &\quad \times \left[\frac{\partial \mathcal{J}_T(x, \pi x)}{\partial m} + \mathcal{J}_T(x, \pi x) \left(\frac{nd}{2m} - \frac{\|x - \pi x\|^2}{2T} \right) \right] dx \end{aligned}$$

and

$$\frac{\partial \mathcal{J}_T(x, \pi x)}{\partial m} = -\mathbb{E} \left[\exp\left(-\int_0^T V(X_{0,x}^{T,\pi x}(t)) dt\right) \int_0^T \nabla V(X_{0,x}^{T,\pi x}(t)) \cdot \frac{d}{dm} X_{0,x}^{T,\pi x}(t) dt \right].$$

Let $Q(t) = \frac{d}{dm} X_{0,x}^{T,\pi x}(t)$, that is, the derivative of the solution to (2.73) w.r.t. the parameter m . This process satisfies the SDE (see [GS72])

$$dQ = -\frac{Q}{T-t} dt - \frac{1}{2\sqrt{m^3}} dw(t), \quad 0 \leq t < T, \quad Q(0) = 0.$$

Clearly,

$$\begin{aligned} Q(t) &= -X_{0,0}^{T,0}(t)/(2m) \\ &= -\left(X_{0,x}^{T,\pi x}(t) - \frac{x}{T}(T-t) - \frac{\pi x}{T}t\right)/(2m). \end{aligned}$$

Thus one obtains (see also [DT10]):

$$\langle K \rangle = \frac{m}{T\mathcal{Z}} \mathcal{K}, \tag{2.76}$$

where

$$\begin{aligned} \mathcal{K} &= \int_{\mathbb{R}^{nd}} \left[\sum_{\pi \in \Pi_n} \varphi_{T/m}(x, \pi x) \mathbb{E} \left[\exp\left(-\int_0^T V\left(X_{0,x}^{T,\pi x}(t)\right) dt\right) \right. \right. \\ &\quad \left. \left. \times \left(\frac{nd}{2m} - \frac{\|x - \pi x\|^2}{2T} + \frac{1}{2m} \int_0^T \nabla V\left(X_{0,x}^{T,\pi x}(t)\right) \cdot \left(X_{0,x}^{T,\pi x}(t) - \frac{x}{T}(T-t) - \frac{\pi x}{T}t\right) dt \right) \right] \right] dx. \end{aligned} \tag{2.77}$$

Here ∇V is an nd -dimensional vector. We note that this expression for the kinetic energy is different to the ones exploited in [Cep95, TI84]. (As was pointed out in [Cep95], it is desirable for computational purposes to have various representations of the kinetic energy.)

2.3 Using noninteracting copies to approximate states of definite energy

We shall now devote further attention to the subject matter of Subsection 2.2.2, in order to demonstrate a result which, in effect, permits simulations of a quantum system that is in a state of definite energy. The problem of simulating excited states has previously been addressed by at least two different methods, known as Correlation Function Monte Carlo [CB88, Cep96ii] and Thermo Field Monte Carlo [Suz86]. Given a suitable method for performing zero-temperature simulations, the results of this section give rise to a new, alternative approach to simulating excited states.

Recalling (2.64)-(2.66), it is clear that

$$\mathcal{I}_T(x_0, x_T) = \frac{n!}{\sum_{\pi \in \Pi_n} \varphi_T(x_0, \pi x_T)} \sum_{m=0}^{\infty} \exp(-E_m T) \phi_m(x_0) \phi_m^*(x_T) \quad .$$

In fact, we may obtain the densities $\phi_m \phi_m^*$ as limits of expressions involving $\mathcal{I}_T(x, x)$, and this shall now be discussed. It is immediately obvious that as T tends to ∞ , the terms with eigenvalues equal to E_0 dominate and thus q_T converges weakly to $\phi_0 \phi_0^*$ as long as $E_0 \neq E_1$. By implication, in particular if we wish to find a statistic such as the expected value in the ground state of an observable A , then

$$\langle A \rangle_0 = \int_{\mathbb{R}^{nd}} A(x) \phi_0 \phi_0^*(x) dx = \lim_{T \rightarrow \infty} \int_{\mathbb{R}^{nd}} A(x) q_T(x, x) dx = \lim_{T \rightarrow \infty} \frac{\int_{\mathbb{R}^{nd}} A(x) \mathcal{I}_T(x, x) dx}{\int_{\mathbb{R}^{nd}} \mathcal{I}_T(x, x) dx} \quad (2.78)$$

At temperature zero, the system is effectively confined to the ground state(s) (see [FH65, Fey72]). As shall be explained in Chapter 4, Formula (2.78) effectively tells us how to collect approximate ground state statistics using Path Integral Monte Carlo, if a sufficiently large T can be used that adequate convergence is achieved, because the integrals and indeed $\mathcal{I}_T(x, x)$, for a given x , are quantities which can be obtained as the limiting values from simulations. As is explained in [Fey72], the average over ground states will be obtained in the case of ground state degeneracy.

However, it is also possible to construct a controlled approximation to the densities for

the excited states of the system. We shall now prove several results which indicate a way to express the sum of the first k eigenstates in terms of functional integrals, with the caveat that there is a specific type of averaging over eigenstates if $1, \dots, k$ should happen to include only some of the eigenstates corresponding to E_k . The approach taken here is to consider first the case of distinguishable particles, and then to proceed to considering bosons and fermions.

2.3.1 Results on how to obtain sums of eigenstates

Let us consider what happens when we have k noninteracting copies of the system of n distinguishable particles inhabiting a d -dimensional space. Since the copies are noninteracting, we think of the potential for the knd -dimensional coordinates as being the sum of the potentials at each copy system position. We shall use x for the full system coordinate vector $x = (x_1, \dots, x_k) \in \mathbb{R}^{knd}$. We shall overload $V : \mathbb{R}^{knd} \rightarrow \mathbb{R}$ to be given by $V(x) = V(x_1, x_2, \dots, x_k) = V(x_1) + V(x_2) + \dots + V(x_k)$. This means then that, e.g.

$$\exp \left(\int_0^T -V(x_1(t), x_2(t)) dt \right) = \exp \left(\int_0^T -V(x_1(t)) dt \right) \exp \left(\int_0^T -V(x_2(t)) dt \right) , \quad (2.79)$$

and hence where copies have sources (x_1, \dots, x_k) and destinations (y_1, \dots, y_k) , and X_i are for the paths followed by each system copy,

$$\begin{aligned} \mathcal{J}_T(x, y) &= \int_{\mathcal{C}_{0,x}^{T,y}} \prod_{j=1}^k \exp \left(\int_0^T -V(X_j(t)) dt \right) dw_{0,x}^{T,y} \\ &= \prod_{j=1}^k \int_{\mathcal{C}_{0,x_j}^{T,y_j}} \exp \left(\int_0^T -V(X_j(t)) dt \right) dw_{0,x_j}^{T,y_j} \\ &= \mathcal{J}_T(x_1, y_1) \mathcal{J}_T(x_2, y_2) \dots \mathcal{J}_T(x_k, y_k) \end{aligned} \quad (2.80)$$

Remark 2.3.1 *In this case (2.35) becomes*

$$\frac{1}{2} \sum_{j=1}^k \sum_{i=1}^{nd} \frac{\partial^2 r_m^{(k)}}{\partial x_{ji}^2}(x) - \sum_{j=1}^k V(x_j) r_m^{(k)}(x) = -F_m r_m(x) \quad (2.81)$$

and then the eigenstates for the joint system with the k independent copies, which we shall denote $r_m^{(k)} : \mathbb{R}^{knd} \rightarrow \mathbb{R}$, are given by ordered products of k elements from the r_m , with

replacement. Clearly $r_{m_1} r_{m_2} \dots r_{m_k}$ has eigenvalue $E_{m_1} + E_{m_2} + \dots + E_{m_k}$. The fact that these are the only eigenfunctions follows from (2.80). Then for $x, y \in \mathbb{R}^{knd}$, there is a sequence of eigenvalues F_m and corresponding eigenfunctions $r_m^{(k)}$ s.t.

$$\begin{aligned} \mathcal{J}_T(x, y) &= \frac{1}{\varphi_T(x, y)} \sum_{m=0}^{\infty} \exp(-F_m T) r_m^{(k)}(x) r_m^{(k)}(y) \\ &= \frac{1}{\prod_{j=1}^k \varphi_T(x_j, y_j)} \sum_{m_1=0}^{\infty} \dots \sum_{m_k=0}^{\infty} \left[\exp\left(-\sum_{j=1}^k E_{m_j} T\right) \prod_{j=1}^k r_{m_j}(x_j) r_{m_j}(y_j) \right] \end{aligned} \quad (2.82)$$

For example, with $k = 2$, we have, where $F_m = E_{m_1} + E_{m_2}$ for $m = 0, 1, \dots$:

$$\begin{aligned} \mathcal{J}_T(x, y) &= \frac{1}{\varphi_T(x, y)} \sum_{m=0}^{\infty} \exp(-F_m T) r_m^{(2)}(x) r_m^{(2)}(y) \\ &= \frac{1}{\varphi_T(x_1, y_1) \varphi_T(x_2, y_2)} \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} \exp(-(E_{m_1} + E_{m_2}) T) r_{m_1}(x_1) r_{m_2}(x_2) r_{m_1}(y_1) r_{m_2}(y_2) \end{aligned}$$

Theorem 2.3.2 (sum to kth state of densities for distinguishable particles) Let $J_{T,k} :$

$\mathbb{R}^{knd} \rightarrow \mathbb{R}$ be given by

$$J_{T,k}(x) = \sum_{\zeta \in \Pi_k} \text{sgn}(\zeta) \varphi_T(x, \zeta x) \mathcal{J}_T(x, \zeta x)$$

where ζx indicates that systems are permuted (with particle indexing maintained within). Let the sequence of multiplicities for (2.81) be M_0, M_1, \dots . Let $M^* = M_0 + M_1 + \dots + M_{p-1}$, and let $k = M^* + k^*$ with $k^* \leq M_p$. Then

$$\sum_{m=0}^{M^*-1} r_m^2(x_1) + \frac{k^*}{M_p} \sum_{m=M^*}^{M^*+M_p-1} r_m^2(x_1) = k \lim_{T \rightarrow \infty} \frac{\int_{\mathbb{R}^{(k-1)nd}} J_{T,k}(x) dx_2 \dots dx_k}{\int_{\mathbb{R}^{knd}} J_{T,k}(x) dx_1 \dots dx_k} \quad (2.83)$$

$$= k \lim_{T \rightarrow \infty} Z_{T,k}^{-1} \int_{\mathbb{R}^{(k-1)nd}} dx_2 \dots dx_k \left[\sum_{\zeta \in \Pi_k} \text{sgn}(\zeta) \sum_{m=0}^{\infty} \exp(-F_m T) r_m^{(k)}(x) r_m^{(k)}(\zeta x) \right] \quad (2.84)$$

with $Z_{T,k} = \int_{\mathbb{R}^{knd}} J_{T,k}(x) dx_1 \dots dx_k$.

Proof. Recognise from (2.82) that

$$\sum_{m=0}^{\infty} \exp(-F_m T) r_m^{(k)}(x) r_m^{(k)}(\zeta x) = \sum_{m_1=0}^{\infty} \dots \sum_{m_k=0}^{\infty} \left[\exp\left(-\sum_{j=1}^k E_{m_j} T\right) \prod_{j=1}^k r_{m_j}(x_j) r_{m_j}(x_{\zeta(j)}) \right] \quad (2.85)$$

Recall that we numbered the eigenstates so that $E_m \leq E_{m+1}$. The proof is based on the idea that in the limit as $T \rightarrow \infty$, the dominant terms in (2.84) will be those with the minimum sum of eigenvalues.

We need to start by showing that all the terms containing repeated choices of eigenstates $m_j = m_l$, $j \neq l$, are zero. So suppose that we pick any term with $m_j = m_l$, some $j < l$. It is clear that Π_k can be partitioned into disjoint subsets, each of two elements which are bijective under the exchange of elements j and l . This is clear because given any $\zeta \in \Pi_k$, applying this operation will yield a different ζ , but this operation is self-inverse. However, if we exchange the order of summations, we shall have

$$J_{T,k}(x) = \sum_{m=0}^{\infty} \exp(-F_m T) \sum_{\zeta \in \Pi_k} \text{sgn}(\zeta) r_m^{(k)}(x) r_m^{(k)}(\zeta x)$$

and then we may note that each sum $\sum_{\zeta \in \Pi_k}$ here is given by the sum of the added contributions from each of these pairs of permutations. However, for our chosen term, the added contributions from each pair is zero because for all x , the permutations in each pair give equal values to $r_m^{(k)}(x) r_m^{(k)}(\zeta x)$ but they have opposite signs for $\text{sgn}(\zeta)$. (Note that this is regardless of whether $E_j = E_l$.) Therefore the only terms which have a nonzero contribution to $J_{T,k}$ are indeed those where $m_j \neq m_l$ for $j \neq l$.

To find the contribution to $\int_{\mathbb{R}^{(k-1)nd}} J_{T,k}(x) dx_2 \cdots dx_k$ from these terms, consider what will happen for permutations other than the identity. Firstly note that

$$r_m(\zeta x) = \prod_{j=1}^k r_{m_{\zeta^{-1}(j)}}(x_j) \quad (2.86)$$

and suppose that for some system copy x_l we have $\zeta^{-1}(l) \neq l$ for some $l > 1$. Then notice that

$$\int_{\mathbb{R}^{(k-1)nd}} \prod_{j=1}^n r_{m_j}(x_j) r_{m_{\zeta^{-1}(j)}}(x_j) dx_2 \cdots dx_n$$

can be separated; ie let us consider the integral w.r.t. dx_l to be performed last, treating the result from integrating over the other variables as constant. Due to the orthogonality of the r_m it then follows, however, that the result must be zero. Therefore in (2.84) the only

contribution to any term with no repeated eigenstate indices is from $\zeta = id$. Thus,

$$\begin{aligned}
\int_{\mathbb{R}^{(k-1)nd}} J_{T,k}(x) dx_2 \dots dx_n &= \\
\int_{\mathbb{R}^{(k-1)nd}} dx_2 \dots dx_n &\left[\sum_{\substack{m_j \in \{0,1,\dots\} \text{ for } j=\{1,\dots,k\}; \\ m_j \neq m_l \text{ for } j \neq l}} \exp \left(-\sum_{j=1}^k E_{m_j} T \right) \prod_{j=1}^k r_{m_j}^2(x_j) \right] = \\
\sum_{m_1=0}^{\infty} &\left[\exp(-E_{m_1} T) r_{m_1}^2(x_1) \sum_{\substack{m_j \in \{0,1,\dots\} \text{ for } j=\{2,\dots,k\}; \\ m_j \neq m_l \text{ for } j \neq l, l \in \{1,\dots,k\}}} \exp \left(-\sum_{j=2}^k E_{m_j} T \right) \int_{\mathbb{R}^{(k-1)nd}} \prod_{j=2}^k r_{m_j}^2(x_j) dx_2 \dots dx_n \right] \\
&= \sum_{m_1=0}^{\infty} \left[\exp(-E_{m_1} T) r_{m_1}^2(x_1) \sum_{\substack{m_j \in \{0,1,\dots\} \text{ for } j=\{2,\dots,k\}; \\ m_j \neq m_l \text{ for } j \neq l, l \in \{1,\dots,k\}}} \exp \left(-\sum_{j=2}^k E_{m_j} T \right) \right] \quad (2.87)
\end{aligned}$$

If we consider that the summands are equal for reorderings of $\{m_2, \dots, m_k\}$ then counting through all $\zeta \in \Pi_{k-1}$ we shall find

$$\int_{\mathbb{R}^{(k-1)nd}} J_{T,k}(x) dx_2 \dots dx_n = (k-1)! \sum_{m_1=0}^{\infty} \sum_{\substack{m_k > m_{k-1} > \dots > m_2 \geq 0; \\ m_j \neq m_l, j \in \{2, \dots, k\}}} \exp \left(-\sum_{j=1}^k E_{m_j} T \right) r_{m_1}^2(x_1) \quad (2.88)$$

and meanwhile

$$\int_{\mathbb{R}^{knd}} J_{T,k}(x) dx_1 \dots dx_n = (k-1)! \sum_{m_1=0}^{\infty} \sum_{\substack{m_k > m_{k-1} > \dots > m_2 \geq 0; \\ m_j \neq m_l, j \in \{2, \dots, k\}}} \exp \left(-\sum_{j=1}^k E_{m_j} T \right) \quad (2.89)$$

$$= k! \sum_{m_k > m_{k-1} > \dots > m_1 \geq 0} \exp \left(-\sum_{j=1}^k E_{m_j} T \right). \quad (2.90)$$

Now let us ask what terms in (2.88) give the minimum value to $\sum_{j=1}^k E_{m_j}$. Clearly these consist of choosing all the values up to $M^* - 1$, and also k^* values from $\{M^*, \dots, M^* + M_p - 1\}$. Thus if we consider the full unordered set of m_j to be sampled without replacement over $\{1, \dots, M^* + M_p - 1\}$ with $\{1, \dots, M^* - 1\}$ fully populated, we are then allowing m_1 to range over this set, in order to count all the dominant terms in (2.88). (This also counts the dominant terms in (2.90), of which there are then $\binom{M_p}{k^*} k$.) Thus, there are $\binom{M_p}{k^*}$ choices which put m_1 equal to any particular value in $\{0, \dots, M^* - 1\}$ but $\binom{M_p - 1}{k^* - 1}$ choices which make m_1

equal to any particular value in $\{M^*, \dots, M^* + M_p - 1\}$, and $\binom{M_p-1}{k^*-1} / \binom{M_p}{k^*} = k / M_p$ so

$$\lim_{T \rightarrow \infty} \frac{\int_{\mathbb{R}^{(k-1)nd}} J_{T,k}(x) dx_2 \cdots dx_k}{\int_{\mathbb{R}^{knd}} J_{T,k}(x) dx_1 \cdots dx_k} = \frac{1}{k} \left(\sum_{m=0}^{M^*-1} r_m^2(x_1) + \frac{k^*}{M_p} \sum_{m=M^*}^{M^*+M_p-1} r_m^2(x_1) \right)$$

as required. ■

In particular, this means that if the sequence of multiplicities is known then we are able to find the average over eigenstates up to a certain energy level:

Corollary 2.3.3 *Let $J_{T,k} : \mathbb{R}^{knd} \rightarrow \mathbb{R}$ be given by*

$$J_{T,k}(x) = \sum_{\zeta \in \Pi_k} \text{sgn}(\zeta) \varphi_T(x, \zeta x) \mathcal{J}_T(x, \zeta x)$$

where ζx indicates that systems are permuted (with particle indexing maintained within). Let the sequence of multiplicities be M_0, M_1, \dots and let $k = M_0 + M_1 + \dots + M_p$. Then

$$\sum_{m=0}^{k-1} r_m^2(x_1) = k \lim_{T \rightarrow \infty} \frac{\int_{\mathbb{R}^{(k-1)nd}} J_{T,k}(x) dx_2 \cdots dx_k}{\int_{\mathbb{R}^{knd}} J_{T,k}(x) dx_1 \cdots dx_k}. \quad (2.91)$$

Proof. Follows immediately with this choice of k . ■

Since we therefore know how to approximate $\sum_{m=0}^{k-1} r_m^2(x_1)$ for any k (without introducing any additional bias) such that exactly the first $p+1$ energy levels are populated, we can take $\frac{1}{M_p} \left(\sum_{m=0}^{k-1} r_m^2(x_1) - \sum_{m=0}^{k-1-M_p} r_m^2(x_1) \right)$ to find the average of the set of eigenstates that share the $(p+1)$ th eigenvalue.

We shall now consider the case of indistinguishable particles and see that while there is not a similarly elegant expression, we can nonetheless still achieve an analogous result via quantities defined in terms of functional integrals.

Theorem 2.3.4 (approximation to n th state for indistinguishable particles) *Let $I_{T,k} :$*

$\mathbb{R}^{knd} \rightarrow \mathbb{R}$ *be given by*

$$I_{T,k}(x) = \sum_{\zeta \in \Pi_k} \text{sgn}(\zeta) \prod_{j=1}^k q(x_j, x_{\zeta(j)}) = \sum_{\zeta \in \Pi_k} \text{sgn}(\zeta) \prod_{j=1}^k \frac{\sum_{\pi \in \Pi_n} \varphi_T(x_j, \pi x_{\zeta(j)})}{n!} \mathcal{I}_T(x_j, x_{\zeta(j)}) \quad .$$

Then

$$\sum_{m=0}^{M^*-1} |\phi_m|^2(x_1) + \frac{k^*}{M_p} \sum_{m=M^*}^{M^*+M_p-1} |\phi_m|^2(x_1) = k \lim_{T \rightarrow \infty} \frac{\int_{\mathbb{R}^{(k-1)nd}} I_{T,k}(x) dx_2 \cdots dx_k}{\int_{\mathbb{R}^{knd}} I_{T,k}(x) dx_1 \cdots dx_k} \quad . \quad (2.92)$$

Proof. We shall work backwards using the logic of the proof of the previous theorem: we shall seek an $I_{T,k}(x)$ giving (2.92); it shall turn out to be as claimed.

Firstly consider that if we can achieve the equivalent of (2.87), ie

$$\int_{\mathbb{R}^{(k-1)nd}} I_{T,k}(x) dx_2 \dots dx_n = \int_{\mathbb{R}^{(k-1)nd}} dx_2 \dots dx_n \left[\sum_{\substack{m_j \in \{0,1,\dots\} \text{ for } j \in \{1,\dots,k\}; \\ m_j \neq m_l \text{ for } j \neq l}} \exp \left(- \sum_{j=1}^k E_{m_j} T \right) \prod_{j=1}^k \phi_{m_j}(x_j) \phi_{m_j}^*(x_j) \right] \quad (2.93)$$

then the same logic as given above will lead to the desired conclusion. In order to achieve (2.93), however, it shall be seen that it is sufficient to have

$$I_{T,k}(x) = \sum_{\zeta \in \Pi_k} \text{sgn}(\zeta) \sum_{\substack{m_j \in \{0,1,\dots\} \text{ for } j \in \{1,\dots,k\}}} \exp \left(- \sum_{j=1}^k E_{m_j} T \right) \prod_{j=1}^k \phi_{m_j}(x_j) \phi_{m_j}^*(x_{\zeta(j)}) \quad (2.94)$$

because then the same reasoning will apply as before, first to get rid of terms with repeated energy indices and to then get rid of contributions from non-identity permutations in the remaining terms. However, (2.94) clearly means that

$$I_{T,k}(x) = \sum_{\zeta \in \Pi_k} \text{sgn}(\zeta) \prod_{j=1}^k q(x_j, x_{\zeta(j)})$$

as claimed. ■

It should be noted that there are therefore marked differences between the way that the copied system must be treated, and the way that a system of fermions is treated in order to get the density. Some permutations of particles will occur with an opposite sign in $I_{T,k}$ for n fermions, from what they would have for \mathcal{I}_T with kn fermions, and others that occur in \mathcal{I}_T with kn fermions will not occur at all in $I_{T,k}$ with n fermions. For the reader to recognise this immediately, Figure 2.1 shows what paths are collected for $I_{T,2}$ for a system of 2 1D fermions (the particles in x_1 are marked in green and those in x_2 are marked in pink).

Remark 2.3.5 *We have presented Theorem 2.3.4 as useful in conjunction with an asymptotic approximation to system ground states via Path Integral Monte Carlo. However, it could well also apply when using a zero-temperature simulation method, provided that it is*

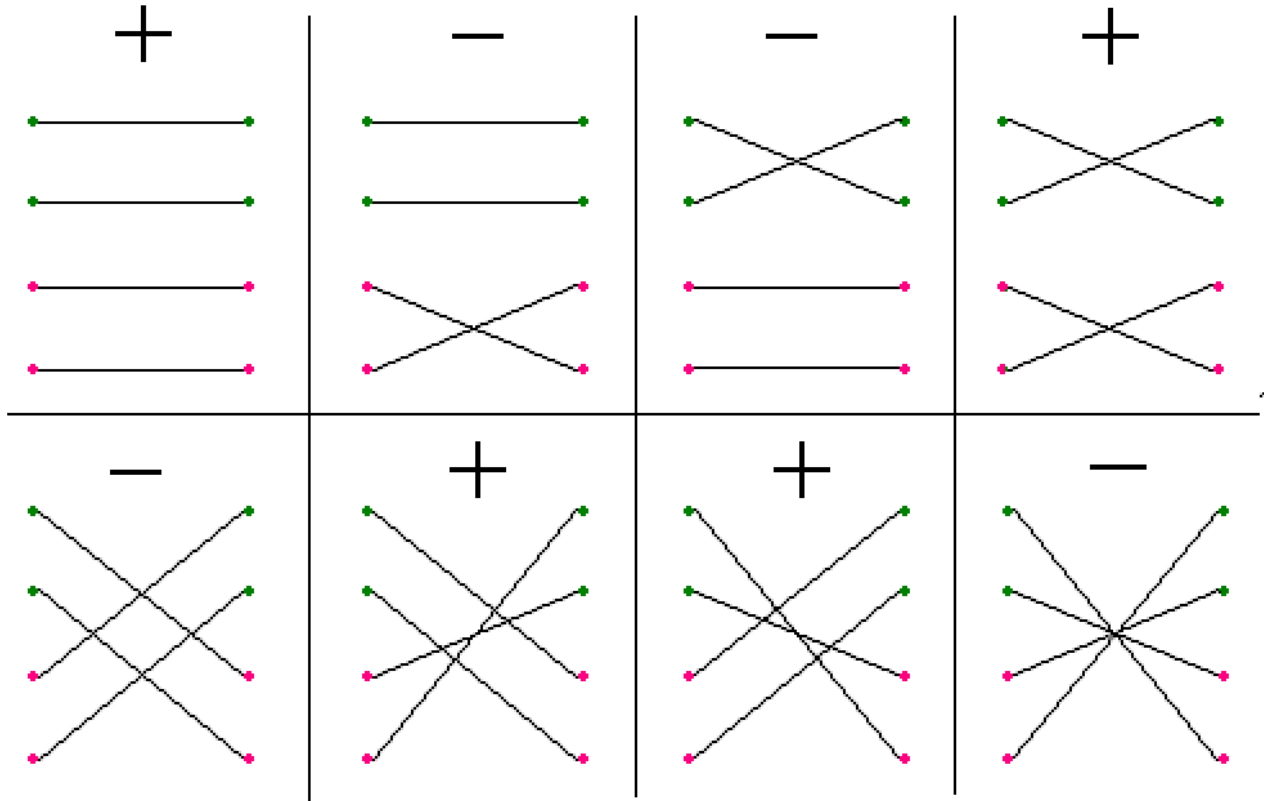


Figure 2.1: Contributions to $I_{T,2}$ for 2 1D fermions

possible to use such a method to make an estimate of $\lim_{T \rightarrow \infty} I_{T,k}(x)$, and this would be a promising direction of further inquiry.

Chapter 3

On numerical integration methods for conditional Wiener integrals

In the preceding chapters, we have discussed at length some of the reasons why we would like to simulate conditional Wiener integrals, which we shall write as

$$\mathcal{U} = \int_{\mathcal{C}_{0,x_0}^{T,x_T}} F(x) dw_{0,x_0}^{T,x_T}(x) = \mathbb{E}F(X) \quad (3.1)$$

for some $F : \mathcal{C}_{0,x_0}^{T,x_T} \rightarrow \mathbb{R}$, where $X \sim w_{0,x_0}^{T,x_T}$. Consequently, as mentioned in Chapter 1, it is desirable to find a random variable for which draws can be made, whose expectation $\overline{\mathcal{U}}$ is close to \mathcal{U} . In particular, in our work we shall focus on the case that this random variable is $F(X^h)$, for another random variable X^h with the same codomain as X , measurable with respect to a set of discretization points $X(t_0), \dots, X(t_N)$ whose maximum spacing is $h := \max_{0 \leq k \leq N-1} (t_{k+1} - t_k)$.

Definition 3.0.6 *In these conditions, when*

$$|\mathcal{U} - \overline{\mathcal{U}}| = |\mathbb{E}F(X) - \mathbb{E}F(X^h)| \leq Kh^p$$

*for a constant K independent of h , we say that the method $\overline{\mathcal{U}}$ has **weak order of convergence** p .*

The key result of this chapter is that a particular piecewise constant method has weak order of convergence 2 in the case of integration w.r.t. the conditional Wiener measure, for

a rather general class of functionals. It is immediately clear that the same must then hold true when the integration measure is $w_{0,x_0}^{T,\Pi_n x_T}$.

We shall begin, in Section 3.1, by recalling the Fréchet derivative and Taylor's Theorem for functionals; this shall be the main tool in proving Theorem 3.2.1. We shall then specify precisely a set of assumptions on the functional being integrated which are sufficient to make the result valid, and then proceed to state and briefly discuss the result; this constitutes Section 3.2. The proof of Theorem 3.2.1 has its own section devoted to it.

3.1 Fréchet derivatives and Taylor's Theorem

In this section we begin by recalling necessary background concepts, namely the Fréchet derivative and Taylor's theorem for functionals. We shall also specify the class of functionals for which the corresponding convergence theorem shall shortly be proved. This is done via the formal assumptions listed in Subsection 3.1.1. Then in Subsection 3.1.2, we give some examples from this class of functionals.

Fréchet Derivatives

Let $A[0, T]$ be the space of right-continuous functions $x : [0, T] \rightarrow \mathbb{R}^d$ which have no essential discontinuities. Suppose $F : A[0, T] \rightarrow \mathbb{R}$. Then if it exists, the first **Fréchet derivative** of F at x in the direction $\delta \in A[0, T]$ is given by (see [KF57]):

$$F^{(1)}(x)(\delta) = \lim_{\varepsilon \rightarrow 0} \frac{F(x(\cdot) + \varepsilon \delta(\cdot)) - F(x(\cdot))}{\varepsilon} \quad (3.2)$$

and further Fréchet derivatives are defined according to the necessary pattern, that the $(n+1)th$ derivative in the directions $(\delta_1, \delta_2, \dots, \delta_n, \delta_{n+1})$ is simply the derivative of the nth derivative in direction $(\delta_1, \dots, \delta_n)$, in the direction δ_{n+1} .

Taylor's theorem

The following result is proven in works such as [KF57].

Theorem 3.1.1 (Taylor's theorem for functionals) *For any $n \in \mathbb{N}$, we can expand $F(x + \delta)$ about $F(x)$ and have an expression for the remainder term:*

$$\begin{aligned} F(x + \delta) = & F(x) + F^{(1)}(x)(\delta) + \cdots + \frac{1}{n!} F^{(n)}(x)(\delta, \dots, \delta) \\ & + \frac{1}{(n+1)!} F^{(n+1)}(x + \lambda\delta)(\delta, \dots, \delta), \quad 0 < \lambda < 1. \end{aligned} \quad (3.3)$$

3.1.1 The class of functionals to be considered

Let us consider functionals $F(x)$ defined on the space $A[0, T]$ of right-continuous d -dimensional vector-functions $x(t)$ on the interval $[0, T]$ without discontinuities of the second kind, i.e., consider functionals on a larger space than $C_{0,a}^{T,b}$. We impose the following assumptions on F .

(FA) Assumptions.

1. *Let $0 < \theta_1 < \cdots < \theta_i < \cdots < \theta_n < T$. Introduce the measure ν_r on $[0, T]^r$ which is the sum of r -dimensional Lebesgue measure on $[0, T]^r$, $(r-1)$ -dimensional Lebesgue measure on the hyperplanes $\{(s_1, \dots, s_r) \in [0, T]^r : s_j = \theta_i\}$, $i = 1, \dots, n$, $j = 1, \dots, r$, and on the diagonal hyperplanes $\{(s_1, \dots, s_r) \in [0, T]^r : s_i = s_j\}$, $(r-2)$ -dimensional Lebesgue measure on $(r-2)$ -dimensional hyperplanes $\{(s_1, \dots, s_r) \in [0, T]^r : s_k = \theta_i$ and $s_l = \theta_j, k \neq l\}$ and $\{(s_1, \dots, s_r) \in [0, T]^r : s_i = s_j \text{ and } s_k = s_l\}$, and so on, including the one-dimensional Lebesgue measure on the lines $\{s_1 = \theta_{i_1}, \dots, s_{r-1} = \theta_{i_{r-1}}\}$, $i_j \in \{1, \dots, n\}$, and on the diagonal $\{s_1 = s_2 = \cdots = s_r\}$ plus the unit measures concentrated on the points $(\theta_{i_1}, \dots, \theta_{i_r})$, $i_j \in \{1, \dots, n\}$.*
2. *We assume that the functional $F(x)$ is six times Fréchet differentiable and that its r -th derivative has the following form:*

$$\begin{aligned} F^{(r)}(x)(\delta_1, \dots, \delta_r) = & \int_{[0, T]^r} v^{(r)}(x; s_1, \dots, s_r) \delta_1(s_1) \cdots \delta_r(s_r) \nu_r(ds_1 \cdots ds_r), \\ & r = 1, \dots, 6, \end{aligned} \quad (3.4)$$

where $\delta_i \in A[0, T]$ and the vector-functions $v^{(r)}(x; s_1, \dots, s_r)$ are symmetric in the arguments s_1, \dots, s_r and uniformly bounded for $x \in A[0, T]$, $s_i \in [0, T]$.

3. For any function $x \in A[0, T]$ constant on a semi-interval $[c_0, c^0) \subset [0, T]$, there are continuous derivatives

$$\begin{aligned} \frac{d}{ds}v^{(1)}(x; s); \quad \frac{\partial}{\partial s_1}v^{(2)}(x; s_1, s_2), \quad s_1 \neq s_2, \quad s_j \neq \theta_i; \quad \frac{d}{ds}v^{(2)}(x; s, s); \\ \frac{d}{ds}v^{(2)}(x; s, \theta_i), \quad i = 1, \dots, n; \end{aligned}$$

which are bounded by a constant independent of $[c_0, c^0)$ and $x \in A[0, T]$.

We recall (see, e.g. [KF57]) that $F^{(r)}(x)(\delta_1, \dots, \delta_r)$ are r -linear functionals. Under Assumptions (FA) we prove a convergence theorem (Theorem 3.2.1) for the method proposed in Section 3.2.2. We emphasize that the method is applicable much more widely. The reason we need to treat up to the sixth Fréchet derivative is that we are going to need to ensure that the local error is third-order in the timestep, and for the sixth derivative it is possible to show that the remainder, when applying Taylor's Theorem is third-order, as we shall see.

Roughly speaking, one might say that we consider functionals of the general form on $A[0, T]$ which satisfy some conditions on smoothness and boundedness. As is usual for any numerical methods, if we weaken the assumptions about the smoothness then, as a rule, the convergence order of the considered method becomes lower than the optimal one. In physical applications, the smoothness part of Assumptions (FA) is not particularly restrictive since it is usually satisfied. The assumption on boundedness of derivatives of functionals can be, to some extent, weakened without loss of convergence order but this would significantly complicate the proof of the convergence theorem. At the same time, the common computational practice in quantum statistical mechanics is to curtail potentials so that they and their derivatives remain bounded which usually implies boundedness of derivatives of functionals. Alternatively, the concept of rejecting exploding trajectories from [MT05] could be exploited here. That is, we might choose not to take into account those trajectories which leave a

bounded domain \mathcal{S} during the time T . The domain \mathcal{S} is chosen so that the boundedness condition is satisfied when $x(\cdot) \in \mathcal{S}$.

3.1.2 Examples of functionals

To illustrate the class of functionals satisfying Assumptions (FA), we give two particular examples here, although many more can be immediately constructed.

1. We start with the integral-type functionals (see the functional needed to compute the correlation function (2.72)):

$$F(x(\cdot)) = \varphi \left(x(\theta), \int_0^T f(t, x(t)) dt \right), \quad 0 \leq \theta \leq T, \quad x \in C_{0,a;T,b}^d. \quad (3.5)$$

One can check that if the functions $f(t, x)$ and $\varphi(x, z)$ have continuous and bounded derivatives up to a sufficiently high order then Assumptions (FA) hold. In particular, the Fréchet derivatives (3.4) have the form here:

$$F^{(1)}(x)(\delta_1) = \int_{[0,T]} v^{(1)}(x; s_1) \delta_1(s_1) \nu_1(ds_1)$$

with

$$\begin{aligned} v^{(1)}(x; s_1) \delta_1(s_1) &= \frac{\partial \varphi}{\partial z} \nabla_x f(s_1, x(s_1)) \cdot \delta_1(s_1), \quad s_1 \neq \theta; \\ v^{(1)}(x; \theta) \delta_1(\theta) &= \nabla_x \varphi \cdot \delta_1(\theta); \end{aligned}$$

and the measure ν_1 being the sum of the Lebesgue measure on $[0, T]$ and the unit measure concentrated at the point θ ;

$$F^{(2)}(x)(\delta_1, \delta_2) = \int_{[0,T]^2} v^{(2)}(x; s_1, s_2) \delta_1(s_1) \delta_2(s_2) \nu_2(ds_1 ds_2)$$

with

$$v^{(2)}(x; s_1, s_2) \delta_1(s_1) \delta_2(s_2) = \frac{\partial^2 \varphi}{\partial z^2} \nabla_x f(s_1, x(s_1)) \cdot \delta_1(s_1) \nabla_x f(s_2, x(s_2)) \cdot \delta_2(s_2),$$

$$s_1 \neq s_2, \quad s_i \neq \theta;$$

$$v^{(2)}(x; s, \theta) \delta_1(s) \delta_2(\theta) = \sum_{i=1}^d \frac{\partial^2 \varphi}{\partial z \partial x^i} \nabla_x f(s, x(s)) \cdot \delta_1(s) \delta_2^i(\theta), \quad s \neq \theta;$$

$$v^{(2)}(x; s, s) \delta_1(s) \delta_2(s) = \frac{\partial \varphi}{\partial z} \sum_{i,j=1}^d \frac{\partial^2 f}{\partial x^i \partial x^j}(s, x(s)) \delta_1^i(s) \delta_2^j(s), \quad s \neq \theta;$$

$$v^{(2)}(x; \theta, \theta) \delta_1(\theta) \delta_2(\theta) = \sum_{i,j=1}^d \frac{\partial^2 \varphi}{\partial x^i \partial x^j} \delta_1^i(\theta) \delta_2^j(\theta);$$

and the measure ν_2 being the sum of the two-dimensional Lebesgue measure on $[0, T]^2$, the one-dimensional Lebesgue measures on the lines $\{s_1 = \theta\}$ and $\{s_2 = \theta\}$ and on the diagonal $\{s_1 = s_2\}$, and the unit measure concentrated at the point (θ, θ) ; the other derivatives can be written analogously. In the above formulas the derivatives of the function φ are taken at the point $\left(x(\theta), \int_0^T f(t, x(t)) dt\right)$ and the dot \cdot means the usual scalar product of vectors.

2. Let functions $f(t, x)$, $g(t, x)$, and $\varphi(z)$ have continuous and bounded derivatives up to a sufficiently high order. Then the functional

$$F(x(\cdot)) = \varphi \left(\int_0^T \int_0^t f(s, x(s)) g(t, x(t)) ds dt \right)$$

satisfies Assumptions (FA).

3.2 The piecewise constant method

In this Section, first, in Subsection 3.2.1, some background details are offered. Then, in Subsection 3.2.2, the precise statement of the main result of this chapter is made.

3.2.1 Background: orders of convergence

When a functional F is given by $F(x(\cdot)) = \exp \left(\int_0^T f(t, x(t)) dt \right)$, we say that F is an **exponential-type functional**. A case of particular interest is when F is the exponentiated

action functional (cf (1.2)):

$$Y(x(\cdot)) = \exp \left(- \int_0^T V(x(t)) dt \right) = \exp(-S(x)).$$

In [GM84, VGM84] (see also [MT04]), the probabilistic approach was used for computing Wiener integrals with respect to the unconditional Wiener measure, and it was shown that a piecewise constant method gives rise to a bias that is $O(h^2)$ where h is the timestep of the method, for a relatively general class of functionals. In [MT04ii] (see also [MT04]) the same approach was exploited to compute conditional Wiener integrals of exponential-type functionals. We may also note that there are a large number of methods and results (see, e.g. [MT04] and references therein) for approximating simple functionals $f(X(T))$, where f is a function from a sufficiently wide class and $X(t)$, $t_0 < t < T$, is a solution of SDEs. But not much attention (except, e.g. [Mac97, MT04, VGM84]) has been paid to approximating general functionals depending on trajectories of the SDE solution.

There is a body of work in the physics literature that is concerned with the special case of integrating the exponentiated action functional. The Markov property of the conditional Wiener measure appears there as the fact that the thermal density matrix may be regarded as the product of "high-temperature" density matrices. One possible way to approximate this product is to treat the commutator of the kinetic and potential energy operators as zero, and this is called the primitive approximation to the action, or the primitive action [Cep95]. This is essentially equivalent to the piecewise constant method that is to be considered in what follows, and comes down to using the trapezoidal rule for $S(x)$. In [Tro59] it was proven that if the Hamiltonian operator is self-adjoint then in the limit of a small timestep, convergence to (2.69) is achieved. The fact that the trapezoidal rule is of order h^2 , in the special case of the action functional, has been addressed in works such as [Suz91, Suz94, IT01].

Thus, it has not to date been proven that for the fairly general class of functionals satisfying (FA), the piecewise constant method has bias of order h^2 when using the conditional Wiener measure. It is readily seen that in applications it can be important to be able to deal with functionals which are not exponential-type, however. As in the case of (2.77), they may

arise naturally in the ‘diagonalization’ of functional integrals corresponding to observables whose corresponding operators are diagonal in the momentum representation.

Higher order methods

Results from [Suz91] state that no method based on quadrature of $S(x)$ can be of higher order than 2 even for the action functional. For exponential-type functionals generally, an explicit Runge-Kutta method of order 4 is described in [MT04]. However, in experiments with the order 4 method, it was found to be rather unstable for longer times T . Specifically, there is a tendency for the Runge-Kutta multiplicand to take values greater than 1, and subsequently, the estimator of $\exp\left(-\int_0^t V(X(s))ds\right)$ will head for $+\infty$. Apparently, if we consider evaluating V at points which are not on the sampled trajectory $X(s)$ then it is possible to attain a higher-order numerical method for the action functional in this way [PD03, Pre04]; in these works only the local error is discussed.

An alternative, which has been used for most of the simulation work in this thesis, is to use Simpson’s Rule for $S(X)$, which has bias of order h^2 , but such that the constant on the h^2 term is considerably less than that encountered using the Trapezoidal Rule.

3.2.2 Definition of the piecewise constant method

Here, on the one hand, we deal with a more complicated system than in [GM84, VGM84], since the SDEs involved in the method are singular. On the other hand, we consider a much wider class of functionals than in [MT04ii]. The proposed method is new in comparison with the ones available in [MT04ii] and it is analogous to the one used in the case of the usual Wiener measure [VGM84].

We introduce a discretization of the time interval $[0, T]$

$$0 = t_0 < t_1 < \cdots < t_N = T$$

so that the points θ_i , $i = 1, \dots, n$, belong to the set $\{t_0, t_1, \dots, t_N\}$. Let

$$h := \max_{0 \leq k \leq N-1} (t_{k+1} - t_k).$$

and $t_{k+1/2} := (t_{k+1} + t_k) / 2$, $k = 0, \dots, N-1$. Let us introduce a piecewise constant function $X^h(t)$, $t \in [0, T]$, given by:

$$X^h(t) := a, \quad t \in [0, t_{1/2}); \quad (3.6)$$

$$X^h(t) := X(t_k), \quad t \in [t_{k-1/2}, t_{k+1/2}), \quad k = 1, \dots, N-1;$$

$$X^h(t) := b, \quad t \in [t_{N-1/2}, T].$$

Clearly, trajectories $X^h(\omega)$ belong to the space $A[0, T]$.

We define the piecewise constant approximation of the conditional Wiener integral \mathcal{U} as follows:

$$\mathcal{U} = \mathbb{E}F(X) \approx \overline{\mathcal{U}} = \mathbb{E}F(X^h). \quad (3.7)$$

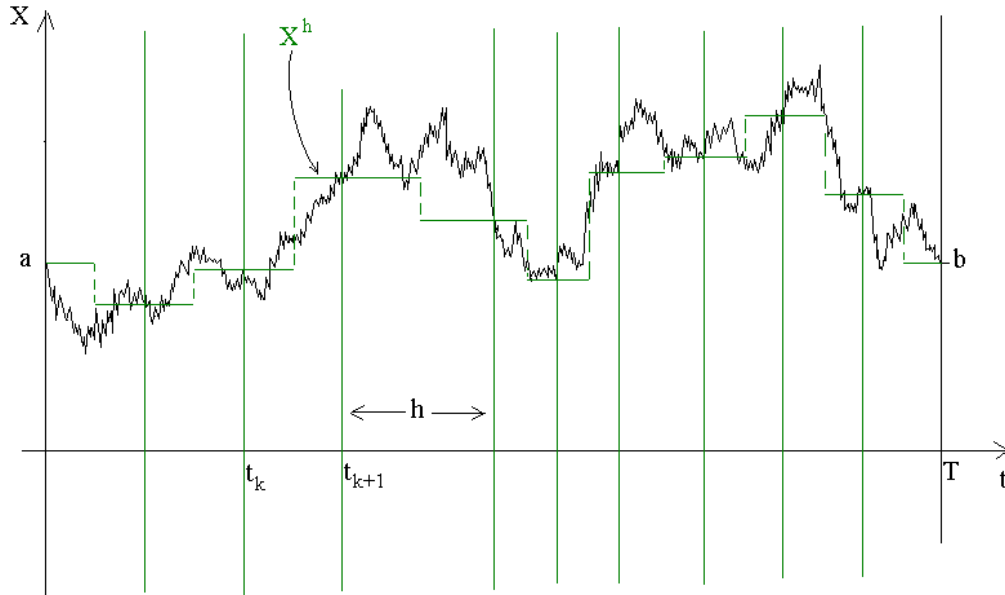


Figure 3.1: Illustration of piecewise constant method X^h .

This method is analogous to the one used in the case of the usual (unconditional) Wiener measure [VGM84] (see also [MT04]). The key result of this chapter is that the method

(3.7), (3.6) applied to evaluation of the conditional Wiener integral (3.1) is of second order of accuracy, i.e.,

$$|\mathcal{U} - \overline{\mathcal{U}}| = |\mathbb{E}F(X) - \mathbb{E}F(X^h)| \leq Kh^2, \quad (3.8)$$

where the constant K is independent of h (although naturally it may vary depending on the structure of timesteps used and depending upon F).

We prove the following convergence theorem.

Theorem 3.2.1 *Assume that Assumptions (FA) hold. The method (3.7),(3.6) applied to evaluation of the Wiener integral (3.1) is of second order of accuracy, i.e.,*

$$|\mathcal{U} - \overline{\mathcal{U}}| = |\mathbb{E}F(X) - \mathbb{E}F(X^h)| \leq Kh^2, \quad (3.9)$$

where the constant K is independent of h .

The proof of the theorem is given in the next section. In Section 4.3, some numerical experiment results are provided which provide some illustration of (3.9).

Remark 3.2.2 *The method (3.7),(3.6) is exact (i.e., there is no integration error) on the class of functionals which depend only on the value of the function $x(t)$ at a finite number of points θ_i , $i = 1, \dots, n$.*

The method (3.7), (3.6) together with the Monte Carlo technique gives an effective algorithm for computing conditional Wiener integrals, which is very simple to realize in practice. The method (3.7), (3.6) can be interpreted as a trapezoidal scheme. This interpretation becomes obvious in the case of integral-type functionals (see (3.38), (3.39)).

Now consider the Euler method, i.e., introduce the piecewise constant function $X_E^h(t)$, $t \in [0, T]$:

$$X_E^h(t) := X(t_k), \quad t \in [t_k, t_{k+1}), \quad k = 0, \dots, N-1; \quad X_E^h(T) := b. \quad (3.10)$$

Theorem 3.2.3 *Assume that Assumptions (FA).1 and (FA).3 hold and (FA).2 holds with $r = 1, 2, 3, 4$ in (3.4). Then*

$$\overset{EULER}{\mathcal{U}} = \mathbb{E}F(X_E^h) \quad (3.11)$$

approximates \mathcal{U} with the first order of accuracy.

The proof of this theorem is easier than that of Theorem 3.2.1 and it is omitted here.

3.3 Proof of the convergence theorem

Here we exploit some constructions from [VGM84], although the singularity of the drift in (2.21) as t approaches T causes additional difficulties, which are overcome by adopting ideas from [MT04ii]. For simplicity and legibility, let us prove the theorem in the one-dimensional case $d = 1$. No additional ideas are required to carry it over to an arbitrary dimension d (see however Remark 3.3.2 at the end of this section). Note that in this section we shall use the letter K to denote various constants which are independent of k and h .

We would like to break down the global error $EF(X) - EF(X^h)$ into a sum of contributions from so-called local errors. With a view to this, we shall now introduce an auxiliary processes $X_k(t)$, $k = 0, \dots, N$:

$$X_k(t) := X(t)\chi_{[0, t_k)}(t) + X(t_k)\chi_{[t_k, T]}(t) + \sum_{j=k}^{N-1} \Delta_j X \chi_{[t_{j+1/2}, T]}(t), \quad (3.12)$$

$$\Delta_j X := X(t_{j+1}) - X(t_j);$$

We shall need to make a careful choice of path to use as an expansion point when applying Taylor's Theorem, and this is supplied by introducing a further auxiliary process, which we shall call $\bar{X}_k(t)$, $k = 0, \dots, N - 1$:

$$\begin{aligned} \bar{X}_k(t) &:= X(t)\chi_{[0, t_k)}(t) + X(t_k)\chi_{[t_k, T]}(t) \\ &+ \sum_{j=k+1}^{N-1} \left(\Delta_j X + (t_{j+1} - t_j) \int_{t_k}^{t_{j+1}} \frac{dW(s')}{T - s'} \right) \chi_{[t_{j+1/2}, T]}(t). \end{aligned} \quad (3.13)$$

We note that $\bar{X}_k(t) = X(t_k)$ for $t \in [t_k, t_{k+3/2}) \cap [0, T]$, i.e., the random function $\bar{X}_k(t)$ is constant on the interval $[t_k, t_{k+3/2}) \cap [0, T]$.

One can see that $X_N(t) = X(t)$ and $X_0(t) = X^h(t)$. We rewrite the global error in the

form:

$$\begin{aligned} \mathbb{E}F(X) - \mathbb{E}F(X^h) &= \mathbb{E}F(X_N) - \mathbb{E}F(X_0) \\ &= \sum_{k=0}^{N-1} [\mathbb{E}F(X_{k+1}) - \mathbb{E}F(X_k)]. \end{aligned} \quad (3.14)$$

Thus, we need to analyze the difference

$$\rho_k := \mathbb{E}F(X_{k+1}) - \mathbb{E}F(X_k). \quad (3.15)$$

Consider the Taylor formula for functionals (3.3) applied with $n = 6$:

$$\begin{aligned} F(x + \delta) &= F(x) + F^{(1)}(x)(\delta) + \cdots + \frac{1}{5!}F^{(5)}(x)(\delta, \dots, \delta) \\ &\quad + \frac{1}{6!}F^{(6)}(x + \lambda\delta)(\delta, \dots, \delta), \quad 0 < \lambda < 1. \end{aligned}$$

We expand $F(X_{k+1})$ and $F(X_k)$ at \bar{X}_k :

$$\begin{aligned} F(X_{k+i}) &= F(\bar{X}_k) + \int_{[0,T]} v^{(1)}(\bar{X}_k; s_1) \delta_{k,i}(s_1) \nu_1(ds_1) + \cdots \\ &\quad + \frac{1}{5!} \int_{[0,T]^5} v^{(5)}(\bar{X}_k; s_1, \dots, s_5) \delta_{k,i}(s_1) \cdots \delta_{k,i}(s_5) \nu_5(ds_1 \cdots ds_5) \\ &\quad + \frac{1}{6!} \int_{[0,T]^6} v^{(6)}(\bar{X}_k + \lambda_i \delta_{k,i}; s_1, \dots, s_6) \delta_{k,i}(s_1) \cdots \delta_{k,i}(s_6) \nu_6(ds_1 \cdots ds_6), \\ &\quad 0 < \lambda_i < 1, \quad i = 0, 1, \end{aligned} \quad (3.16)$$

where

$$\begin{aligned} \delta_{k,0}(s) &= X_k(s) - \bar{X}_k(s) = \Delta_k X \chi_{[t_{k+1/2}, T]}(s) - \int_{t_k}^{t_{k+1}} \frac{dW(s')}{T - s'} \sum_{j=k+1}^{N-1} (t_{j+1} - t_j) \chi_{[t_{j+1/2}, T]}(s) \\ &= \chi_{[t_{k+1/2}, T]}(s) \left[(t_{k+1} - t_k) \frac{b - X(t_k)}{T - t_k} + (T - t_{k+1}) \int_{t_k}^{t_{k+1}} \frac{dW(s')}{T - s'} \right] \\ &\quad - \int_{t_k}^{t_{k+1}} \frac{dW(s')}{T - s'} \sum_{j=k+1}^{N-1} (t_{j+1} - t_j) \chi_{[t_{j+1/2}, T]}(s), \end{aligned} \quad (3.17)$$

$$\begin{aligned}
\delta_{k,1}(s) &= X_{k+1}(s) - \bar{X}_k(s) = (X(s) - X(t_k)) \chi_{[t_k, t_{k+1})}(s) + \Delta_k X \chi_{[t_{k+1}, T]}(s) \\
&\quad - \int_{t_k}^{t_{k+1}} \frac{dW(s')}{T - s'} \sum_{j=k+1}^{N-1} (t_{j+1} - t_j) \chi_{[t_{j+1/2}, T]}(s) \\
&= \chi_{[t_k, t_{k+1})}(s) \left[(s - t_k) \frac{b - X(t_k)}{T - t_k} + (T - s) \int_{t_k}^s \frac{dW(s')}{T - s'} \right] \\
&\quad + \chi_{[t_{k+1}, T]}(s) \left[(t_{k+1} - t_k) \frac{b - X(t_k)}{T - t_k} + (T - t_{k+1}) \int_{t_k}^{t_{k+1}} \frac{dW(s')}{T - s'} \right] \\
&\quad - \int_{t_k}^{t_{k+1}} \frac{dW(s')}{T - s'} \sum_{j=k+1}^{N-1} (t_{j+1} - t_j) \chi_{[t_{j+1/2}, T]}(s).
\end{aligned}$$

It is clear that $\delta_{k,0}(s) = \delta_{k,1}(s)$ for $s \notin (t_k, t_{k+1})$. It can also be seen that the measure ν_r , $r = 1, \dots, 6$, of the set $S_k^{(r)}$ on which the difference $\prod_{j=1}^r \delta_{k,1}(s_j) - \prod_{j=1}^r \delta_{k,0}(s_j)$ is different from zero has order $O(h)$. Indeed, $S_k^{(r)} = \bigcup_{j=1}^r \{(s_1, \dots, s_r) : s_j \in (t_k, t_{k+1})\}$ and hence $\nu_r(S_k^{(r)}) < r\nu_r(\{(s_1, \dots, s_r) : s_1 \in (t_k, t_{k+1})\})$, which is of order $O(h)$. Further, it is not difficult to verify that the integral $\int_{t_k}^s \frac{dW(s')}{T - s'}$, $t_k \leq s \leq t_{k+1}$, and \bar{X}_k are independent by showing that $E\left[\bar{X}_k(t) \int_{t_k}^s \frac{dW(s')}{T - s'}\right] = 0$ for any $0 \leq t \leq T$ and $t_k \leq s \leq t_{k+1}$. In what follows these properties are used in analysis of the parts of ρ_k . We shall also exploit the inequality (see, e.g. [MT04ii, Lemma A.4]) for any $p \geq 1$

$$E|b - X(t_k)|^{2p} \leq K(T - t_k)^p. \quad (3.18)$$

We have from (3.15) and (3.16):

$$\begin{aligned}
\rho_k &= E \int_{[0, T]} v^{(1)}(\bar{X}_k; s_1) [\delta_{k,1}(s_1) - \delta_{k,0}(s_1)] \nu_1(ds_1) \\
&\quad + \frac{1}{2} E \int_{[0, T]^2} v^{(2)}(\bar{X}_k; s_1, s_2) [\delta_{k,1}(s_1)\delta_{k,1}(s_2) - \delta_{k,0}(s_1)\delta_{k,0}(s_2)] \nu_2(ds_1 ds_2) + \dots \\
&\quad + \frac{1}{5!} E \int_{[0, T]^5} v^{(5)}(\bar{X}_k; s_1, \dots, s_5) \left[\prod_{j=1}^5 \delta_{k,1}(s_j) - \prod_{j=1}^5 \delta_{k,0}(s_j) \right] \nu_5(ds_1 \dots ds_5) \\
&\quad + \frac{1}{6!} E \int_{[0, T]^6} [v^{(6)}(\bar{X}_k + \lambda_1 \delta_{k,1}; s_1, \dots, s_6) \prod_{j=1}^6 \delta_{k,1}(s_j) \\
&\quad - v^{(6)}(\bar{X}_k + \lambda_0 \delta_{k,0}; s_1, \dots, s_6) \prod_{j=1}^6 \delta_{k,0}(s_j)] \nu_6(ds_1 \dots ds_6).
\end{aligned} \quad (3.19)$$

Before we start with analysis of ρ_k , we state the lemma which will be used in estimating the second term of (3.19) and which is proved at the end of this section.

Lemma 3.3.1 *Let $U_s(x) := v^{(2)}(x; s, s)$. The following estimate holds*

$$\left| \mathbb{E} U_{t_k}(\bar{X}_k) \left[\frac{(b - X(t_k))^2}{(T - t_k)^2} - \frac{1}{T - t_k} \right] \right| \leq \frac{K}{\sqrt{T - t_k}},$$

where $K > 0$ is a constant independent of k and h .

Now we analyze the terms forming ρ_k in (3.19). Introduce the indicator $I_k = I_{\{\theta_1, \dots, \theta_n\}}(t_k)$.

We obtain for the first term in (3.19):

$$\begin{aligned} r_k^{(1)} &:= \mathbb{E} \int_{[0, T]} v^{(1)}(\bar{X}_k; s_1) [\delta_{k,1}(s_1) - \delta_{k,0}(s_1)] \nu_1(ds_1) \\ &= \mathbb{E} \int_{t_k}^{t_{k+1}} v^{(1)}(\bar{X}_k; s_1) [\delta_{k,1}(s_1) - \delta_{k,0}(s_1)] ds_1 \\ &\quad + v^{(1)}(\bar{X}_k; t_k) [\delta_{k,1}(t_k) - \delta_{k,0}(t_k)] I_k + v^{(1)}(\bar{X}_k; t_{k+1}) [\delta_{k,1}(t_{k+1}) - \delta_{k,0}(t_{k+1})] I_{k+1} \\ &= \mathbb{E} \int_{t_k}^{t_{k+1}} v^{(1)}(\bar{X}_k; s_1) \left[(s_1 - t_k) \frac{b - X(t_k)}{T - t_k} + (T - s_1) \int_{t_k}^{s_1} \frac{dW(s')}{T - s'} \right] ds_1 \\ &\quad - \mathbb{E} \int_{t_{k+1/2}}^{t_{k+1}} v^{(1)}(\bar{X}_k; s_1) \left[(t_{k+1} - t_k) \frac{b - X(t_k)}{T - t_k} + (T - t_{k+1}) \int_{t_k}^{t_{k+1}} \frac{dW(s')}{T - s'} \right] ds_1 \\ &= \mathbb{E} \frac{b - X(t_k)}{T - t_k} \left[\int_{t_k}^{t_{k+1}} v^{(1)}(\bar{X}_k; s_1) (s_1 - t_k) ds_1 - (t_{k+1} - t_k) \int_{t_{k+1/2}}^{t_{k+1}} v^{(1)}(\bar{X}_k; s_1) ds_1 \right] \\ &= \mathbb{E} \frac{b - X(t_k)}{T - t_k} \left[\int_{t_k}^{t_{k+1/2}} v^{(1)}(\bar{X}_k; s_1) (s_1 - t_k) ds_1 - \int_{t_{k+1/2}}^{t_{k+1}} v^{(1)}(\bar{X}_k; s_1) (t_{k+1} - s_1) ds_1 \right]. \end{aligned}$$

Integrating by parts, we get

$$\begin{aligned} r_k^{(1)} &= \mathbb{E} \frac{b - X(t_k)}{T - t_k} \left[v^{(1)}(\bar{X}_k; t_{k+1/2}) \frac{(t_{k+1} - t_k)^2}{8} - \int_{t_k}^{t_{k+1/2}} \frac{d}{ds_1} v^{(1)}(\bar{X}_k; s_1) \frac{(s_1 - t_k)^2}{2} ds_1 \right. \\ &\quad \left. - v^{(1)}(\bar{X}_k; t_{k+1/2}) \frac{(t_{k+1} - t_k)^2}{8} - \int_{t_{k+1/2}}^{t_{k+1}} \frac{d}{ds_1} v^{(1)}(\bar{X}_k; s_1) \frac{(t_{k+1} - s_1)^2}{2} ds_1 \right] \\ &= -\mathbb{E} \frac{b - X(t_k)}{T - t_k} \left[\int_{t_k}^{t_{k+1/2}} \frac{d}{ds_1} v^{(1)}(\bar{X}_k; s_1) \frac{(s_1 - t_k)^2}{2} ds_1 \right. \\ &\quad \left. + \int_{t_{k+1/2}}^{t_{k+1}} \frac{d}{ds_1} v^{(1)}(\bar{X}_k; s_1) \frac{(t_{k+1} - s_1)^2}{2} ds_1 \right]. \end{aligned}$$

It follows from here and the inequality (3.18) that

$$|r_k^{(1)}| \leq \frac{Kh^3}{\sqrt{T - t_k}}, \quad k = 0, \dots, N - 1. \quad (3.20)$$

Now consider the second term in (3.19). We obtain

$$\begin{aligned}
r_k^{(2)} &:= \frac{1}{2} \mathbb{E} \int_{[0,T]^2} v^{(2)}(\bar{X}_k; s_1, s_2) [\delta_{k,1}(s_1)\delta_{k,1}(s_2) - \delta_{k,0}(s_1)\delta_{k,0}(s_2)] \nu_2(ds_1 ds_2) \\
&= \frac{1}{2} \mathbb{E} \int_{[0,T]^2} v^{(2)}(\bar{X}_k; s_1, s_2) \\
&\quad \times \left\{ \left[(s_1 - t_k)(s_2 - t_k) \frac{(b - X(t_k))^2}{(T - t_k)^2} + (s_1 \wedge s_2 - t_k) \frac{T - s_1 \vee s_2}{T - t_k} \right] \right. \\
&\quad \times \chi_{[t_k, t_{k+1})}(s_1) \chi_{[t_k, t_{k+1})}(s_2) \\
&\quad + 2 \left[(t_{k+1} - t_k)(s_1 - t_k) \frac{(b - X(t_k))^2}{(T - t_k)^2} + (s_1 - t_k) \frac{T - t_{k+1}}{T - t_k} \right] \chi_{[t_k, t_{k+1})}(s_1) \chi_{[t_{k+1}, T]}(s_2) \\
&\quad + \left[(t_{k+1} - t_k)^2 \frac{(b - X(t_k))^2}{(T - t_k)^2} + (t_{k+1} - t_k) \frac{T - t_{k+1}}{T - t_k} \right] \\
&\quad \times \left(\chi_{[t_{k+1}, T]}(s_1) \chi_{[t_{k+1}, T]}(s_2) - \chi_{[t_{k+1}/2, T]}(s_1) \chi_{[t_{k+1}/2, T]}(s_2) \right) \\
&\quad - \frac{2}{T - t_k} \sum_{j=k+1}^{N-1} (t_{j+1} - t_j) \chi_{[t_{j+1}/2, T]}(s_2) \\
&\quad \times \left. \left[(s_1 - t_k) \chi_{[t_k, t_{k+1})}(s_1) - (t_{k+1} - t_k) \chi_{[t_{k+1}/2, t_{k+1})}(s_1) \right] \right\} \nu_2(ds_1 ds_2).
\end{aligned} \tag{3.21}$$

We decompose the integral from (3.21) and estimate each part separately. We have

$$\begin{aligned}
A_{1k} &:= \mathbb{E} \int_{[0,T]^2} v^{(2)}(\bar{X}_k; s_1, s_2) \\
&\quad \times \left[(s_1 - t_k)(s_2 - t_k) \frac{(b - X(t_k))^2}{(T - t_k)^2} + (s_1 \wedge s_2 - t_k) \frac{T - s_1 \vee s_2}{T - t_k} \right] \\
&\quad \times \chi_{[t_k, t_{k+1})}(s_1) \chi_{[t_k, t_{k+1})}(s_2) \nu_2(ds_1 ds_2) \\
&= \mathbb{E} \int_{t_k}^{t_{k+1}} v^{(2)}(\bar{X}_k; s, s) \left[(s - t_k)^2 \frac{(b - X(t_k))^2}{(T - t_k)^2} + (s - t_k) \frac{T - s}{T - t_k} \right] ds \\
&\quad + \mathbb{E} \int_{t_k}^{t_{k+1}} \int_{t_k}^{t_{k+1}} v^{(2)}(\bar{X}_k; s_1, s_2) \\
&\quad \times \left[(s_1 - t_k)(s_2 - t_k) \frac{(b - X(t_k))^2}{(T - t_k)^2} + (s_1 \wedge s_2 - t_k) \frac{T - s_1 \vee s_2}{T - t_k} \right] ds_1 ds_2,
\end{aligned} \tag{3.22}$$

where the last integral is estimated by Kh^3 by observing that $\sup |v^{(2)}|$ is bounded (see Assumptions (FA)) and using (3.18) to get

$$\mathbb{E} \frac{(b - X(t_k))^2}{(T - t_k)^2} \leq \frac{K}{T - t_k} \leq \frac{K}{h}.$$

Also note that in (3.22) we omit the integrals over the measure concentrated on the lines $s = t_k$ and $s = t_{k+1}$ and over the unit measures since it is obvious that they are equal to zero. Further, since $v^{(2)}(\bar{X}_k; s, s) = v^{(2)}(\bar{X}_k; t_k, t_k) + \int_{t_k}^s \frac{d}{ds'} v^{(2)}(\bar{X}_k; s', s') ds'$, the first integral in the right-hand side of (3.22) can be written as

$$\begin{aligned} & \mathbb{E} \int_{t_k}^{t_{k+1}} v^{(2)}(\bar{X}_k; s, s) \left[(s - t_k)^2 \frac{(b - X(t_k))^2}{(T - t_k)^2} + (s - t_k) \frac{T - s}{T - t_k} \right] ds \\ &= \mathbb{E} v^{(2)}(\bar{X}_k; t_k, t_k) \left[\frac{(b - X(t_k))^2}{(T - t_k)^2} \int_{t_k}^{t_{k+1}} (s - t_k)^2 ds + \int_{t_k}^{t_{k+1}} (s - t_k) \frac{T - s}{T - t_k} ds \right] \\ &+ \mathbb{E} \int_{t_k}^{t_{k+1}} \int_{t_k}^s \frac{d}{ds'} v^{(2)}(\bar{X}_k; s', s') \left[(s - t_k)^2 \frac{(b - X(t_k))^2}{(T - t_k)^2} + (s - t_k) \frac{T - s}{T - t_k} \right] ds' ds, \end{aligned}$$

where the second integral is estimated by Kh^3 using the same arguments as in (3.22). So,

$$\begin{aligned} A_{1k} &= \mathbb{E} v^{(2)}(\bar{X}_k; t_k, t_k) \left[\frac{(b - X(t_k))^2}{(T - t_k)^2} \frac{(t_{k+1} - t_k)^3}{3} + \frac{(t_{k+1} - t_k)^2}{2} \frac{T - t_{k+1} + (t_{k+1} - t_k)/3}{T - t_k} \right] \\ &+ O(h^3) \end{aligned}$$

with $|O(h^3)| \leq Kh^3$. The next part of (3.21) can be written as

$$\begin{aligned} A_{2k} &:= 2\mathbb{E} \int_{[0, T]^2} v^{(2)}(\bar{X}_k; s_1, s_2) \left[(t_{k+1} - t_k)(s_1 - t_k) \frac{(b - X(t_k))^2}{(T - t_k)^2} + (s_1 - t_k) \frac{T - t_{k+1}}{T - t_k} \right] \\ &\quad \times \chi_{[t_k, t_{k+1})}(s_1) \chi_{[t_{k+1}, T]}(s_2) \nu_2(ds_1 ds_2) \\ &= 2\mathbb{E} \int_{t_{k+1}}^T \int_{t_k}^{t_{k+1}} v^{(2)}(\bar{X}_k; s_1, s_2) \left[(t_{k+1} - t_k)(s_1 - t_k) \frac{(b - X(t_k))^2}{(T - t_k)^2} \right. \\ &\quad \left. + (s_1 - t_k) \frac{T - t_{k+1}}{T - t_k} \right] ds_1 ds_2 \\ &+ \sum_{i=1}^n I_{\theta_i > t_k} \mathbb{E} \int_{t_k}^{t_{k+1}} v^{(2)}(\bar{X}_k; s_1, \theta_i) \left[(t_{k+1} - t_k)(s_1 - t_k) \frac{(b - X(t_k))^2}{(T - t_k)^2} + (s_1 - t_k) \frac{T - t_{k+1}}{T - t_k} \right] ds_1 \\ &= 2\mathbb{E} \left[\frac{(t_{k+1} - t_k)^3}{2} \frac{(b - X(t_k))^2}{(T - t_k)^2} + \frac{(t_{k+1} - t_k)^2}{2} \frac{T - t_{k+1}}{T - t_k} \right] \\ &\quad \times \left[\int_{t_{k+1}}^T v^{(2)}(\bar{X}_k; t_k, s_2) ds_2 + \sum_{i=1}^n I_{\theta_i > t_k} v^{(2)}(\bar{X}_k; t_k, \theta_i) \right] + O(h^3). \end{aligned}$$

The third part of (3.21) is

$$\begin{aligned} A_{3k} &:= \mathbb{E} \left[(t_{k+1} - t_k)^2 \frac{(b - X(t_k))^2}{(T - t_k)^2} + (t_{k+1} - t_k) \frac{T - t_{k+1}}{T - t_k} \right] \int_{[0, T]^2} v^{(2)}(\bar{X}_k; s_1, s_2) \\ &\quad \times \left[\chi_{[t_{k+1}, T]}(s_1) \chi_{[t_{k+1}, T]}(s_2) - \chi_{[t_{k+1}/2, T]}(s_1) \chi_{[t_{k+1}/2, T]}(s_2) \right] \nu_2(ds_1 ds_2). \end{aligned}$$

We have for the integral in A_{3k} :

$$\begin{aligned}
& \int_{[0,T]^2} v^{(2)}(\bar{X}_k; s_1, s_2) \left[\chi_{[t_{k+1}, T]}(s_1) \chi_{[t_{k+1}, T]}(s_2) - \chi_{[t_{k+1/2}, T]}(s_1) \chi_{[t_{k+1/2}, T]}(s_2) \right] \nu_2(ds_1 ds_2) \\
&= \int_{t_{k+1}}^T v^{(2)}(\bar{X}_k; s, s) ds - \int_{t_{k+1/2}}^T v^{(2)}(\bar{X}_k; s, s) ds \\
&+ 2 \sum_{i=1}^n I_{\theta_i > t_k} \left[\int_{t_{k+1}}^T v^{(2)}(\bar{X}_k; s, \theta_i) ds - \int_{t_{k+1/2}}^T v^{(2)}(\bar{X}_k; s, \theta_i) ds \right] \\
&+ \int_{t_{k+1}}^T \int_{t_{k+1}}^T v^{(2)}(\bar{X}_k; s_1, s_2) ds_1 ds_2 - \int_{t_{k+1/2}}^T \int_{t_{k+1/2}}^T v^{(2)}(\bar{X}_k; s_1, s_2) ds_1 ds_2 \\
&= -\frac{(t_{k+1} - t_k)}{2} v^{(2)}(\bar{X}_k; t_k, t_k) - (t_{k+1} - t_k) \sum_{i=1}^n I_{\theta_i > t_k} v^{(2)}(\bar{X}_k; t_k, \theta_i) + O(h^2) \\
&- 2 \int_{t_{k+1/2}}^T \int_{t_{k+1/2}}^{t_{k+1}} v^{(2)}(\bar{X}_k; s_1, s_2) ds_1 ds_2 + \int_{t_{k+1/2}}^{t_{k+1}} \int_{t_{k+1/2}}^{t_{k+1}} v^{(2)}(\bar{X}_k; s_1, s_2) ds_1 ds_2 \\
&= -\frac{(t_{k+1} - t_k)}{2} v^{(2)}(\bar{X}_k; t_k, t_k) - (t_{k+1} - t_k) \sum_{i=1}^n I_{\theta_i > t_k} v^{(2)}(\bar{X}_k; t_k, \theta_i) \\
&- (t_{k+1} - t_k) \int_{t_{k+1}}^T v^{(2)}(\bar{X}_k; t_k, s_2) ds_2 + O(h^2).
\end{aligned}$$

Then

$$\begin{aligned}
A_{3k} &= -\mathbb{E} \left[\left(v^{(2)}(\bar{X}_k; t_k, t_k) + 2 \sum_{i=1}^n I_{\theta_i > t_k} v^{(2)}(\bar{X}_k; t_k, \theta_i) \right) \right. \\
&\quad \times \left(\frac{(t_{k+1} - t_k)^3}{2} \frac{(b - X(t_k))^2}{(T - t_k)^2} + \frac{(t_{k+1} - t_k)^2}{2} \frac{T - t_{k+1}}{T - t_k} \right) \Big] \\
&- \mathbb{E} \left[(t_{k+1} - t_k)^3 \frac{(b - X(t_k))^2}{(T - t_k)^2} + (t_{k+1} - t_k)^2 \frac{T - t_{k+1}}{T - t_k} \right] \int_{t_{k+1}}^T v^{(2)}(\bar{X}_k; t_k, s_2) ds_2 \\
&+ O(h^3).
\end{aligned}$$

The last part of (3.21) is

$$\begin{aligned}
A_{4k} &:= -\frac{2}{T - t_k} \mathbb{E} \int_{[0,T]^2} v^{(2)}(\bar{X}_k; s_1, s_2) \sum_{j=k+1}^{N-1} (t_{j+1} - t_j) \chi_{[t_{j+1/2}, T]}(s_2) \\
&\quad \times \left[(s_1 - t_k) \chi_{[t_k, t_{k+1})}(s_1) - (t_{k+1} - t_k) \chi_{[t_{k+1/2}, t_{k+1})}(s_1) \right] \nu_2(ds_1 ds_2) \\
&= -\frac{2}{T - t_k} \mathbb{E} \int_{[0,T]^2} v^{(2)}(\bar{X}_k; s_1, s_2) \sum_{j=k+1}^{N-1} (t_{j+1} - t_j) \chi_{[t_{j+1/2}, T]}(s_2) \\
&\quad \times \left[(s_1 - t_k) \chi_{[t_k, t_{k+1/2})}(s_1) - (t_{k+1} - s_1) \chi_{[t_{k+1/2}, t_{k+1})}(s_1) \right] \nu_2(ds_1 ds_2)
\end{aligned}$$

$$\begin{aligned}
&= -\frac{2}{T-t_k} \mathbb{E} \left[\int_{t_{k+3/2}}^T \sum_{j=k+1}^{N-1} (t_{j+1} - t_j) \chi_{[t_{j+1/2}, T]}(s_2) \right. \\
&\quad \times \left(\int_{t_k}^{t_{k+1/2}} v^{(2)}(\bar{X}_k; s_1, s_2) (s_1 - t_k) ds_1 - \int_{t_{k+1/2}}^{t_{k+1}} v^{(2)}(\bar{X}_k; s_1, s_2) (t_{k+1} - s_1) ds_1 \right) ds_2 \\
&\quad + \sum_{i=1}^n I_{\theta_i > t_{k+1}} (\theta_i - t_{k+1}) \\
&\quad \times \left. \left(\int_{t_k}^{t_{k+1/2}} v^{(2)}(\bar{X}_k; s_1, \theta_i) (s_1 - t_k) ds_1 - \int_{t_{k+1/2}}^{t_{k+1}} v^{(2)}(\bar{X}_k; s_1, \theta_i) (t_{k+1} - s_1) ds_1 \right) \right].
\end{aligned}$$

Exploiting arguments similar to the ones used before, it is not difficult to get that $A_{4k} = O(h^3)$.

As a result, we obtain

$$\begin{aligned}
r_k^{(2)} &= \frac{1}{2} (A_{1k} + A_{2k} + A_{3k} + A_{4k}) \\
&= -\frac{(t_{k+1} - t_k)^3}{12} \mathbb{E} v^{(2)}(\bar{X}_k; t_k, t_k) \left[\frac{(b - X(t_k))^2}{(T - t_k)^2} - \frac{1}{T - t_k} \right] + O(h^3).
\end{aligned} \tag{3.23}$$

Applying Lemma 3.3.1, we get

$$|r_k^{(2)}| \leq \frac{Kh^3}{\sqrt{T - t_k}}. \tag{3.24}$$

Now we estimate the remaining terms in (3.19). We obtain from (3.17):

$$\delta_{k,0}(s) = \chi_{[t_{k+1/2}, T]}(s) (t_{k+1} - t_k) \frac{b - X(t_k)}{T - t_k} + \int_{t_k}^{t_{k+1}} \frac{dW(s')}{T - s'} \sum_{j=k+1}^{N-1} (t_{j+1} - t_j) \chi_{[t_{k+1/2}, t_{j+1/2})}(s).$$

Then

$$\begin{aligned}
&\mathbb{E} v^{(3)}(\bar{X}_k; s_1, s_2, s_3) \prod_{i=1}^3 \delta_{k,0}(s_i) \\
&= \mathbb{E} v^{(3)}(\bar{X}_k; s_1, s_2, s_3) \\
&\quad \times \prod_{i=1}^3 \left((t_{k+1} - t_k) \frac{b - X(t_k)}{T - t_k} \chi_{[t_{k+1/2}, T]}(s_i) + \int_{t_k}^{t_{k+1}} \frac{dW(s')}{T - s'} \sum_{j=k+1}^{N-1} (t_{j+1} - t_j) \chi_{[t_{k+1/2}, t_{j+1/2})}(s_i) \right) \\
&= \mathbb{E} v^{(3)}(\bar{X}_k; s_1, s_2, s_3) (t_{k+1} - t_k)^2 \frac{b - X(t_k)}{(T - t_k)^2} \left((t_{k+1} - t_k) \frac{(b - X(t_k))^2}{T - t_k} \prod_{i=1}^3 \chi_{[t_{k+1/2}, T]}(s_i) \right. \\
&\quad \left. + \sum_{i=1}^3 \frac{\chi_{[t_{k+1/2}, T]}(s_i)}{T - t_{k+1}} \prod_{l \neq i} \sum_{j=k+1}^{N-1} (t_{j+1} - t_j) \chi_{[t_{k+1/2}, t_{j+1/2})}(s_l) \right).
\end{aligned}$$

From here, we get the estimate

$$\left| \mathbb{E}[v^{(3)}(\bar{X}_k; s_1, s_2, s_3) \prod_{j=1}^3 \delta_{k,0}(s_j)] \right| \leq \frac{Kh^2}{\sqrt{T-t_k}}.$$

Analogously, we obtain

$$\left| \mathbb{E}[v^{(3)}(\bar{X}_k; s_1, s_2, s_3) \prod_{j=1}^3 \delta_{k,1}(s_j)] \right| \leq \frac{Kh^2}{\sqrt{T-t_k}}.$$

Then, also taking into account that the measure ν_3 of the set $S_k^{(3)}$ on which the difference $\prod_{j=1}^3 \delta_{k,1}(s_j) - \prod_{j=1}^3 \delta_{k,0}(s_j)$ is different from zero has order $O(h)$, we arrive at

$$\begin{aligned} & \left| \frac{1}{6} \mathbb{E} \int_{[0,T]^3} v^{(3)}(\bar{X}_k; s_1, s_2, s_3) \left[\prod_{j=1}^3 \delta_{k,1}(s_j) - \prod_{j=1}^3 \delta_{k,0}(s_j) \right] \nu_3(ds_1 ds_2 ds_3) \right| \quad (3.25) \\ &= \left| \frac{1}{6} \mathbb{E} \int_{[0,T]^3} I_{S_k^{(3)}}(s_1, s_2, s_3) v^{(3)}(\bar{X}_k; s_1, s_2, s_3) \left[\prod_{j=1}^3 \delta_{k,1}(s_j) - \prod_{j=1}^3 \delta_{k,0}(s_j) \right] \nu_3(ds_1 ds_2 ds_3) \right| \\ &\leq \frac{1}{6} \int_{[0,T]^3} I_{S_k^{(3)}}(s_1, s_2, s_3) \left[|\mathbb{E} v^{(3)}(\bar{X}_k; s_1, s_2, s_3) \prod_{j=1}^3 \delta_{k,1}(s_j)| \right. \\ &\quad \left. + |\mathbb{E} v^{(3)}(\bar{X}_k; s_1, s_2, s_3) \prod_{j=1}^3 \delta_{k,0}(s_j)| \right] \nu_3(ds_1 ds_2 ds_3) \leq \frac{Kh^3}{\sqrt{T-t_k}}. \end{aligned}$$

Since we have for the terms in (3.17)

$$\begin{aligned} \mathbb{E}(\Delta_k X)^4 &\leq Kh^2, \quad \mathbb{E}(X(s) - X(t_k))^4 \chi_{[t_k, t_{k+1})}(s) \leq Kh^2, \\ \mathbb{E} \left(\int_{t_k}^{t_{k+1}} \frac{dW(s')}{T-s'} \sum_{j=k+1}^{N-1} (t_{j+1} - t_j) \chi_{[t_{j+1/2}, T]}(s) \right)^4 &\leq Kh^2, \end{aligned}$$

and the measure ν_4 of the set $S_k^{(4)}$ on which the difference $\prod_{j=1}^4 \delta_{k,1}(s_j) - \prod_{j=1}^4 \delta_{k,0}(s_j)$ is different from zero has order $O(h)$, we obtain

$$\begin{aligned} & \left| \frac{1}{4!} \mathbb{E} \int_{[0,T]^4} v^{(4)}(\bar{X}_k; s_1, \dots, s_4) \left[\prod_{j=1}^4 \delta_{k,1}(s_j) - \prod_{j=1}^4 \delta_{k,0}(s_j) \right] \nu_4(ds_1 \dots ds_4) \right| \quad (3.26) \\ &\leq \frac{1}{4!} \sup |v^{(4)}| \int_{[0,T]^4} I_{S_k^{(4)}}(s_1, \dots, s_4) \left(\mathbb{E} \left| \prod_{j=1}^4 \delta_{k,1}(s_j) \right| + \mathbb{E} \left| \prod_{j=1}^4 \delta_{k,0}(s_j) \right| \right) \nu_4(ds_1 \dots ds_4) \\ &\leq Kh^3. \end{aligned}$$

By analogous arguments, we get

$$\begin{aligned}
& \left| \frac{1}{5!} \mathbb{E} \int_{[0,T]^5} v^{(5)}(\bar{X}_k; s_1, \dots, s_5) \left[\prod_{j=1}^5 \delta_{k,1}(s_j) - \prod_{j=1}^5 \delta_{k,0}(s_j) \right] \nu_5(ds_1 \cdots ds_5) \right| \\
& \leq \frac{1}{5!} \sup |v^{(5)}| \int_{[0,T]^5} \mathbb{E} \left| \prod_{j=1}^5 \delta_{k,1}(s_j) - \prod_{j=1}^5 \delta_{k,0}(s_j) \right| \nu_5(ds_1 \cdots ds_5) \\
& \leq Kh^{7/2}.
\end{aligned} \tag{3.27}$$

Since $E \prod_{j=1}^6 |\delta_{k,i}(s_j)| \leq Kh^3$, the last term in (3.19) is estimated as

$$\begin{aligned}
& \left| \frac{1}{6!} \mathbb{E} \int_{[0,T]^6} [v^{(6)}(\bar{X}_k + \lambda_1 \delta_{k,i}; s_1, \dots, s_6) \prod_{j=1}^6 \delta_{k,1}(s_j) \right. \\
& \quad \left. - v^{(6)}(\bar{X}_k + \lambda_0 \delta_{k,i}; s_1, \dots, s_6) \prod_{j=1}^6 \delta_{k,0}(s_j)] \nu_6(ds_1 \cdots ds_6) \right| \\
& \leq \frac{1}{6!} \sup |v^{(6)}| \left| \int_{[0,T]^6} \mathbb{E} \left[\prod_{j=1}^6 |\delta_{k,1}(s_j)| + \prod_{j=1}^6 |\delta_{k,0}(s_j)| \right] \nu_6(ds_1 \cdots ds_6) \right| \\
& \leq Kh^3.
\end{aligned} \tag{3.28}$$

Substituting (3.20), (3.24)-(3.28) in (3.19), we get

$$|\rho_k| \leq \frac{Kh^3}{\sqrt{T - t_k}}, \quad k = 0, \dots, N-1,$$

which together with (3.14)-(3.15) implies (3.9). Theorem 3.2.1 is proved. ■

Proof of Lemma 3.3.1. Assumptions (FA) ensure that for a fixed $\tau \in [0, T]$ the functional $U_\tau(x) = v^{(2)}(x; \tau, \tau)$ is Fréchet differentiable and its derivative has the form:

$$U_\tau^{(1)}(x)(\delta) = \int_0^T u^{(1)}(x; s) \delta(s) ds + u^{(1)}(x; \tau) \delta(\tau) + \sum_{i=1}^n u^{(1)}(x; \theta_i) \delta(\theta_i),$$

where $u^{(1)}(x; s)$ is uniformly bounded for $x \in A[0, T]$, $s \in [0, T]$.

We also note [MT04ii, Corollary A.1] that

$$\psi(t_l) := \frac{(b - X(t_l))^2}{(T - t_l)^2} - \frac{1}{T - t_l}, \quad l = 0, \dots, N-1,$$

is a martingale.

Introduce the auxiliary processes $\bar{X}_k^{(0)}(t)$, $k = 0, \dots, N-1$:

$$\bar{X}_k^{(0)}(t) := \bar{X}_k(t)\chi_{[0,t_k)}(t) + b\chi_{[t_k,T]}(t).$$

Using the Taylor formula for functionals, we get

$$\begin{aligned} U_{t_k}(\bar{X}_k) &= U_{t_k}(\bar{X}_k^{(0)}) + \int_0^T u^{(1)}(\bar{X}_k^{(0)} + \lambda\delta; s)\delta(s)ds + u^{(1)}(\bar{X}_k^{(0)} + \lambda\delta; t_k)\delta(t_k) \\ &\quad + \sum_{i=1}^n u^{(1)}(\bar{X}_k^{(0)} + \lambda\delta; \theta_i)\delta(\theta_i), \end{aligned}$$

where

$$\begin{aligned} \delta(s) &= \bar{X}_k(s) - \bar{X}_k^{(0)}(s) \\ &= \left[X(t_k) + \sum_{j=k+1}^{N-1} \left(\Delta_j X + (t_{j+1} - t_j) \int_{t_k}^{t_{k+1}} \frac{dW(s')}{T - s'} \right) \chi_{[t_{j+1/2}, T]}(s) - b \right] \chi_{[t_k, T]}(s) \end{aligned}$$

and $0 < \lambda < 1$.

We have

$$\begin{aligned} |EU_{t_k}(\bar{X}_k)\psi(t_k)| &\leq |EU_{t_k}(\bar{X}_k^{(0)})\psi(t_k)| + \left| E\psi(t_k) \int_{t_k}^T u^{(1)}(\bar{X}_k^{(0)} + \lambda\delta; s) \right. \\ &\quad \times \left[X(t_k) + \sum_{j=k+1}^{N-1} \left(\Delta_j X + (t_{j+1} - t_j) \int_{t_k}^{t_{k+1}} \frac{dW(s')}{T - s'} \right) \chi_{[t_{j+1/2}, T]}(s) - b \right] ds \left. \right| \\ &\quad + \left| E\psi(t_k)u^{(1)}(\bar{X}_k^{(0)} + \lambda\delta; t_k)(X(t_k) - b) \right| + \sum_{i=1}^n I_{\theta_i > t_k} \left| E\psi(t_k)u^{(1)}(\bar{X}_k^{(0)} + \lambda\delta; \theta_i)\delta(\theta_i) \right|. \end{aligned} \quad (3.29)$$

It is not difficult to see that the second term in the right-hand side of (3.29) is bounded by a constant and the third and fourth terms are bounded by $K/\sqrt{T - t_k}$. Thus,

$$|EU_{t_k}(\bar{X}_k)\psi(t_k)| \leq |EU_{t_k}(\bar{X}_k^{(0)})\psi(t_k)| + \frac{K}{\sqrt{T - t_k}}. \quad (3.30)$$

Now introduce the auxiliary processes $\bar{X}_k^{(j)}(t)$, $j = 1, \dots, k$, $k = 0, \dots, N-1$:

$$\bar{X}_k^{(j)}(t) := \bar{X}_k^{(j-1)}(t)\chi_{[0,t_{k-j})}(t) + b\chi_{[t_{k-j}, T]}(t).$$

We have

$$U_{t_k}(\bar{X}_k^{(j-1)}) = U_{t_k}(\bar{X}_k^{(j)}) + \int_0^T u^{(1)}(\bar{X}_k^{(j)} + \lambda\delta; s)\delta(s)ds + \sum_{i=1}^n u^{(1)}(\bar{X}_k^{(j)} + \lambda\delta; \theta_i)\delta(\theta_i),$$

where

$$\delta(s) = \bar{X}_k^{(j-1)}(s) - \bar{X}_k^{(j)}(s) = (X(s) - b) \chi_{[t_{k-j}, t_{k-j+1})}(s).$$

Then (as before, $I_k = I_{\{\theta_1, \dots, \theta_n\}}(t_k)$) :

$$\begin{aligned} U_{t_k}(\bar{X}_k^{(j-1)}) &= U_{t_k}(\bar{X}_k^{(j)}) + \int_{t_{k-j}}^{t_{k-j+1}} u^{(1)}(\bar{X}_k^{(j)} + \lambda\delta; s) [X(s) - b] ds \\ &\quad + I_{k-j} u^{(1)}(\bar{X}_k^{(j)} + \lambda\delta; t_{k-j}) [X(t_{k-j}) - b]. \end{aligned} \quad (3.31)$$

Recalling that $\psi(t_l)$, $l = 0, \dots, N-1$, is a martingale and observing that $U_{t_k}(\bar{X}_k^{(j)})$ is $\mathcal{F}_{t_{k-j}}$ -measurable, we get that

$$\left| \mathbb{E} U_{t_k}(\bar{X}_k^{(j)}) \psi(t_{k-j+1}) \right| = \left| \mathbb{E} U_{t_k}(\bar{X}_k^{(j)}) \psi(t_{k-j}) \right|. \quad (3.32)$$

It follows from (3.31)-(3.32) that

$$\begin{aligned} \left| \mathbb{E} U_{t_k}(\bar{X}_k^{(j-1)}) \psi(t_{k-j+1}) \right| &\leq \left| \mathbb{E} U_{t_k}(\bar{X}_k^{(j)}) \psi(t_{k-j}) \right| \\ &\quad + \left| \mathbb{E} \psi(t_{k-j+1}) \int_{t_{k-j}}^{t_{k-j+1}} u^{(1)}(\bar{X}_k^{(j)} + \lambda\delta; s) [X(s) - b] ds \right| \\ &\quad + I_{k-j} \left| \mathbb{E} \psi(t_{k-j+1}) u^{(1)}(\bar{X}_k^{(j)} + \lambda\delta; t_{k-j}) [X(t_{k-j}) - b] \right|. \end{aligned} \quad (3.33)$$

The second term in the right-hand side of (3.33) is estimated as

$$\begin{aligned} &\left| \mathbb{E} \psi(t_{k-j+1}) \int_{t_{k-j}}^{t_{k-j+1}} u^{(1)}(\bar{X}_k^{(j)} + \lambda\delta; s) [X(s) - b] ds \right| \\ &\leq \sup |u^{(1)}| \int_{t_{k-j}}^{t_{k-j+1}} \sqrt{\mathbb{E} \psi^2(t_{k-j+1})} \sqrt{\mathbb{E} [X(s) - b]^2} ds \\ &\leq \frac{K}{T - t_{k-j+1}} \int_{t_{k-j}}^{t_{k-j+1}} \sqrt{T - s} ds \\ &\leq \frac{K}{\sqrt{T - t_{k-j+1}}} (t_{k-j+1} - t_{k-j}). \end{aligned}$$

The third term in the right-hand side of (3.33) is estimated as $K I_{k-j} / \sqrt{T - t_{k-j+1}}$. Then

$$\begin{aligned} \left| \mathbb{E} U_{t_k}(\bar{X}_k^{(j-1)}) \psi(t_{k-j+1}) \right| &\leq \left| \mathbb{E} U_{t_k}(\bar{X}_k^{(j)}) \psi(t_{k-j}) \right| + \frac{K}{\sqrt{T - t_{k-j+1}}} (t_{k-j+1} - t_{k-j}) \\ &\quad + \frac{K I_{k-j}}{\sqrt{T - t_{k-j+1}}}, \quad j = 1, \dots, k. \end{aligned} \quad (3.34)$$

It follows from (3.30), (3.34), and the evident inequality $\left| \mathbb{E} U_{t_k}(\bar{X}_k^{(k)}) \psi(0) \right| \leq K$ that

$$\left| \mathbb{E} U_{t_k}(\bar{X}_k) \psi(t_k) \right| \leq \frac{K}{\sqrt{T-t_k}} + K \sum_{j=1}^k \frac{(t_{k-j+1} - t_{k-j})}{\sqrt{T-t_{k-j+1}}} + K \sum_{j=1}^k \frac{I_{k-j}}{\sqrt{T-t_{k-j+1}}}.$$

Recalling that the number of points θ_i is equal to the fixed n , we get $\sum_{j=1}^k I_{k-j} \leq n$. Finally, we obtain

$$\begin{aligned} \left| \mathbb{E} U_{t_k}(\bar{X}_k) \psi(t_k) \right| &\leq \frac{K}{\sqrt{T-t_k}} + \frac{K}{\sqrt{T-t_k}} \sum_{j=1}^k (t_{k-j+1} - t_{k-j}) + \frac{K}{\sqrt{T-t_k}} \sum_{j=1}^k I_{k-j} \\ &\leq \frac{K}{\sqrt{T-t_k}}. \end{aligned}$$

Lemma 3.3.1 is proved. ■

Remark 3.3.2 *It is notable that in the multidimensional case ($d > 1$), the integrand of (3.21) contains cross-terms in all coordinate pairs i, j , viz $\delta_{k,1}^i(s_1) \delta_{k,1}^j(s_2) - \delta_{k,0}^i(s_1) \delta_{k,0}^j(s_2)$. The terms corresponding to $i = j$ are estimated in the same way as in the considered one-dimensional case. For $i \neq j$, the contribution from all stochastic integral terms is zero and the right-hand side of (3.21) has terms with $(b^i - X^i(t_k))(b^j - X^j(t_k))/(T - t_k)^2$, which are martingales [MT04ii, Corollary A.1] and their further estimation yields $O(h^3/\sqrt{T-t_k})$ again. In (3.23) it should be understood that the term $1/(T-t_k)$ only appears for $i = j$.*

3.4 Integral-type functionals

In this section we consider conditional Wiener integrals of integral-type functionals:

$$F(x(\cdot)) = \varphi \left(x(\theta), \int_0^T f(t, x(t)) dt \right), \quad 0 < \theta < T, \quad x \in C_{0,a;T,b}^d. \quad (3.35)$$

Introduce the scalar process $Z(t)$ satisfying the equation

$$dZ = f(t, X(t))dt, \quad Z(0) = 0, \quad (3.36)$$

where $X(t)$ is the solution of (2.21). Clearly, the conditional Wiener integral \mathcal{U} from (3.1) of the functional (3.35) is equal to the expectation

$$\mathcal{U} = \mathbb{E} \varphi(X(\theta), Z(T)). \quad (3.37)$$

The approximation (3.7), (3.6) applied to (3.1), (3.35) results in the trapezoidal method for Z :

$$\mathcal{U} \approx \bar{\mathcal{U}} = \mathbb{E}\varphi(X(\theta), Z_N), \quad (3.38)$$

where

$$\begin{aligned} Z_0 &= 0, \\ Z_{k+1} &= Z_k + \frac{t_{k+1} - t_k}{2} [f(t_k, X(t_k)) + f(t_{k+1}, X(t_{k+1}))], \quad k = 0, \dots, N-1. \end{aligned} \quad (3.39)$$

Recall that the time discretization used here is so that $\theta \in \{t_0, t_1, \dots, t_N\}$.

If we assume that $\varphi(x, z)$ and $f(t, x)$ have bounded derivatives up to a sufficiently high order, it follows from the general Theorem 3.2.1 that the method (3.38), (3.39) for (3.1), (3.35) has the second order of accuracy; i.e., the estimate (3.9) is valid for it. The other set of assumptions under which the theorem is valid are that $f(t, x)$ and its derivatives up to a sufficiently high order are bounded and $\varphi(x, z)$ is sufficiently smooth. We note that in the case of integral-type functionals, the convergence theorem can be proved more simply, exploiting a more standard technique used in the weak-sense approximation of SDEs [MT04] (see its application in the case of conditional Wiener integrals of exponential-type functionals in [MT04ii] and in the case of usual Wiener integrals in [VGM84]). It is interesting that no method of the form

$$Z_{k+1} = Z_k + (t_{k+1} - t_k) \sum_{i=1}^3 \alpha_i f(t_k + \beta_i, X(t_k + \beta_i)), \quad \alpha_i \in \mathbb{R}, \quad \beta_i \in [0, t_{k+1} - t_k],$$

has order of accuracy higher than two (in the case of usual Wiener integrals, see a similar comment in [VGM84]). At the same time, in the case of integral-type functionals of a particular form – the exponential-type functionals $F(x(\cdot)) = \exp[\int_0^T f(t, x(t)) dt]$, a fourth-order Runge-Kutta method was constructed in [MT04ii].

We made a computational comparison between (3.39) and the fourth-order Runge-Kutta method in computing the potential energy of one particle in a 1D harmonic oscillator. Despite being of lower order, the method (3.39) turns out to be preferable due to its stability properties. These follow from preservation by (3.39) of such structural properties of

exponential-type functionals as positivity and monotonicity, which can be broken down in the case of the fourth-order Runge-Kutta method from [MT04ii] (see similar observations although in a different context in [MT09]). Further, instead of the trapezoidal rule (3.39), we can use Simpson's rule:

$$\begin{aligned} Z_0 &= 0, \\ Z_{k+1} &= Z_k + \frac{t_{k+1} - t_k}{6} \left[f(t_k, X(t_k)) + 4f(t_{k+1/2}, X(t_{k+1/2})) + f(t_{k+1}, X(t_{k+1})) \right], \\ k &= 0, \dots, N-1. \end{aligned} \tag{3.40}$$

Although both methods (3.39) and (3.40) are of order two, the method (3.40) had much smaller bias in our experiments than the method (3.39) and thus was computationally more effective. The methods (3.38), (3.39) and (3.38), (3.40) extend the arsenal of numerical tools considered in [MT04ii, MT04] for computing action integrals given by (2.32).

Chapter 4

Simulations of Boltzmann and bosonic statistics

In Chapter 1, we explained why it was important to develop a weak approximation to our functional integral, in order to then apply the Monte Carlo technique. In the previous chapter, we introduced such a weak approximation method. In this chapter, we shall begin by offering a discussion of simulation methods, in Sections 4.1 and 4.2. In Section 4.3, we then provide some illustrative experimental results, which are in agreement with the theoretical predictions of the previous chapter. The reader should understand that no particularly novel ideas are introduced in this Chapter; rather, it is included to provide a sound platform for what follows.

4.1 The probabilistic approach to finding expectations of observables

In Chapter 1, we already introduced the Monte Carlo method for finding information about conditional Wiener integrals (cf (3.1)):

$$\mathcal{U} = \int_{\mathcal{C}_{0,x_0}^{T,x_T}} F(x) dw_{0,x_0}^{T,x_T}$$

but in order to find expectations of observables, we must make use of (2.69) or (2.71), and both of these take the form of one functional integral divided by another, this latter being the partition function \mathcal{Z} , given by (2.55). Considering (2.77) and (2.72), it is clear that when (2.71) leads to an expression suitable for practical implementation, the denominator \mathcal{Z} is often retained and the numerator still involves integration with respect to $w_{0,y}^{T,y}$ or $w_{0,y}^{T,\Pi_n y}$, or another conditional Wiener measure. Consequently it is of interest to consider how to simulate a ratio of two "functional integrals" \mathcal{A}/\mathcal{Z} where for some functional F and for the action functional Y ,

$$\mathcal{A} = \frac{1}{n!} \sum_{\pi \in \Pi_n} c(\pi) \int_{\mathbb{R}^{nd}} \int_{\mathcal{C}_{0,y}^{T,\pi y}} \varphi_T(y, \pi y) F(x) dw_{0,y}^{T,\pi y}(x) dy \quad (4.1)$$

$$\mathcal{Z} = \frac{1}{n!} \sum_{\pi \in \Pi_n} c(\pi) \int_{\mathbb{R}^{nd}} \int_{\mathcal{C}_{0,y}^{T,\pi y}} \varphi_T(y, \pi y) Y(x) dw_{0,y}^{T,\pi y}(x) dy. \quad (4.2)$$

Here for concision we define $c(\pi)$ for the case of distinguishable particles to be $n!$ for the identity and 0 otherwise. We still let $c(\pi) = 1$ for bosons and $c(\pi) = \text{sgn}(\pi)$ for fermions. There are a number of rearrangements of (4.1)-(4.2) based on exchanging the sequence of the sum and integrals, and this shall be discussed further in Subsection 4.2.3.

We should also note that if rather than (4.1), (2.71) were to lead to an expression with numerator

$$\mathcal{A} = \frac{1}{n!} \sum_{\pi \in \Pi_n} c(\pi) \int_{\mathbb{R}^{nd} \times \mathbb{R}^{nd}} \int_{\mathcal{C}_{0,y_1}^{T,\pi y_2}} \varphi_T(y_1, \pi y_2) F(x) dw_{0,y_1}^{T,\pi y_2}(x) dy_1 dy_2$$

then it should be clear how to extend the discussion of this chapter to that case.

4.1.1 Integrating in the space of loops

Let us first consider how information about quantities such as \mathcal{A} (including \mathcal{Z}) may be obtained. We refer to these as functional integrals because we may regard them as integrals in the space of loops $\mathcal{C}_O = \{x \in \mathcal{C}^{[0,T]^{nd}} : x(0) = x(T)\}$, as follows. Let $\ell(y, \pi y) : [0, T] \rightarrow \mathbb{R}^{nd}$ be given by $\ell_{y,\pi y}(s) = (1 - \frac{s}{T})y + \frac{s}{T}\pi y$. Let us use the notation $\Lambda \times w_{0,y}^{T,y}$ for the measure on \mathcal{C}_O induced by the law of $z + \ell_{y,y}$ when $(y, z) \sim \Lambda \times w_{0,0}^{T,0}$, ie the product of Lebesgue measure

with the conditional Wiener measure $w_{0,0}^{T,0}$. Then using the linear translation property of the conditional Wiener measure, which was outlined in Subsection 2.1.3,

$$\begin{aligned}
\mathcal{A} &= \frac{1}{n!} \int_{\mathbb{R}^{nd}} \int_{\mathcal{C}_{0,0}^{T,0}} \sum_{\pi \in \Pi_n} c(\pi) \varphi_T(y, \pi y) F(z + \ell_{y,\pi y}) dw_{0,0}^{T,0}(z) dy \\
&= \frac{1}{n!} \int_{\mathcal{C}_O} \sum_{\pi \in \Pi_n} c(\pi) \varphi_T(y, \pi y) F(x - \ell_{y,y} + \ell_{y,\pi y}) d\left(\Lambda \times w_{0,y}^{T,y}\right)(y, x) \\
&: = \int_{\mathcal{C}_O} G(x) d\left(\Lambda \times w_{0,y}^{T,y}\right)(y, x) .
\end{aligned} \tag{4.3}$$

To facilitate exposition, for the remainder of this section and up until Subsection 4.2.3, we shall discuss how to estimate \mathcal{A}_D and \mathcal{Z}_D , the functional integrals relevant for distinguishable particles (boltzmannons). Discarding $\varphi_T(y, y)$,

$$\mathcal{A}_D = \int_{\mathcal{C}_O} F(x) d\left(\Lambda \times w_{0,y}^{T,y}\right)(y, x) . \tag{4.4}$$

The reader is asked to bear in mind, however, that the following discussion applies readily to the case of integrating G rather than F and that thus, as we explain in Subsection 4.2.3, it is relevant to at least two ways of approximating \mathcal{A}/\mathcal{Z} .

Clearly $\Lambda \times w_{0,0}^{T,0}$ is not a probability measure, but using the Radon-Nikodym theorem (see e.g. [Coh80]) we may nonetheless rewrite (4.4) in a form that is accessible for probabilistic methods. Where μ is a Lebesgue-equivalent probability measure over $\mathcal{B}(\mathbb{R}^{nd})$ and $f : \mathbb{R}^{nd} \rightarrow \mathbb{R}$ is a measurable function,

$$\int_{\mathbb{R}^{nd}} f(x) d\Lambda(x) = \int_{\mathbb{R}^{nd}} f(x) \frac{d\Lambda}{d\mu}(x) d\mu(x) .$$

Consequently, for any Lebesgue-equivalent probability measure μ over $\mathcal{B}(\mathbb{R}^{nd})$,

$$\mathcal{A}_D = \mathbb{E}_{\xi \sim \mu} \left[\mathbb{E}_{X \sim w_{0,\xi}^{T,\xi}} [F(X)] \frac{d\Lambda}{d\mu}(\xi) \right]$$

and we should keep in mind that a similar fact holds with regard to $\widetilde{\mathcal{A}}_D$, ie the approximation to \mathcal{A}_D that arises from substituting an approximating functional \widetilde{F} for F ; and also with regard to $\widetilde{\mathcal{Z}}_D$. Then if we proceed to define a set of i.i.d. random variables $\{\xi_i\}_{i=1}^M$ with $\xi_i \sim \mu$, and let ${}_i X_{\xi_i}$ be a set of i.i.d. random variables with codomain $\mathbb{R}^{(N+1)nd}$ and the

law of ${}_iX_{i\xi}$ induced from $w_{0,i\xi}^{T,i\xi}$, and (recognising that \tilde{F} , although defined on $A[0, T]$, is measurable w.r.t. $X(h), X(2h), \dots, X(T)$) we define a random variable

$$\tilde{\mathcal{A}}^{\text{MC}} = \frac{1}{M} \sum_{i=1}^M \left[\tilde{F}({}_iX_{i\xi}) \frac{d\Lambda}{d\mu}({}_i\xi) \right]$$

then it is clear that according to the perspective advanced in Chapter 1, if the drawn value of $\tilde{\mathcal{A}}^{\text{MC}}$ is x then the posterior distribution of $\tilde{\mathcal{A}}_D$ is approximately Gaussian $(x, \text{Var}(\tilde{\mathcal{A}}^{\text{MC}}))$.

The question arises of what choice of μ is expedient; basic intuition tells us that we would like μ not to ascribe a high measure to regions where $\mathbb{E}_{X_\xi \sim w_{0,\xi}^{T,\xi}}[F(X_\xi)]$ is near zero, and to ascribe adequate measure to regions where it is not. This is complemented by recalling the standard conditional variance formula (e.g. [Wil01]), which yields

$$\begin{aligned} \text{Var} \left[\begin{array}{l} \xi \sim \mu \\ X_\xi \sim w_{0,\xi}^{T,\xi} \end{array} \right] \left(\tilde{F}(X_\xi) \frac{d\Lambda}{d\mu}(\xi) \right) &= \\ &= \int_{\mathbb{R}^{nd}} \text{Var} \left(\tilde{F}(X_y) \frac{d\Lambda}{d\mu}(y) \right) d\mu(y) + \text{Var}_{\xi \sim \mu} \mathbb{E} \left[\tilde{F}(X_\xi) \frac{d\Lambda}{d\mu}(\xi) \right] \\ &= \int_{\mathbb{R}^{nd}} \text{Var} \left(\tilde{F}(X_y) \right) \frac{d\Lambda}{d\mu}(y)^2 dy - \tilde{\mathcal{A}}_D^2 + \int_{\mathbb{R}^{nd}} \mathbb{E} \left[\tilde{F}^2(X_y) \frac{d\Lambda}{d\mu}(y)^2 \right] d\mu(y) \\ &= \int_{\mathbb{R}^{nd}} \left(\text{Var} \left(\tilde{F}(X_y) \right) + \mathbb{E} \left[\tilde{F}^2(X_y) \right] \right) \frac{d\Lambda}{d\mu}(y)^2 dy - \tilde{\mathcal{A}}_D^2 \end{aligned}$$

Therefore where we set $p = \frac{d\mu}{d\Lambda}$, optimal choice of p (and hence μ) makes the function of y given by

$$\ddagger(y) = \frac{d}{dp(y)} \left(\left(\text{Var} \left(\tilde{F}(X_y) \right) + \mathbb{E} \left[\tilde{F}^2(X_y) \right] \right) / p(y) \right)$$

a constant function; that is to say, $\ddagger(y_1) = \ddagger(y_2)$ for all $y_1, y_2 \in \mathbb{R}^{nd}$. It follows that for the optimal μ ,

$$\frac{d\mu}{d\Lambda}(y) \propto \left(\text{Var} \left(\tilde{F}(X_y) \right) + \mathbb{E} \left[\tilde{F}^2(X_y) \right] \right)^{1/2}.$$

If we guess that $\text{Var} \left(\tilde{F}(X_y) \right)$ may be somewhat proportional to $\mathbb{E} \tilde{F}(X_y)$ then one good choice of μ would be to make $\frac{d\mu}{d\Lambda}$ proportional to $\left(\mathbb{E} \tilde{F}(X_y) \right)^\alpha$ for some optimal $\alpha \in (1/2, 1)$. It seems conceivable that $\mathbb{E} \left[\tilde{F}^2(X_y) \right]$ often dominates; for example if the conditional distribution of $\tilde{F}(X_y)$ were $\text{Uniform}[s, t]$ for some $0 \leq s < t < 1$ then $\mathbb{E} \left[\tilde{F}(X_y)^2 \right] = \frac{(t-s)^2}{3} + ts$ whereas $\text{Var} \left(\tilde{F}(X_y) \right) = \frac{(t-s)^2}{12}$. Thus we may imagine that the optimal α is towards 1.

In simple experiments, using the Gibbs measure for V , ie, $\frac{d\mu}{d\Lambda}(y) \propto \exp(-TV(y))$, may achieve similar effects to optimizing for \mathcal{Z}_D , as long as T is sufficiently small; in some cases may be possible to sample according to this μ directly. In most of the simulations performed in this chapter, we simply use a Gaussian μ , with the covariance matrix optimized by trial and error in order to achieve a lower variance for $\tilde{\mathcal{A}}^{\text{MC}}, \tilde{\mathcal{Z}}^{\text{MC}}$, and hence for the posterior distributions of $\tilde{\mathcal{A}}_D, \tilde{\mathcal{Z}}_D$.

Estimation of the ratio of functional integrals

We drop the D on \mathcal{A}_D since the following discussion carries over with no significant changes for indistinguishable particles.

We have already assumed that we cannot compute quantities such as \mathcal{A} and \mathcal{Z} directly, and instead use numerical methods $\tilde{\mathcal{A}} \approx \mathcal{A}$ and $\tilde{\mathcal{Z}} \approx \mathcal{Z}$. Let us denote the biases by $\varepsilon_1, \varepsilon_2 \in \mathbb{R}$ so that $\tilde{\mathcal{A}} = \mathcal{A} + \varepsilon_1$, $\tilde{\mathcal{Z}} = \mathcal{Z} + \varepsilon_2$. The question immediately arises of in what sense the posterior distribution of $\tilde{\mathcal{A}}, \tilde{\mathcal{Z}}$ gives rise to information about \mathcal{A}/\mathcal{Z} . If we are content to use the simple heuristic of considering a Taylor expansion of the function x^{-1} about \mathcal{Z} for $(\mathcal{Z} + \varepsilon_2)^{-1}$ then in fact this yields a most fortuitous result:

$$\langle \widetilde{\mathcal{A}} \rangle := \frac{\tilde{\mathcal{A}}}{\tilde{\mathcal{Z}}} = \frac{\mathcal{A}}{\mathcal{Z}} \left(1 - \frac{\varepsilon_2}{\mathcal{Z}} + \frac{\varepsilon_1}{\mathcal{A}} + \text{H.O.T.} \right) \quad (4.5)$$

where H.O.T. indicates terms of order 2 and above in $\varepsilon_1, \varepsilon_2$. Therefore we may note in passing that if it happens that $\varepsilon_1/\varepsilon_2$ is close to \mathcal{A}/\mathcal{Z} , then it appears that this will have a favourable effect on the bias for the ratio. More generally, it is clear that the ratio bias is the same order in the time-step h as $\varepsilon_1, \varepsilon_2$.

Recalling the Bayesian framework of Chapter 1, if we think of $\tilde{\mathcal{A}}, \tilde{\mathcal{Z}}$ and $\tilde{\mathcal{A}}^{\text{MC}}, \tilde{\mathcal{Z}}^{\text{MC}}$ as modelled by random variables θ_1, θ_2 and ϑ_1, ϑ_2 respectively, with a joint pdf, then we speak of the (marginal) joint pdf of (θ_1, θ_2) as the prior pdf for $(\tilde{\mathcal{A}}, \tilde{\mathcal{Z}})$, and given a realisation of $(\tilde{\mathcal{A}}^{\text{MC}}, \tilde{\mathcal{Z}}^{\text{MC}}) = (x_1, x_2)$, we speak of the conditional pdf of (θ_1, θ_2) given $(\vartheta_1, \vartheta_2) = (x_1, x_2)$ as the posterior pdf for $(\tilde{\mathcal{A}}, \tilde{\mathcal{Z}})$; likewise for $\tilde{\mathcal{A}}/\tilde{\mathcal{Z}}$. If we collect enough information that the posterior distribution of $(\tilde{\mathcal{A}}, \tilde{\mathcal{Z}})$ is (approximately) bivariate Gaussian about $(\tilde{\mathcal{A}}^{\text{MC}}, \tilde{\mathcal{Z}}^{\text{MC}})$,

with $\tilde{\mathcal{A}} = \tilde{\mathcal{A}}^{\text{MC}} + \xi_1, \tilde{\mathcal{Z}} = \tilde{\mathcal{Z}}^{\text{MC}} + \xi_2$ then writing

$$\frac{1}{\tilde{\mathcal{Z}}^{\text{MC}} + \xi_2} = \frac{1}{\tilde{\mathcal{Z}}^{\text{MC}}} - \frac{\xi_2}{\tilde{\mathcal{Z}}^{\text{MC}^2}} + \frac{\xi_2^2}{\tilde{\mathcal{Z}}^{\text{MC}^3}} + \dots$$

it follows that conditional on $(\tilde{\mathcal{A}}^{\text{MC}}, \tilde{\mathcal{Z}}^{\text{MC}})$, where we use $E_{\frac{\tilde{\mathcal{A}}}{\tilde{\mathcal{Z}}}}$ to indicate the posterior expectation of $\frac{\tilde{\mathcal{A}}}{\tilde{\mathcal{Z}}}$, neglecting the contribution from priors we have

$$E_{\frac{\tilde{\mathcal{A}}}{\tilde{\mathcal{Z}}}} = \frac{\tilde{\mathcal{A}}^{\text{MC}}}{\tilde{\mathcal{Z}}^{\text{MC}}} - \frac{\text{Cov}(\xi_1, \xi_2)}{\tilde{\mathcal{Z}}^{\text{MC}^2}} + \frac{\tilde{\mathcal{A}}^{\text{MC}} \text{Var}(\xi_2)}{\tilde{\mathcal{Z}}^{\text{MC}^3}} + \text{H.O.T.}$$

where H.O.T. indicates expectations of terms of order 3 and above in $\xi_{1,2}/\tilde{\mathcal{Z}}^{\text{MC}}$. If all the other terms can be safely neglected then we shall think that

$$E_{\frac{\tilde{\mathcal{A}}}{\tilde{\mathcal{Z}}}} \approx \frac{\tilde{\mathcal{A}}^{\text{MC}}}{\tilde{\mathcal{Z}}^{\text{MC}}} \quad . \quad (4.6)$$

Neglecting the contribution from priors, the posterior variance of $\frac{\tilde{\mathcal{A}}}{\tilde{\mathcal{Z}}}$ is approximately equal to $\text{Var}\left(\frac{\tilde{\mathcal{A}}^{\text{MC}}}{\tilde{\mathcal{Z}}^{\text{MC}}}\right)$ (we shall further discuss the posterior distribution of $\frac{\tilde{\mathcal{A}}}{\tilde{\mathcal{Z}}}$ shortly). By expanding $\tilde{\mathcal{Z}}^{\text{MC}^{-1}}$ as before, roughly speaking we have:

$$\text{Var}\left(\frac{\tilde{\mathcal{A}}^{\text{MC}}}{\tilde{\mathcal{Z}}^{\text{MC}}}\right) \approx \frac{\text{Var}\left(\tilde{\mathcal{A}}^{\text{MC}}\right)}{E\left[\tilde{\mathcal{Z}}^{\text{MC}}\right]^2} - 2 \frac{E\left[\tilde{\mathcal{A}}^{\text{MC}}\right] \text{Cov}\left(\tilde{\mathcal{A}}^{\text{MC}}, \tilde{\mathcal{Z}}^{\text{MC}}\right)}{E\left[\tilde{\mathcal{Z}}^{\text{MC}}\right]^3} + \frac{E\left[\tilde{\mathcal{A}}^{\text{MC}}\right]^2 \text{Var}\left(\tilde{\mathcal{Z}}^{\text{MC}}\right)}{E\left[\tilde{\mathcal{Z}}^{\text{MC}}\right]^4} \quad (4.7)$$

so clearly, sampling $\tilde{\mathcal{A}}^{\text{MC}}, \tilde{\mathcal{Z}}^{\text{MC}}$ independently would be inefficient. It is advantageous to obtain as high a covariance between $\tilde{\mathcal{A}}^{\text{MC}}$ and $\tilde{\mathcal{Z}}^{\text{MC}}$ as possible. A simple approach is to use the same Brownian bridges for both, and let the initial points be perfectly correlated:

$$\tilde{\mathcal{A}}^{\text{MC}} = \frac{1}{M} \sum_{i=1}^M \left[\tilde{F}({}_iX_{(1)}) \frac{d\Lambda}{d\mu_1}({}_i\xi) \right] \quad (4.8)$$

$$\tilde{\mathcal{Z}}^{\text{MC}} = \frac{1}{M} \sum_{i=1}^M \left[\tilde{Y}({}_iX_{(2)}) \frac{d\Lambda}{d\mu_2} \left(\frac{\sigma_2}{\sigma_1} {}_i\xi \right) \right] \quad (4.9)$$

where $\{{}_i\xi\}_{i=1}^M$ is a set of i.i.d. r.v.'s with ${}_i\xi \sim \text{Gaussian}(0, \sigma_1^2)$, and $\{{}_iX_{(1)}\}_{i=1}^M$ is a set of i.i.d. r.v.'s with codomain $\mathbb{R}^{(N+1)nd}$, s.t. the law of ${}_iX_{(1)}$ induces $w_{0, {}_i\xi}^{T, {}_i\xi}$; and where ${}_iX_{(2)} = {}_iX_{(1)} - {}_i\xi + \frac{\sigma_2}{\sigma_1} {}_i\xi$. (In the case of indistinguishable particles, one approach would be, for example, to replace F with G , cf (4.3), and likewise for Y .)

Discretisation sampling methods

Some works, such as [Cep95], describe sampling Brownian bridges using the Bisection method due to Lévy [Lev39]. However, it is sometimes more convenient to apply the incremental construction of the Brownian bridge as introduced at (2.25). The cost to achieve particular Monte Carlo errors should be identical to that obtained from using the Lévy construction, or indeed from using any other method of sampling the Gaussian finite-dimensional distribution of the discretised path points. In Chapter 3, we used the concept of incremental sampling in order to prove Theorem 3.2.1. However, it is clear that the method of sampling the discretisation has no impact on the properties of a numerical method for which the sampled discretisation serves as input; this shall be emphasised again in Subsection 4.2.4.

4.1.2 Making interval estimates for a ratio of bivariate Gaussian variables

Clearly it is desirable to be able to report a credibility interval for $\tilde{\mathcal{A}}/\tilde{\mathcal{Z}}$, given a bivariate Gaussian posterior density for $\tilde{\mathcal{A}}, \tilde{\mathcal{Z}}$. We could simply employ (4.7), assuming that we shall use the sample covariance matrix of $\tilde{\mathcal{A}}^{\text{MC}}, \tilde{\mathcal{Z}}^{\text{MC}}$ as if it were the actual covariance matrix. We then proceed to approximate the distribution of $\tilde{\mathcal{A}}/\tilde{\mathcal{Z}}$ as Gaussian and voilà! However, there is no guarantee that the interval thus obtained genuinely has probability at least p . Approximating the ratio distribution as Gaussian is unjustified; in fact the ratio distribution is not symmetric, and indeed for some parameters is not unipolar. A better alternative in this direction would be to exploit an approximation result such as [Hin69].

Bivariate Gaussian contour method

It is usually preferable to find an interval (r_1, r_2) guaranteed to be such that *a posteriori*, $P\left(\tilde{\mathcal{A}}/\tilde{\mathcal{Z}} \in (r_1, r_2)\right) > p$. One approach is to consider the contour of likelihood bounding a set of measure p , and form a sector from 0 bounded by rays which are tangent to this ellipse. Solving equations yields four candidate points which must be checked for low and

high extreme values of x/y . Where the sampled value for $(\tilde{\mathcal{A}}^{\text{MC}}, \tilde{\mathcal{Z}}^{\text{MC}})$ is (x_0, y_0) , and where we let $\text{Var}(\tilde{\mathcal{A}}^{\text{MC}}) = \sigma_x^2$, $\text{Var}(\tilde{\mathcal{Z}}^{\text{MC}}) = \sigma_y^2$ and denote the correlation by ρ , and where we set

$$K_0 = \frac{x_0^2}{\sigma_x^2} - 2\rho \frac{x_0 y_0}{\sigma_x \sigma_y} + \frac{y_0^2}{\sigma_y^2}$$

$$K = -2(1 - \rho^2) \ln \left((1 - \rho) \sqrt{1 - \rho^2} \right)$$

the possible solutions (points that may extremise x/y) have

$$y = \left(1 - \frac{K}{K_0}\right) y_0 \pm \sqrt{K \left(1 - \frac{K}{K_0}\right) \left(\frac{\sigma_y^2}{1 - \rho^2} - \frac{y_0^2}{K_0}\right)}$$

and for each of these values,

$$x = x_0 + \sigma_x \left(\rho \frac{(y - y_0)}{\sigma_y} \pm \sqrt{K - \frac{(y - y_0)^2}{\sigma_y^2} (1 - \rho^2)} \right).$$

The main limitation of this procedure is that it delivers intervals which become increasingly wide relative to the true ones as ρ gets close to 1.

Bivariate Gaussian Monte Carlo method

If one wishes to find a more accurate interval, happily this may be accomplished by a straightforward expedient: run a post-process Monte Carlo. We let $\{B_i\}_{i=1}^{N_{\text{boxes}}}$ be a partition of some interval (a, b) such that we are happy to assume *a priori* that $\tilde{\mathcal{A}}/\tilde{\mathcal{Z}}$ lies within (a, b) . (For example, we might choose, say, $N_{\text{boxes}} = 10^4$.) We could obtain (a, b) using one of the above methods, or by trial and error. Then by making a large number M_{ratio} (say, 10^9) of bivariate Gaussian draws (again, treating the collected sample covariance matrix as the true covariance matrix) we may store an array with the sampled frequencies that the ratio lies within each B_i . A collection of the B_i which form a Highest Probability Density (HPD) interval can then be ascertained from the array of frequencies. To get a satisfactory estimate this way is not instant, but typically takes less than 10 minutes using a 2.0 GHz machine. The question of exactly how the observed HPD interval differs from the true HPD interval depends on the posterior distribution of $\tilde{\mathcal{A}}/\tilde{\mathcal{Z}}$, and is not addressed here.

4.2 Importance sampling and Markov Chain Monte Carlo

In this section we discuss importance sampling, explain that Markov Chain Monte Carlo (MCMC) is used in order to achieve this, discuss optimal sampling, and introduce the action measure, which has a derivative w.r.t. $\Lambda \times w_{0,y}^{T,y}$ that is proportional to $Y(x)$. We also explain how we may extend our MCMC to include sampling of initial points and permutations. The MCMC approach [Met53] is well-known and was first introduced in the path integral context in [FC81]; a comprehensive and modern explanation is offered in [Cep95, Cep03]. This section draws on material from [Cep95], and in particular, Section 5F of that work explains the Multilevel Metropolis method as it applies to bosons. However, our focus is on showing that MCMC is the logical development of the approach discussed in the previous section. A paper which explains a related idea and also makes some other interesting points is [GM98]. A good resource on the general topic of MCMC is [GRS96]; see also [Mur07, Nea93].

4.2.1 Importance sampling and the action measure

In some cases it will be undesirable to use the conditional Wiener measure to sample paths. Intuitively, if there are, for instance, many repelling particles, then most Brownian bridges will have a near-zero contribution to $\tilde{\mathcal{A}}^{\text{MC}}, \tilde{\mathcal{Z}}^{\text{MC}}$. Consequently, it is desirable to reduce the Monte Carlo variance by using *importance sampling* (cf [MT04, p.123]); that is, by using a different sampling measure and applying the Radon-Nikodym theorem. (See also [HF04, p.253] for a discussion of importance sampling and Monte Carlo in a slightly different context.)

To simplify the discussion, we shall consider the finite-dimensional object that is actually evaluated. Let $w^* = \Lambda \times w_{0,y}^{T,y}$. Let us write $\tilde{\mathcal{B}} = \mathcal{B}(X(0), X(h), X(2h), \dots, X(T))$ and let \tilde{w}^* be the measure on $\tilde{\mathcal{B}}$ induced by the finite-dimensional distribution corresponding to w^* . Rather than using (4.8)-(4.9), let us consider sampling ${}_iX$ with respect to some $\tilde{\nu}$ which is

equivalent with \tilde{w}^* and using this same ${}_iX$ for both $\tilde{\mathcal{A}}^{\text{MC}}$ and $\tilde{\mathcal{Z}}^{\text{MC}}$:

$$\tilde{\mathcal{A}}^{\text{MC}} = \frac{1}{M} \sum_{i=1}^M \left[\tilde{F}({}_iX) \frac{d\tilde{w}^*}{d\tilde{\nu}}({}_iX) \right] \quad (4.10)$$

$$\tilde{\mathcal{Z}}^{\text{MC}} = \frac{1}{M} \sum_{i=1}^M \left[\tilde{Y}({}_iX) \frac{d\tilde{w}^*}{d\tilde{\nu}}({}_iX) \right] . \quad (4.11)$$

This has the advantage that we may decide to let $\frac{d\tilde{\nu}}{d\tilde{w}^*}$ be proportional to any function $f : \mathbb{R}^{Nnd} \rightarrow \mathbb{R}$, without needing to know the integral of f with respect to \tilde{w}^* in order to evaluate contributions, if we are interested only in the ratio $\tilde{\mathcal{A}}^{\text{MC}} / \tilde{\mathcal{Z}}^{\text{MC}}$, because we may write

$$\frac{\tilde{\mathcal{A}}^{\text{MC}}}{\tilde{\mathcal{Z}}^{\text{MC}}} = \frac{\frac{1}{M} \sum_{i=1}^M \left[\tilde{F}({}_iX) \frac{1}{f({}_iX)} \right]}{\frac{1}{M} \sum_{i=1}^M \left[\tilde{Y}({}_iX) \frac{1}{f({}_iX)} \right]} . \quad (4.12)$$

Again employing the rough approximation (4.7) and noting that under (4.10)-(4.11),

$$\begin{aligned} Var(\tilde{\mathcal{Z}}^{\text{MC}}) &= \frac{1}{M} \left(\int_{\mathbb{R}^{Nnd}} \left[\tilde{Y}(x)^2 \frac{d\tilde{w}^*}{d\tilde{\nu}}(x)^2 \right] d\tilde{\nu}(x) - \tilde{\mathcal{Z}}^2 \right) \\ Var(\tilde{\mathcal{A}}^{\text{MC}}) &= \frac{1}{M} \left(\int_{\mathbb{R}^{Nnd}} \left[\tilde{F}(x)^2 \frac{d\tilde{w}^*}{d\tilde{\nu}}(x)^2 \right] d\tilde{\nu}(x) - \tilde{\mathcal{A}}^2 \right) \\ Cov(\tilde{\mathcal{A}}^{\text{MC}}, \tilde{\mathcal{Z}}^{\text{MC}}) &= \frac{1}{M} \left(\int_{\mathbb{R}^{Nnd}} \left[\tilde{F}(x) \tilde{Y}(x) \frac{d\tilde{w}^*}{d\tilde{\nu}}(x)^2 \right] d\tilde{\nu}(x) - \tilde{\mathcal{A}} \tilde{\mathcal{Z}} \right) \end{aligned}$$

we find that

$$Var\left(\frac{\tilde{\mathcal{A}}^{\text{MC}}}{\tilde{\mathcal{Z}}^{\text{MC}}}\right) \approx \frac{1}{M \tilde{\mathcal{Z}}^2} \left(\int_{\mathbb{R}^{Nnd}} \left[\left(\tilde{F}(x) - \tilde{Y}(x) \frac{\tilde{\mathcal{A}}}{\tilde{\mathcal{Z}}} \right)^2 \frac{d\tilde{w}^*}{d\tilde{\nu}}(x)^2 \right] d\tilde{\nu}(x) - K \right)$$

where $K \in \mathbb{R}$ is a constant. We assumed $\tilde{\nu}$ to be equivalent with \tilde{w}^* , so let us write p for

$\frac{d\tilde{\nu}}{d\Lambda} = \frac{d\tilde{\nu}}{d\tilde{w}^*} \frac{d\tilde{w}^*}{d\Lambda}$, a function from \mathbb{R}^{Nnd} to \mathbb{R}^+ . Then subject to

$$\int_{\mathbb{R}^{Nnd}} p(x) dx = 1$$

we may proceed by trying to minimize

$$\begin{aligned}
\int_{\mathbb{R}^{Nd}} \left[\left(\tilde{F}(x) - \tilde{Y}(x) \frac{\tilde{\mathcal{A}}}{\tilde{\mathcal{Z}}} \right)^2 \frac{d\tilde{w}^*}{d\tilde{\nu}}(x) \right] d\tilde{w}(x) &= \\
&= \int_{\mathbb{R}^{Nd}} \left[\left(\tilde{F}(x) - \tilde{Y}(x) \frac{\tilde{\mathcal{A}}}{\tilde{\mathcal{Z}}} \right)^2 \frac{d\tilde{w}^*}{d\tilde{\nu}}(x) \frac{d\tilde{w}^*}{d\Lambda}(x) \right] dx \\
&= \int_{\mathbb{R}^{Nd}} \left[\left(\tilde{F}(x) - \tilde{Y}(x) \frac{\tilde{\mathcal{A}}}{\tilde{\mathcal{Z}}} \right)^2 \frac{d\tilde{w}^*}{d\Lambda}(x) \frac{d\Lambda}{d\tilde{\nu}}(x) \frac{d\tilde{w}^*}{d\Lambda}(x) \right] dx \\
&= \int_{\mathbb{R}^{Nd}} \left[\left(\tilde{F}(x) - \tilde{Y}(x) \frac{\tilde{\mathcal{A}}}{\tilde{\mathcal{Z}}} \right)^2 \frac{d\tilde{w}^*}{d\Lambda}(x)^2 \frac{1}{p(x)} \right] dx
\end{aligned}$$

but then it is clear that the optimum is to make $\frac{d}{dp(x)} \left(\left(\tilde{F}(x) - \tilde{Y}(x) \frac{\tilde{\mathcal{A}}}{\tilde{\mathcal{Z}}} \right)^2 \frac{d\tilde{w}^*}{d\Lambda}(x)^2 \right) / p(x)$ equal at all $x \in \mathbb{R}^{Nd}$. It follows that

$$p(x) \propto \left| \left(\tilde{F}(x) - \tilde{Y}(x) \frac{\tilde{\mathcal{A}}}{\tilde{\mathcal{Z}}} \right) \frac{d\tilde{w}^*}{d\Lambda}(x) \right|$$

and thus that $\frac{d\tilde{\nu}}{d\tilde{w}}(x)$ is proportional to $\left| \tilde{F}(x) - \tilde{Y}(x) \frac{\tilde{\mathcal{A}}}{\tilde{\mathcal{Z}}} \right|$. Since $\frac{\tilde{\mathcal{A}}}{\tilde{\mathcal{Z}}}$ is not known a priori (in fact it is what we are trying to find), we would have to guess a value Υ for this in order to get a reasonable measure with respect to which we might integrate; perhaps a suitable value would be furnished by information from a short non-optimized simulation. We may then use (4.12):

$$\frac{\tilde{\mathcal{A}}^{\text{MC}}}{\tilde{\mathcal{Z}}^{\text{MC}}} = \frac{\frac{1}{M} \sum_{i=1}^M \left[\tilde{F}(iX) \frac{1}{|\tilde{F}(iX) - \Upsilon \tilde{Y}(iX)|} \right]}{\frac{1}{M} \sum_{i=1}^M \left[\tilde{Y}(iX) \frac{1}{|\tilde{F}(iX) - \Upsilon \tilde{Y}(iX)|} \right]}.$$

and no error is introduced here. Using a suboptimal measure (inaccurate Υ) affects only the variance obtained.

In the work [Cep96], the case is discussed of a quantity that is diagonal in the position representation, so that $F = AY$ and an intuitive discussion of the optimal sampling measure in that case is offered.

Using the action measure gets rid of the need for a normalizing constant

The author of [Cep96] then argues that it is preferable to instead let $\frac{d\tilde{\nu}}{dw}(x)$ be proportional to $|\tilde{Y}(x)|$. In the case of boltzmannons (and bosons), we shall then find that the denominator integral disappears completely in (4.12):

$$\frac{\tilde{\mathcal{A}}^{\text{MC}}}{\tilde{\mathcal{Z}}^{\text{MC}}} = \frac{1}{M} \sum_{i=1}^M \left[\frac{\tilde{F}(iX)}{\tilde{Y}(iX)} \right] \quad (4.13)$$

and in particular if $F = AY$ then the contributions are just $\tilde{A}(iX)$. Since we usually think of $F(x)$ as being positively correlated with $Y(x)$ under w^* , it seems that in general, this (discretized) **action measure** $\tilde{\nu}$ makes a sensible choice of measure. In [Cep96, p.5] several other arguments against using the optimal sampling measure are advanced.

In fact we could have performed all of the above discussion using the space of loops, although being able to utilize Lebesgue measure was clearly more convenient here. If we now introduce the (true) action measure ν , defined on $\mathcal{B}(\mathcal{C}_O)$ via $\frac{d\nu}{dw^*}(x) \propto Y(x)$; ie, to spell it out, for any $B \in \mathcal{B}(\mathcal{C}_O)$,

$$\nu(B) = \frac{\int_B Y(x) d\left(\Lambda \times w_{0,y}^{T,y}\right)(y, x)}{\int_{\mathcal{C}_O} Y(x) d\left(\Lambda \times w_{0,y}^{T,y}\right)(y, x)}. \quad (4.14)$$

then it is already evident that $\tilde{\nu}$ is the measure induced by the finite-dimensional distribution of ν . We shall loosely refer to either $\tilde{\nu}$ or ν as the action measure, in the case of distinguishable particles, without much cause for confusion. It is clear that

$$\langle A \rangle = \int_{\mathcal{C}_O} \frac{F(x)}{Y(x)} d\nu(x)$$

and we have introduced an approximation,

$$\langle \widetilde{A} \rangle = \int_{\mathbb{R}^{Nd}} \frac{\tilde{F}(x)}{\tilde{Y}(x)} d\tilde{\nu}(x). \quad (4.15)$$

4.2.2 Sampling according to the action measure via Markov Chain Monte Carlo

The question is then begged, however, of how sampling according to the action measure is to be performed. The answer is that since we know the relative probability of two points

in the sample space (ie, two discretisations), we may use Metropolis Markov Chain Monte Carlo [Met53, FC81, Cep03]. We form a Markov chain with a transition density such that when ergodicity is attained, the invariant measure is the measure that we seek. In order to obtain the correct transition density, we use a two-step algorithm consisting of (1) drawing a "proposal" or "candidate" move; (2) accepting or rejecting this move. Sampling according to the action measure and sampling via Markov Chain Monte Carlo (MCMC) are separate concepts which should not be intellectually confounded, although in practice it is necessary to run an MCMC in order to sample according to the action measure. The MCMC candidate move could in principle follow any distribution conditional on the current point in (discretised) path space.

In Path Integral Monte Carlo, a popular strategy for proposing candidate moves is known as the Multilevel Metropolis method [Cep95, Cep03] (note that despite the name, there is no connection with hierarchical modelling). In order to create a candidate for the MCMC step, first a subset A of the particles and a subset $\{i, \dots, i+2^j-1\}$, with counting modulo N , of the time-slices are sampled, according to whatever law we see fit. Outside of $A \times \{i, \dots, i+2^j-1\}$, the candidate will be left equal to the current position. Within $A \times \{i, \dots, i+2^j-1\}$, we sample "free-particle" moves, i.e. Brownian bridges with altered variance, to be the candidate moves. We now use Eq. (5.32) from [Cep95]. We shall use μ for the pdf that corresponds to the finite-dimensional distribution of the conditional Wiener measure. Let $s \in \mathbb{R}^{nd(2^j-1)}$ represent the current position and $s' \in \mathbb{R}^{nd(2^j-1)}$ the candidate position, and let $s_k \in \mathbb{R}^{nd(2^k-1)}$ denote the first k levels of s ; likewise s'_k . Let $p_k : \mathbb{R}^{nd(2^k-1)} \rightarrow \mathbb{R}$ denote likelihood in the target (that is, the pdf of the finite-dimensional distribution of $\tilde{\nu}$) for the first k levels. Meanwhile let $q_k : \mathbb{R}^{nd(2^k-1)} \rightarrow \mathbb{R}$ denote likelihood in the candidate for the first k levels. Then

$$p_k(s) = \mu(s_k) \tilde{Y}(s_k). \quad (4.16)$$

The acceptance probability for level k is then given by

$$P_k(s'_k) = \min \left[1, \frac{q_k(s_k | s'_{k-1}) p_k(s'_k) p_{k-1}(s_{k-1})}{q_k(s'_k | s'_{k-1}) p_k(s_k) p_{k-1}(s'_{k-1})} \right]. \quad (4.17)$$

Therefore if q_k is taken equal to μ , we have

$$\frac{q_k(s_k)p_k(s')p_{k-1}(s)}{q_k(s'_k)p_k(s)p_{k-1}(s')} = \frac{\mu(s_k|s_{k-1})\mu(s'_{k-1})\mu(s'_k|s'_{k-1})\mu(s_{k-1})}{\mu(s'_k|s'_{k-1})\mu(s_{k-1})\mu(s_k|s_{k-1})\mu(s'_{k-1})} \frac{\tilde{Y}(s'_k)\tilde{Y}(s_{k-1})}{\tilde{Y}(s'_{k-1})\tilde{Y}(s_k)} \quad (4.18)$$

$$= \frac{\tilde{Y}(s'_k)\tilde{Y}(s_{k-1})}{\tilde{Y}(s'_{k-1})\tilde{Y}(s_k)}. \quad (4.19)$$

Thus, distributing the candidate via Brownian bridge and then not introducing any compensating factors for this gives a correct formula. In effect if we take the candidate as per the Brownian bridge then we can accept or reject based on \tilde{Y} . This then yields sampling according to $\tilde{\nu}$. The Multilevel approach involves the additional steps to spread the chance of rejection over the levels, ie to include a rejection step at each iteration of the bisection method, since computing \tilde{Y} for the candidate path is assumed to be the most expensive part of the algorithm. In practice, when a section of the path is being resampled, often a cheap approximation to \tilde{Y} may be used for the earlier steps. It is possible to then weight back for this as further candidate path points $\{x(t_k)\}$ are chosen.

4.2.3 Including permutations in the random sampling

We have deliberately treated the case of distinguishable particles first but let us now reconsider how to treat bosons (and by extension, fermions). Recalling (4.3),

$$\mathcal{A} = \frac{1}{n!} \sum_{\pi \in \Pi_n} c(\pi) \int_{\mathcal{C}_O} \varphi_T(y, \pi y) F(x) d\left(\Lambda \times w_{0,y}^{T,\pi y}\right)(y, x) \quad (4.20)$$

$$= \int_{\mathcal{C}_O} G(x) d\left(\Lambda \times w_{0,y}^{T,y}\right)(y, x) \quad (4.21)$$

$$= \int_{\Pi_n \times \mathcal{C}_O} c(\pi) \varphi_T(y, \pi y) F(x) d\left(U \times \Lambda \times w_{0,y}^{T,y}\right)(\pi, y, x) \quad (4.22)$$

where U indicates the discrete uniform measure on the finite set Π_n . Furthermore, if we set $\mathcal{C}_{\Pi} = \{X \in \mathcal{C}_0^T : X(T) = \pi X(0) \text{ for some } \pi \in \Pi_n\}$ and define $w^{\star\Pi} = \Lambda \times \alpha(y)w_{0,y}^{T,\Pi_n y}$ where $\alpha(y) = \sum_{\pi \in \Pi_n} \varphi_T(y, \pi y)$,

$$\mathcal{A} = \int_{\mathcal{C}_{\Pi}} c\left(\frac{x(T)}{y}\right) F(x) dw^{\star\Pi}(y, x) \quad (4.23)$$

In general, collecting all $n!$ summands in (4.20) is inefficient, since they do not provide equally important contributions to the integral, and sampling within Π_n is desirable. To

sample according to a normalised $w_{0,y}^{T,\Pi_n y}$, ie to use a mixing measure where permutations receive Gaussian weightings, would also be inefficient.

As already discussed, the popular approach to the problem is to avoid having to estimate a denominator integral separately (cf (4.12)), which here means that we sample within $\Pi_n \times \mathcal{C}_O$ so that $|c(\pi)\varphi_T(y, \pi y)Y(x)| = \varphi_T(y, \pi y)Y(x)$ is the Radon-Nikodym derivative between the sampling measure and $U \times \Lambda \times w_{0,y}^{T,y}$. To look at this another way, we could say that we are sampling X within \mathcal{C}_Π so that $Y(x)$ is the Radon-Nikodym derivative between the sampling measure and $w^{\star\Pi}$. We then shall find that (4.13) again applies. Let us therefore define, now for indistinguishable particles, the **action measure** ν , defined on the Borel σ -algebra $\mathcal{B}(\mathcal{C}_\Pi)$ and given by

$$\nu(B) = \frac{\int_B Y(x) dw^{\star\Pi}(y, x)}{\int_{\mathcal{C}_\Pi} Y(x) dw^{\star\Pi}(y, x)}. \quad (4.24)$$

It is immediately evident from the definition of ν that for bosons,

$$\langle A \rangle = \int_{\mathcal{C}_\Pi} \frac{F(x)}{Y(x)} d\nu(x).$$

In particular, if $F(x) = A(x)Y(x)$ then

$$\langle A \rangle = \int_{\mathcal{C}_\Pi} A(x) d\nu(x). \quad (4.25)$$

In order to perform the sampling according to the action measure via the Multilevel Metropolis approach in the case of bosons, we proceed much as described in Subsection 4.2.2, except that during our MC step we allow the possibility that some particles are permuted between the first and last time-slices of the move. (In [Cep95] the suggested method involves including cyclic permutations in each candidate move.) To justify this procedure, intuitively it is sufficient to recognise that if we consider times 0 and T to be identified, then by symmetry we might as well allow the permutation to happen at any time-slice. (In fact, it is partly in this that the elegance of using the action measure lies: the initial system position, once identified with the terminal system position, is robbed of any special significance and is treated the same way to the position at any other time-slice.)

We can, more rigorously, regard the problem from the perspective of using the state space \mathbb{R}^{nd}/Π_n . Let us use \mathcal{C}^Π for the set of loops $\{X : [0, T] \rightarrow \mathbb{R}^{nd}/\Pi_n \text{ s.t. } X(0) = X(T)\}$. Then as long as F has permutational symmetry, we have for bosons (cf (2.68))

$$\mathcal{A} = \frac{1}{n!} \int_{\mathcal{C}^\Pi} F(x) d\left(\Lambda \times \alpha(y) w_{0, \Pi_n y}^{T, \Pi_n y}\right)(y, x) . \quad (4.26)$$

Imagine that we intended to sample (in \mathcal{C}^Π) according to $\Lambda \times w_{0, \Pi_n y}^{T, \Pi_n y}$, ie the projection of w^* on to \mathcal{C}^Π . Let us resample intermediate discretisation points in \mathbb{R}^{nd}/Π_n ; say that in projection, $X_{i-1} = \Pi_n x_1$ and $X_{i+2j} = \Pi_n x_2$. We have said that we can obtain $w_{t_{i-1}, \Pi_n x_1}^{t_{i+2j}, \Pi_n x_2}$ via $w_{t_{i-1}, x_1}^{t_{i+2j}, \Pi_n x_2}$ (see Subsection 2.1.4), so it follows that it is valid to generate points in \mathbb{R}^{nd} according to $\tilde{w}_{t_{i-1}, x_1}^{t_{i+2j}, \Pi_n x_2}$, and then consider the equivalent points in \mathbb{R}^{nd}/Π_n to have been sampled.

Again, to develop an intuitive understanding of this, consider what happens if we are trying to integrate with respect to w_{0, x_0}^{T, x_T} and we resample the section of the path over $(0, t)$, but according to $w_{0, x_0}^{t, \Pi_n x}$ where x is the $X(t)$ already obtained. It is a fact that the points $\Pi_n x$ have the same relative likelihood under $w_{0, x_0}^{T, \Pi_n x_T}$ as their relative probability under $w_{0, x_0}^{t, \Pi_n x}$; this is clear if one considers that the measures for $X(t)$ supplied by w_{0, x_0}^{T, x_T} and $w_{0, x_0}^{T, \gamma x_T}$ are conjugated by γ . Meanwhile $X_{(t, T]}$ has been sampled according to $w_{t, x}^{T, x_T}$ but in projection to paths with state space \mathbb{R}^{nd}/Π_n this is the same as $w_{t, \gamma x}^{T, x_T}$ for any $\gamma \in \Pi_n$. Thus the projection of the resampled path is being sampled according to $w_{0, \Pi_n x_0}^{T, \Pi_n x_T}$.

What this really means is that in effect we have a procedure, involving allowing permutations at intermediate times, for sampling in \mathcal{C}_Π according to w^* . But we may therefore employ the usual Metropolis approach, using this type of sampling in w^* for the candidate at each step, to create samples according to ν .

An alternative procedure is to also consider permutations between each step of the resampled section, as they are generated. This could be advantageous both from the perspective of consistency and from the perspective of being congenial to good progress through the full space of paths.

Permutation sampling for fermions Although our main discussion of fermions is in the next chapter, it is pertinent to note at this point why exactly the same sampling procedure is valid in the fermion case, if one wishes to attain the action measure. The difference is then that we are no longer integrating a functional which is measurable with respect to the projection of X into \mathcal{C}^Π . Rather, we integrate such a functional multiplied by the sign of the overall permutation (ie, that permutation which is obtained by travelling along the path X through a full circumference of time from any t back to itself). However, it is easily seen that where $\pi = \pi_1\pi_2$, $\text{sgn}(\pi) = \text{sgn}(\pi_1)\text{sgn}(\pi_2)$ and consequently if we multiply our functional by the signature of the permutation obtained every time that we perform a resampling, then the overall sign will be maintained equal to the signature of the overall permutation.

4.2.4 Numerical results still apply under importance sampling

It should be emphasised that to this point, there can be no interaction between the effectiveness of numerical and simulation methods. In Chapter 1, we assumed that we would try to find a weak approximation $\tilde{\mathcal{U}}$ to a functional integral \mathcal{U} , and then estimate the value of $\tilde{\mathcal{U}}$ via Monte Carlo; the sampling measure and other aspects of the Monte Carlo method cannot affect the relationship between \mathcal{U} and $\tilde{\mathcal{U}}$. In this Chapter, in Subsection 4.1.1, we have introduced the same idea for estimating $\langle A \rangle$; in particular, we identified a weak approximation to $\langle A \rangle$ (cf (4.5),(4.15)), and have also discussed how to best use Monte Carlo to provide information about this $\widetilde{\langle A \rangle}$. To show that the method defined in Subsection 3.2.2 is of second order when using the action measure, as it is for direct path sampling, amounts only to observing that the expectation of the Monte Carlo estimator of $\widetilde{\langle A \rangle}$ is the same in both cases. Consequently, we should write

$$\widetilde{\langle A \rangle} = \int_{\mathcal{C}_\Pi} \tilde{A}(x) d\tilde{\nu}(x).$$

To say that there is no interaction at all between the numerical method and the simulation method is to elide a subtle point: the efficiency of the Monte Carlo may of course be different for a different value of the time-step h , since every h gives rise to a different simulation. This

shall actually become important later on, but for bosons probably does not matter at all.

The piecewise constant method as "primitive action"

Interestingly, it was anticipated in [Cep95] that for many so-called 'static' quantities, the numerical method introduced in Subsection 3.2.2 should be second-order. In the physics literature, using this method to estimate the action functional (which then gives rise to sampling via the action measure) is referred to as using the *symmetrized primitive action*.

The author of [Cep95] writes:

We have symmetrized ... with respect to R_m and R_{m-1} [system positions at the end and start of the time interval] since one knows the exact density matrix is symmetric and thus the symmetrized form is more accurate.

(The reader should note that more generally, our $\bar{\nu}$ is not universally taken to be implied by the 'primitive action'; e.g. as in [Sim05] this term may refer to using the Euler method (3.11).) As mentioned, it was already known for the action functional itself [Suz91] that the method of Subsection 3.2.2 is second-order. Our result expands on this in a way which apparently vindicates the intuition held by physicists.

4.3 Numerical results for boltzmannons and bosons

In this section we present some experiment results intended to provide illustrations of various facts. Firstly we demonstrate the simulation of a correlation function, for a 1-dimensional system containing 1 particle, both for the harmonic potential and an quartic potential, using the piecewise constant method from Subsection 3.2.2 in conjunction with direct path sampling. We then compare this method with the Euler method in an example where the two are different, finding the kinetic energy of four bosons in a 1-dimensional system. The use of the action measure is then demonstrated, in an experiment which recreates that of [CP84]. We then return to considering the case of a 1-dimensional harmonic

oscillator system with one particle, to graph the simulated position density and to compare with using the numerical method given by applying Simpson's Rule to the action, S .

4.3.1 Experiment results: Correlation function

Let $d = 1, n = 1$ and consider the correlation function $\Gamma(\theta)$, $0 \leq \theta \leq T$ (see (2.72)):

$$\begin{aligned} \Gamma(\theta) &= \langle x(0)x(\theta) \rangle \\ &= \frac{1}{\mathcal{Z}(T)} \int_{-\infty}^{\infty} \int_{C_{0,y;T,y}} x(0) x(\theta) Y(x) dw_{0,y}^{T,y}(x) dy = \frac{\int_{-\infty}^{\infty} \mathcal{R}(y, y) dy}{\int_{-\infty}^{\infty} \mathcal{J}(y, y) dy}, \end{aligned} \quad (4.27)$$

where

$$\mathcal{R}(y, y) = \int_{C_{0,y}^{T,y}} x(0)x(\theta)Y(x)dw_{0,y}^{T,y}(x) \quad (4.28)$$

Using (4.8-4.9), (calling our numerator $\overline{\mathcal{R}}^{\text{MC}}$) we evaluate (4.27) for the harmonic potential

$$V(x) = \frac{\omega^2}{2}x^2 \quad (4.29)$$

and for the quartic potential

$$V(x) = \frac{\omega^2}{2}x^4. \quad (4.30)$$

In the case of the harmonic potential (4.29), the correlation function is equal to [Kle95, Chapter 3]:

$$\Gamma(\theta) = \frac{1}{2\omega} \frac{\cosh \omega(\theta - T/2)}{\sinh(\omega T/2)}, \quad 0 \leq \theta \leq T. \quad (4.31)$$

Recall (see Section 3.2.2) that the discretization of the time interval $[0, T]$ should be such that the point θ belongs to the set of discretization points $\{t_0, t_1, \dots, t_N\}$.

The results of the experiments are presented in Table 4.1 and in Fig. 4.1. In (4.8-4.9), μ is taken to be Gaussian and the parameters σ_1 and σ_2 are taken to be 1.2 and 0.8, respectively, in order to give low variance to $\overline{\mathcal{R}}^{\text{MC}}, \overline{\mathcal{Z}}^{\text{MC}}$. As above, in Table 4.1 the values before “ \pm ” are estimates of the bias, computed as the difference between the exact $\Gamma(1)$ and its sampled approximations, while the values after “ \pm ” give half of the size of the confidence interval for the corresponding estimator with probability 0.95. To compute the bias, the exact value $\Gamma(1) \doteq 0.1840098$ obtained from (4.31) was used. The number of Monte Carlo runs M is

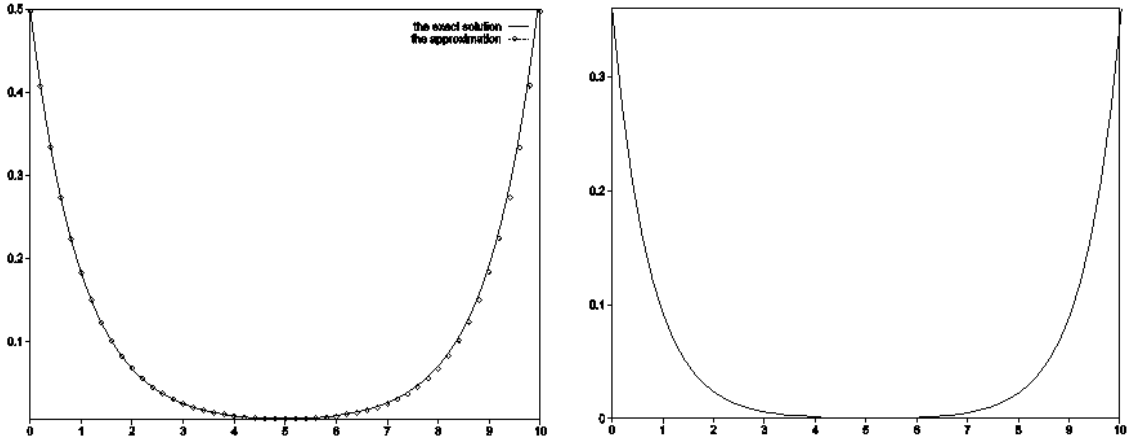


Figure 4.1: *Correlation function.* The dependence of the correlation function $\Gamma(\theta)$ from (4.27) on θ simulated with $h = 0.2$ and $M = 10^8$ for $T = 10$. The left figure corresponds to the harmonic potential (4.29) and the right figure – to the anharmonic potential (4.30), both with $\omega = 1$.

chosen here so that the Monte Carlo error is small in comparison with the bias. It is not difficult to see that the experiment illustrates second-order convergence of the method. We note that weighted least squares fitting of Ch^2 to the data of 4.1 yields $C \doteq 0.015$, with the maximum absolute value of the residuals being equal to 3×10^{-5} .

Table 4.1: *Correlation function.* The error in evaluating the correlation function $\Gamma(\theta)$ from (4.27) in the case of the harmonic potential (4.29) with $\omega = 1$, $T = 10$ and $\theta = 1$.

h	M	error
0.250	10^9	$9.78 \times 10^{-4} \pm 0.72 \times 10^{-4}$
0.200	10^9	$6.18 \times 10^{-4} \pm 0.72 \times 10^{-4}$
0.125	10^{10}	$2.45 \times 10^{-4} \pm 0.23 \times 10^{-4}$
0.100	5×10^{10}	$1.46 \times 10^{-4} \pm 0.10 \times 10^{-4}$

In Fig. 4.1 (left) the results of simulation of $\Gamma(\theta)$ with $h = 0.2$ are compared with the exact curve from (4.31). Thanks to the second-order of accuracy of the proposed numerical method, these curves visually coincide even for this relatively large time step. Figure 4.1 (right) demonstrates behaviour of the correlation function in the case of the quartic potential (4.30). The presented curve is obtained with the time step $h = 0.2$ and it visually coincides with the one simulated with $h = 0.05$. These experiments give further confirmation of our

theoretical results.

4.3.2 Experiment results: Kinetic energy of 4 1D bosons

We now consider a system of 4 bosons in a 1-dimensional space, ($n = 4, d = 1$) with mass $m = 1$ in the harmonic potential

$$V(x_1, \dots, x_n) = \frac{x_1^2}{2} + \dots + \frac{x_n^2}{2} \quad (4.32)$$

at inverse temperature $T = 1.2$. It is known (see, e.g. [TI84]) that in the case that $d = 1$, the kinetic energy is equal to

$$\langle K \rangle = \frac{1}{4} \sum_{l=1}^n l \coth \left(\frac{lT}{2} \right) - \frac{n(n-1)}{8}.$$

and therefore for these parameters, the exact value of the kinetic energy is $\langle K \rangle \doteq 1.3740081$. Again we use (4.8-4.9) in conjunction with the approximation scheme (3.7). We take μ to be Gaussian about zero, with $\sigma_1 = \sigma_2 = 2$.

Since the particles are noninteracting, we can decompose $\mathcal{U}(x, \pi x)$ and $\mathcal{I}(x, \pi x)$ to permanents as follows (see a similar idea in [TI84]). Let $U : \mathbb{R} \rightarrow \mathbb{R}$ be such that $V(x) = \sum_{i=1}^r U(x_i)$ and let

$$\begin{aligned} \mathcal{J}_1(x_i, x_j) &= (2\pi T/m)^{-1/2} \exp \left(-\frac{(x_i - x_j)^2}{2T/m} \right) \mathbb{E} \left[\exp \left(-\int_0^T U \left(X_{0,x_i}^{T,x_j}(t) \right) dt \right) \right. \\ &\quad \times \left. \left(\frac{1}{2m} - \frac{(x_i - x_j)^2}{2T} + \frac{1}{2m} \int_0^T U' \left(X_{0,x_i}^{T,x_j}(t) \right) \left(X_{0,x_i}^{T,x_j}(t) - \frac{x_i}{T}(T-t) - \frac{x_j}{T}t \right) dt \right) \right], \\ \mathcal{J}_2(x_i, x_j) &= (2\pi T/m)^{-1/2} \exp \left(-\frac{(x_i - x_j)^2}{2T/m} \right) \mathbb{E} \exp \left(-\int_0^T U \left(X_{0,x_i}^{T,x_j}(t) \right) dt \right). \end{aligned}$$

It is not difficult to show that

$$\mathcal{U}(y, \pi y) = \sum_{\pi \in \Pi_n} \sum_{l=1}^n \mathcal{J}_1(y_l, (\pi y)_l) \prod_{k \in \{1, \dots, n\} \setminus \{l\}} \mathcal{J}_2(y_k, (\pi y)_k), \quad \mathcal{I}(y, \pi y) = \sum_{\pi \in \Pi_n} \prod_{k=1}^n \mathcal{J}_2(y_k, (\pi y)_k). \quad (4.33)$$

and that consequently a similar statement holds for $\bar{\mathcal{U}}, \bar{\mathcal{I}}$ and indeed $\mathcal{U}^{\text{EULER}}, \mathcal{I}^{\text{EULER}} \dots$

We remark that although we illustrate the above decomposition into permanents in order to compute \mathcal{K} and \mathcal{Z} for the case of particles in a 1-dimensional space, its generalization for noninteracting particles in any real space is straightforward.

Table 4.2: *Kinetic energy of bosons.* The errors in evaluating the kinetic energy E_{kin} of the system of four bosons (2.76) in the case of the harmonic potential (4.32) with $T = 1.2$, $n = 4$, and $m = 1$. The number of Monte Carlo runs $M = 10^9$.

h	Euler method	Method (3.7), (3.6)
0.20	$0.236 \pm 0.55 \times 10^{-4}$	$0.533 \times 10^{-2} \pm 0.75 \times 10^{-4}$
0.15	$0.175 \pm 0.61 \times 10^{-4}$	$0.300 \times 10^{-2} \pm 0.75 \times 10^{-4}$
0.10	$0.116 \pm 0.66 \times 10^{-4}$	$0.128 \times 10^{-2} \pm 0.76 \times 10^{-4}$
0.05	$0.057 \pm 0.71 \times 10^{-4}$	$0.035 \times 10^{-2} \pm 0.75 \times 10^{-4}$

We analyze two methods: the method (3.7), (3.6) and the Euler method (3.11), (3.10). The results are presented in Table 4.3, which gives the errors of the two methods. As in the previous examples, the Monte Carlo error was made relatively small in order to be able to analyze the bias. It is clearly seen from the data that the method (3.7), (3.6) converges with order two while the Euler method exhibits the first order convergence as expected (see Theorems 3.2.1 and 3.2.3).

4.3.3 Experiment results: Potential energy of 64 boltzmannons

To illustrate that the result of Section (3.2) applies equally when performing an MCMC, we shall present an example involving the potential energy of a system of many distinguishable particles. The experiment here recreates that of [CP84]. We consider 64 particles interacting via the Lennard-Jones potential

$$V(r) = \frac{4\varepsilon}{r^{12}} - \frac{4\varepsilon}{r^6}, \quad (4.34)$$

where the value used for well depth ε was 10.22. In all experiments, $T = 0.195695$, so that $\varepsilon T = 2.0$, but the variance of Brownian bridges was inflated by a factor of $1.856 \approx 0.1816\varepsilon$ relative to the standard conditional Wiener measure. This corresponds to the choice of parameters in [CP84], that one unit of length represents 2.556 \AA and $\frac{\hbar^2}{2m} = 6.0596$.

The simulation uses a periodic boundary condition such that the particles inhabit a cube about 0 with side length $L = 5.6$. To generate candidate moves, first a candidate move is

generated in \mathbb{R}^{3N} , then by identifying the boundaries of the cube, this is mapped to a point in $[-L/2, L/2]^{3N}$. To evaluate the interaction potential between any pair of particles, we consider the shortest distance between them on the cube with identified boundaries, which means that V is set to zero for $r > L/2$.

The initial configuration was generated by taking a uniform grid of particles fixed in place for all $t \in [0, T]$, and running a burn-in of 10^5 attempted MCMC moves. The number of points used in each candidate move was between 2 and 12, with a lower number more likely. We sample uniformly a start position i for the set of time-slices in which the path for these points is to be resampled, and an index j is sampled uniformly between 1 and 6. Then the time-slices used are $\{i + 1, \dots, i + 2^j - 1\}$, with modular counting on $\{1, \dots, N\}$.

Variance estimation

A windowing method was used in order to gather sampling variances. The windows are non-overlapping but results are autocorrelated due to the fact that they are part of the same MCMC chain. We let $\widehat{\langle V \rangle}$ signify the estimate of $\overline{\langle V \rangle}$ obtained using the average of V samples over the Markov Chain, and let $\widehat{\langle V \rangle}_n$ represent the average value of V samples over the n th window. Due to autocorrelation, to obtain a credibility interval for $\overline{\langle V \rangle}$, the following formula, which follows from Eq. (5.6) of [Cep95], was used: where σ^2 denotes the true sampling variance of the average over windows of $\widehat{\langle V \rangle}$, σ_{0i} denotes the covariance of $\widehat{\langle V \rangle}_n$ with $\widehat{\langle V \rangle}_{n-i}$, and N_W is a sufficiently large number of windows,

$$\sigma^2 \approx \frac{1}{N_W} \left(\sigma_{00} + 2 \sum_{i=0}^{N_W-1} \sigma_{0i} \right) \quad (4.35)$$

The number of attempted MCMC moves per sample of V was 100 and the number of samples of V per window was 1000. The number of windows was 10000 for $M = 512$ and greater in the other experiments. Between windows, there was a burn-on of 10000 attempted MCMC moves with no sampling of V . A depth of 5 autocovariances was used in approximating (4.35), as autocovariance was observed to diminish over this interval and to be negligible thereafter.

Results of MCMC

The results of the experiments are given in Table 4.3. For M up to 256, the bias is here estimated by comparison with the reference value obtained with $M = 512$. The variance for the posterior distribution of the “bias” is then taken to be the sum of the relevant $\widehat{\langle V \rangle}$ variances, and the covariance between biases is then the variance of the $M = 512$ result. Equal-tails 95% credibility intervals for the ratios of bias were obtained by Monte Carlo with 10^8 BVN draws. The entry in the table for $M = 16$ represents the ratio of bias between $M = 16$ and $M = 32$, and so on; clearly it is only possible to report this up to the ratio between $M = 128$ and $M = 256$ so the remaining cells are left blank.

Table 4.3: Results of Markov Chain Monte Carlo to simulate 64 distinguishable particles interacting via Lennard-Jones potential and collect an estimate of potential energy.

M	$\widehat{\langle V \rangle}$	s.d.	“Bias”	s.d.	Ratio	Low(95%)	High(95%)
16	-1.9373086	9.531E-05	-7.918E-02	3.103E-04	2.690037	2.652768	2.728818
32	-1.8875645	9.673E-05	-2.943E-02	3.107E-04	3.167031	3.023195	3.33071
64	-1.8674245	1.039E-04	-9.294E-03	3.130E-04	3.899932	3.2325	5.0453
128	-1.8605138	1.427E-04	-2.383E-03	3.280E-04	2.928521	1.388	10.165
256	-1.8589445	2.292E-04	-8.137E-04	3.738E-04			
512	-1.8581307	2.953E-04					

The remarkable learning from these results is that for this simulation it takes a very small value of the time-step h for the quadratic decrease of the bias to become evident, even in this example which ostensibly should lend itself to a nice result. Convergence towards a quadratic decrease is observed here only as the number of time steps is increased past 128 and this simulation already has a very long running time (the last experiment here took over 39 days on a 2.0 GHz machine) in order for this quadratic decrease to be distinguished.

4.3.4 Experiment results: 1D harmonic oscillator with 1 particle

Theoretically it has been established that the method (3.7) has bias of order h^2 , and it has been pointed out that if we let $\tilde{\mathcal{I}}$ be found by using Simpson’s Rule (3.40) for $S(x)$,

rather than the Trapezoidal Rule (3.39), then this also has bias of order h^2 . However, there is reason to believe that the bias using Simpson's Rule should be smaller, and this is tested in the following example.

Analytical solution

Consider the one-dimensional harmonic oscillator with force constant 1, ie let 1 particle be subject to a potential $V : \mathbb{R} \rightarrow \mathbb{R}$ given by $V(x) = \frac{1}{2}x^2$. For convenience, let the mass of the particle be \hbar^2 . We recall the analytical solution for this system in order to make comparisons with simulation results. It is well-known (see [AF97]) that the eigenstates of the Hamiltonian are given by

$$E_m = m + \frac{1}{2}, \quad \psi_m(x) = (2^m m! \pi^{1/2})^{-1/2} H_m(x) e^{-\frac{1}{2}x^2}, \quad (4.36)$$

where H_m denotes the m th Hermite polynomial, given by $H_0(x) = 1, H_1(x) = 2x$ and the recurrence relation

$$H_{m+1} = 2xH_m - 2mH_{m-1}.$$

Therefore

$$\mathcal{J}(x, x) = \frac{1}{\varphi_T(x, x)} \sum_{m=0}^{\infty} \exp\left(-\left(m + \frac{1}{2}\right)T\right) \frac{H_m^2(x) \exp(-x^2)}{2^m m! \pi^{1/2}} \quad (4.37)$$

However, the solution to the functional integral can also be found analytically in closed form.

Considering the equation (1.9.7) of [BS02], where $W \sim w_{0,x_0}^T$ is a 1-dimensional Brownian motion, and where μ signifies the measure on $\mathcal{B}(\mathbb{R})$ induced by the law of $W(T)$,

$$\begin{aligned} \mathbb{E} \left[\exp\left(-\frac{\gamma^2}{2} \int_0^T W(s)^2 ds\right) \middle| W(T) = x_T \right] \frac{d\mu}{d\Lambda}(x_T) = \\ \left(\frac{\gamma}{2\pi \sinh(T\gamma)} \right)^{1/2} \exp\left(-\frac{(x_0^2 + x_T^2) \gamma \cosh(T\gamma) - 2\gamma x_0 x_T}{2 \sinh(T\gamma)}\right) \end{aligned}$$

and it therefore follows that

$$\begin{aligned} \mathbb{E} \left[\exp\left(-\frac{\gamma^2}{2} \int_0^T W(s)^2 ds\right) \middle| W(T) = x_T \right] = \\ (\gamma T \operatorname{csch}(\gamma T))^{1/2} \exp\left(\frac{(x_T - x_0)^2}{2T} - \frac{1}{2}\gamma(x_0^2 + x_T^2) \coth(\gamma T) + \gamma x_0 x_T \operatorname{csch}(\gamma T)\right). \quad (4.38) \end{aligned}$$

Clearly $\mathbb{E} \left[\exp\left(-\frac{1}{2} \int_0^T W(s)^2 ds\right) \middle| W(T) = x_T \right] = \mathcal{J}(x_0, x_T)$.

Results

To visually illustrate how well the simulations reproduce the actual system, we graph the position density $\rho(x)$. We take a partition of \mathbb{R} into intervals B_i and approximate $E1_{B_i}$ for each i , simultaneously and being careful to keep an eye on the covariance matrix. We use the approximation $P(B_i)/\Lambda(B_i) \approx \frac{dP}{d\Lambda}(x_i)$ for x_i the midpoint of B_i in order to provide comparison with the pdf $\rho(x)$. Here $\Lambda(B_i) = 0.04$. The results are displayed in Figures 4.2-4.4. The error bars in these figures show 95% credibility intervals for the simulation values. For these simulations, Simpson's Rule (cf (3.40)) was used to approximate S in order to apply (4.8)-(4.9). It can immediately be seen that there is a very good fit between the

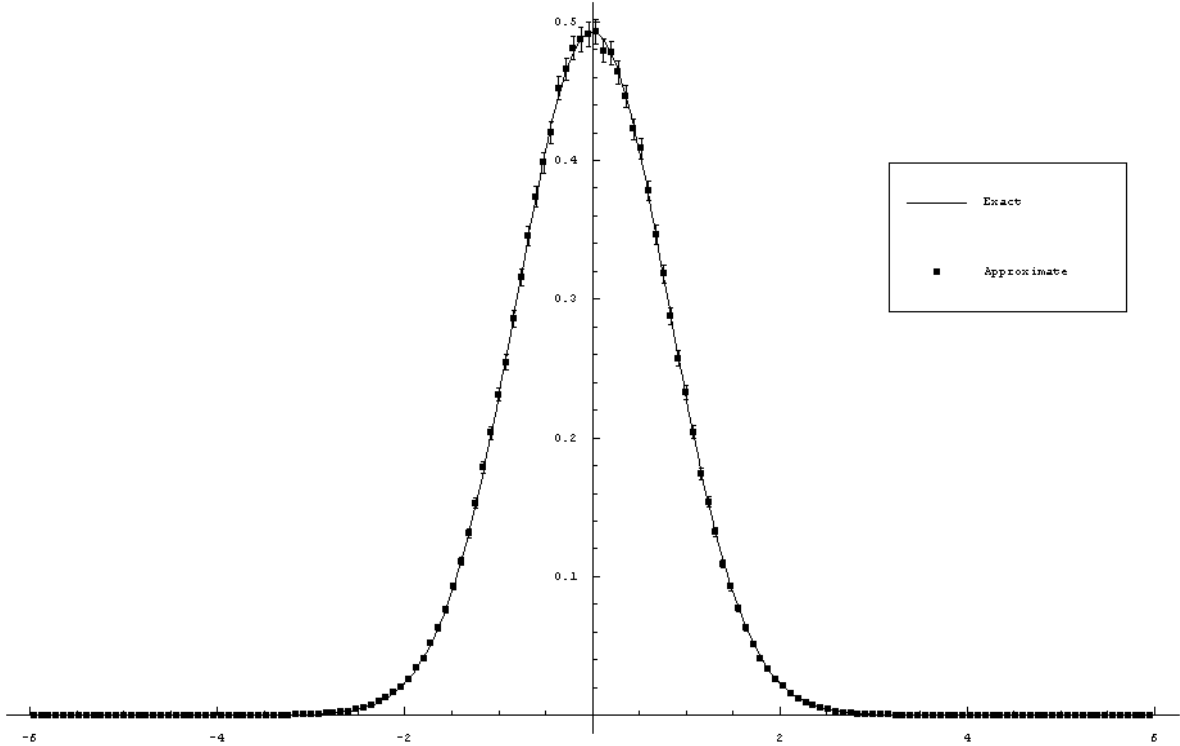


Figure 4.2: Simulated density, for 1 particle in H.O., $T = 2, h = 0.05, M = 10^6$, against solution

points generated by simulation and the true density, so we may have some confidence that our algorithm is converging to the correct density (as $M \rightarrow \infty$ and $h \rightarrow 0$) and that we may proceed to discuss it accordingly. We can see that our reported Monte Carlo error easily accounts for the difference between our experimental results and the true answer. It appears

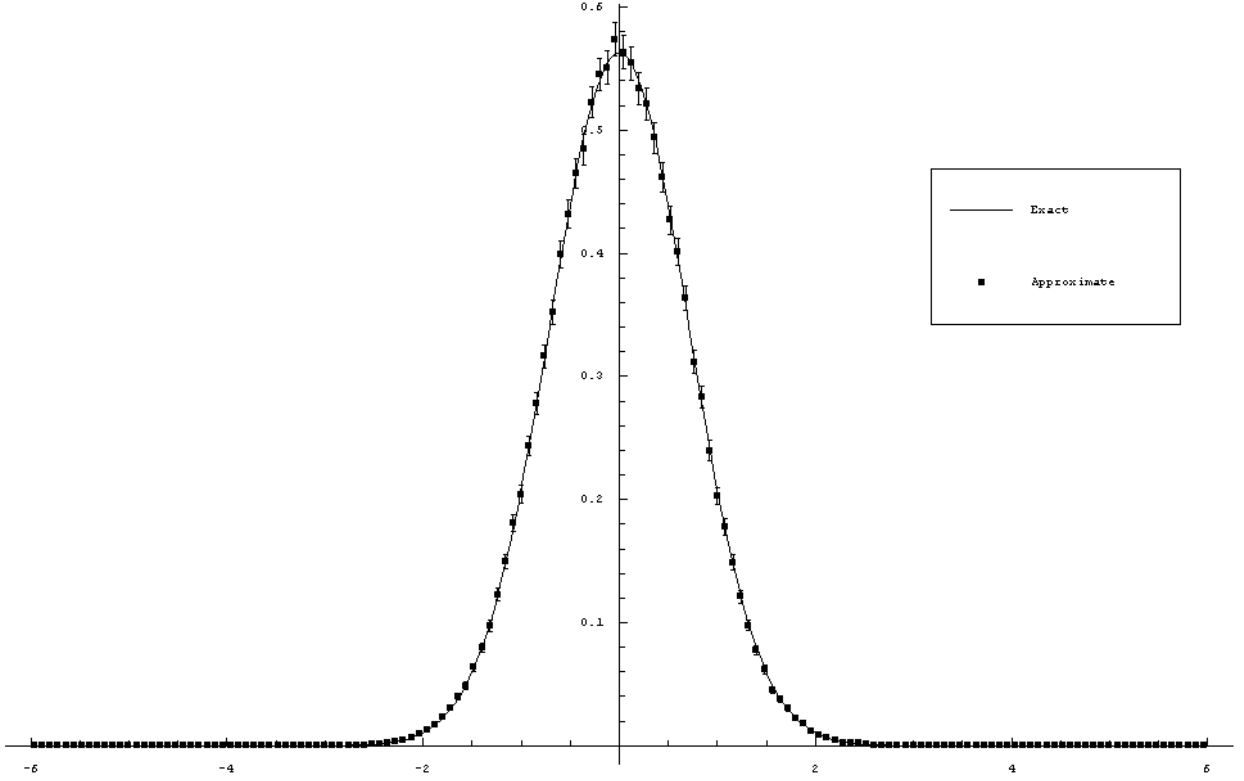


Figure 4.3: Simulated density for 1 particle in H.O. with $T = 6, h = 0.1, M = 10^6$, against solution

that in these examples bias is negligible by comparison with Monte Carlo error. It can be seen that at $T = 10$, we can still obtain a very close visual fit to the position density using 10^8 paths.

We let $\widetilde{\langle V \rangle}$ indicate the approximation to the average potential energy $\langle V \rangle$ given by using Simpson's Rule for S . Setting $F = V(x_0)Y(x)$, the value of $\widetilde{\langle V \rangle}$ was estimated using (4.8-4.9). We use $\widehat{\langle V \rangle}$ to denote $\frac{\widetilde{F}_{\text{MC}}}{\widetilde{Z}_{\text{MC}}}$, and realisations of this are tabulated in Table 4.4. The correct value is $\langle V \rangle \doteq 0.3282588$. A credibility interval for $\widetilde{\langle V \rangle}$ is provided using the Bivariate Gaussian contour method.

It can be seen that the variance of $\widehat{\langle V \rangle}$ diminishes approximately linearly as the number of trajectories is increased, which is consistent with (4.7). It can also be seen that the results exhibit the expected convergence, and that bias begins to surface at 10^9 trajectories.

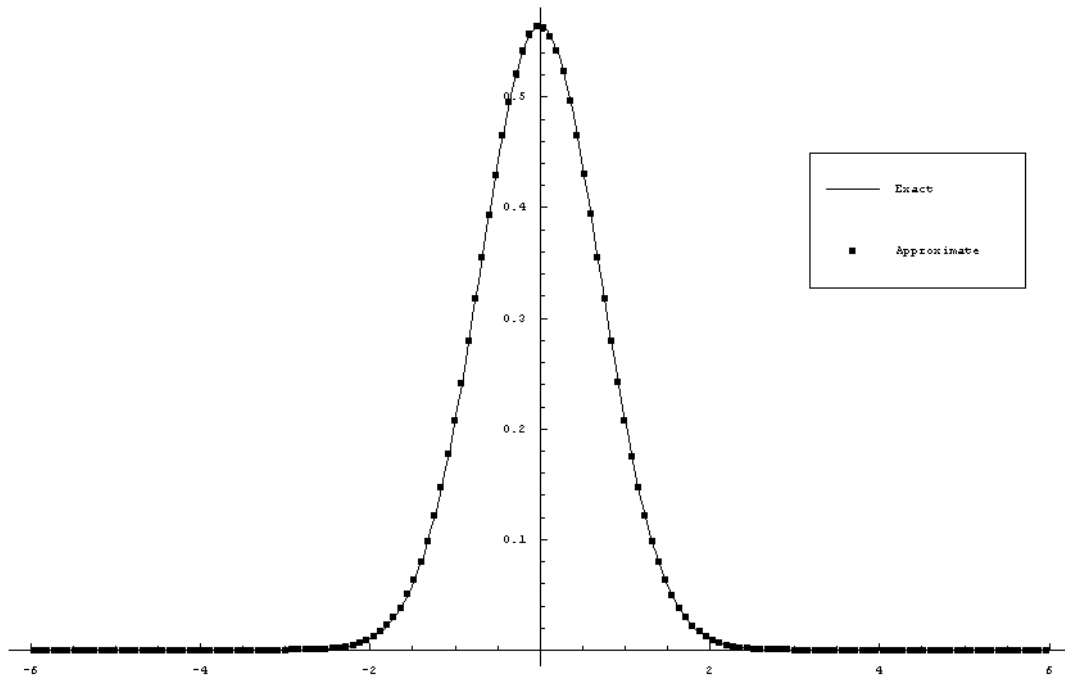


Figure 4.4: Simulated density for 1 particle in H.O., $T = 10, h = 0.2, M = 10^8$, against solution

Table 4.4: *Potential energy of 1 particle in a harmonic potential; to 9 d.p.* The results were created by using Simpson's Rule for S and direct path sampling. The reference value is 0.3282588.

Method	T	h	M	$\widehat{\langle V \rangle}$	Min $\widehat{\langle V \rangle}$ (95%)	Max $\widehat{\langle V \rangle}$ (95%)
Simpson's Rule	2	0.1	10^6	0.328675017	0.327607571	0.329744697
Simpson's Rule	2	0.1	10^7	0.328157911	0.327820619	0.328495426
Simpson's Rule	2	0.1	10^8	0.328336051	0.328229332	0.328442793
Simpson's Rule	2	0.1	10^9	0.328322286	0.328288539	0.328356036

Comparison of Simpson's Rule with Trapezoidal Rule

As mentioned, for integral-type functionals the piecewise constant numerical integration method described in Section 3.2 becomes the Trapezoidal Rule. Setting $F = V(x_0)Y(x)$, the average potential energy $\langle V \rangle$ was estimated using (4.8)-(4.9) in conjunction with, variously, the Trapezoidal Rule or Simpson's Rule for S , which from theory we believe to both have bias of order h^2 . Throughout this experiment, we took $T = 4$ and $M = 10^{10}$. The results are shown in Table 4.5. The correct value of the average potential energy is $\langle V \rangle \doteq 0.259328680$.

A credibility interval for $\widehat{\langle V \rangle}$, or $\overline{\langle V \rangle}$ in the case of the Trapezoidal Rule, is calculated using the Bivariate Gaussian contour method.

Table 4.5: *Potential energy for 1 particle in a harmonic potential; to 6 d.p.* The results show a comparison between using Simpson's Rule or the Trapezoidal Rule for S . The reference value is 0.259328680. For all results, $T = 4$ and $M = 10^{10}$.

Method	h	$\widehat{\langle V \rangle}$	$\widehat{\langle V \rangle}$ [or $\overline{\langle V \rangle}$]		$\widehat{\langle V \rangle} - \langle V \rangle$	Std. Diff [to 2 d.p.]
			Min (95%)	Max (95%)		
Simpson's Rule	0.1	0.259365	0.259354	0.259377	0.000036507	6.3
Simpson's Rule	0.2	0.259498	0.259486	0.259509	0.00016883	29.1
Simpson's Rule	0.4	0.259967	0.259955	0.259978	0.000637902	109.47
Trapezoid Rule	0.1	0.259026	0.259015	0.259038	0.000302438	52.32
Trapezoid Rule	0.2	0.258110	0.258099	0.258121	0.00121844	212.11
Trapezoid Rule	0.4	0.254532	0.254521	0.254543	0.00479634	855.55

The figures for "Std. Diff", which appear in the last column, are calculated by dividing $\langle V \rangle - Min$ by 1.96 to obtain a "standard error" and then dividing $\widehat{\langle V \rangle} - \langle V \rangle$ by this number. This gives a flavour of how certain we can be about the exact extent of bias (as opposed to Monte Carlo error) in contributing to $\widehat{\langle V \rangle} - \langle V \rangle$.

We can, loosely speaking, observe from the results given in Table 4.5 for $T = 4$ that as determined theoretically in Chapter 3, both methods are of order 2 in h . If we consider the posterior distribution for the ratio of bias between $h = 0.1$ and $h = 0.2$, it is clear that 4 is a very plausible value. The same thing is even more evident for the ratio of bias between $h = 0.2$ and $h = 0.4$. It is remarkable that the bias for Simpson's Rule is positive while that for the Trapezoidal Rule is negative, for this potential.

Let $\widehat{\langle V \rangle} = \langle V \rangle + \varepsilon_{0.1}$ indicate the limit of $\widehat{\langle V \rangle}$ with $h = 0.1$. It would be possible to get the joint posterior distribution of the ratios $\varepsilon_{0.4}/\varepsilon_{0.2}, \varepsilon_{0.2}/\varepsilon_{0.1}$ by performing a Monte Carlo, making draws from the posterior distribution of $\tilde{\mathcal{A}}, \tilde{\mathcal{Z}}$ for $h = 0.1, 0.2, 0.4$ simultaneously, recording the corresponding values of $\varepsilon_{0.4}/\varepsilon_{0.2}, \varepsilon_{0.2}/\varepsilon_{0.1}$. However, we can form a very rough ">95%" bound on the ratios of bias by simply taking the ratios of the extreme values of the bias corresponding to the intervals given in Table 4.5. This is shown in Table 4.6.

Table 4.6: *Credibility intervals for ratios between biases incurred using Simpson's Rule and Trapezoidal Rule; to 6 d.p.* Calculated simply by comparing extreme values from intervals for biases.

Method		Min (95%) ratio of bias	Max (95%) ratio of bias
Simpson's	from $h=0.2$ to $h=0.1$	3.289627	7.165374
Simpson's	from $h=0.4$ to $h=0.2$	3.476549	4.123792
Trapezoid	from $h=0.2$ to $h=0.1$	3.847378	4.224209
Trapezoid	from $h=0.4$ to $h=0.2$	3.891472	3.982261

Since these will be overestimates of the true intervals, it is clear that for the Trapezoidal Rule, convergence to a ratio of 4 has not taken place at $h = 0.4$, but does thereafter; for Simpson's Rule the most plausible value of the ratio is also 4. It is evident that both methods display results consistent with a bias of order h^2 , but that nonetheless, the gain from using Simpson's Rule rather than the Trapezoidal Rule to approximate $\int_0^T V(X(t))dt$ is considerable. In fact the impact, for this particular potential, is comparable to (but less than) the impact on Monte Carlo error of increasing the number of trajectories by a factor of 10. Considering results for different T , it can be seen that the bias is at the 4th place for $h = 0.2$, for all these times, and is < 0.0002 . So we may conclude that using $h = 0.1$ will yield a bias at the 5th place.

Chapter 5

Fermionic simulations and the fermion sign problem

5.1 Simulating systems of fermions

5.1.1 Functional integrals for fermion systems

Direct path sampling

Recalling (??), for fermions

$$\langle A \rangle = \frac{\mathcal{A}}{\mathcal{Z}} = \frac{\int_{\mathcal{C}_{\Pi}} \operatorname{sgn} \left(\frac{x(T)}{x(0)} \right) F(x) dw^{\star\Pi}(x)}{\int_{\mathcal{C}_{\Pi}} \operatorname{sgn} \left(\frac{x(T)}{x(0)} \right) Y(x) dw^{\star\Pi}(x)}.$$

We have a choice about how to incorporate the sign $\operatorname{sgn} \left(\frac{x(T)}{x(0)} \right)$: we can either suppose that we have the integral of a positive functional with respect to a signed measure, or the integral of a signed functional with respect to a positive measure. For the time being we take the latter point of view.

The alternative ways of writing the functional integrals (4.20)-(??) lead to several different ways of performing direct path sampling. The most naïve is that corresponding to (4.20) or (4.22); meanwhile, (4.21) corresponds to using linear translation to generate permuted paths, which is somewhat more beneficial here because of the negative covariance between

the countersigned summands thus created.

Using the action measure

Direct path sampling simulations of fermions are useful for simple experiments, but fail rapidly for the same reasons as direct path sampling simulations of bosons (cf [Cep95, Cep96]). Therefore the action measure is again often used. Recalling (4.10)-(4.12) and letting $\frac{d\tilde{\nu}}{d\tilde{w}^{\star\Pi}}(x)$ be proportional to $\tilde{Y}(x)$, which is the absolute value of the denominator functional, we see that where $X^i \sim \tilde{\nu}$, we must take

$$\widehat{\langle A \rangle} = \frac{\sum_{i=1}^M \left[\operatorname{sgn} \left(\frac{X^i(T)}{X^i(0)} \right) \frac{\tilde{F}(X^i)}{\tilde{Y}(X^i)} \right]}{\sum_{i=1}^M \left[\operatorname{sgn} \left(\frac{X^i(T)}{X^i(0)} \right) \right]}. \quad (5.1)$$

It is impossible to avoid incurring a denominator term when the particles are fermions, since we cannot sample according to a signed measure. The $\hbar \rightarrow 0$ limit (ie, the limit obtained with an indefinitely fine discretisation of time) of the expectation of the denominator in (5.1) is called the **average sign** and shall sometimes be designated $\langle \pm \rangle$:

$$\begin{aligned} \langle \pm \rangle &= \frac{\mathcal{Z}_{\text{FERMI}}}{\mathcal{Z}_{\text{BOSE}}} = \frac{\int_{\mathcal{C}_{\Pi}} \operatorname{sgn} \left(\frac{x(T)}{x(0)} \right) Y(x) dw^{\star\Pi}(x)}{\int_{\mathcal{C}_{\Pi}} Y(x) dw^{\star\Pi}(x)} \\ &= \frac{\int_{\mathbb{R}^{nd}} \sum \operatorname{sgn}(\pi) \varphi_T(y, \pi y) \mathcal{J}(y, \pi y) dy}{\int_{\mathbb{R}^{nd}} \sum \varphi_T(y, \pi y) \mathcal{J}(y, \pi y) dy} = \int_{\mathcal{C}_{\Pi}} \operatorname{sgn} \left(\frac{x(T)}{x(0)} \right) d\nu. \end{aligned}$$

While this is the average sign in a simulation which relies on using the action functional directly, in other simulations the average sign may be different, as we describe below.

5.1.2 Introduction to the fermion sign problem

In this subsection we shall endeavour to define the fermion sign problem and to superficially explain how it arises.

Definition 5.1.1 *An integral with (real or integer) parameter θ is said to have a sign problem (for θ) in the case that*

$$\lim_{\theta \rightarrow \infty} \frac{\int F_{\theta} d\mu_{\theta}}{\int |F_{\theta}| d\mu_{\theta}} = 0 \quad (5.2)$$

Remark 5.1.2 *It is obvious from this definition that for $F_\theta > 0$, no sign problem can occur.*

1D real example Let a family of real-valued random variables X_θ be defined on some (Ω, \mathcal{F}, P) and have pdf given by $f_\theta(x) = \varphi_1(\frac{1}{\theta}, x)$. Then $\int X_\theta dP$ has a sign problem for θ .

We shall casually say that $\int F_\theta d\mu$ "has a sign problem" in the case that θ is T .

The reason for the fermion sign problem

Recall from (2.62) that for fermions,

$$\left(\sum_{\pi \in \Pi_n} \varphi_T(x, \pi x) \right) \mathcal{I}_T(x, x) = \sum_{\pi \in \Pi_n} \text{sgn}(\pi) \varphi_T(x, \pi x) \mathcal{J}_T(x, \pi x). \quad (5.3)$$

Let us imagine first of all that we had to independently estimate the integrals $\mathcal{J}_T(x, \pi x)$. The variance of the estimate of the unnormalized density given by (5.3) must then be the same as if $\text{sgn}(\pi)$ were not present, but the presence of $\text{sgn}(\pi)$ means that the quantity being estimated is much smaller than otherwise. This explains how the problem arises in the case of direct path sampling based on (4.20). It shall be observed that the variance increases relative to the expectation, as T is increased - much more rapidly than in the boltzmannon or boson case.

The efficiency of a simulation

We shall now consider the variance of $\widehat{\langle A \rangle}$ given by a formula similar to (5.1) using a measure which makes the denominator equal to the observed average sign. We restrict our attention to observables that are diagonal (although by setting $\tilde{A} = \frac{\tilde{F}}{I}$ an extension is not difficult). However, we shall consider a more general case where we have some functional I (rather than, hitherto, Y) such that we take

$$\widehat{\langle A \rangle} = \frac{\frac{1}{M} \sum_{i=1}^M [C(X^i) \tilde{A}(X^i)]}{\frac{1}{M} \sum_{i=1}^M [C(X^i)]},$$

where A (and hence \tilde{A}) is assumed to be a positive functional and C is a functional which takes values in $\{-1, 1\}$, and sampling is with respect to ν with $\frac{d\nu}{dw^*} \propto \tilde{I}$ for a positive

functional I . Here we call $E_\nu C$ the average sign corresponding to I , noting that

$$E_\nu C = \int_{\mathcal{C}_\Pi} C(x) d\nu = \frac{\int_{\mathcal{C}_\Pi} C(x) \tilde{I}(x) dw^{\star\Pi}}{\int_{\mathcal{C}_\Pi} \tilde{I}(x) dw^{\star\Pi}}.$$

It is particularly of interest to discuss the relationship between the variance of this quantity and the equivalent that is found in the unsigned case, ie by assuming that $C = 1$ always. (In the case that $I = Y$ and $C(X^i) = c \left(\frac{X(T)}{x_T} \right)$ this means that we are comparing the variance obtained in simulating fermions using (5.1) with that obtained in simulating bosons the same way.) Recalling (4.7), discarding terms in $\frac{1}{M^2}$ yields that (where all moments are taken under ν and where we suspend the argument iX)

$$Var \left(\widehat{\langle A \rangle} \right) \approx \frac{1}{M (E[C])^2} \left(Var \left(C\tilde{A} \right) - 2 \frac{E[C\tilde{A}]}{E[C]} Cov \left(C\tilde{A}, C \right) + \frac{(E[C\tilde{A}])^2}{(E[C])^2} Var(C) \right).$$

But

$$Var \left(C\tilde{A} \right) = E \left[\tilde{A}^2 \right] - (E[C\tilde{A}])^2,$$

and since $C^2 = 1$ everywhere,

$$\begin{aligned} Cov \left(C\tilde{A}, C \right) &= E \left[C^2 \tilde{A} \right] - (E[C]) E[C\tilde{A}] \\ &= E \left[\tilde{A} \right] - E[C] E[C\tilde{A}], \end{aligned}$$

Then

$$\frac{E[C\tilde{A}]}{E[C]} Cov \left(C\tilde{A}, C \right) = E \left[\tilde{A} \right] \frac{E[C\tilde{A}]}{E[C]} - (E[C\tilde{A}])^2,$$

and

$$\frac{(E[C\tilde{A}])^2}{(E[C])^2} Var(C) = \frac{(E[C\tilde{A}])^2}{(E[C])^2} - (E[C\tilde{A}])^2.$$

Consequently,

$$\begin{aligned} Var \left(\widehat{\langle A \rangle} \right) &\approx \frac{1}{M (E[C])^2} \left((E[\tilde{A}])^2 - 2E[\tilde{A}] \frac{E[C\tilde{A}]}{E[C]} + \frac{(E[C\tilde{A}])^2}{(E[C])^2} \right) \\ &= \frac{1}{M (E[C])^2} \left(Var(\tilde{A}) + \left(E[\tilde{A}] - \frac{E[C\tilde{A}]}{E[C]} \right)^2 \right) \end{aligned} \quad (5.4)$$

In the case of the action-based simulation with $I = Y$, the term inside the squared parentheses in (5.4) is the difference between $\widetilde{\langle A \rangle}$ for bosons and $\widetilde{\langle A \rangle}$ for fermions; meanwhile the variance for bosons is $Var(\tilde{A})/M$. It is clear from (5.4) that if we now introduce, following [Cep96], the **efficiency** \mathcal{E} , defined to be the square of the average sign, then the relative cost of a simulation (to obtain the same variance) due to the presence of signs is governed by \mathcal{E} and is always more than \mathcal{E} times greater.

Thus, in the case of sampling according to the action measure and using (5.1), the simulation cost to achieve a particular Monte Carlo error threshold scales with \mathcal{E} , relative to the corresponding cost for bosons plus the square of the difference of the observable's expectations. Since $\langle \pm \rangle \rightarrow 0$ both as $T \rightarrow 0$, and as $n \rightarrow \infty$, this gives rise to the **fermion sign problem**: although the cost scaling for bosons is polynomial, the cost scaling for fermions is exponential in T and in n , if we use the methods introduced in Chapter 4. In the words of [Cep96] (p. 4), there is "an exponentially vanishing signal-to-noise ratio".

The fact that $\langle \pm \rangle \rightarrow 0$ is immediately evident in a simple case, such as a pair of particles starting from $(a, b) \in \mathbb{R}^{2d}$. The relative likelihood that they will arrive at (a, b) or (b, a) under the Wiener measure on paths is becoming close to 1.

$$\begin{aligned} \left(1 + \exp\left(-\frac{(a-b)^2}{T}\right)\right) \int_{\mathcal{C}_{0,(a,b)}^{T,\Pi_2(a,b)}} \text{sgn}\left(\frac{X(T)}{(a,b)}\right) Y(X) dw_{0,(a,b)}^{T,\Pi_2(a,b)} \\ = \mathcal{J}_T((a,b), (a,b)) - \exp\left(-\frac{(a-b)^2}{T}\right) \mathcal{J}_T((a,b), (b,a)) \end{aligned}$$

Meanwhile, intuitively it appears that for large T , $\mathcal{J}_T((a,b), (a,b))$ is close to $\mathcal{J}_T((a,b), (b,a))$ because paths which visit the same parts of space have a similar likelihood under the two measures. This very rough intuition shall be developed into something much more meaningful in Chapter 7. Meanwhile, in general there are $n!$ negative and positive terms in \mathcal{I}_T and as T increases these terms are all becoming closer together in value. This leads to an exponential decrease in the efficiency. Using (4.38), it is not difficult to find numerically that for two noninteracting fermions in a 1-dimensional harmonic oscillator, the efficiency is exactly $\mathcal{E} = \exp(-2T)$.

The sign problem is also discussed in [FH65, Cep92, LL70], amongst other works.

To summarise the conclusions from this and the previous chapter, then, for direct path sampling, the observed behaviour, even without a sign problem, will usually be an exponential increase in the cost to achieve a particular variance, as T is increased. However, using the action measure will, in the absence of a sign problem, allow a polynomial increase of cost [Cep95], whereas in the presence of a sign problem, the increase of cost is exponential. When a sign problem exists, the cost increase under direct path sampling is also exponential, but with a much higher exponent than in the unsigned case.

5.1.3 The fermion sign problem is insoluble

A very instructive contribution to the study of the fermion sign problem was made in [TW05], which essentially demonstrates that the problem is insoluble. It is shown that any exact algorithm (ie, one without uncontrolled approximations) has to scale exponentially with n , unless the so-called " $P=NP$ " condition (see e.g. [Sip92]) holds, and this is generally thought to not be the case. That the fermion sign problem should turn out to be insoluble, on a computer using classical logic, is redolent of Feynman's earlier scepticism about the ability of such computers to perform quantum simulations effectively - precisely for the reason that in quantum mechanics "negative probabilities" occur (see [Fey83, p.480]). In [Cep95ii], an heuristic suggestion is made for why the Troyer-Wiese theorem should hold:

Suppose we have a system which has a probability p of having a positive contribution and a probability $q = 1 - p$ of having a negative contribution. Now the efficiency, or signal-to-noise ratio is simply the integral divided by the total number of samplings: $(p-q)/(p+q)$. Now put N of these systems together. Using the binomial theorem, the signal-to-noise ratio is now: $(p-q)^N/(p+q)^N = e^{-cN}$ where $c = -\ln(1-2q) \approx 2q$. Thus no matter how small q may be, one gets exponential scaling.

Works such as [TW05, Cep96] frequently refer to a notional equivalence between increasing the number of particles and decreasing the temperature, because the relevant variable, where the sign problem is concerned, is the propensity for fermion exchanges to take place.

Consequently it is common to loosely define the sign problem as being the manifestation of exponential cost scaling with T and n . The result we are going to prove in Subsection 5.3.2 also suggests that for noninteracting particles there is some connection between being able to simulate for larger T and for larger n . But as we shall see, there are reasons to be careful about taking the equivalence at face value. While the case of large n has been addressed through a variety of approximate techniques in physics, an exact method is not available, and until the advent of quantum computers, probably never will be. On the other hand, it is possible to perform zero-temperature simulations of few-body systems of fermions without having to use a special strategy to avoid the sign problem, using other Quantum Monte Carlo methods. So in particular, it is not necessary to believe that solving the sign problem w.r.t. T , for fixed n , is impossible, just because it is impossible to solve it w.r.t. n with fixed T .

5.2 Extant practical approaches to fermion simulations

The very notoriety of the fermion sign problem seems to have stimulated a variety of research activity in the area over the last 20 years. In this section, we shall give some brief explanation of several particular popular approaches. There is not room to offer either in-depth analysis or an exhaustive catalogue of attempted solutions; the intention is to give some broad indications of the present state of research on this topic. With this aim in view, we now offer brief descriptions of two notable methods, Restricted Path Integral Monte Carlo and the Multilevel Blocking approach. Other well-known approaches, which, like these, would be best described as partially effective, include Gaussian Quantum Monte Carlo [CD04, Cor08] and the high-temperature "Direct" Path Integral Monte Carlo [ZNF77, She05]. A recent combinatoric approach based on using the Slater determinant formalism is treated in [AT05, AT06].

5.2.1 Restricted Path Integral Monte Carlo

As we shall see in Chapter 6, in the special case of $d = 1$, the fermion sign problem has a solution, which consists in restricting the simulation to the set of paths such that the particles do not ever collide (see Theorem 6.2.1). Using the Restricted Path Integral Monte Carlo (RPIMC) method, described in e.g. [Cep95ii, Cep92, Cep96] (see also [Cep91]), this solution appears as a special case; the general principle is that the nodes of the thermal density matrix (ie the surfaces where it is 0) are avoided. A justification for the method appears in works such as [Cep95ii, Cep96]. It rests upon the idea that, to quote from [Cep96],

The flux of positive paths at any spot on the nodal surface exactly cancels the flux of negative paths because the gradient of the density matrix is continuous across the node.

This is therefore apparently an exact method, if the nodes are known, in that restricting the simulated paths to not cross any of these nodal surfaces should give the correct estimate of \mathcal{Z}_T . However, since knowing the location of the nodes is tantamount to knowing the solution, it is necessary to use an approximation, or expert knowledge, in order to determine their probable location. This is the main problem for the RPIMC method: an imperfect choice of inputted wavefunction nodes results in an imperfect output. The bias incurred is usually described as uncontrolled below a certain temperature threshold, known as the Fermi temperature (see e.g. [Cep00]), but the sign problem is, at any rate, avoided. Wavefunction node estimates may be obtained approximately by using so-called Variational Monte Carlo, and this was used in [CM00, Mil00] to perform an RPIMC calculation for the Hydrogen hugoniot (shock wave). RPIMC has also been used to analyse nuclear fusion [PM04].

RPIMC is analogous to the fixed-node Diffusion Monte Carlo (DMC) method which applies in the zero-temperature case [And76, CA80, NKPTR02]; see also [CJL06]. As the 'fixed-node' moniker suggests, in the zero-temperature case a method exists which allows nodes to be adapted while more information about the solution is computed, and this is called the release-node method [CA84]. Is it inconceivable that a similar exact approach could exist in the case of finite-temperature PIMC? However, release-node DMC has costs

that scale exponentially with n (see e.g. [Tow06]); in view of [TW05] this is of course unsurprising.

A relatively recent theoretical contribution to RPIMC was made by [Ste99].

5.2.2 Multilevel Blocking

Antisymmetrizing on each time slice

Using (2.68), it is clear, from the comparison with using the state space \mathbb{R}^{nd}/Π_n , that we may consider particle exchange at intermediate times. Taking account of positive and negative contributions from this at each time-slice leads to a different measure on discretisations in which there is some amount of sign cancellation, compared with taking account of exchange only over the whole path, as we do if we use $w_{0,x}^{T,\Pi_n x}$.

This idea is well-known [TI84, NK92] and in the physics literature this is described as "antisymmetrizing" on every time slice. The benefits of doing only this were further investigated numerically in [Lyu05] and it is seen that on its own, this improvement does not make extensive headway against the sign problem. It is apparently incompatible with the Restricted Path Integral Monte Carlo (cf [Cep95ii]).

Multilevel blocking

The Multilevel Blocking approach is defined in [EMWG98], building on other work such as [Mak92, GS89]. It is an exact approach. It relies on a sampling strategy which creates 'blocks' of paths which are, apparently, in effect sampled together. This basic principle (but not the Multilevel Blocking approach itself) can be applied to directly give the exact solution in 1 dimension, as shall be seen in the following chapter.

The authors of [EMWG98] begin by stating their preference for 'antisymmetrizing on each time-slice'. The Multilevel Blocking technique is a way of taking the sign cancellations of that approach further. It involves constructing a sequence of levels (as with the Multilevel Metropolis method discussed in Section 4.2.2) so that the first consists of 1, 3, 5, 7.. ; the

second consists of $2, 6, 10, \dots$ and so on to L levels where $N = 2^L + 1$. Within each of these levels, working upwards, they create K samples of moves according to the antisymmetrized action measure. Each of these moves carries either a positive or negative sign, but by taking them all into account in forming the next level of moves, the average sign from the K samples is used. The claim is made in [EMWG98] that a complete solution to the sign problem is achieved by this expedient, although with caveats such as that there is a systematic error dependent on K .

In works such as [DM01], numerical experiments have been presented with the principal intention of demonstrating that the required K grows relatively slowly. It turns out that even for small systems, K must be taken to be more than about 200. It seems that the cost scaling of the algorithm relative to straightforward PIMC must be at least K , and the extra systematic error that is introduced apparently scales with $K^{-1/2}$ [DM01]. Given the Troyer-Wiese theorem, unless $P=NP$ then the conclusion that the necessary value of K explodes with n is unavoidable. The algorithm is said to work best at low temperatures [DM01].

The MLB approach was originally developed in connection with so-called real-time path integrals; that is, in the context of quantum dynamics. Although sadly beyond the scope of this thesis, this is an important application; see works such as [MH89] for an introductory treatment.

5.3 Two simple algorithms: linear translation and expansion

In this section, we introduce two more straightforward approaches which help to develop intuition about the sign problem and illustrate its resilience.

5.3.1 Linear translation

Let us recall from (2.62) that

$$\mathcal{I}_T(x_0, x_T) = \sum_{\pi \in \Pi_n} c(\pi) \frac{\varphi_T(x_0, \pi x_T)}{\sum_{\pi' \in \Pi_n} \varphi_T(x_0, \pi' x_T)} \mathcal{J}_T(x_0, \pi x_T).$$

If we think of the $\mathcal{J}_T(x_0, \pi x_T)$ as being the subject of separate simulations, then it is clear that to reduce the variance of our Monte Carlo, it is desirable to increase the covariance between their estimators. We have already noted that linear translation, as a bijection between $\mathcal{C}_{0,a_1}^{T,b_1}$ and $\mathcal{C}_{0,a_2}^{T,b_2}$, conjugates w_{0,a_1}^{T,b_1} with w_{0,a_2}^{T,b_2} , and that in terms of SDEs this simply corresponds to using the same realisation of $W(t)$ for both bridges. Therefore this gives rise to a method for generating covariances between estimators of the $\mathcal{J}_T(x_0, \pi x_T)$. Write

$$\sum_{\pi' \in \Pi_n} \varphi_T(x_0, \pi' x_T) \mathcal{I}_T(x_0, x_T) = \int_{\mathcal{C}_{0,0}^{T,0}} \sum_{\pi \in \Pi_n} c(\pi) \varphi_T(x_0, \pi x_T) \exp(-S(X + \ell_{x_0, \pi x_T})) dw_{0,0}^{T,0} \quad (5.5)$$

Of course, this different functional integral then gives rise to a different importance sampling which could be implemented via an MCMC. An alternative way of describing the situation (5.5) is that a set of $n!$ processes $X^{(i)}$ are specified by $X^{(i)}(0) = x_0$ and the SDEs

$$dX^{(i)} = \frac{\pi_i x_T - X^{(i)}}{T - t} dt + dW(t)$$

for the same nd -dimensional Brownian motion W , where π_i is a labelling of the members of Π_n .

Clearly the covariances between positive and negative contributions will grow when T increases. We shall call this the **linear translation** algorithm.

Favourable sampling

We can take the same line of thinking further. Noting that the difference between paths is deterministic, we may consider sampling paths according to a measure other than the integration measure in which paths are actually brought closer together, away from the endpoints, in a covariance-enhancing swoop. This requires us to invoke only the Radon-Nikodym theorem; we may regard it as applying in the space of paths, or indeed merely in

finite-dimensional real space since it is sufficient to think about methods for $\tilde{\mathcal{A}}$. Being forced to include the Radon-Nikodym derivative will increase the Monte Carlo variance, but the pain of this is related to the rate of change of the path difference. Therefore for large T it is clear that the optimum is to bring the paths close together for most of time, since the gain in covariance should eventually outweigh the extra cost.

This approach was not highly developed since this kind of thinking does not lead to a solution to the sign problem. But in some basic tests it was seen that a covariance swoop brought about a moderate improvement on linear translation.

5.3.2 Rearranging to use expanded paths, for noninteracting particles

It is possible to improve on the linear translation algorithm by altering the expression for \mathcal{I} so that we regard its summands (positive and negative) as arising from looping paths with the same endpoints but of different lengths. This potentially allows a greater amount of covariance to be achieved in their estimation. In this subsection we shall need to introduce some extra notation: $\mathcal{G}_{T,n}$ shall indicate \mathcal{G}_T for a system with n noninteracting fermions. We again let d be the dimension of the system space. A partial description of the position density of the system is given by the particle density $\varpi : \mathbb{R}^d \rightarrow \mathbb{R}$, the marginal density (with full measure n) obtained by integrating out the positions of all but one particle:

$$\begin{aligned}\varpi(x_1) &= \int_{\mathbb{R}^{(n-1)d}} \rho(x) dx_2 \cdots dx_n \\ &= \frac{\int_{\mathbb{R}^{(n-1)d}} \mathcal{G}_T(x, x) dx_2 \cdots dx_n}{\int_{\mathbb{R}^{nd}} \mathcal{G}_T(x', x') dx'}.\end{aligned}$$

The particle density is enough to tell us about statistics which are diagonal in the position representation and not dependent on the relative positions of particles, such as the potential energy.

Theorem 5.3.1 (Expanding Paths Theorem)

$$\int_{\mathbb{R}^{(n-1)d}} \mathcal{G}_T(x, x) dx_2 \cdots dx_n = \sum_{k=1}^n c(k) \frac{(n-1)!}{(n-k)!} (2\pi kT)^{-d/2} \mathcal{J}_{kT}(x_1, x_1) \int_{\mathbb{R}^{(n-k)d}} \mathcal{G}_{T, n-k}(x, x) dx \quad (5.6)$$

where $c(k) = 1$ for k odd, -1 for k even and where we set $\int_{\mathbb{R}^{nd}} \mathcal{G}_{T,0}(x, x) dx = 1$.

Proof. Considering the finite-dimensional distributions of the Brownian bridge, we may break down a looping path of length nT into n segments. Where we let x_0 and x_n denote the same point:

$$\mathcal{J}_{nT}(x_1, x_1) = \int_{\mathbb{R}^{(n-1)d}} \frac{(2\pi nT)^{d/2}}{(2\pi T)^{nd/2}} \exp \left(-\frac{1}{2T} \sum_{j=1}^n \|x_j - x_{j-1}\|^2 \right) \prod_{j=1}^n \mathcal{J}_T(x_{j-1}, x_j) dx_2 \cdots dx_n. \quad (5.7)$$

Moreover, where π is any cyclic permutation of the n particle coordinate vectors, since none of the variables with respect to which we integrate are distinguished from one another, we may write

$$\begin{aligned} \mathcal{J}_{nT}(x_1, x_1) &= \int_{\mathbb{R}^{(n-1)d}} \frac{(2\pi nT)^{d/2}}{(2\pi T)^{nd/2}} \exp \left(-\frac{1}{2T} \sum_{j=1}^n \|x_j - x_{\pi(j)}\|^2 \right) \prod_{j=1}^n \mathcal{J}_T(x_j, x_{\pi(j)}) dx_2 \cdots dx_n \\ &= \int_{\mathbb{R}^{(n-1)d}} \frac{(2\pi nT)^{d/2}}{(2\pi T)^{nd/2}} \exp \left(-\frac{\|x - \pi x\|^2}{2T} \right) \mathcal{J}_T(x, \pi x) dx_2 \cdots dx_n. \end{aligned} \quad (5.8)$$

Categorise the elements of Π_n according to the length of the cycle in which is found the distinguished coordinate, viz x_1 . Also name C_k the cyclic permutation group of order k .

Then

$$\begin{aligned} \int_{\mathbb{R}^{(n-1)d}} \mathcal{G}_{T,n}(x, x) dx_2 \cdots dx_n &= \int_{\mathbb{R}^{(n-1)d}} \sum_{\pi \in \Pi_n} c(\pi) (2\pi T)^{-nd/2} \exp \left(-\frac{\|x - \pi x\|^2}{2T} \right) \mathcal{J}_T(x, \pi x) dx_2 \cdots dx_n \\ &= \int_{\mathbb{R}^{(n-1)d}} \sum_{k=1}^n \sum_{\phi \in \Phi_k} \left(\sum_{\pi \in C_k} c(\pi) (2\pi T)^{-kd/2} \exp \left(-\frac{\|x' - \pi x'\|^2}{2T} \right) \mathcal{J}_T(x', \pi x') \right) \times \\ &\quad \left(\sum_{\pi \in \Pi_{n-k}} c(\pi) (2\pi T)^{-(n-k)d/2} \exp \left(-\frac{\|x'' - \pi x''\|^2}{2T} \right) \mathcal{J}_T(x'', \pi x'') \right) dx_2 \cdots dx_n, \end{aligned}$$

where ϕ signifies a relabelling of the n particles *excluding* x_1 , and Φ_k signifies a particular set of such relabellings, so constituted that each possible partition into the first $k-1$ and the

succeeding $n - k - 1$ variables is represented exactly once. Thus Φ_k has $\binom{n-1}{k-1}$ elements and represents the ways of choosing the elements in the same cycle as x_1 . Meanwhile ϕ intentionally does not appear in (5.9) except as a counter, but we let x' signify (x_1, \dots, x_k) and let x'' signify (x_{k+1}, \dots, x_n) . When $k = n$ the righthand bracket in (5.9) is construed to be 1. Then, reusing (5.8),

$$\begin{aligned}
& \int_{\mathbb{R}^{(n-1)d}} \mathcal{G}_{T,n}(x, x) dx_2 \cdots dx_n = \\
& \sum_{k=1}^n \left[\int_{\mathbb{R}^{(k-1)d}} \sum_{\phi \in \Phi_k} \left(\sum_{\pi \in C_k} c(\pi) (2\pi T)^{-kd/2} \exp\left(-\frac{\|x' - \pi x'\|^2}{2T}\right) \mathcal{J}_T(x', \pi x') \right) dx_2 \cdots dx_k \right. \\
& \quad \left. \times \int_{\mathbb{R}^{(n-k)d}} \mathcal{G}_{T,n-k}(x, x) dx \right] \\
& = \sum_{k=1}^n c(k) \sum_{\phi \in \Phi_k} \sum_{\pi \in C_k} (2\pi k T)^{-d/2} \mathcal{J}_{kT}(x_1, x_1) \int_{\mathbb{R}^{(n-k)d}} \mathcal{G}_{T,n-k}(x, x) dx \\
& = \sum_{k=1}^n c(k) |\Phi_k| |C_k| (2\pi k T)^{-d/2} \mathcal{J}_{kT}(x_1, x_1) \int_{\mathbb{R}^{(n-k)d}} \mathcal{G}_{T,n-k}(x, x) dx \quad . \quad (5.9)
\end{aligned}$$

But then since

$$|\Phi_k| |C_k| = \binom{n-1}{k-1} (k-1)! = \frac{(n-1)!}{(n-k)!}, \quad (5.10)$$

the result follows. ■

It is notable that in view of the results proven in Section 2.3, Theorem 5.3.1 gives us, in particular, an alternative method of collecting statistics about the n th quantum state of a single particle with a background potential.

The case of a noninteracting pair of fermions

It follows from (5.6) that the particle density for two noninteracting fermions is given by

$$\varpi(x) = \frac{\mathcal{J}_T(x, x) \int_{\mathbb{R}^d} \mathcal{J}_T(y, y) dy - \sqrt{\pi T} \mathcal{J}_{2T}(x, x)}{\frac{1}{2} \int_{\mathbb{R}^d} \left[\mathcal{J}_T(x', x') \int_{\mathbb{R}^d} \mathcal{J}_T(y, y) dy - \sqrt{\pi T} \mathcal{J}_{2T}(x', x') \right] dx'} \quad . \quad (5.11)$$

This means that the expectation of a diagonal statistic A is found via

$$\langle A \rangle = \int_{\mathbb{R}^d} A(x) \varpi(x) dx = \frac{\int_{\mathbb{R}^d} \mathcal{J}_T(y, y) dy \int_{\mathbb{R}^d} A(x) \mathcal{J}_T(x) dx - \sqrt{\pi T} \int_{\mathbb{R}^d} A(x) \mathcal{J}_{2T}(x) dx}{\frac{1}{2} \left(\left(\int_{\mathbb{R}^d} \mathcal{J}_T(x, x) dx \right)^2 - \sqrt{\pi T} \int_{\mathbb{R}^d} \mathcal{J}_{2T}(x, x) dx \right)} \quad . \quad (5.12)$$

Therefore our simulation procedure has to recover estimates of 4 quantities (and an estimate to use for the covariance matrix of the estimators); these are $\mathcal{Z}_{T,1}, \mathcal{A}_{T,1}, \mathcal{Z}_{2T,1}, \mathcal{A}_{2T,1}$, where

$$\begin{aligned}\mathcal{Z}_{T,1} &= \int_{\mathbb{R}^d} \mathcal{J}_T(x, x) dx \\ \mathcal{A}_{T,1} &= \int_{\mathbb{R}^d} A(x) \mathcal{J}_T(x, x) dx .\end{aligned}$$

The most naïve way to calculate first state statistics via (5.12) is that we separately collect the statistics for time (ie, inverse temperature) T and for time $2T$. However, clearly the variance of the posterior for $\langle A \rangle$ can be reduced, if we can sample so as to give a positive covariance between the estimators for time T and those for time $2T$. Therefore instead we proceed by first choosing an initial point x_0 , sampling a discretised path $X_1 \sim \tilde{w}_{0,x_0}^{T,x_0}$, and creating another path X_2 via a linear mapping; viz, for all $t \in [0, 2T]$, we set

$$X_2(t) - x_0 = \sqrt{2} \left(X_1 \left(\frac{t}{2} \right) - x_0 \right) .$$

This linear bijection conjugates w_{0,x_0}^{T,x_0} with w_{0,x_0}^{2T,x_0} and therefore conjugates $\tilde{w}_{0,x_0}^{T,x_0}$ with $\tilde{w}_{0,x_0}^{2T,x_0}$.

Covariance correction

In the following there is some advantage to recognising that it is the Bayesian posterior distribution of the true statistics that we would like to apprehend. Owing to the fact that we are interested in the expectation of a nonlinear function of the quantities which we can approach via Monte Carlo, we need to introduce a small correction for covariance. Suppose that we know that the approximate quantities $\widetilde{\mathcal{Z}}_{T,1}, \widetilde{\mathcal{A}}_{T,1}, \widetilde{\mathcal{Z}}_{2T,1}, \widetilde{\mathcal{A}}_{2T,1}$ have a posterior distribution that is multivariate Gaussian about our Monte Carlo estimators $\widetilde{\mathcal{Z}}_{T,1}^{\text{MC}}, \widetilde{\mathcal{A}}_{T,1}^{\text{MC}}, \widetilde{\mathcal{Z}}_{2T,1}^{\text{MC}}, \widetilde{\mathcal{A}}_{2T,1}^{\text{MC}}$. Then

$$\begin{aligned}\mathbb{E} \left[\widetilde{\mathcal{Z}}_{T,1} \widetilde{\mathcal{A}}_{T,1} - \sqrt{\pi T} \widetilde{\mathcal{A}}_{2T,1} \right] &= \mathbb{E} \widetilde{\mathcal{Z}}_{T,1} \mathbb{E} \widetilde{\mathcal{A}}_{T,1} + \text{Cov} \left(\widetilde{\mathcal{Z}}_{T,1}, \widetilde{\mathcal{A}}_{T,1} \right) - \sqrt{\pi T} \mathbb{E} \widetilde{\mathcal{A}}_{2T,1} \\ &= \widetilde{\mathcal{Z}}_{T,1}^{\text{MC}} \widetilde{\mathcal{A}}_{T,1}^{\text{MC}} + \text{Cov} \left(\widetilde{\mathcal{Z}}_{T,1}, \widetilde{\mathcal{A}}_{T,1} \right) - \sqrt{\pi T} \widetilde{\mathcal{A}}_{2T,1}^{\text{MC}}\end{aligned}$$

and likewise

$$\left[\widetilde{\mathcal{Z}}_{T,1} \widetilde{\mathcal{Z}}_{T,1} - \sqrt{\pi T} \widetilde{\mathcal{Z}}_{2T,1} \right] = \widetilde{\mathcal{Z}}_{T,1}^{\text{MC}} \widetilde{\mathcal{Z}}_{T,1}^{\text{MC}} + \text{Var} \left(\widetilde{\mathcal{Z}}_{T,1} \right) - \sqrt{\pi T} \widetilde{\mathcal{Z}}_{2T,1}^{\text{MC}} .$$

5.4 Numerical results from linear translation and expanding paths algorithms, for 2 1D fermions

We concern ourselves, in the following, with the particle density and potential energy of a system with 2 noninteracting fermions in a 1-dimensional space. From the result in Section 2.3 it follows that for large T , simulating this system gives a route to approximating the first excited state of a single fermion with a background potential.

We shall begin by providing some graphs of approximations to this 2-fermion particle density so as to provide a visual demonstration of the phenomena that are observed. It

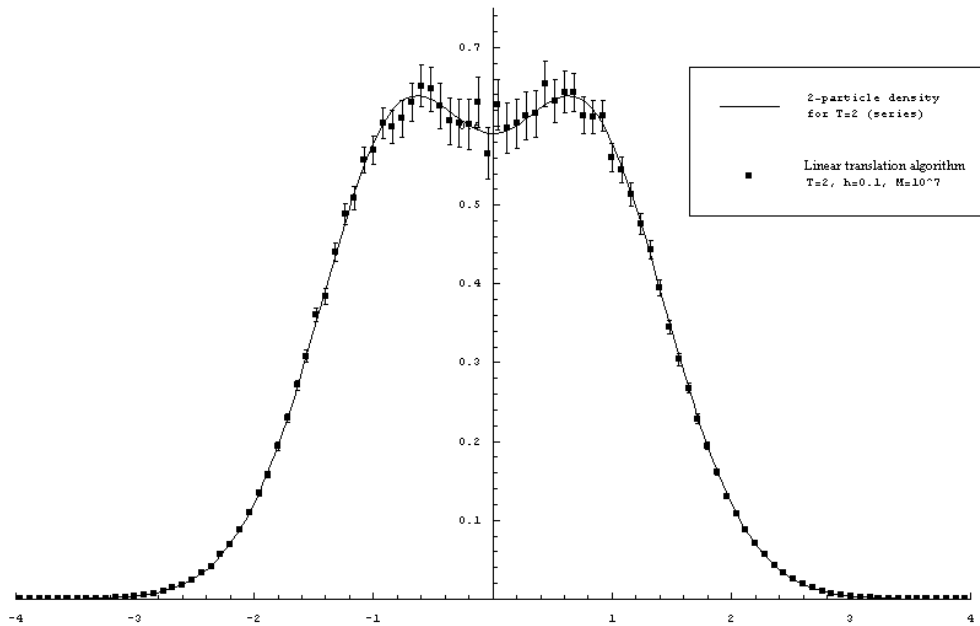


Figure 5.1: Particle density using linear translation algorithm, $T = 2, h = 0.1, M = 10^7$.

is immediately obvious that the fit is poorer for the 2-particle simulation using the linear translation algorithm than for the 1-particle simulation. Indeed, at 10^7 sampled paths (see Figure 5.1), the errors are greater than those for the 1-particle simulation with 10^6 . However it can be observed that the simulation does produce sensible results. It can immediately be seen that the expanding algorithm is more efficient than the linear translation algorithm, in that the graph in Figure 5.2 is produced with only 10^6 paths. From Figure 5.3, it can be seen that

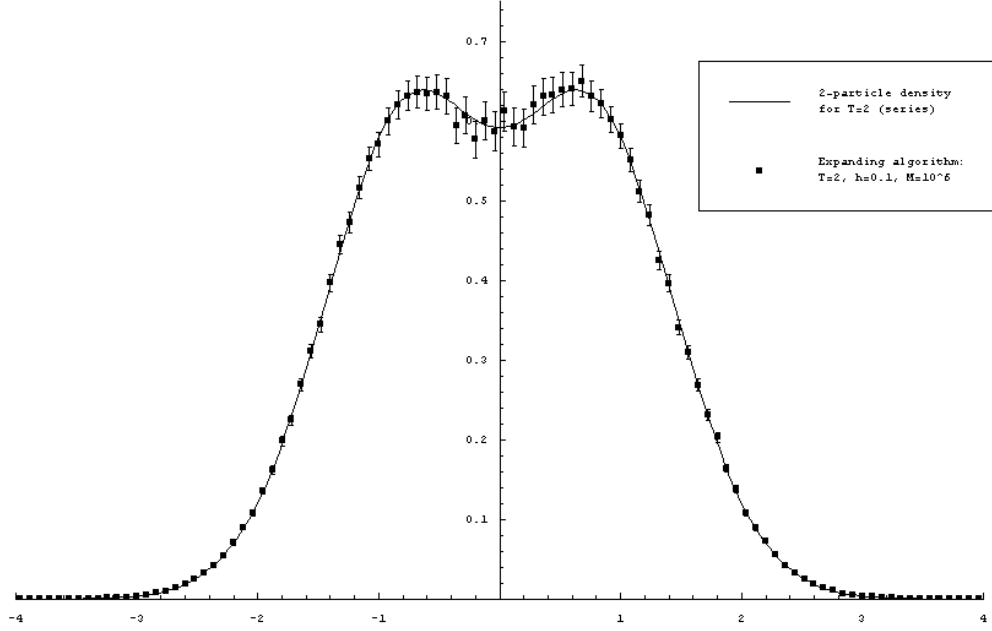


Figure 5.2: Particle density using expanding algorithm, $T = 2, h = 0.1, M = 10^6$.

a visually good fit is obtained by using 10^7 paths in the expanding algorithm. In fact here the Monte Carlo converges to a narrow range and an apparently upward bias is marginally visible. Now if we increase T to 4 then immediately the variance is very much greater, to the point where in Figure 5.4, a fairly poor but recognisable curve is obtained only at 10^8 sampled paths. Just based on visual comparisons, it seems that 10^9 paths, rather than just 10^7 paths, are needed for a good fit at $T = 4$, rather than $T = 2$ (see Figure 5.5). This already takes about 36 hours on a 2.0 GHz machine. One would surmise from the foregoing visual results that while it can be seen that the algorithm works, it is becoming increasingly difficult for the simulations to converge as time increases. For $T=6$, in Figure 5.6 it can be seen that for the linear translation algorithm, even at 10^{10} paths the fit is not particularly good; at a stretch we could claim that the curve is recognisable as being similar.

The expanding algorithm performs better (Figure 5.7), although it is already apparent that this is considerably worse than the fit at 10^9 samples for $T = 4$. The rapid increase in variance is due to the fermion sign problem. In both cases that we increased T by 2, we have had to increase M by 2 orders of magnitude to achieve similar results. Simulations for

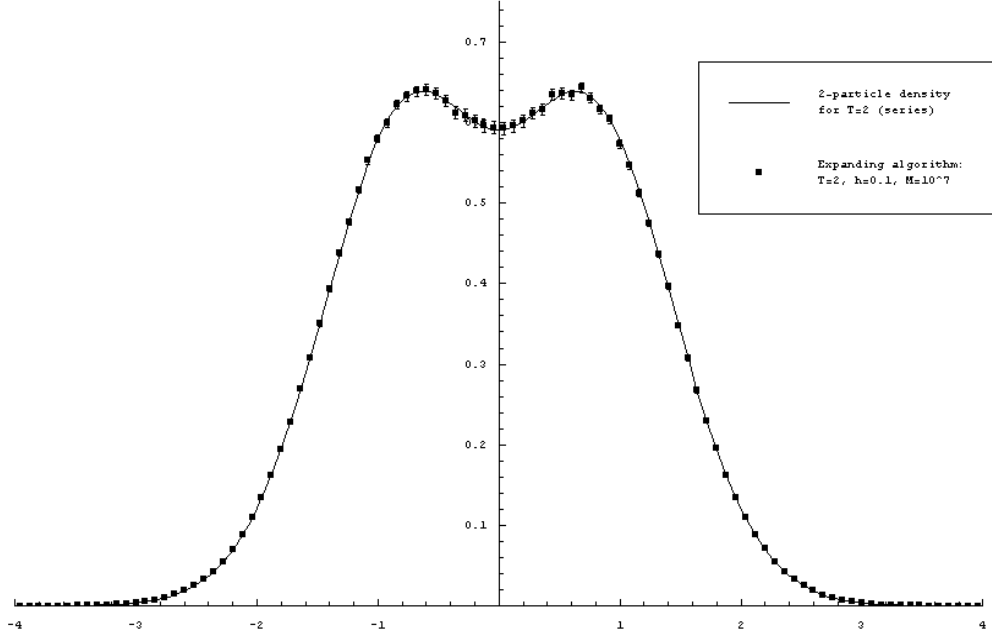


Figure 5.3: Particle density using expanding algorithm, $T = 2, h = 0.1, M = 10^7$.

$T = 8$ and $T = 10$ were also performed and were seen to consistently require an exponentially greater value of M .

In Table 5.1, we provide the results from estimating $\langle V \rangle$ for this system. It can again be seen that for the expanding paths algorithm, Monte Carlo variance is lower, but still grows very rapidly with T . It should be noted that even though $\widetilde{\langle V \rangle}$ is nonlinearly related to the quantities being sampled, its variance still eventually scales inversely with the number of trajectories. That is to say, if we calculate a "standard error" ("S.E.") by taking one side of the credibility interval for $\widetilde{\langle V \rangle}$ and dividing by 1.96 then the ratio between these quantities for $M = 10^9$ and $M = 10^{10}$ is close to $10^{1/2}$. (The calculation is not performed for the $T = 8$ results since it would clearly be meaningless.)

In order to achieve a similar bias, a timestep of half the size must be used with the Expanding algorithm, and taking account of this doubling in cost, we are able to find from Table 5.1 that for the same cost, the reduction in Monte Carlo variance is by a factor of about 2.5.

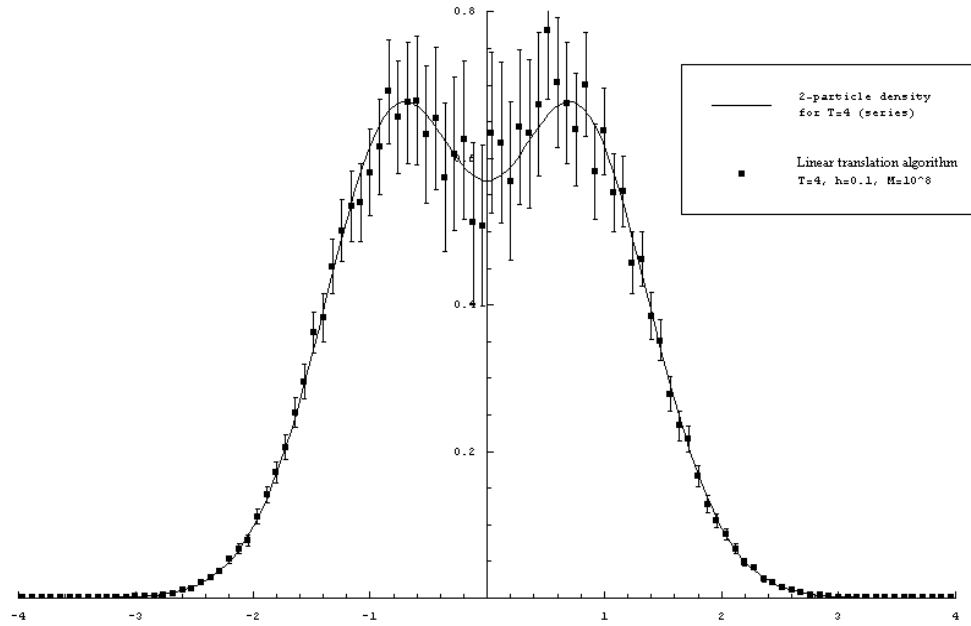


Figure 5.4: Particle density using linear translation algorithm, $T = 4$, $h = 0.1$, $M = 10^8$.

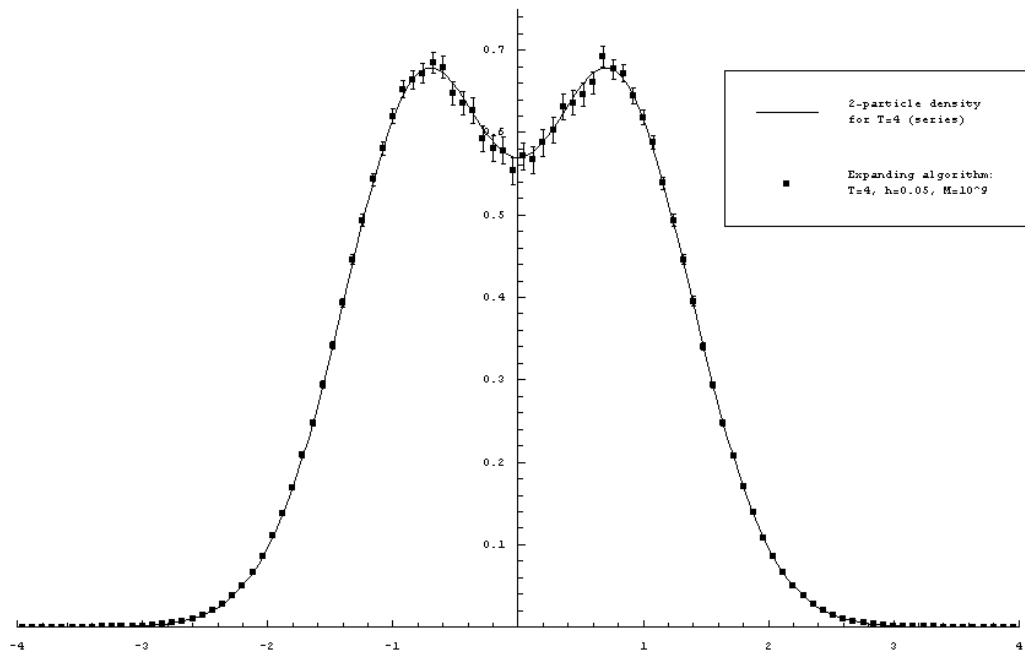


Figure 5.5: Particle density using expanding algorithm, $T = 4$, $h = 0.05$, $M = 10^9$.

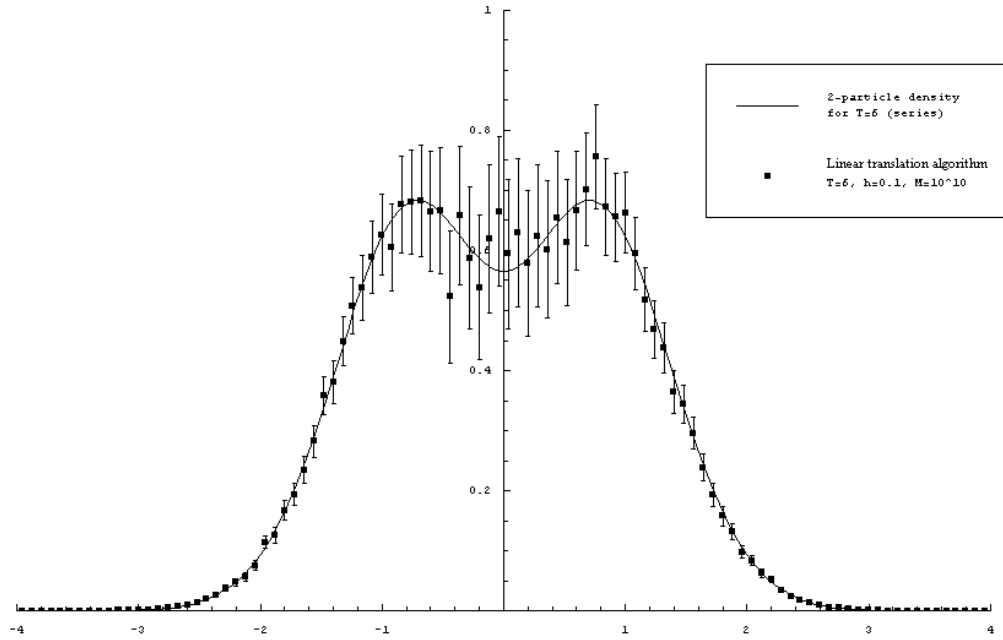


Figure 5.6: Particle density using linear translation algorithm, $T = 6$, $h = 0.1$, $M = 10^{10}$.

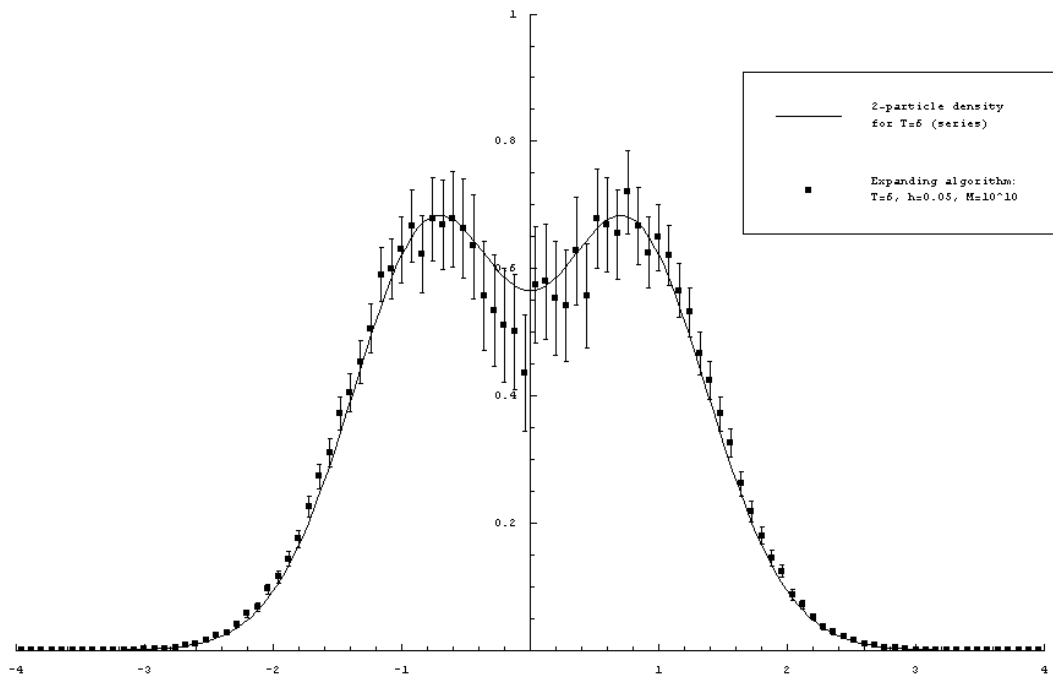


Figure 5.7: Particle density from expanding algorithm, $T = 6$, $h = 0.05$, $M = 10^{10}$.

Table 5.1: *Potential energy of 2 noninteracting 1D fermions in a harmonic oscillator.* Comparing results using linear translation and expanding paths algorithms.

Algorithm	M	h	T	$\widehat{\langle V \rangle}$	Low $\widetilde{\langle V \rangle}$ (95%)	High $\widetilde{\langle V \rangle}$ (95%)	"S.E."
Linear	1E+08	0.1	2	1.096326	1.09449	1.098169	9.37E-04
Expanding	1E+08	0.05	2	1.096618	1.095799	1.097437	4.18E-04
Linear	1E+09	0.1	4	1.007462	1.001784	1.013209	1.34E-03
Expanding	1E+09	0.05	4	1.014332	1.011706	1.01697	2.91E-03
Linear	1E+09	0.1	6	1.007543	0.947482	1.076177	3.28E-02
Expanding	1E+09	0.05	6	1.039808	0.996951	1.085848	2.27E-02
Linear	1E+09	0.1	8	1.077427	0.573328	34.4002	
Expanding	1E+09	0.05	8	-1.72861	-2.50015	3.82476	

Chapter 6

Solution to the fermion sign problem in 1 dimension

In the previous chapter, we discussed the fermion sign problem which, as many authors agree [Cep03, FH65], is a significant obstacle to performing effective fermion simulations. In one dimension, it is possible to solve the problem in the sense that, as we shall see, we can partition the path space such that for one part of space we need only to take positive contributions and in another, positive and negative contributions cancel exactly. It is immediately evident that if the functional being sampled is nonnegative then (5.2) cannot hold. This means that MCMC simulations can be constructed which do not suffer from an exponential explosion of variance as T increases. In this Chapter we shall construct the partition, show how to sample from the positive-contribution subspace, and demonstrate the effectiveness of an algorithm based on this approach. This solution to the problem is already known, and in the literature it normally appears as a special case of the RPIMC method already discussed in Subsection 5.2.1. In our approach, on the other hand, the main focus shall be the more general idea of partitioning the space of paths into subsets so that we can make progress with cancellation between positive and negative contributions to the estimate of an observable's expectation. We shall see in the succeeding chapter that this concept generalises to a somewhat different method than that of RPIMC. The importance of

this chapter's exposition is twofold: firstly, we are able to discuss the solution to the 1D sign problem mathematically, without needing to invoke concepts from physics, and secondly, it provides us with the foundation needed for a novel, exact approach to the sign problem in higher dimensions, advanced in the chapter that follows.

We shall begin by treating the case of two particles in a 1-dimensional space and then explain how this is generalised to n particles in a 1-dimensional space.

6.1 The Partitioning Theorem for two fermions in 1D

Recall that we are interested in collecting a Monte Carlo estimate of $\mathcal{I}_T(\cdot)$ defined by (2.62). In this section, we particularly consider $\mathcal{I}_T((a, b), (a, b))$ where $a, b \in \mathbb{R}$ are the hypothetical positions of two fermions. We develop the idea that it is possible to construct a partition so that we can find $\mathcal{I}_T((a, b), (a, b))$ by taking a positive contribution on one part of $\mathcal{C}_{0,(a,b)}^{T,(a,b)}$ and an unweighted sum of positive and negative contributions on the other. This latter turns out to be zero, eliminating the need to collect any negative contributions in Monte Carlo.

Note to begin with that

$$\mathcal{I}_T((a, b), (a, b)) = \frac{1}{1 + \exp\left(-\frac{(a-b)^2}{T}\right)} \left(\mathcal{J}_T((a, b), (a, b)) - \exp\left(-\frac{(a-b)^2}{T}\right) \mathcal{J}_T((a, b), (b, a)) \right). \quad (6.1)$$

As stated, we shall show that this is equal to the integral of a particular nonnegative functional. We shall begin by demonstrating the fact that the coefficient $\exp\left(-\frac{(a-b)^2}{T}\right)$ is equal to the probability of the two particles meeting under $w_{0,(a,b)}^{T,(a,b)}$, and then we shall see that the cancellation of positive and negative contributions becomes possible because the measure on first crossing times is the same under both $w_{0,(a,b)}^{T,(b,a)}$ and under $w_{0,(a,b)}^{T,(a,b)}$ conditioned on crossing.

We shall use $v_{0,\Delta_0}^{T,\Delta_T}$ to denote the measure on $\mathcal{C}_{0,\Delta_0}^{T,\Delta_T}$ induced by the distribution of a difference of Brownian bridges. That is to say, suppose that a 2-dimensional $X : \Omega \rightarrow \mathcal{C}_{0,x_0}^{T,x_T}$ is distributed according to w_{0,x_0}^{T,x_T} , so that $X^{(1)} : \Omega \rightarrow \mathcal{C}_{0,x_0}^{T,x_T^{(1)}}$ is distributed according to $w_{0,x_0}^{T,x_T^{(1)}}$ and

$X^{(2)} : \Omega \rightarrow \mathcal{C}_{0,x_0^{(2)}}^{T,x_T^{(2)}}$ is distributed independently according to $w_{0,x_0^{(2)}}^{T,x_T^{(2)}}$. Let $x_0^{(1)} - x_0^{(2)} = \Delta_0$ and $x_T^{(1)} - x_T^{(2)} = \Delta_T$. Then for $A \in \mathcal{B}(\mathcal{C}_{0,\Delta_0}^{T,\Delta_T})$, $v_{0,\Delta_0}^{T,\Delta_T}(A) := w_{0,x_0}^{T,x_T} \left(\left\{ X \in \mathcal{C}_{0,x_0}^{T,x_T} \mid X^{(2)} - X^{(1)} \in A \right\} \right)$.

Proposition 6.1.1 (The measure of a difference of Brownian Bridges) *Given*

$$\Delta_0, \Delta_T > 0,$$

$$\forall A \in \mathcal{B}(\mathcal{C}_{0,\Delta_0}^{T,\Delta_T}) : v_{0,\Delta_0}^{T,\Delta_T}(A) = w_{0,\Delta_0}^{2T,\Delta_T}(2A) \quad (6.2)$$

where we define $\lambda A = \{Z^* \in \mathcal{C}_{0,\Delta_0}^{\lambda T,\Delta_T} : Z^*(\lambda t) = Z(t) \mid Z \in A\}$.

Proof. We can obtain the measure $v_{0,\Delta_0}^{T,\Delta_T}$ on $\mathcal{C}_{0,\Delta_0}^{T,\Delta_T}$ which is induced by the distribution of a difference of bridges by considering iteratively the finite-dimensional distributions (in a manner originally due to P.Lévy [Lev39]), as follows. Suppose that we consider first the distribution for one intermediate point-pair at $T/2$. According to (2.4),

$$X^{(1)}(T/2) \sim \text{Gaussian} \left(X^{(1)}(0)/2 + X^{(1)}(T)/2, T/4 \right) \quad (6.3)$$

$$X^{(2)}(T/2) \sim \text{Gaussian} \left(X^{(2)}(0)/2 + X^{(2)}(T)/2, T/4 \right) \quad (6.4)$$

Therefore

$$\begin{aligned} X^{(1)}(T/2) - X^{(2)}(T/2) &\sim \\ \text{Gaussian} \left((X^{(1)}(0) - X^{(2)}(0)) / 2 + (X^{(1)}(T) - X^{(2)}(T)) / 2, T/2 \right) &. \end{aligned} \quad (6.5)$$

But let us now consider subsequent points. Suppose that we already fixed a set of points at times $hi/2^r$ where $0 \leq i \leq 2^r$. By a standard result, for $t = Ti/2^r + T/2^{r+1}$ for $i < 2^r$,

$$X^{(1)}(t) \sim \text{Gaussian} \left(X^{(1)}(t - T/2^{r+1}) / 2 + X^{(1)}(t + T/2^{r+1}) / 2, T/2^{r+2} \right),$$

$$X^{(2)}(t) \sim \text{Gaussian} \left(X^{(2)}(t - T/2^{r+1}) / 2 + X^{(2)}(t + T/2^{r+1}) / 2, T/2^{r+2} \right),$$

but then

$$\begin{aligned} X^{(1)}(t) - X^{(2)}(t) &\sim \\ \text{Gaussian} \left(\begin{aligned} &(X^{(1)}(t - T/2^{r+1}) - X^{(2)}(t - T/2^{r+1})) / 2 \\ &+ (X^{(1)}(t + T/2^{r+1}) - X^{(2)}(t + T/2^{r+1})) / 2 \end{aligned}, T/2^{r+1} \right) &. \end{aligned}$$

It follows by induction that for any r , the distribution of $X^{(1)} - X^{(2)}$ at points at times $hi/2^r$ is the same as the distribution of points at times $hi/2^{r-1}$ on a Brownian bridge of length $2T$. The result follows. ■

It is germane to our whole discussion to know the probability of two Brownian bridges crossing. We have now arrived at an important point because we can proceed to calculate such probabilities by applying a boundary-crossing theorem for the Brownian bridge.

Lemma 6.1.2 (The probability of bridges crossing) *For $\Delta_0, \Delta_T > 0$, the probability of two Brownian bridges crossing is given by*

$$v_{0,\Delta_0}^{T,\Delta_T} \left(\left\{ Z \in \mathcal{C}_{0,\Delta_0}^{T,\Delta_T} \mid \min_{[0,T]} Z(t) \leq 0 \right\} \right) = \exp \left(-\frac{\Delta_0 \Delta_T}{T} \right) \quad (6.6)$$

Proof. From the preceding Proposition,

$$v_{0,\Delta_0}^{T,\Delta_T} \left(\left\{ Z \in \mathcal{C}_{0,\Delta_0}^{T,\Delta_T} \mid \min_{[0,h]} Z(t) \leq 0 \right\} \right) = w_{0,\Delta_0}^{2T,\Delta_T} \left(\left\{ Z \in \mathcal{C}_{0,\Delta_0}^{2T,\Delta_T} \mid \min_{[0,2T]} Z(t) \leq 0 \right\} \right) \quad (6.7)$$

However,

$$\begin{aligned} w_{0,\Delta_0}^{2T,\Delta_T} \left(\left\{ Z \in \mathcal{C}_{0,\Delta_0}^{2T,\Delta_T} \mid \min_{[0,2T]} Z(t) \leq 0 \right\} \right) \\ = w_{0,\Delta_0}^{2T,\Delta_T} \left(\left\{ Z \in \mathcal{C}_{0,0}^{2T, -(\Delta_0 - \Delta_T)} \mid \min_{[0,2T]} Z(t) \leq -\Delta_0 \right\} \right) \\ = w_{0,\Delta_0}^{2T,\Delta_T} \left(\left\{ Z \in \mathcal{C}_{0,0}^{2T,\Delta_0 - \Delta_T} \mid \max_{[0,2T]} Z(t) \geq \Delta_0 \right\} \right) \end{aligned}$$

using the translation and symmetry properties of the conditional Wiener measure described in Subsection 2.1.1. However, according to a Theorem stated on pages 264-265 of [KS98],

$$w_{0,\Delta_0}^{2T,\Delta_T} \left(\left\{ Z \in \mathcal{C}_{0,0}^{2T,\Delta_0 - \Delta_T} \mid \max_{[0,2T]} Z(t) \geq \Delta_0 \right\} \right) = \exp \left(-\frac{2\Delta_0 \Delta_T}{2T} \right) = \exp \left(-\frac{\Delta_0 \Delta_T}{T} \right) \quad (6.8)$$

[in their notation $\Delta_0 = \beta$ and $\Delta_0 - \Delta_T = a$]. Hence

$$v_{0,\Delta_0}^{T,\Delta_T} \left(\left\{ Z \in \mathcal{C}_{0,\Delta_0}^{T,\Delta_T} \mid \min_{[0,T]} Z(t) \leq 0 \right\} \right) = \exp \left(-\frac{\Delta_0 \Delta_T}{T} \right). \quad (6.9)$$

■

The event that paths coincide We shall divide $\mathcal{C}_{0,x_0}^{T,x_T}$ into two classes: those bridges where the coordinates (particle positions) coincide at least once (therefore at least twice, and indeed an infinite number of times) and those where they do not. Let $a > b$ and let

$$E_0 = \left\{ X \in \mathcal{C}_{0,(a,b)}^{T,(a,b)} \mid X^{(2)}(t) < X^{(1)}(t) \ \forall t \in [0, T] \right\}.$$

Then by continuity of the paths,

$$E_0^c = \left\{ X \in \mathcal{C}_{0,(a,b)}^{T,(a,b)} \mid X^{(2)}(t) = X^{(1)}(t), \text{ some } t \in [0, T] \right\}.$$

It is immediately evident that these are events because it is obvious how we might construct a sequence of cylinders with E_0 as its limit.

The first crossing time $\tau_1(X)$ Recalling Definition 2.1.14, we shall now let τ_1 denote the **first crossing time** $\tau_1 : E_0^c \rightarrow [0, T]$:

$$\tau_1(X) = \min \{ s \in [0, T] : X^{(1)}(s) = X^{(2)}(s) \}$$

We shall now compute the pdf of this random variable, which we shall variously refer to as τ_1 or $\tau_1(X)$ depending on the context.

Lemma 6.1.3 (The pdf of the first crossing time) *Let $X \sim w_{0,x_0}^{T,x_T}$ and define Δ_0, Δ_T as on p.127. Denote the measure induced on $\mathcal{B}([0, T])$ by the distribution of $\tau_1(X)$ by η . Then*

$$\frac{d\eta}{d\Lambda} = \frac{\Delta_0}{s} \frac{1}{\sqrt{2\pi} \sqrt{\frac{2s(T-s)}{T}}} \exp \left(-\frac{((T-s)\Delta_0 + s\Delta_T)^2}{4Ts(T-s)} \right).$$

(Note that $\eta([0, T])$ is then the probability of crossing, $\exp \left(-\frac{\Delta_0 \Delta_T}{T} \right)$.)

Proof. We know from Proposition 6.1.1 that the finite-dimensional distribution of a difference of bridges, with the difference at the initial time Δ_0 and the difference at the terminal time Δ_T , over a time interval $[0, T]$, is identical to that of a Brownian bridge from Δ_0 to Δ_T over a time interval $[0, 2T]$, with a linear scaling on time. We shall use the following standard method for finding the first crossing time pdf: we use the fact that where $f(x, y)$ is

a joint pdf, $g(x)$ is the unconditional pdf of x and h is that of y , and $f(y|x)$ has its standard meaning, $f(x, y) = g(x)f(y|x) = h(y)f(x|y)$. The equation (2.0.2) of [BS02] states that for Brownian motion with $X(0) = \Delta_0$, where τ is the first hitting time to zero,

$$\frac{d}{ds}w_{0,\Delta_0}(\tau \in (0, s)) = \frac{|\Delta_0|}{\sqrt{2\pi}s^{3/2}} \exp\left(-\frac{\Delta_0^2}{2s}\right) \quad (6.10)$$

and it follows that the joint pdf for τ and the terminal value $X(T)$ is given by

$$\frac{\partial^2}{\partial s \partial z}w_{0,\Delta_0}(\tau \in (0, s), X(T) \leq z) = \frac{|\Delta_0|}{\sqrt{2\pi}s^{3/2}} \exp\left(-\frac{\Delta_0^2}{2s}\right) \frac{1}{\sqrt{2\pi}\sqrt{T-s}} \exp\left(-\frac{z^2}{2(T-s)}\right) \quad (6.11)$$

but from this we can deduce that conditional on the terminal value, the pdf of the first hitting time to zero is

$$\begin{aligned} \frac{d}{ds}w_{0,\Delta_0}^{T,z}(\tau \in (0, s)) &= \frac{d}{ds}w_{0,\Delta_0}(t \in (0, s)|X(T) = z) = \\ &= \frac{|\Delta_0|}{\sqrt{2\pi}s^{3/2}} \exp\left(-\frac{\Delta_0^2}{2s}\right) \frac{1}{\sqrt{2\pi}\sqrt{T-s}} \exp\left(-\frac{z^2}{2(T-s)}\right) \sqrt{2\pi}\sqrt{T} \exp\left(+\frac{(z-\Delta_0)^2}{2T}\right) \\ &= \frac{\Delta_0}{s} \frac{1}{\sqrt{2\pi}\sqrt{\frac{s(T-s)}{T}}} \exp\left(-\Delta_0^2\left(\frac{1}{2s} - \frac{1}{2T}\right) - z^2\left(\frac{1}{2(T-s)} - \frac{1}{2T}\right) - 2\Delta_0 z\left(\frac{1}{2T}\right)\right) \end{aligned} \quad (6.12)$$

by dividing the previous formula by the unconditional density of terminal values. Now let us consider that $v_{0,\Delta_0}^{T,z}(t \leq s) = w_{0,\Delta_0}^{2T,z}(t \leq 2s)$ and this means

$$\frac{d}{ds}v_{0,\Delta_0}^{T,z}(t \leq s) = \frac{d}{ds}w_{0,\Delta_0}^{2T,z}(t \leq 2s) = 2\frac{d}{d(2s)}w_{0,\Delta_0}^{2T,z}(t \leq 2s) \quad .$$

Thus, to get the pdf for first hitting time to zero of a difference of Brownian bridges rather than a Brownian bridge, we need to double all the times on the right-hand side in (6.12) and then double the result to reflect $\frac{d(2s)}{ds}$, ie:

$$\begin{aligned} \frac{d}{ds}v_{0,\Delta_0}^{T,\Delta_T}(\tau \in (0, s)) &= \\ &= \frac{\Delta_0}{s} \frac{1}{\sqrt{2\pi}\sqrt{\frac{2s(T-s)}{T}}} \exp\left(-\Delta_0^2\left(\frac{1}{4s} - \frac{1}{4T}\right) - \Delta_T^2\left(\frac{1}{4(T-s)} - \frac{1}{4T}\right) - 2\Delta_0\Delta_T\left(\frac{1}{4T}\right)\right) \end{aligned} \quad (6.13)$$

and the result follows. ■

We shall define the **first crossing point** $x_1(X)$ to be the value $X^{(1)}(\tau_1(X))$. It is clear that $x_1 \circ X$ is Borel-measurable, since if we condition on τ_1 then for any interval U , $\{x_1 \in U\}$ is Borel.

We are now in a position to prove the main result of this section. Recalling the notation for measure conditioning introduced at the start of Section 2.2,

Theorem 6.1.4 (Partitioning Theorem for two 1D fermions) *For $a, b \in \mathbb{R}$,*

$$\mathcal{I}_T((a, b), (a, b)) = \frac{1 - \exp\left(-\frac{(a-b)^2}{T}\right)}{1 + \exp\left(-\frac{(a-b)^2}{T}\right)} \mathcal{J}_T^{E_0}((a, b), (a, b)) \quad (6.14)$$

Proof. It is immediately obvious that all bridges from (a, b) to (b, a) must have the coordinates swap at least once, so τ_1 is defined for $X \in \mathcal{C}_{0,(a,b)}^{T,(b,a)}$. Let η_1^* signify the measure on $[0, T] \times \mathbb{R}$ induced by the distribution of (τ_1, x_1) for $X \in E_0^c$ under $w_{0,(a,b)}^{T,(a,b)} \Big|_{E_0^c}$ and let η_2^* signify the measure on $[0, T] \times \mathbb{R}$ induced by the distribution of (τ_1, x_1) for $X \in \mathcal{C}_{0,(a,b)}^{T,(b,a)}$ under $w_{0,(a,b)}^{T,(b,a)}$. That is to say, for $B_1 \in \mathcal{B}([0, T]), B_2 \in \mathcal{B}(\mathbb{R})$:

$$\eta_1^*(B_1 \times B_2) = w_{0,(a,b)}^{T,(a,b)} \Big|_{E_0^c} (\{X \in E_0^c : \tau_1(X) \in B_1, x_1(X) \in B_2\}) , \quad (6.15)$$

$$\eta_2^*(B_1 \times B_2) = w_{0,(a,b)}^{T,(b,a)} \left(\left\{ X \in \mathcal{C}_{0,(a,b)}^{T,(b,a)} : \tau_1(X) \in B_1, x_1(X) \in B_2 \right\} \right) . \quad (6.16)$$

We shall use the notation (this matches the definition of $E(t, x)$ on p.26):

$$E(t, x) = \{X \in E_0^c : \tau_1(X) = t, x_1(X) = x\} , \quad (6.17)$$

$$E^{\text{SW}}(t, x) = \left\{ X \in \mathcal{C}_{0,(a,b)}^{T,(b,a)} : \tau_1(X) = t, x_1(X) = x \right\} . \quad (6.18)$$

Therefore we may rewrite $\mathcal{I}_T((a, b))$ as follows:

$$\begin{aligned} \left(1 + \exp\left(-\frac{(a-b)^2}{T}\right)\right) \mathcal{I}_T((a, b), (a, b)) &= w_{0,(a,b)}^{T,(a,b)}(E_0) \mathcal{J}_T^{E_0}((a, b), (a, b)) \\ &\quad + w_{0,(a,b)}^{T,(a,b)}(E_0^c) \int_{[0,T] \times \mathbb{R}} \mathcal{J}_T^{E(t,x)}((a, b), (a, b)) d\eta_1^*(t, x) \\ &\quad - \exp\left(-\frac{(a-b)^2}{T}\right) \int_{[0,T] \times \mathbb{R}} \mathcal{J}_T^{E^{\text{SW}}(t,x)}((a, b), (b, a)) d\eta_2^*(t, x) . \end{aligned}$$

However, recalling (2.26) it is clear that

$$\begin{aligned}\mathcal{J}_T((a, b), (a, b); E(t, x)) &= \mathcal{J}_t^{E(t, x)}((a, b), (x, x)) \mathcal{J}_{T-t}^{E(t, x)}((x, x), (a, b)) \\ &= \mathcal{J}_t^{E(t, x)}((a, b), (x, x)) \mathcal{J}_{T-t}((x, x), (a, b))\end{aligned}$$

but for the same reasons,

$$\mathcal{J}_T^{E^{\text{SW}}(t, x)}((a, b), (b, a)) = \mathcal{J}_t^{E(t, x)}((a, b), (x, x)) \mathcal{J}_{T-t}((x, x), (b, a)) \quad .$$

However, due to permutational symmetry of the potential,

$$\mathcal{J}_{T-t}((x, x), (a, b)) = \mathcal{J}_{T-t}((x, x), (b, a))$$

Now note from Lemma 6.1.2 that the probability of independent Brownian bridges in $\mathcal{C}_{0, (a, b)}^{T, (a, b)}$ crossing is $w_{0, (a, b)}^{T, (a, b)}(E_0^c) = \exp\left(-\frac{(a-b)^2}{T}\right)$. Thus,

$$\begin{aligned}w_{0, (a, b)}^{T, (a, b)}(E_0^c) &\int_{[0, T] \times \mathbb{R}} \mathcal{J}_T^{E(t, x)}((a, b), (a, b)) d\eta_1^*(t, x) \\ &\quad - \exp\left(-\frac{(a-b)^2}{T}\right) \int_{[0, T] \times \mathbb{R}} \mathcal{J}_T^{E^{\text{SW}}(t, x)}((a, b), (b, a)) d\eta_2^*(t, x) \\ &= \exp\left(-\frac{(a-b)^2}{T}\right) \int_{[0, T] \times \mathbb{R}} \mathcal{J}_t^{E(t, x)}((a, b), (x, x)) \mathcal{J}_{T-t}((x, x), (a, b)) d(\eta_1^* - \eta_2^*)(t, x) \quad .\end{aligned}\tag{6.19}$$

Now suppose $\tau_1 = t$ is given and consider the conditional distribution of x_1 , in the case of either $X \in E_0^c$ or $X \in \mathcal{C}_{0, (a, b)}^{T, (b, a)}$. As we noted in Subsection 2.1.3, $\frac{X^{(1)} - X^{(2)}}{\sqrt{2}}$ and $\frac{X^{(1)} + X^{(2)}}{\sqrt{2}}$ can be regarded as independent Brownian bridges; and indeed, $\frac{X^{(1)} + X^{(2)}}{\sqrt{2}}$ is in $\mathcal{C}_{0, (a+b)/\sqrt{2}}^{T, (a+b)/\sqrt{2}}$ whether the terminal point of X is (a, b) or (b, a) . Therefore in both cases it is clear that the distribution of the value of $\frac{X^{(1)} + X^{(2)}}{\sqrt{2}}$ at t , corresponding to the usual finite-dimensional distribution for a Brownian bridge, is Gaussian with mean $(a+b)/\sqrt{2}$ and variance $t(T-t)/T$. However, we know that $X^{(1)}(t) = X^{(2)}(t) = \frac{1}{\sqrt{2}} \frac{X^{(1)}(t) + X^{(2)}(t)}{\sqrt{2}}$ and therefore in both cases, $x_1(X)$ is Gaussian with mean $(a+b)/2$ and variance $t(T-t)/(2T)$. We shall let μ denote the measure induced by this distribution. We let η_1 signify the measure on $[0, T]$ induced by the distribution of τ_1 for $X \in E_0^c$ under $w_{0, (a, b)}^{T, (a, b)} \Big|_{E_0^c}$ and let η_2 signify the measure on $[0, T]$ induced by the

distribution of τ_1 for $\mathcal{C}_{0,(a,b)}^{T,(b,a)}$ under $w_{0,(a,b)}^{T,(b,a)}$. We let $\Delta_0 = a - b$. Then we may write the formula of (6.19) as

$$\exp\left(-\frac{\Delta_0^2}{T}\right) \int_{[0,T]} \int_{\mathbb{R}} \left[\mathcal{J}_t^{E(t,x)}((a,b), (x,x)) \mathcal{J}_{T-t}((x,x), (a,b)) \right] d\mu(x) d(\eta_1 - \eta_2)(t) . \quad (6.20)$$

Hence our next step is to identify η_1 and η_2 , and we shall then see that they are equal. According to Lemma 6.1.3, where η is for the measure on $[0, T]$ induced by the distribution of τ_1 for $X \in \mathcal{C}_{0,(a,b)}^{T,(a,b)}$ under $w_{0,(a,b)}^{T,(a,b)}$,

$$\frac{d\eta}{d\Lambda} = \frac{\Delta_0}{s} \frac{1}{\sqrt{2\pi} \sqrt{\frac{2s(T-s)}{T}}} \exp\left(-\frac{T\Delta_0^2}{4s(T-s)}\right)$$

and therefore, dividing by $w_{0,(a,b)}^{T,(a,b)}(E_0^c)$,

$$\begin{aligned} \frac{d\eta_1}{d\Lambda} &= \frac{\Delta_0}{s} \frac{1}{\sqrt{2\pi} \sqrt{\frac{2s(T-s)}{T}}} \exp\left(\frac{\Delta_0^2}{T} - \frac{T\Delta_0^2}{4s(T-s)}\right) \\ &= \frac{\Delta_0}{s} \frac{1}{\sqrt{2\pi} \sqrt{\frac{2s(T-s)}{T}}} \exp\left(\frac{(4s(T-s) - T^2)\Delta_0^2}{4Ts(T-s)}\right) \end{aligned}$$

whereas Lemma 6.1.3 also yields that

$$\frac{d\eta_2}{d\Lambda} = \frac{\Delta_0}{s} \frac{1}{\sqrt{2\pi} \sqrt{\frac{2s(T-s)}{T}}} \exp\left(-\frac{((T-2s)\Delta_0)^2}{4Ts(T-s)}\right) .$$

However, it is then apparent that, perhaps surprisingly, $\frac{d\eta_1}{d\Lambda} = \frac{d\eta_2}{d\Lambda}$. Hence there is perfect cancellation between the positive contributions to \mathcal{I}_T from paths in E_0^c and the negative contributions from paths in $\mathcal{C}_{0,(a,b)}^{T,(b,a)}$. Therefore

$$\begin{aligned} \left(1 + \exp\left(-\frac{\Delta_0^2}{T}\right)\right) \mathcal{I}_T((a,b), (a,b)) &= w_{0,(a,b)}^{T,(a,b)}(E_0) \mathcal{J}_T^{E_0}((a,b), (a,b)) = \\ &= \left(1 - \exp\left(-\frac{\Delta_0^2}{T}\right)\right) \mathcal{J}_T^{E_0}((a,b), (a,b)) \quad . \quad (6.21) \end{aligned}$$

■

The quantity $I_{T,2}$, defined as per Theorem 2.3.4, is relevant to finding the first excited state for one particle, and is equal to \mathcal{I}_T for 2 particles. Therefore a similar result to Theorem 6.1.4 clearly holds w.r.t. $I_{T,2}$.

6.2 The Partitioning Theorem for n fermions in 1D

The approach of the previous section brings out the importance of the distribution of first crossing times for $w_{0,(a,b)}^{T,(a,b)}$ and $w_{0,(a,b)}^{T,(b,a)}$. However, there is an alternative, more geometrical approach which allows us to treat an indefinite number of particles. This shall now be explained.

Theorem 6.2.1 (Partitioning Theorem for n 1D fermions) *Let $E_0 \subseteq \mathcal{C}_{0,x}^{T,\Pi_n x}$ signify the event that there are no crossings, ie no values of t, i, j for which $X^{(i)}(t) = X^{(j)}(t)$ with $i \neq j$. Then*

$$\mathcal{I}_T(x, x) = w_{0,x}^{T,\Pi_n x}(E_0) \mathcal{J}_T^{E_0}(x, x) \quad (6.22)$$

Proof. As was noted at (2.67),

$$\begin{aligned} \mathcal{I}_T(x, x) &= \int_{\mathcal{C}_{0,x}^{T,\Pi_n x}} \left[c \left(\frac{X(T)}{x} \right) \exp \left(\int_0^T -V(X(s)) ds \right) \right] dw_{0,x}^{T,\Pi_n x} \\ &= \int_{E_0} \exp \left(\int_0^T -V(X(s)) ds \right) dw_{0,x}^{T,\Pi_n x} \\ &+ \int_{[0,T] \times \mathbb{R}^n} \int_{\mathcal{C}_{0,x}^{T,\Pi_n x}} \left[c \left(\frac{X(T)}{x} \right) \exp \left(\int_0^T -V(X(s)) ds \right) \right] d w_{0,x}^{T,\Pi_n x} \Big|_{E(t,x_\sharp)} d\eta'(t, x_\sharp) \quad (6.23) \end{aligned}$$

where $E(t, x_\sharp)$ is the event that the first crossing of any pair of particles happens at time t with system position x_\sharp and η' is the measure induced by the joint distribution of the first crossing time and system point, with full measure $w_{0,x}^{T,\Pi_n x}(E_0^c)$. However, we shall now see that the second integral in (6.23) is zero. Let i_1, i_2 be the indices of the particles that meet at t . Recall, as was noted in Chapter 1, that Π_n may be partitioned into pairs which are closed under premultiplication by the pair-exchange of i_1, i_2 ; let us call the even elements in these pairs π_k and the odd elements π'_k with $k = 1, \dots, n!/2$. Then since

$$w_{0,x}^{T,\Pi_n x} \Big|_{E(t,x_\sharp)} = w_{0,x}^{t,x_\sharp} \Big|_{E(t,x_\sharp)} \times w_{t,x_\sharp}^{T,\Pi_n x} \text{ (due to the strong Markov property for } w_{0,x}^{T,\Pi_n x}),$$

$$\begin{aligned} \int_{\mathcal{C}_{0,x}^{T,\Pi_n x}} \left[c \left(\frac{X(T)}{x} \right) \exp \int_0^T -V(X(s))ds \right] d w_{0,x}^{T,\Pi_n x} \Big|_{E(t,x_\sharp)} = \\ \frac{\sum_{k=1}^{n!} \varphi_{T-t}(x_\sharp, \pi_k x)}{\sum_{k=1}^{n!} \varphi_T(x, \pi_k x)} \int_{\mathcal{C}_{0,x}^{t,x_\sharp}} \exp \left(\int_0^T -V(X(s))ds \right) d w_{0,x}^{t,x_\sharp} \Big|_{E(t,x_\sharp)} \times \\ \int_{\mathcal{C}_{t,x_\sharp}^{T,\Pi_n x}} \left[c \left(\frac{X(T)}{x} \right) \exp \left(\int_0^T -V(X(s))ds \right) \right] d w_{t,x_\sharp}^{T,\Pi_n x}, \end{aligned}$$

but

$$\begin{aligned} \int_{\mathcal{C}_{t,x_\sharp}^{T,\Pi_n x}} \left[c \left(\frac{X(T)}{x} \right) \exp \int_0^T -V(X(s))ds \right] d w_{t,x_\sharp}^{T,\Pi_n x} = \\ \frac{1}{\sum_{k=1}^{n!} \varphi_{T-t}(x_\sharp, \pi_k x)} \sum_{k=1}^{n!/2} [\varphi_{T-t}(x_\sharp, \pi_k x) \mathcal{J}_{T-t}(x_\sharp, \pi_k x) - \varphi_{T-t}(x_\sharp, \pi'_k x) \mathcal{J}_{T-t}(x_\sharp, \pi'_k x)] = 0 \end{aligned}$$

since for every k , $\varphi_{T-t}(x_\sharp, \pi_k x) = \varphi_{T-t}(x_\sharp, \pi'_k x)$ and $\mathcal{J}_{T-t}(x_\sharp, \pi_k x) = \mathcal{J}_{T-t}(x_\sharp, \pi'_k x)$. However, it is clear that

$$\begin{aligned} \int_{E_0} \exp \left(\int_0^T -V(X(s))ds \right) d w_{0,x}^{T,\Pi_n x} &= \frac{1}{\sum_{k=1}^{n!} \varphi_T(x, \pi_k x)} \int_{E_0} \exp \left(\int_0^T -V(X(s))ds \right) d w_{0,x}^{T,x} \\ &= w_{0,x}^{T,\Pi_n x}(E_0) \mathcal{J}_T^{E_0}(x, x) \end{aligned}$$

and the result follows. ■

Remark 6.2.2 *It is clear that a similar result also applies in the case of other exponential-type functionals than $\exp \left(- \int_0^T V(X(s))ds \right)$ since the only special property of this functional that has been used is its multiplicative property.*

6.3 Simulation methods for 1-dimensional non-crossing bridges

The preceding results mean that the fermion sign problem is solved in 1 dimension, as long as we can perform the necessary simulations. The main task is to perform integrals with respect to the non-crossing bridge measure $w_0^* = \Lambda \times w_{0,x_0}^{T,x_0} \Big|_{E_0}$. We have already noted that

for the case of two fermions, we can find the probability of a crossing in an interval under $w_{t_{k_1}, x_{k_1}}^{t_{k_2}, x_{k_2}}$, and by using the Karlin-McGregor theorem [KM59] in conjunction with Bayes Rule, it should be straightforward to do the same for the case of $n > 2$ fermions. From this point it is possible to proceed in two ways depending on the type of simulation in view. In the case of direct path sampling, the most appealing approach is to identify the finite-dimensional distributions of $w_{0, x_0}^{T, x_0} \Big|_{E_0}$ explicitly; since this is a Markov measure, we can construct the transition density; it will then remain to develop a sampling method for this density. In this Section, we follow out the details of this approach for the case of 2 fermions. As described in Chapter 4, we may then create an algorithm where the expectation is also taken over values of the initial point $x_0 = (a, b)$ so that estimates of quantities such as (2.69) can be achieved.

Alternatively, if we were to construct a Markov Chain Monte Carlo simulation to implement importance sampling, taking $\frac{d\nu}{dw_0^*} = Y$, then the problem of using non-crossing bridges becomes substantially easier. In this case, intuition suggests that a simple rejection method would almost certainly be adequate to sample according to $w_{t_{k_1}, x_{k_1}}^{t_{k_2}, x_{k_2}} \Big|_{E_0}$, given that x_{k_1}, x_{k_2} will already be such that samples according to $w_{t_{k_1}, x_{k_1}}^{t_{k_2}, x_{k_2}}$ should already have a reasonable probability of acceptance, especially if the number of time intervals between k_1 and k_2 is sufficiently small. Thus, in practical applications this would be a far more efficient way to proceed. Since the probability of not crossing during each time interval, as established using the Karlin-McGregor theorem, may be a sum involving a large number of terms which are expensive to evaluate, some truncation based on particle proximity might be needed. Simulations based on this MCMC approach are almost certainly *more* efficient than the simulations performed in Section 6.5, even before the benefit of importance sampling is taken into account. However, understanding the details of the direct path sampling approach, outlined in what follows, is certainly helpful and relevant for understanding how to implement either approach.

Using direct path sampling, we do have the option to sample discretisation points according to a different pdf than $\widetilde{w_{0, x_0}^{T, x_0} \Big|_{E_0}}$ and apply a Radon-Nikodym reweighting - or, equiva-

lently, to reject paths with an appropriate probability. This latter is a distinct concept from using a naïve rejection method for sampling from the non-crossing fdd by taking $w_{t_k, x_k}^{t_{k+1}, x_{k+1}}$ as the candidate measure. We may infer, though, from the difficulties encountered using that method, that there would be similar (or worse) difficulties inherent in a method dependent upon rejecting or reweighting entire Brownian bridges. It appears to be a characteristic of an effective simulation, whether using MCMC or direct path sampling, that the sampling measure should itself be concentrated on E_0 .

In Subsection 6.3.1, the conditional pdf to sample one point (ie the transition density) for $\widetilde{w_{0, x_0}^{T, x_0}} \Big|_{E_0}$ is derived explicitly. In Subsection 6.3.2 then methods of sampling from this pdf are considered. Three methods were implemented and one of them was found to be substantially faster than the others. Finally in Subsection 6.3.3, we discuss the initial point distribution.

6.3.1 The pdf for a point on a non-crossing Brownian bridge

We need to know how to sample discretisations of 2-coordinate paths $X = (X^{(1)}, X^{(2)})$ according to $w_{0, (a, b)}^{T, (a, b)} \Big|_{E_0}$. It is assumed that it is adequate to sample incrementally (ie we collect the values in order of time) at equidistant points. That is, set $t_k = kh$, for $k = 1, \dots, N-1$, with $h = T/N$. Fixing k , we shall assume that the values of $X(t_1), \dots, X(t_{k-1})$ have been drawn and consider how to sample the value for $X(t_k)$. In other words, our route to the joint pdf of $X(t_1), \dots, X(t_{N-1})$ is to consider the conditional pdf for $X(t_k)$ given $X(t_{k-1})$, since as we shall see, $w_{0, (a, b)}^{T, (a, b)} \Big|_{E_0}$ retains conditional independence. One way to do this would be to use the Karlin-McGregor theorem [KM59] which states that to find the transition density of a process where all coordinates are constrained to not cross, one must take a particular determinant; namely, one must sum over the transition densities to each permutation of the points in $X(t_k)$ and apply $\text{sgn}(\pi)$ as the cofactor. (This makes a curious parallel to (6.22).) However, we shall prefer a more direct approach using Bayes' Rule.

We shall use $x_l^{(1)}, x_l^{(2)}$ for the drawn values of the particle positions at time l . We can

access the desired conditional pdf as follows.

Lemma 6.3.1 *Let D be a Borel subset of $\{\{x, y\} \in \mathbb{R}^2 | x > y\}$ and let $A(D)$ be the event that $X(t_k) \in D$. Let $E_0 = E_1 \times E_2$ with $E_1 \subseteq \mathcal{C}_{0,(a,b)}^{t_{k-1}}$ and $E_2 \subseteq \mathcal{C}_{t_{k-1}}^{T,(a,b)}$. Let μ_w denote the measure for $X(t)$ under $w_{t_{k-1},x_{k-1}}^{T,(a,b)}$. Then*

$$\begin{aligned} & w_{0,(a,b)}^{T,(a,b)} \Big|_{E_0, X(t_{k-1})=x_{k-1}} (A(D)) \\ &= \frac{\int_D \left(1 - \exp \left(-\frac{(x^{(1)}-x^{(2)})(a-b)}{(N-k)h} \right) \right) \left(1 - \exp \left(-\frac{(x_{k-1}^{(1)}-x_{k-1}^{(2)})(x^{(1)}-x^{(2)})}{h} \right) \right) d\mu_w(x)}{w_{t_{k-1},x_{k-1}}^{T,(a,b)}(E_2)} \end{aligned} \quad (6.24)$$

Proof. Recall that conditional independence is inherited when conditioning on a set that is not of measure zero (cf (2.19)). Therefore

$$w_{0,(a,b)}^{T,(a,b)} \Big|_{E_0, X(t_{k-1})=x_{k-1}} = w_{0,(a,b)}^{t_{k-1},x_{k-1}} \Big|_{E_1} \times w_{t_{k-1},x_{k-1}}^{T,(a,b)} \Big|_{E_2}.$$

Then since $A(D) = \mathcal{C}_{0,(a,b)}^{t_{k-1}} \times A_2(D)$ where $A_2(D)$ is the event in $\mathcal{B}(\mathcal{C}_{t_{k-1}}^{T,(a,b)})$ that $X(t_k) \in D$,

$$\begin{aligned} w_{0,(a,b)}^{T,(a,b)} \Big|_{E_0, X(t_{k-1})=x_{k-1}} (A(D)) &= w_{0,(a,b)}^{t_{k-1},x_{k-1}} \Big|_{E_1} \left(\mathcal{C}_{0,(a,b)}^{t_{k-1}} \right) w_{t_{k-1},x_{k-1}}^{T,(a,b)} \Big|_{E_2} (A_2(D)) \\ &= w_{t_{k-1},x_{k-1}}^{T,(a,b)} \Big|_{E_2} (A_2(D)) \end{aligned}$$

Heuristically speaking, since $X(t_k)$ is conditionally independent of the values taken prior to time t_{k-1} , in fact any specified path up to time t_{k-1} , whether it confers membership of E_1 or not, will be associated with the same conditional distribution of $X(t_k)$. Now since $w_{t_{k-1},x_{k-1}}^{T,(a,b)}(E_2) \neq 0$,

$$w_{t_{k-1},x_{k-1}}^{T,(a,b)} \Big|_{E_2} (A_2(D)) = \frac{w_{t_{k-1},x_{k-1}}^{T,(a,b)}(A_2(D) \cap E_2)}{w_{t_{k-1},x_{k-1}}^{T,(a,b)}(E_2)}.$$

But,

$$\begin{aligned} w_{t_{k-1},x_{k-1}}^{T,(a,b)}(A_2(D) \cap E_2) &= \int_D w_{t_{k-1},x_{k-1}}^{T,(a,b)} \Big|_{X(t)=x} (A \cap E_2) d\mu(x) \\ &= \int_D w_{t_{k-1},x_{k-1}}^{T,(a,b)} \Big|_{X(t)=x} (E_2) d\mu(x) \end{aligned}$$

since membership of $A_2(D)$ will apply for any path with $X(t) \in D$. However, where $E_2 = E_3 \times E_4$ with $E_3 \subseteq \mathcal{C}_{t_{k-1}}^{t_k}$, $E_4 \subseteq \mathcal{C}_{t_k}^{T,(a,b)}$,

$$w_{t_{k-1},x_{k-1}}^{T,(a,b)} \Big|_{X(t)=x} (E_2) = w_{t_{k-1},x_{k-1}}^{t_k,x} (E_3) w_{t_k,x}^{T,(a,b)} (E_4). \quad (6.25)$$

According to Lemma 6.1.2,

$$w_{t_{k-1}, x_{k-1}}^{t_k, x}(E_3) = 1 - \exp\left(-\frac{(x^{(1)} - x^{(2)})(a - b)}{(N - k)h}\right)$$

$$w_{t_k, x}^{T, (a, b)}(E_4) = 1 - \exp\left(-\frac{(x_{k-1}^{(1)} - x_{k-1}^{(2)})(x^{(1)} - x^{(2)})}{h}\right)$$

and the result follows. ■

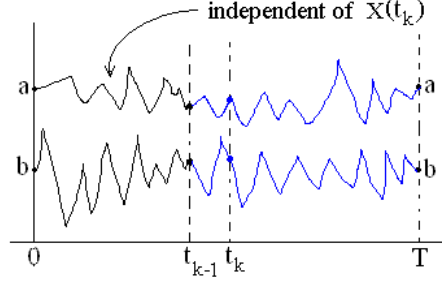


Figure 6.1: Schematic of how the pdf of $X(t_k)$ is determined when conditioning on E_0

Considering the finite-dimensional distributions of the conditional Wiener measure (2.4), it can be recognised that μ_w is induced by the independent bivariate Gaussian distribution with centre

$$\left(\left(ha + (N - k)hx_{k-1}^{(1)} \right) / (N - k + 1)h, \left(hb + (N - k)hx_{k-1}^{(2)} \right) / (N - k + 1)h \right) \quad (6.26)$$

$$= \left(\left(a + (N - k)x_{k-1}^{(1)} \right) / (N - k + 1), \left(b + (N - k)x_{k-1}^{(2)} \right) / (N - k + 1) \right) := (x_\mu^{(1)}, x_\mu^{(2)})$$

and with the variances in both directions given by

$$(N - k)h^2 / ((N - k + 1)h) = (N - k)h / (N - k + 1). \quad (6.27)$$

Hence in order to simulate according to $w_{t_{k-1}, x_{k-1}}^{T, (a, b)} \Big|_{E_2}$ it follows from (6.24) that we need to draw from the distribution with pdf given by

$$f(x^{(1)}, x^{(2)}) = C \left(1 - \exp\left(-\frac{(x^{(1)} - x^{(2)})(a - b)}{(N - k)h}\right) \right) \times$$

$$\left(1 - \exp\left(-\frac{(x_{k-1}^{(1)} - x_{k-1}^{(2)})(x^{(1)} - x^{(2)})}{h}\right) \right) \exp\left(-\frac{(x^{(1)} - x_\mu^{(1)})^2 + (x^{(2)} - x_\mu^{(2)})^2}{2(N - k)h / (N - k + 1)}\right)$$

(6.28)

over the viable region $\{x^{(1)} > x^{(2)}\}$ and 0 elsewhere; here C is a normalizing constant equal to $(2\pi(N-k)h/(N-k+1))^{-1} / w_{t_{k-1}, x_{k-1}}^{T,(a,b)}(E_2)$. Let ν be the measure on \mathbb{R}^2 induced by this distribution. For the example parameters of Figures 6.2-6.3, this function turns out

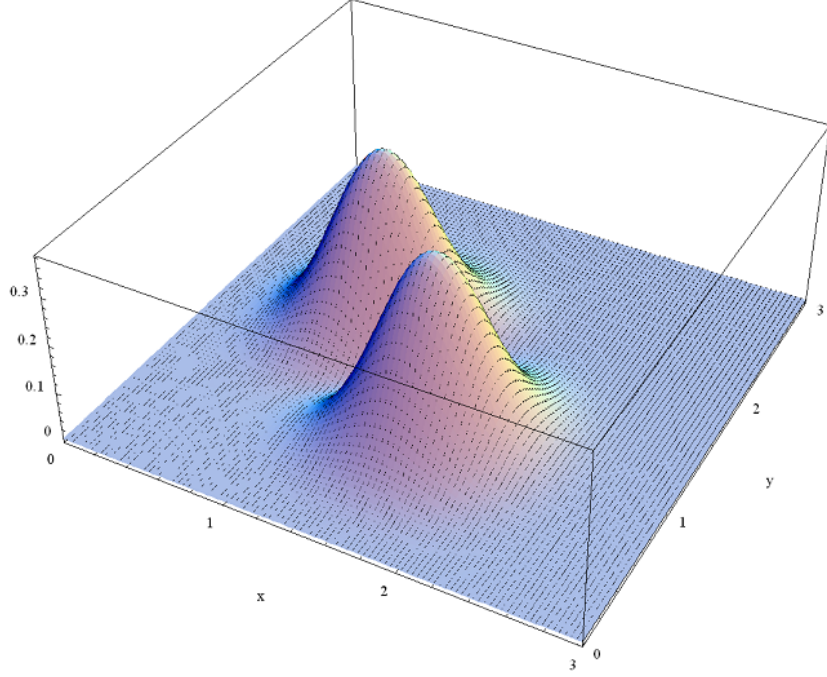


Figure 6.2: $f(x,y)$, unnormalized.

to still be unipolar; f itself is bipolar with $f(x^{(1)}, x^{(2)}) = f(x^{(2)}, x^{(1)})$. Intuition suggests that it is wise to rotate our coordinate basis for the system position, because if we choose a basis rotated by 45 degrees then only one coordinate shall be affected by the constraint for them not to cross, and thus independence of the coordinates shall be maintained. Along with some elementary manipulations, this allows us to find the following useful result. For simplicity in stating the result we shall introduce some additional notation:

$$\begin{aligned}\Delta_{k-1} &= x_{k-1}^{(1)} - x_{k-1}^{(2)} ; \\ \Delta_T &= a - b ; \\ T_2 &= (N - k + 1)h .\end{aligned}$$

Lemma 6.3.2 (a more useful form of the non-crossing point pdf) *Let*

$$X \sim w_{0,(a,b)}^{T,(a,b)} \Big|_{E_0, X(t_{k-1})=x_{k-1}} .$$

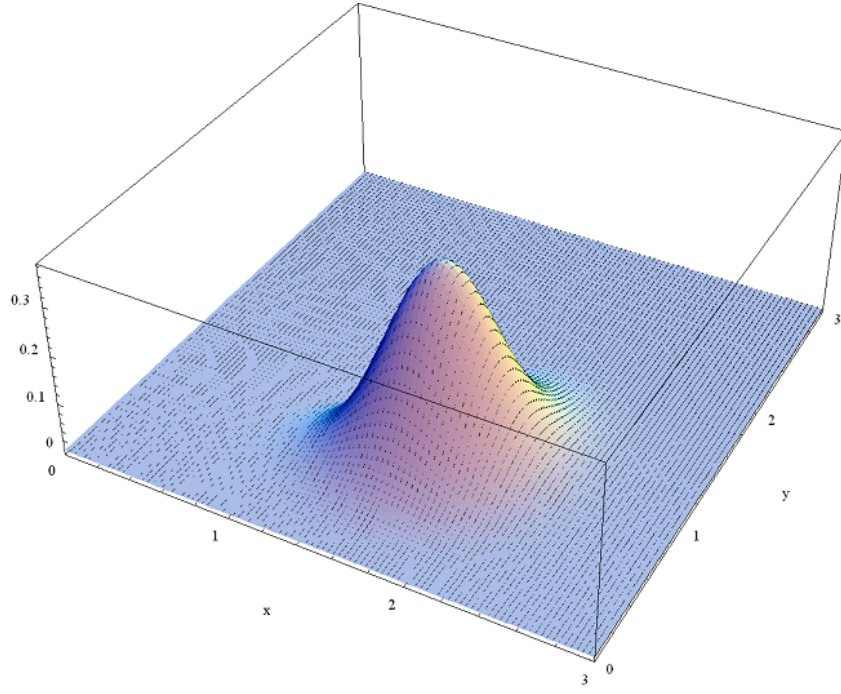


Figure 6.3: unnormalized pdf for $X(t)$ when $X \sim w_{t_{k-1}, x_{k-1}}^{T, (a, b)} \Big|_{E_2}$.

Then $\frac{X^{(1)}(t_k) + X^{(2)}(t_k)}{\sqrt{2}}$ and $\frac{X^{(1)}(t_k) - X^{(2)}(t_k)}{\sqrt{2}}$ are independent and

$$\frac{X^{(1)}(t_k) + X^{(2)}(t_k)}{\sqrt{2}} \sim \text{Gaussian} \left(\frac{(T_2 - h) \left(x_{k-1}^{(1)} + x_{k-1}^{(2)} \right) + h(a + b)}{\sqrt{2}T_2}, \frac{h(T_2 - h)}{T_2} \right) .$$

The pdf for $\frac{X^{(1)}(t_k) - X^{(2)}(t_k)}{\sqrt{2}}$ is, for $y > 0$,

$$\begin{aligned} f(y) = C^{-1} & \left(\exp \left(-\frac{\left(y - \frac{(T_2 - h)\Delta_{k-1} + h\Delta_T}{\sqrt{2}T_2} \right)^2}{2h(T_2 - h)/T_2} \right) + \exp \left(-\frac{\left(y + \frac{(T - h)\Delta_{k-1} + h\Delta_T}{\sqrt{2}T_2} \right)^2}{2h(T_2 - h)/T_2} \right) \right. \\ & - \exp \left(\frac{-\Delta_{k-1} \Delta_T}{T_2} \right) \left(\exp \left(-\frac{\left(y - \frac{(T - h)\Delta_{k-1} - h\Delta_T}{\sqrt{2}T_2} \right)^2}{2h(T_2 - h)/T_2} \right) \right. \\ & \left. \left. + \exp \left(-\frac{\left(y + \frac{(T - h)\Delta_{k-1} - h\Delta_T}{\sqrt{2}T_2} \right)^2}{2h(T_2 - h)/T_2} \right) \right) \right) , \quad (6.29) \end{aligned}$$

with

$$C = \left(1 - \exp \left(\frac{-\Delta_{k-1} \Delta_T}{T_2} \right) \right) \sqrt{2\pi h(T_2 - h)/T_2} .$$

Proof. First we perform some standard manipulations on our pdf formula, breaking it down into a sum of Gaussian pdfs. Expanding and applying the equalities $(N - k + 1)x_\mu^{(1)} =$

$$(N - k)x_{k-1}^{(1)} + a, (N - k + 1)x_{\mu}^{(2)} = (N - k)x_{k-1}^{(2)} + b,$$

$$\begin{aligned} f(x^{(1)}, x^{(2)}) \propto & \exp \left(-\frac{\left(x^{(1)} - x_{\mu}^{(1)}\right)^2 + \left(x^{(2)} - x_{\mu}^{(2)}\right)^2}{2(N - k)h/(N - k + 1)} \right) \\ & - \exp \left(-\left(\frac{\left(x^{(1)} - x^{(2)}\right)(a - b)}{(N - k)h} + \frac{\left(x^{(1)} - x_{\mu}^{(1)}\right)^2 + \left(x^{(2)} - x_{\mu}^{(2)}\right)^2}{2(N - k)h/(N - k + 1)} \right) \right) + \\ & - \exp \left(-\left(\frac{\left(x_{k-1}^{(1)} - x_{k-1}^{(2)}\right)(x^{(1)} - x^{(2)})}{h} + \frac{\left(x^{(1)} - x_{\mu}^{(1)}\right)^2 + \left(x^{(2)} - x_{\mu}^{(2)}\right)^2}{2(N - k)h/(N - k + 1)} \right) \right) + \\ & + \exp \left(-\left(\frac{(N - k + 1)\left(x_{\mu}^{(1)} - x_{\mu}^{(2)}\right)(x^{(1)} - x^{(2)})}{(N - k)h} + \frac{\left(x^{(1)} - x_{\mu}^{(1)}\right)^2 + \left(x^{(2)} - x_{\mu}^{(2)}\right)^2}{2(N - k)h/(N - k + 1)} \right) \right). \end{aligned} \quad (6.30)$$

However, notice that

$$\begin{aligned} & \frac{\left(x^{(1)} - x^{(2)}\right)(a - b)}{(N - k)h} + \frac{\left(x^{(1)} - x_{\mu}^{(1)}\right)^2 + \left(x^{(2)} - x_{\mu}^{(2)}\right)^2}{2(N - k)h/(N - k + 1)} = \\ & \quad \frac{\left(x^{(1)} - x_{\mu}^{(1)} + \frac{a-b}{N-k+1}\right)^2 + \left(x^{(2)} - x_{\mu}^{(2)} - \frac{a-b}{N-k+1}\right)^2}{2(N - k)h/(N - k + 1)} + \\ & \quad \frac{2\frac{a-b}{N-k+1}\left(x_{\mu}^{(1)} - x_{\mu}^{(2)}\right) - 2\left(\frac{a-b}{N-k+1}\right)^2}{2(N - k)h/(N - k + 1)} \\ & \frac{\left(x_{k-1}^{(1)} - x_{k-1}^{(2)}\right)(x^{(1)} - x^{(2)})}{h} + \frac{\left(x^{(1)} - x_{\mu}^{(1)}\right)^2 + \left(x^{(2)} - x_{\mu}^{(2)}\right)^2}{2(N - k)h/(N - k + 1)} = \\ & \quad \frac{\left(x^{(1)} - x_{\mu}^{(1)} + \frac{(N-k)(x_{k-1}^{(1)} - x_{k-1}^{(2)})}{N-k+1}\right)^2 + \left(x^{(2)} - x_{\mu}^{(2)} - \frac{(N-k)(x_{k-1}^{(1)} - x_{k-1}^{(2)})}{N-k+1}\right)^2}{2(N - k)h/(N - k + 1)} + \\ & \quad \frac{2\frac{(N-k)(x_{k-1}^{(1)} - x_{k-1}^{(2)})}{N-k+1}\left(x_{\mu}^{(1)} - x_{\mu}^{(2)}\right) - 2\left(\frac{(N-k)(x_{k-1}^{(1)} - x_{k-1}^{(2)})}{N-k+1}\right)^2}{2(N - k)h/(N - k + 1)} \\ & \frac{\left(x_{\mu}^{(1)} - x_{\mu}^{(2)}\right)(x^{(1)} - x^{(2)})}{(N - k)h/(N - k + 1)} + \frac{\left(x^{(1)} - x_{\mu}^{(1)}\right)^2 + \left(x^{(2)} - x_{\mu}^{(2)}\right)^2}{2(N - k)h/(N - k + 1)} = \frac{\left(x^{(1)} - x_{\mu}^{(2)}\right)^2 + \left(x^{(2)} - x_{\mu}^{(1)}\right)^2}{2(N - k)h/(N - k + 1)}. \end{aligned}$$

Moreover, in fact we have

$$\begin{aligned} x_\mu^{(1)} - \frac{a-b}{N-k+1} &= \frac{b + (N-k)x_{k-1}^{(1)}}{N-k+1} = x_\mu^{(2)} + \frac{(N-k)(x_{k-1}^{(1)} - x_{k-1}^{(2)})}{N-k+1} \\ x_\mu^{(2)} + \frac{a-b}{N-k+1} &= \frac{a + (N-k)x_{k-1}^{(2)}}{N-k+1} = x_\mu^{(1)} - \frac{(N-k)(x_{k-1}^{(1)} - x_{k-1}^{(2)})}{N-k+1} \end{aligned}$$

so that we may rewrite (6.30) as

$$\begin{aligned} f(x^{(1)}, x^{(2)}) &\propto \exp \left(-\frac{(x^{(1)} - x_\mu^{(1)})^2 + (x^{(2)} - x_\mu^{(2)})^2}{2(N-k)h/(N-k+1)} \right) \\ &+ \exp \left(-\frac{(x^{(1)} - x_\mu^{(2)})^2 + (x^{(2)} - x_\mu^{(1)})^2}{2(N-k)h/(N-k+1)} \right) \\ &- \exp \left(-\frac{(x_{k-1}^{(1)} - y_{k-1}^{(2)})(a-b)}{h(N-k+1)} \right) \times \\ &\left(\exp \left(-\frac{(x^{(1)} - x_\mu^{(1)} + \frac{a-b}{N-k+1})^2 + (x^{(2)} - x_\mu^{(2)} - \frac{a-b}{N-k+1})^2}{2(N-k)h/(N-k+1)} \right) + \right. \\ &\left. \exp \left(-\frac{(x^{(1)} - x_\mu^{(2)} - \frac{a-b}{N-k+1})^2 + (x^{(2)} - x_\mu^{(1)} + \frac{a-b}{N-k+1})^2}{2(N-k)h/(N-k+1)} \right) \right). \end{aligned}$$

We are now ready to rotate the basis. Write $y = \frac{x^{(1)} - x^{(2)}}{\sqrt{2}}$ and $x' = \frac{x^{(1)} + x^{(2)}}{\sqrt{2}}$. Notice that for any $\kappa_{1,2} \in \mathbb{R}$:

$$(x^{(1)} - \kappa_1)^2 + (x^{(2)} - \kappa_2)^2 = \left(x' - \frac{\kappa_1 + \kappa_2}{\sqrt{2}} \right)^2 + \left(y - \frac{\kappa_1 - \kappa_2}{\sqrt{2}} \right)^2 \quad (6.31)$$

and therefore

$$\begin{aligned} (x^{(1)} - x_\mu^{(1)})^2 + (x^{(2)} - x_\mu^{(2)})^2 &= \left(x' - \frac{x_\mu^{(1)} + x_\mu^{(2)}}{\sqrt{2}} \right)^2 + \left(y - \frac{x_\mu^{(1)} - x_\mu^{(2)}}{\sqrt{2}} \right)^2 \\ \left(x^{(1)} - x_\mu^{(1)} + \frac{a-b}{N-k+1} \right)^2 + \left(x^{(2)} - x_\mu^{(2)} - \frac{a-b}{N-k+1} \right)^2 \\ &= \left(x' - \frac{x_\mu^{(1)} + x_\mu^{(2)}}{\sqrt{2}} \right)^2 + \left(y - \frac{x_\mu^{(1)} - x_\mu^{(2)} - 2\frac{a-b}{N-k+1}}{\sqrt{2}} \right)^2 \end{aligned}$$

$$\begin{aligned}
& \left(x^{(1)} - x_\mu^{(2)} - \frac{a-b}{N-k+1} \right)^2 + \left(x^{(2)} - x_\mu^{(1)} + \frac{a-b}{N-k+1} \right)^2 \\
&= \left(x' - \frac{x_\mu^{(1)} + x_\mu^{(2)}}{\sqrt{2}} \right)^2 + \left(y - \frac{x_\mu^{(2)} - x_\mu^{(1)} + 2\frac{a-b}{N-k+1}}{\sqrt{2}} \right)^2 \\
& (x^{(1)} - x_\mu^{(2)})^2 + (x^{(2)} - x_\mu^{(1)})^2 = \left(x' - \frac{x_\mu^{(1)} + x_\mu^{(2)}}{\sqrt{2}} \right)^2 + \left(y - \frac{x_\mu^{(2)} - x_\mu^{(1)}}{\sqrt{2}} \right)^2
\end{aligned}$$

The absolute value of the determinant of the Jacobian matrix for the change of coordinates from $(x^{(1)}, x^{(2)})$ to (x', y) is 1. Thus the joint pdf for $\left(\frac{X^{(1)}(t_k) + X^{(2)}(t_k)}{\sqrt{2}}, \frac{X^{(1)}(t_k) - X^{(2)}(t_k)}{\sqrt{2}} \right)$ is given by

$$g(x', y) \propto \exp \left(- \frac{\left(x' - \frac{x_\mu^{(1)} + x_\mu^{(2)}}{\sqrt{2}} \right)^2}{2(T_2 - h)h/T_2} \right) f(y)$$

where for $y > 0$,

$$\begin{aligned}
f(y) \propto & \exp \frac{-\left(y - \frac{x_\mu^{(1)} - x_\mu^{(2)}}{\sqrt{2}} \right)^2}{2(T_2 - h)h/T_2} + \exp \frac{-\left(y - \frac{x_\mu^{(2)} - x_\mu^{(1)}}{\sqrt{2}} \right)^2}{2(T_2 - h)h/T_2} - \\
& \exp \left(\frac{-\left(x_{k-1}^{(1)} - y_{k-1}^{(2)} \right) (a-b)}{h(N-k+1)} \right) \times \\
& \left(\exp \frac{-\left(y - \frac{x_\mu^{(1)} - x_\mu^{(2)} - 2\frac{a-b}{N-k+1}}{\sqrt{2}} \right)^2}{2(T_2 - h)h/T_2} + \exp \frac{-\left(y - \frac{x_\mu^{(2)} - x_\mu^{(1)} + 2\frac{a-b}{N-k+1}}{\sqrt{2}} \right)^2}{2(T_2 - h)h/T_2} \right).
\end{aligned}$$

To get the result as stated, we rewrite this pdf for $\frac{X^{(1)}(t_k) - X^{(2)}(t_k)}{\sqrt{2}}$ using the Δ_0, Δ_T notation, and find the normalizing constant:

$$C = \frac{1}{2} \left(2\sqrt{2\pi h(T_2 - h)/T_2} - 2 \exp \left(\frac{-\Delta_0 \Delta_T}{T_2} \right) \sqrt{2\pi h(T_2 - h)/T_2} \right) \quad (6.32)$$

$$= \left(1 - \exp \left(\frac{-\Delta_0 \Delta_T}{T_2} \right) \right) \sqrt{2\pi h(T_2 - h)/T_2} \quad (6.33)$$

Here the $\frac{1}{2}$ occurs because of the cutoff at zero, since f would otherwise be symmetrical. ■

Let ν denote the measure on \mathbb{R} induced by the distribution of $\frac{X^{(1)}(t_k) - X^{(2)}(t_k)}{\sqrt{2}}$. If for further

brevity we set

$$\begin{aligned} K &: = \exp\left(\frac{-\Delta_0 \Delta_T}{T_2}\right) \\ \sigma^2 &: = h(T_2 - h)/T_2 \\ y_{\nu 1} &= \frac{(T - h) \Delta_0 + h \Delta_T}{\sqrt{2}T} \\ y_{\nu 2} &= \frac{(T - h) \Delta_0 - h \Delta_T}{\sqrt{2}T} \end{aligned}$$

then the pdf for $\frac{X^{(1)}(t_k) - X^{(2)}(t_k)}{\sqrt{2}}$ is

$$\begin{aligned} f(y) = \frac{1}{(1 - K)\sqrt{2\pi}\sigma} &\left(\exp\left(-\frac{(y - y_{\nu 1})^2}{2\sigma^2}\right) + \exp\left(-\frac{(y + y_{\nu 1})^2}{2\sigma^2}\right) \right. \\ &\left. - K \left(\exp\left(-\frac{(y - y_{\nu 2})^2}{2\sigma^2}\right) + \exp\left(-\frac{(y + y_{\nu 2})^2}{2\sigma^2}\right) \right) \right) \quad (6.34) \end{aligned}$$

Thanks to Lemma 6.3.2, we are faced with sampling from a 1-dimensional distribution. It is possible to create decomposition and rejection methods for multi-dimensional distributions, but this 1-dimensionality makes it easier, and means that we could also apply the universal method. It is an experimental observation that when we constrain $y > 0$, this distribution is unipolar. However this has yet to be proven and seems to be analytically intractable. It would be even more expedient to know the peak, but solving this problem appears to be even more intractable, at least by trying to directly solve $f'(y_*) = 0$.

Therefore in order construct simulations, it is helpful to at least know the mean and variance of $\frac{X^{(1)}(t_k) - X^{(2)}(t_k)}{\sqrt{2}}$, and these are given by the following result.

Proposition 6.3.3 (mean and st.dev. of distance coordinate) *Let $X \sim w_{0,(a,b)}^{T,(a,b)} \Big|_{E_0, X(t_{k-1})=x_{k-1}}$.*

Then $\frac{X^{(1)}(t_k) - X^{(2)}(t_k)}{\sqrt{2}}$ has mean and variance

$$\begin{aligned} \mathbb{E} \frac{X^{(1)}(t_k) - X^{(2)}(t_k)}{\sqrt{2}} &= \frac{1}{1 - K} \left(\frac{\sqrt{2}}{\sqrt{\pi}} \sigma \exp\left(\frac{-y_{\nu 1}^2}{2\sigma^2}\right) + y_{\nu 1} (\Phi(y_{\nu 1}/\sigma) - \Phi(-y_{\nu 1}/\sigma)) \right. \\ &\quad \left. - K \left(\frac{\sqrt{2}}{\sqrt{\pi}} \sigma \exp\left(\frac{-y_{\nu 2}^2}{2\sigma^2}\right) + y_{\nu 2} (\Phi(y_{\nu 2}/\sigma) - \Phi(-y_{\nu 2}/\sigma)) \right) \right) \\ \text{Var} \left(\frac{X^{(1)}(t_k) - X^{(2)}(t_k)}{\sqrt{2}} \right) &= \frac{1}{1 - K} (y_{\nu 1}^2 + \sigma^2 - K (y_{\nu 2}^2 + \sigma^2)) - \left(\mathbb{E} \frac{X^{(1)}(t_k) - X^{(2)}(t_k)}{\sqrt{2}} \right)^2 \end{aligned}$$

Proof. We perform an integration by parts: for $\alpha \in \mathbb{R}$,

$$\begin{aligned} \frac{d}{dy} \left(-\sigma^2 \exp \left(\frac{-(y-\alpha)^2}{2\sigma^2} \right) \right) &= (y-\alpha) \exp \left(\frac{-(y-\alpha)^2}{2\sigma^2} \right) \Rightarrow \\ \int_0^\infty y \exp \left(\frac{-(y-\alpha)^2}{2\sigma^2} \right) dy &= \left[-\sigma^2 \exp \left(\frac{-(y-\alpha)^2}{2\sigma^2} \right) \right]_0^\infty + \alpha \int_0^\infty \exp \left(\frac{-(y-\alpha)^2}{2\sigma^2} \right) dy \\ &= \sigma^2 \exp \left(\frac{-\alpha^2}{2\sigma^2} \right) + \alpha \sqrt{2\pi} \sigma \Phi(\alpha/\sigma) \end{aligned}$$

where $\Phi(\alpha/\sigma)$ denotes the probability that a standard normal variable would lie below α/σ .

Therefore

$$\begin{aligned} Ey &= C^{-1} \int_0^\infty y f(y) dy = \frac{1}{(1-K)\sqrt{2\pi}\sigma} \left(\sigma^2 \exp \left(\frac{-y_{\nu 1}^2}{2\sigma^2} \right) + y_{\nu 1} \sqrt{2\pi} \sigma \Phi(y_{\nu 1}/\sigma) \right. \\ &\quad + \sigma^2 \exp \left(\frac{-(-y_{\nu 1})^2}{2\sigma^2} \right) - y_{\nu 1} \sqrt{2\pi} \sigma \Phi(-y_{\nu 1}/\sigma) - K \sigma^2 \exp \left(\frac{-y_{\nu 2}^2}{2\sigma^2} \right) + y_{\nu 2} \sqrt{2\pi} \sigma \Phi(y_{\nu 2}/\sigma) \\ &\quad \left. + K \sigma^2 \exp \left(\frac{-(-y_{\nu 2})^2}{2\sigma^2} \right) - y_{\nu 2} \sqrt{2\pi} \sigma \Phi(-y_{\nu 2}/\sigma) \right) \\ &= \frac{1}{1-K} \left(\frac{\sqrt{2}}{\sqrt{\pi}} \sigma \exp \left(\frac{-y_{\nu 1}^2}{2\sigma^2} \right) + y_{\nu 1} (\Phi(y_{\nu 1}/\sigma) - \Phi(-y_{\nu 1}/\sigma)) \right. \\ &\quad \left. - K \left(\frac{\sqrt{2}}{\sqrt{\pi}} \sigma \exp \left(\frac{-y_{\nu 2}^2}{2\sigma^2} \right) + y_{\nu 2} (\Phi(y_{\nu 2}/\sigma) - \Phi(-y_{\nu 2}/\sigma)) \right) \right) \end{aligned}$$

as required. Meanwhile, we can find $\int_0^\infty y^2 f(y) dy$. Notice that

$$\begin{aligned} \frac{d}{dy} \left(-\sigma^2 y \exp \left(\frac{-(y-\alpha)^2}{2\sigma^2} \right) \right) &= y(y-\alpha) \exp \left(\frac{-(y-\alpha)^2}{2\sigma^2} \right) - \sigma^2 \left(\frac{-(y-\alpha)^2}{2\sigma^2} \right) \Rightarrow \\ \int_0^\infty y^2 \exp \left(\frac{-(y-\alpha)^2}{2\sigma^2} \right) dy &= \left[-\sigma^2 y \exp \left(\frac{-(y-\alpha)^2}{2\sigma^2} \right) \right]_0^\infty \\ &\quad + \alpha \int_0^\infty y \exp \left(\frac{-(y-\alpha)^2}{2\sigma^2} \right) dy + \sigma^2 \int_0^\infty \exp \left(\frac{-(y-\alpha)^2}{2\sigma^2} \right) dy \\ &= 0 + \alpha \sigma^2 \exp \left(\frac{-\alpha^2}{2\sigma^2} \right) + (\alpha^2 + \sigma^2) \sigma \sqrt{2\pi} \Phi(\alpha/\sigma) \end{aligned}$$

Hence

$$\begin{aligned}
\int_0^\infty y^2 f(y) dy &= \frac{1}{1-K} \left(\frac{y_{\nu 1} \sigma}{\sqrt{2\pi}} \exp\left(\frac{-y_{\nu 1}^2}{2\sigma^2}\right) + (y_{\nu 1}^2 + \sigma^2) \Phi(y_{\nu 1}/\sigma) \right. \\
&\quad - \frac{y_{\nu 1} \sigma}{\sqrt{2\pi}} \exp\left(\frac{-(-y_{\nu 1})^2}{2\sigma^2}\right) + ((-y_{\nu 1})^2 + \sigma^2) \Phi(-y_{\nu 1}/\sigma) \\
&\quad - K \left(\frac{y_{\nu 2} \sigma}{\sqrt{2\pi}} \exp\left(\frac{-y_{\nu 2}^2}{2\sigma^2}\right) + (y_{\nu 2}^2 + \sigma^2) \Phi(y_{\nu 2}/\sigma) \right. \\
&\quad \left. \left. - \frac{y_{\nu 2} \sigma}{\sqrt{2\pi}} \exp\left(\frac{-(-y_{\nu 2})^2}{2\sigma^2}\right) + ((-y_{\nu 2})^2 + \sigma^2) \Phi(-y_{\nu 2}/\sigma) \right) \right) \\
&= \frac{1}{1-K} (y_{\nu 1}^2 + \sigma^2 - K (y_{\nu 2}^2 + \sigma^2))
\end{aligned}$$

and the result follows. ■

6.3.2 Drawing a point on a non-crossing Brownian bridge

In view of the above alternative restatements of the joint pdf for a point $X(t_k)$ when $X \sim w_{0,(a,b)}^{T,(a,b)} \Big|_{E_0, X(t_{k-1})=x_{k-1}}$ there are several ways we can approach the problem of sampling from the distribution of $X(t_k)$. The simplest is to draw disregarding the conditioning on E_0 and then reject according to the probability of a crossing during (t_{k-1}, T) . Alternatively, we could make use of Lemma 6.3.2 to avoid having to take multidimensional samples at all, and instead be left with a 1-dimensional sampling problem, ie sampling the distance coordinate. For this case, two well-known 1D sampling methods were compared: a Decomposition method, and a Rejection method (cf [Gen98]). Alternatively one might use the Universal method (see [MT04, Gen98]) but an approximate solution for inversion of the cdf would be needed, and this complication was considered to be best avoided. It was verified that a transformation method (cf [Wil01, p.249]) based on transforming a Gaussian draw is not possible. Clearly, in principle there could be other readily sampled distributions that it might be possible to transform and attain the desired pdf, but no such object was discovered. The details of each attempted sampling method now follow.

Naïve Rejection method

As mentioned, in view of (6.24), we can draw a value x for $X(t_k)$ if we draw from the Gaussian which is induced by the conditional Wiener measure and then reject with probability given by (cf 6.25):

$$1 - \left(1 - \exp \left(- \frac{(x^{(1)} - x^{(2)}) (x_{k-1}^{(1)} - x_{k-1}^{(2)})}{h} \right) \right) \left(1 - \exp \left(- \frac{(x^{(1)} - x^{(2)}) (a - b)}{(N - k)h} \right) \right) \quad (6.35)$$

in the case that for our draw, $x^{(1)} > x^{(2)}$, and with certain rejection otherwise. In the case that we draw $x^{(1)} < x^{(2)}$, we cannot reflect the draw to save on draws, since although the Gaussian candidate measure is symmetric, the centre is not on the line $x = y$.

The greatest virtue of this method is its simplicity, but it is not possible to combine it with just a Gaussian measure for the initial points (a, b) . In experiments of this kind, the situation was soon encountered, during the first or second step, that points are located too close together so that the acceptance probability for the next point is persistently less than 10^{-10} . It is likely that some improvement may be found if a more favourable initial point distribution is used. In this case, however, one has already reduced simplicity somewhat. It is notable, however, that when a Markov Chain Monte Carlo approach is adopted, initial points are no longer distinguished from the rest. This should make the naïve rejection method viable without necessitating severe complications, since the part of discretisation space where crossings are almost inevitable should then be accessed commensurately infrequently.

Reorientation + Decomposition

Since we established that a linear transformation of coordinates makes the new coordinates independent, we could proceed by taking a Gaussian draw for $\frac{X^{(1)}(t_k) + X^{(2)}(t_k)}{\sqrt{2}}$ (since efficient methods are known [MT04] for simulating a Gaussian distribution) and then, we are left with the need to sample $\frac{X^{(1)}(t_k) - X^{(2)}(t_k)}{\sqrt{2}}$. That is, as mentioned, we then need to sample from the 1-dimensional distribution corresponding to the pdf (6.29). One generic exact method for 1-dimensional sampling is volume decomposition.

Volume decomposition depends on the assumption that the target distribution is unipolar, or at least that its modes are all approximately known. It means, in principle, that we rewrite the pdf as the infinite linear combination of drawable pdfs, with positive coefficients. (We shall loosely call a pdf "drawable" when an efficient sampling algorithm is known.) The choice of basis functions is arbitrary as long as they are drawable pdfs. However, for simplicity let us choose each to be the uniform distribution over an interval. Then we are decomposing the area under the curve into boxes: for some set of $P_i > 0$, and some set of x_{i1}, x_{i2} ,

$$f = \sum_{i=1}^{\infty} P_i \frac{I_{[x_{i1}, x_{i2}]}}{x_{i2} - x_{i1}}. \quad (6.36)$$

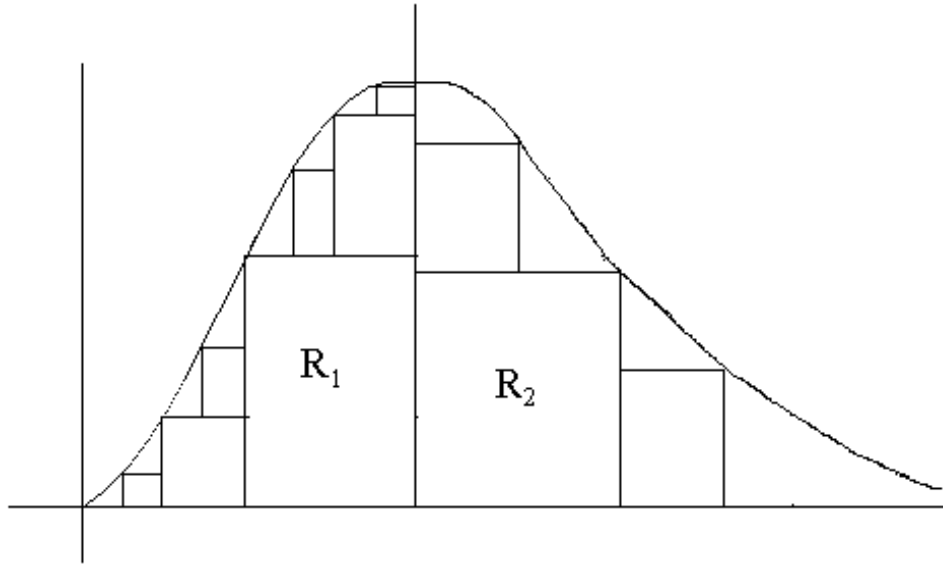


Figure 6.4: Decomposing the volume of probability under the pdf curve

The most conceptually obvious algorithm would seem to be as follows. Let f be the pdf; we assume the distribution is unipolar.

1. Find the maximum likelihood peak of f ; call this y_{ML} . Let $\kappa_0 \in (0, y_{ML})$.
2. Form a subset of the area under the curve by taking $(y_{ML} - \kappa_0, y_{ML}) \times (0, f(y_{ML} - \kappa_0))$.

We shall call this R_1 . If we are trying to sample the area beneath the curve uniformly (cf [Wil01, p.249]) then we should attribute to this $P_1 = f(y_{ML} - \kappa_0)\kappa_0$.

3. Do likewise on the right-hand side of y_0 : take $R_2 = (y_{ML}, y_{ML} + \kappa_0) \times (0, f(y_{ML} + \kappa_0))$, and attribute to it $P_2 = f(y_{ML} + \kappa_0)\kappa_0$.
4. At the leftmost point so far encountered (call y_{LEFT}), create the next $R_i = (y_{LEFT} - \kappa_i, y_{LEFT}) \times (0, f(y_{LEFT} - \kappa_i))$. Or, if $y_{LEFT} - \kappa_i < 0$ then $R_i = \emptyset$.
5. Create a further set of R_i by bisecting existing intervals.
6. At the rightmost point encountered so far (call y_{RIGHT}), create the next $R_i = (y_{RIGHT} + \kappa_i, y_{RIGHT}) \times (0, f(y_{RIGHT} + \kappa_i))$.
7. Return to step 4; or, if 0 is the current leftmost point then return to step 5. Thus we shall obtain values for P_i and x_{i1}, x_{i2} in (6.36).
8. We make a uniform draw $u \sim U[0, 1]$ and this can be used to index both which R_i we lie in and what point to draw from the corresponding interval. For suppose that

$$\sum_{i=1}^{k-1} P_i \leq u < \sum_{i=1}^k P_i.$$

Then our draw is

$$x_{k1} + \left(u - \sum_{i=1}^{k-1} P_i \right) (x_{k2} - x_{k1}).$$

However, in practice a couple of modifications to this schema become necessary. Firstly, it is notable that once u is known, we will know when we have computed k coefficients such that $u < \sum_{i=1}^k P_i$ and we can stop iterating. Obviously to compute an infinite number of terms would take an infinite number of evaluations of f so this is just as well. So it makes sense to draw u first and then check whether $u < \sum_{i=1}^k P_i$ after each (k th) box.

Secondly, in the case of distributions such as that of the distance coordinate for a point on a non-crossing bridge, it is not apparently possible to identify the maximum likelihood point analytically and so we cannot rely on knowing the sign of the derivative. However, as long as the function is unipolar we can still use the fact that for any interval $[x_1, x_2]$, $\min_{[x_1, x_2]} f = \min_{\{x_1, x_2\}} f$. The approximation to the curve after n iterations will look like a set of

intervals demarcated by j_n points where $j_{n+1} = 2j_n + 1$ and $j_1 = 2$ (so that $j_n = 3 * 2^{n-1} - 1$) where the height of the approximation curve in interval $[x_1^*, x_2^*]$ is $\min_{\{x_1^*, x_2^*\}} f$.

In practice, a limitation was encountered, that the maximum number of iterations is bounded by the amount of computer RAM. viz, to run the 25th iteration with the 24th in storage, will probably require about 600 Mb storage if there are 4 bytes in a double precision number, and this amount roughly doubles at each iteration. The solution used was to re-start the procedure (with a new draw of u) if $u > \sum_{i=1}^{j_{24}} P_i$, introducing a small bias. This means that, in effect, the pdf actually being drawn from is a step function approximation to the intended pdf, with j_{24} steps and curtailed at the rightmost point. As it happens, $j_{24} = 25165823$.

For this reason, where σ is the standard deviation of $\frac{X^{(1)}(t_k) - X^{(2)}(t_k)}{\sqrt{2}}$, the distance $\sigma/q^{1/2}$ was chosen for κ in determining the new leftmost and rightmost points at iteration q , rather than say σ/q , since using $q^{1/2}$ means that at the 24th iteration, we can reach 8 s.d. on the right-hand side. A rough numerical analysis indicates the probability of the neglected tail is of the magnitude 10^{-16} . As a conservative estimate, it should therefore be safe to consider the procedure unbiased if the number of samples used in a program is less than about 10^{17} .

Although we cannot find the peak, it is still necessary to pick two sensible initial points to be the endpoints of the interval in iteration 1. Since we can compute the mean and standard deviation of the distributions that we are concerned with, the points used in simulations were $E \frac{X^{(1)}(t_k) - X^{(2)}(t_k)}{\sqrt{2}} - \sigma, E \frac{X^{(1)}(t_k) - X^{(2)}(t_k)}{\sqrt{2}} + \sigma$. The conclusion from the simulations using this method was that it demonstrates one way in which an effective simulation can be achieved, but is unnecessarily expensive. The simulations are 10 times more costly than with the following method.

Reorientation + Rejection

Again relying on Lemma 6.3.2, we may sample from the distribution given by (6.29) using Rejection sampling, another generic exact method. Specifically, we draw from a drawable

"candidate" pdf, and then reject with a probability given by the ratio of the two pdfs, multiplied by some $\alpha^{-1} < 1$. This effectively means that there is a candidate measure μ with full measure α and $\nu \ll \mu$, and our acceptance probability is $d\nu/d\mu$, the Radon-Nikodym derivative between the target measure and the candidate measure. This method, amongst others, is discussed in [Gen98].

There are two popular approaches to Rejection sampling: the ziggurat method and adaptive rejection. When the modes of the distribution are assumed known, the ziggurat method creates a "ziggurat" candidate (ie, a step function everywhere greater than the target pdf f) using a finite initial set of evaluations of f . The adaptive rejection method improves the ziggurat by updating it each time the target pdf is called. Neither approach is applicable here because parameters are different every time that a draw is made, and (6.29) is not such that an easy transformation will turn a draw for one set of parameters into a draw for another set of parameters.

If $d\nu/d\mu > 1$ at some points, then upon sampling such a point, we could apply a reweighting to the contributions to the functional integral. However, this rapidly becomes complicated. A method based on this idea was implemented and gave poor results, so it is considered imperative for there to be no such reweighting. Therefore it is necessary to use a candidate measure whose Radon-Nikodym derivative with respect to Lebesgue (that is, the pdf of the distribution which induces this candidate measure, if it were to be a probability measure) is everywhere greater than or equal to that of the target. Or at least, we need to know that the problematic samples with $d\nu/d\mu > 1$ will occur with a sufficiently low frequency, such as 10^{-15} , that their influence can be disregarded in our simulation.

Various candidate measures were considered and most attempts encountered significant problems. The measures tested were based on the exponential distribution, the hyperbolic distribution, the gamma distribution, and these distributions spliced with the Gaussian, and with each other, horizontally and vertically. (Sampling from a horizontal splicing of pdfs can be accomplished by knowing the relative weight intended for the component parts and

choosing at the beginning from which of the parts we are to sample. Sampling from a vertical splicing means sampling from a linear combination of pdfs, which can be similarly achieved.). In principle, any probability measure μ' with $\nu \ll \mu'$ can be boosted by some α to the point where $\mu(\{d\nu/d\mu > 1\})$ is sufficiently small. It is not clear how we could use any other kind of boost than to multiply by a constant. Therefore the relevant property of a candidate measure is, what α is needed in order for it to be evident that this point has been reached. For most of the measures tried it was clear that this α would be of an unacceptably large magnitude (since α^{-1} must of course be the acceptance ratio).

It turns out that an effective candidate can be produced using a splice of two Gaussians. Nothing useful comes of splicing about the mean. Rather, if we assume that the maximum point of the target is known, it appears that the two parts of the target to either side of it can then be modelled, separately, as Gaussian. (Under changing parameters, the mean strays a variable distance from the maximum point and so a method based on splicing around the mean will fail for some parameters.) However, it was found that estimating y_{ML} by using a quadratic approximation to the target from a Taylor series about the mean seems to give reasonably accurate results. That is, experimentally, where we write $E\Delta$ for $E[X^{(1)}(t_k) - X^{(2)}(t_k)]$,

$$y_{ML} \approx \widehat{y_{ML}} = -\sqrt{\left(\frac{f_2}{f_3}\right)^2 - 2\frac{f_1}{f_3} - \frac{f_2}{f_3} + \frac{E\Delta}{\sqrt{2}}} \quad (6.37)$$

where, recalling the notation defined following Lemma 6.3.2,

$$f_1 = \frac{1}{C} \left(-\frac{\frac{E\Delta}{\sqrt{2}} - y_{\nu 1}}{\sigma^2} \exp\left(-\frac{\left(\frac{E\Delta}{\sqrt{2}} - y_{\nu 1}\right)^2}{2\sigma^2}\right) - \frac{\frac{E\Delta}{\sqrt{2}} + y_{\nu 1}}{\sigma^2} \exp\left(-\frac{\left(\frac{E\Delta}{\sqrt{2}} + y_{\nu 1}\right)^2}{2\sigma^2}\right) \right. \\ \left. + K \left(\frac{\frac{E\Delta}{\sqrt{2}} - y_{\nu 2}}{\sigma^2} \exp\left(-\frac{\left(\frac{E\Delta}{\sqrt{2}} - y_{\nu 2}\right)^2}{2\sigma^2}\right) + \frac{\frac{E\Delta}{\sqrt{2}} + y_{\nu 2}}{\sigma^2} \exp\left(-\frac{\left(\frac{E\Delta}{\sqrt{2}} + y_{\nu 2}\right)^2}{2\sigma^2}\right) \right) \right)$$

$$\begin{aligned}
f_2 = \frac{1}{C} & \left(\left(-\frac{1}{\sigma^2} + \frac{\left(\frac{E\Delta}{\sqrt{2}} - y_{\nu 1}\right)^2}{\sigma^4} \right) \exp \left(-\frac{\left(\frac{E\Delta}{\sqrt{2}} - y_{\nu 1}\right)^2}{2\sigma^2} \right) \right. \\
& + \left(-\frac{1}{\sigma^2} + \frac{\left(\frac{E\Delta}{\sqrt{2}} + y_{\nu 1}\right)^2}{\sigma^4} \right) \exp \left(-\frac{\left(\frac{E\Delta}{\sqrt{2}} + y_{\nu 1}\right)^2}{2\sigma^2} \right) \\
& + K \left(\left(\frac{1}{\sigma^2} - \frac{\left(\frac{E\Delta}{\sqrt{2}} - y_{\nu 2}\right)^2}{\sigma^4} \right) \exp \left(-\frac{\left(\frac{E\Delta}{\sqrt{2}} - y_{\nu 2}\right)^2}{2\sigma^2} \right) \right. \\
& \quad \left. \left. + \left(\frac{1}{\sigma^2} - \frac{\left(\frac{E\Delta}{\sqrt{2}} + y_{\nu 2}\right)^2}{\sigma^4} \right) \exp \left(-\frac{\left(\frac{E\Delta}{\sqrt{2}} + y_{\nu 2}\right)^2}{2\sigma^2} \right) \right) \right)
\end{aligned}$$

$$\begin{aligned}
f_3 = \frac{-3f_1}{\sigma^2} + \frac{1}{C} & \left(-\frac{\left(\frac{E\Delta}{\sqrt{2}} - y_{\nu 1}\right)^3}{\sigma^6} \exp \left(-\frac{\left(\frac{E\Delta}{\sqrt{2}} - y_{\nu 1}\right)^2}{2\sigma^2} \right) \right. \\
& - \frac{\left(\frac{E\Delta}{\sqrt{2}} + y_{\nu 1}\right)^3}{\sigma^6} \exp \left(-\frac{\left(\frac{E\Delta}{\sqrt{2}} + y_{\nu 1}\right)^2}{2\sigma^2} \right) \\
& + K \left(\frac{\left(\frac{E\Delta}{\sqrt{2}} - y_{\nu 2}\right)^3}{\sigma^6} \exp \left(-\frac{\left(\frac{E\Delta}{\sqrt{2}} - y_{\nu 2}\right)^2}{2\sigma^2} \right) \right. \\
& \quad \left. \left. + \frac{\left(\frac{E\Delta}{\sqrt{2}} + y_{\nu 2}\right)^3}{\sigma^6} \exp \left(-\frac{\left(\frac{E\Delta}{\sqrt{2}} + y_{\nu 2}\right)^2}{2\sigma^2} \right) \right) \right)
\end{aligned}$$

Obviously this approximation would give rise to some error if we were trying to model the target exactly, but it is clear visually that it is unimportant when we are just trying to create a candidate measure with full measure > 1 . (It is only the cost of this Taylor approximation that we have to worry about.)

Approximately speaking, the target is somewhat like a Gaussian about its peak with one side compressed so that it meets zero at zero, and intuition of the problem makes it plausible that this should be so. So it is safe to allow the left-hand side of the candidate to have variance $\tau^2 := Var\left(\frac{\Delta}{\sqrt{2}}\right)$, ie the same as the target's overall variance: at the very least we know that the mean square negative deviation away from y_{ML} in the target is less than

this, since it is less than the mean square positive deviation.

Now for the peaks of the candidate components to meet, as seems desirable, the "normalizing constant" on the right-hand component must be $\frac{1}{\sqrt{2\pi\tau}}$ as on the left. This normalizing constant will be relevant in the acceptance probability when we draw a point on the right-hand side. This means that choosing a higher variance than τ on the right-hand side (as we must do) implies a greater candidate measure on this side. The relative probability of the two sides is thus controlled by the variance of the right-hand side, so a natural choice might be

$$\tau_{\text{right}} = \tau \frac{\nu(y_{ML}, \infty)}{\nu(0, y_{ML})} \quad (6.38)$$

so that the candidate gives the same relative probability to the two sides of the peak as the target does. The fact that the left-hand candidate works should then mean that the right-hand candidate must also work, since the right-hand picture is then in some sense a reflection of the left. Of course, we know that

$$\begin{aligned} \nu(0, y_{ML}) = \frac{1}{1-K} & \left(\Phi\left(\frac{y_{ML} - y_{\nu 1}}{\sigma}\right) - \Phi\left(\frac{-y_{ML} - y_{\nu 1}}{\sigma}\right) \right. \\ & \left. - K \left(\Phi\left(\frac{y_{ML} - y_{\nu 2}}{\sigma}\right) - \Phi\left(\frac{-y_{ML} - y_{\nu 2}}{\sigma}\right) \right) \right) \end{aligned}$$

Figures 6.5-6.8 show example curves using the candidate Radon-Nikodym derivative:

$$\begin{aligned} & \frac{1}{\sqrt{2\pi\tau}} \exp\left(\frac{-(y - \widehat{y_{ML}})^2}{2\tau^2}\right), 0 < y \leq \widehat{y_{ML}}; \\ & \frac{1}{\sqrt{2\pi\tau}} \exp\left(\frac{-(y - \widehat{y_{ML}})^2}{2\tau^2 \left(\frac{\nu(\widehat{y_{ML}}, \infty)}{\nu(0, \widehat{y_{ML}})}\right)^2}\right), y > \widehat{y_{ML}} \end{aligned} \quad (6.39)$$

There is no extra boost to the candidate.

In the candidate which was actually used so far, the formula used for the variance of the right-hand side is instead $(3(1 - \nu(-\infty, y_{ML})) + 0.5)^2 \tau^2$. This is arbitrary, but works empirically. (In fact given the range which $\nu(-\infty, y_{ML})$ actually takes, this ranges from being a similar value to the right-hand variance above, and a somewhat greater value.)

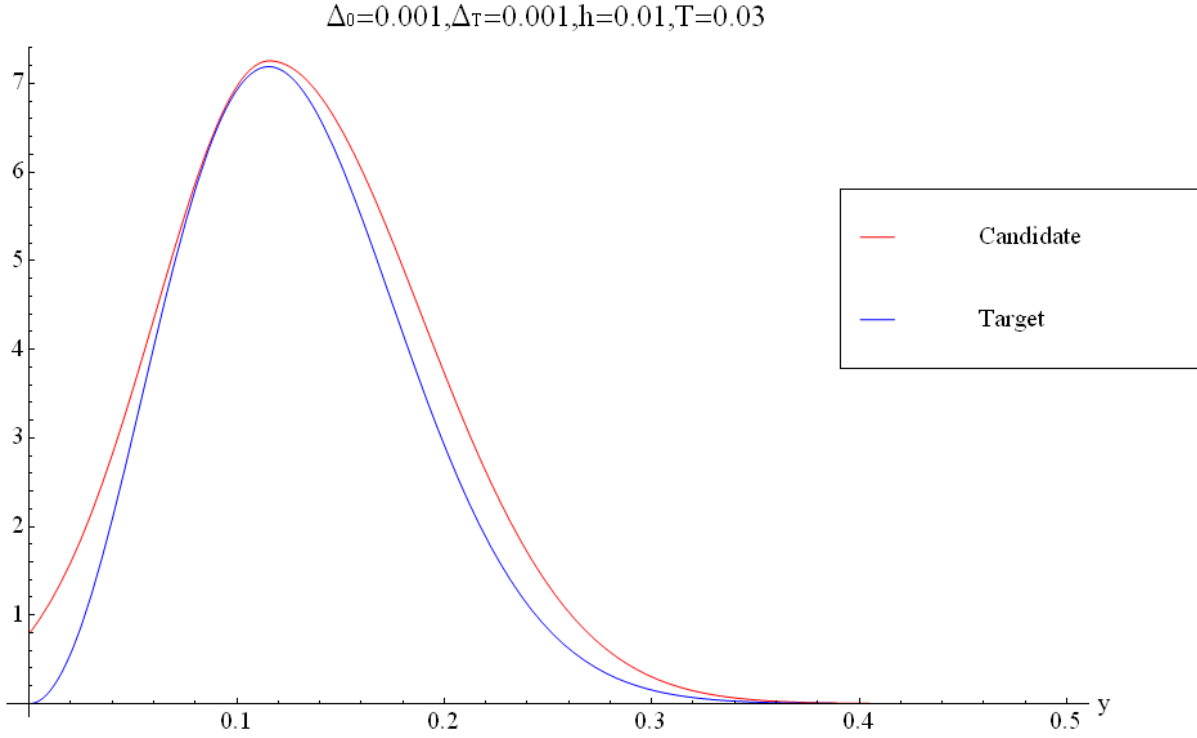


Figure 6.5: RN derivative of candidate measure compared with pdf of target measure. $\Delta_0 = 0.001$, $\Delta_T = 0.001$, $h = 0.01$, $T = 0.03$.

In the Gaussian cdf, the tail after 8 s.d. has probability $< 10^{-15}$. It seems apparent from graphing with different parameters that if there is a point where our candidate meets the target then it is at least this kind of distance from the peak. Actually, modelling the curve this way, we barely would encounter errors with no extra boost. In simulations, an extra boost of 1.1 for the candidate was used, to ensure that there would be no problems and it seems likely that this value is adequate for any foreseeable simulation. Because we used $\nu(y_{ML}, \infty)$ in the formula for our candidate, the candidate measure α is variable (but in general it is not much more than 1). From a run of the simulation with $h = 0.1$, $T = 8$, $M = 10^9$, there were 1197 occasions when $d\nu/d\mu$ was greater than 1, ie a proportion of order 10^{-8} , – but it is expected that this would be much less using $\tau_{\text{right}} = \tau \frac{\nu(y_{ML}, \infty)}{\nu(0, y_{ML})}$.

Because both component distributions are Gaussian, the splicing can be achieved here by simply taking a Gaussian draw and if it is negative, changing its sign with probability $(\frac{\tau_{\text{right}}}{\tau} - 1) / (\frac{\tau_{\text{right}}}{\tau} + 1)$. Naturally we then multiply by the appropriate sd for the side of

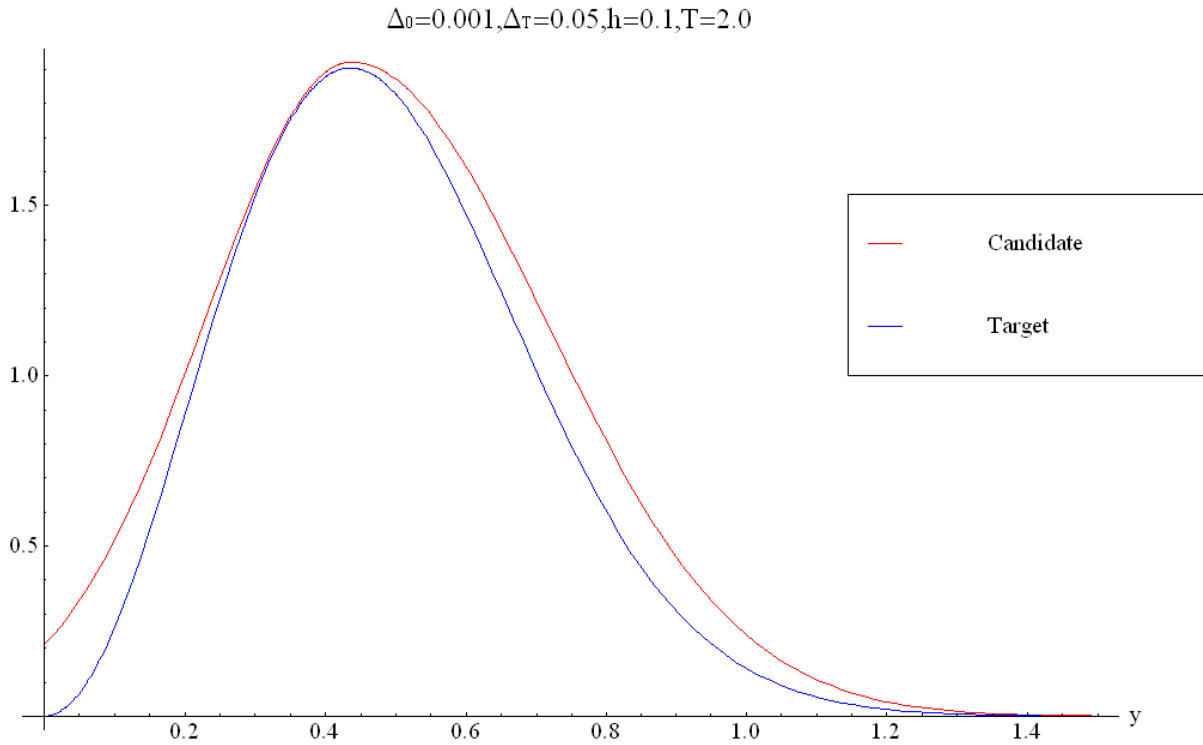


Figure 6.6: RN derivative of candidate measure compared with pdf of target measure. $\Delta_0 = 0.001$, $\Delta_T = 0.05$, $h = 0.1$, $T = 2.0$.

zero that it is on, and add our estimate of y_{ML} , and we have a draw from the candidate.

The method has not been analytically proven to work since it seems problematic to demonstrate rigorously that the candidate selected remains greater than the target for all y or for a set of measure $1 - \varepsilon$ with ε negligibly small. However, we can console ourselves with the fact that it has been empirically tested, and that this will be adequate to know that it will always work in practice, because it is always the same pdf that needs to be sampled, independent of the problem at hand. In practice this method provided an advantage in speed of some 10 times over the Decomposition method explained above; both gave equal results. Even though neither method could be proven to be valid analytically, the fact that the results agree gives definite confirmation that as would appear credible from the graph, $d\nu/d\mu > 1$ bias is not creeping in, and that the assumption of unipolarity was not significantly violated. No "variance factor" \hbar^2/m was used but this should be unimportant as whenever such a factor is introduced we should just adapt the candidate commensurately.

It seems that if we were to have had any concerns about the complexity or viability of

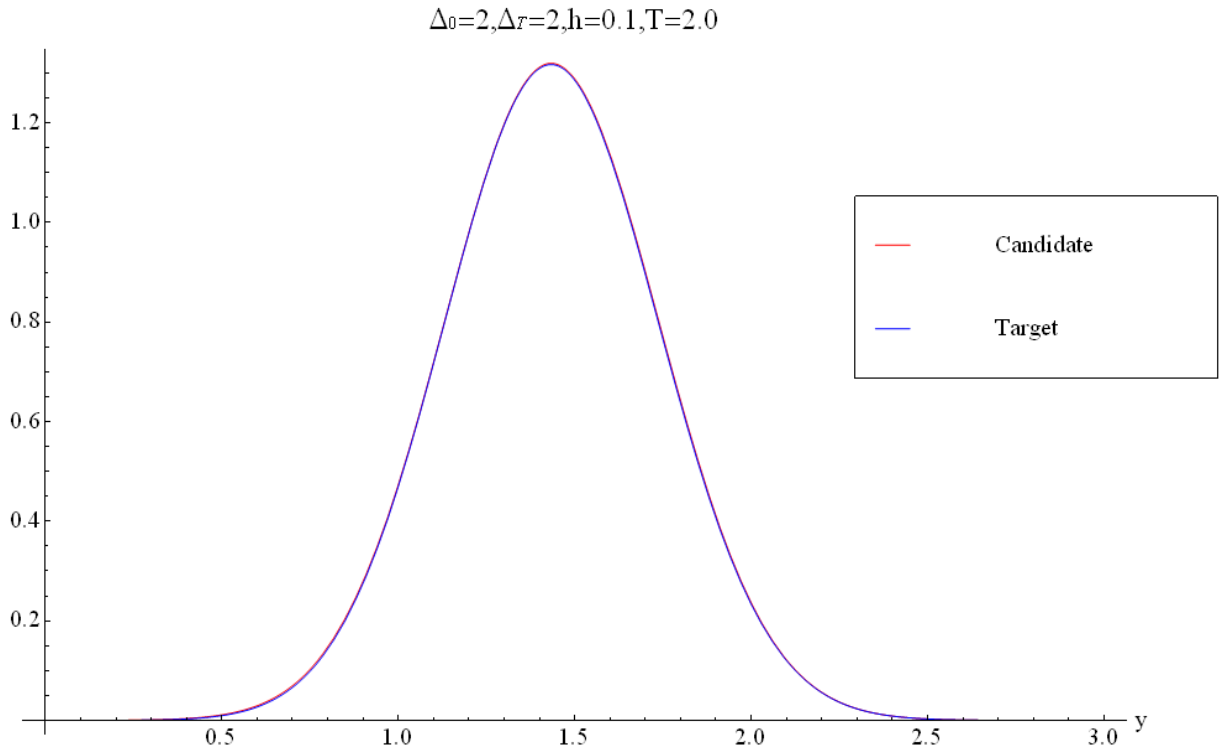


Figure 6.7: RN derivative of candidate measure compared with pdf of target measure. $\Delta_0 = 2.0$, $\Delta_T = 2.0$, $h = 0.1$, $T = 2.0$.

sampling from the non-crossing bridge fdd pdf, in order to exploit the formulation of the problem that arises from the Theorem 6.2.1 above, then these concerns must be laid to rest. Sampling is not difficult to program. The total cost per sample is equivalent to about 30 calls to the exponential function, but as we shall see, this was low enough for a simulation of the harmonic oscillator to be run with success.

6.3.3 The initial point distribution

In earlier attempts, a Gaussian distribution of initial points was used. Since we collect an approximation to $w_{0,(a,b)}^{T,(a,b)}(E_0)\mathcal{J}_T^{E_0}((a,b),(a,b))$ at each point, this gives rise to the sampling of many pairs of close points which have a very small contribution to the overall integral,

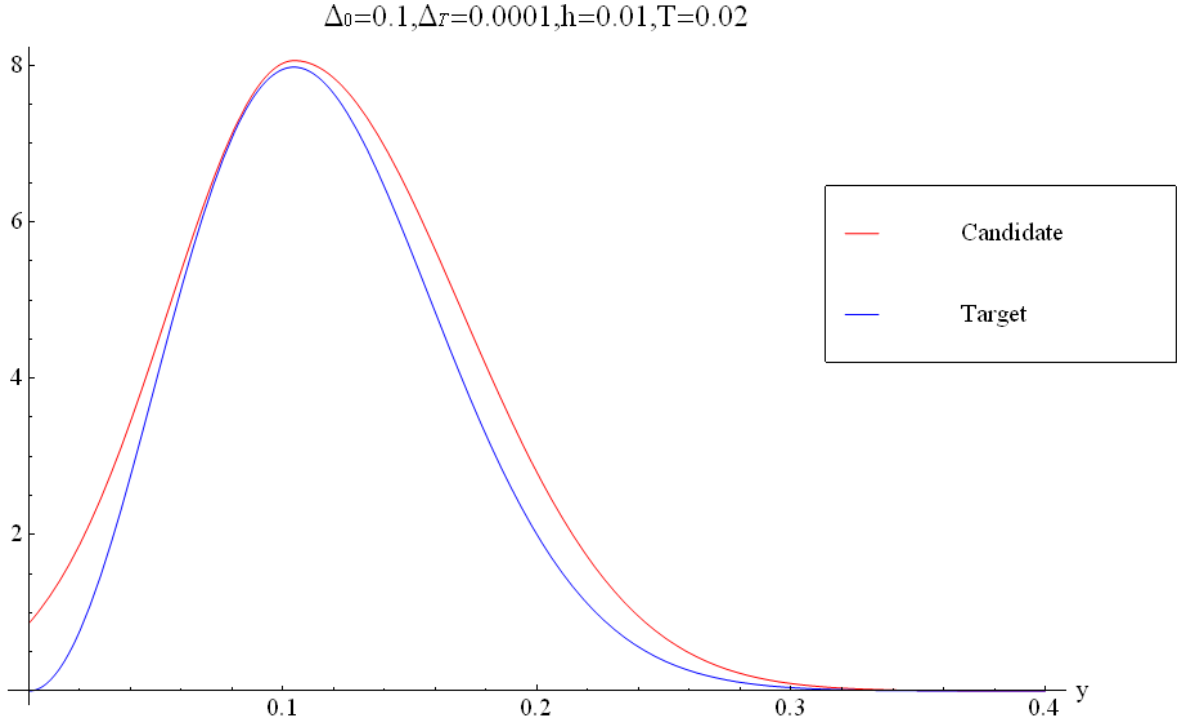


Figure 6.8: RN derivative of candidate measure compared with pdf of target measure. $\Delta_0 = 0.1$, $\Delta_T = 0.0001$, $h = 0.01$, $T = 0.02$.

due to the high probability of bridges crossing. Therefore a different pdf was used:

$$\begin{aligned} g(a, b) &\propto \exp\left(-\frac{a^2 + b^2}{2\sigma^2}\right) \left(1 - \exp\left(-\frac{(a - b)^2}{T}\right)\right) \\ &= \exp\left(-\frac{a^2 + b^2}{2\sigma^2}\right) - \exp\left(-\frac{a^2 + b^2}{2\sigma^2} - \frac{(a - b)^2}{T}\right). \end{aligned} \quad (6.40)$$

We make the transformation to $(a', b') = \left(\frac{a+b}{\sqrt{2}}, \frac{a-b}{\sqrt{2}}\right)$ and noting again that the modulus of the determinant of the Jacobian is 1,

$$\begin{aligned} g(a', b') &= \exp\left(-\frac{a^2 + b^2}{2\sigma^2}\right) - \exp\left(-\frac{a^2 + b^2}{2\sigma^2} - \frac{(a - b)^2}{T}\right) \\ &= \exp\left(-\frac{\left(\frac{a+b}{\sqrt{2}}\right)^2 + \left(\frac{a-b}{\sqrt{2}}\right)^2}{2\sigma^2}\right) - \exp\left(-\frac{\left(\frac{a+b}{\sqrt{2}}\right)^2 + \left(\frac{a-b}{\sqrt{2}}\right)^2}{2\sigma^2} - \frac{(a - b)^2}{T}\right) \\ &= \exp\left(-\frac{a'^2 + b'^2}{2\sigma^2}\right) - \exp\left(-\frac{a'^2 + b'^2}{2\sigma^2} - \frac{b'^2}{T/2}\right) \\ &= \exp\left(-\frac{a'^2}{2\sigma^2}\right) \left(\exp\left(-\frac{b'^2}{2\sigma^2}\right) - \exp\left(-b'^2 \left(\frac{1}{T/2} + \frac{1}{2\sigma^2}\right)\right)\right). \end{aligned} \quad (6.41)$$

So it is clear that these coordinates are independently distributed. We can draw a'

Gaussian and then obtain b' from decomposition or another method. Note that

$$\frac{1}{T/2} + \frac{1}{2\sigma^2} = \frac{T/2 + 2\sigma^2}{T\sigma^2} = \frac{1}{2T\sigma^2/(T + 4\sigma^2)} \quad (6.42)$$

It turns out that for decomposition purposes, we can find the peaks of $g(b')$ and prove that it is unipolar when we consider only positive b' , for

$$\frac{dg}{db'} = \frac{-b'}{\sigma^2} \exp\left(-\frac{b'^2}{2\sigma^2}\right) + \frac{b'}{T\sigma^2/(T + 4\sigma^2)} \exp\left(-\frac{b'^2}{2T\sigma^2/(T + 4\sigma^2)}\right) \quad (6.43)$$

Hence

$$\frac{dg}{db'}(b'_0) = 0, b'_0 \neq 0$$

and thus

$$\frac{1}{\sigma^2} \exp\left(-\frac{b'^2}{2\sigma^2}\right) = \frac{1}{T\sigma^2/(T + 4\sigma^2)} \exp\left(-\frac{b'^2}{2T\sigma^2/(T + 4\sigma^2)}\right) .$$

Therefore

$$\frac{T}{T + 4\sigma^2} = \exp\left(-\frac{b'^2}{2T\sigma^2/(T + 4\sigma^2)} + \frac{b'^2}{2\sigma^2}\right) .$$

Therefore

$$b'^2 = \left(\ln \frac{T}{T + 4\sigma^2}\right) / \left(\frac{-1}{2T\sigma^2/(T + 4\sigma^2)} + \frac{1}{2\sigma^2}\right) .$$

It is convenient, for decomposition, also to know the variance, and in order to get this we will find the mean and the mean square of the distribution. Let us first compute the normalizing factor C :

$$C = \frac{\sqrt{2\pi}\sigma - \sqrt{2\pi}\sqrt{\frac{T\sigma^2}{T+4\sigma^2}}}{2} = \sqrt{\frac{\pi}{2}} \left(\sigma - \sqrt{\frac{T\sigma^2}{T + 4\sigma^2}}\right) \quad (6.44)$$

where here the factor of $1/2$ enters because of the cutoff at zero, about which g is symmetric.

Now the mean is given by

$$\begin{aligned} \int_0^\infty b' g(b') db' &= C^{-1} \int_0^\infty b' \left(\exp\left(-\frac{b'^2}{2\sigma^2}\right) - \exp\left(-\frac{b'^2}{2T\sigma^2/(T + 4\sigma^2)}\right) \right) db' \\ &= C^{-1} \left(\left[-\sigma^2 \exp\left(-\frac{b'^2}{2\sigma^2}\right) \right]_0^\infty - \left[-\frac{T\sigma^2}{T + 4\sigma^2} \exp\left(-\frac{b'^2}{2T\sigma^2/(T + 4\sigma^2)}\right) \right]_0^\infty \right) \\ &= C^{-1} \left(\sigma^2 - \frac{T\sigma^2}{T + 4\sigma^2} \right) . \end{aligned} \quad (6.45)$$

Now notice that applying $E(X^2) = Var X + (EX)^2$,

$$\int_{-\infty}^{\infty} b'^2 \exp\left(-\frac{b'^2}{2\sigma^2}\right) db' = \sqrt{2\pi}\sigma^3 \quad (6.46)$$

but the integrand is symmetric about zero so it follows that

$$\int_0^{\infty} b^2 \exp\left(-\frac{b^2}{2\sigma^2}\right) db = \frac{\sqrt{\pi}}{\sqrt{2}}\sigma^3 \quad . \quad (6.47)$$

Therefore

$$\begin{aligned} \int_0^{\infty} b^2 g(b) db &= C^{-1} \int_0^{\infty} b^2 \left(\exp\left(-\frac{b^2}{2\sigma^2}\right) - \exp\left(-\frac{b^2}{2T\sigma^2/(T+4\sigma^2)}\right) \right) db \\ &= C^{-1} \left(\frac{\sqrt{\pi}}{\sqrt{2}}\sigma^3 - \frac{\sqrt{\pi}}{\sqrt{2}} \left(\frac{T\sigma^2}{T+4\sigma^2} \right)^{3/2} \right) \\ &= \left(\sigma^3 - \left(\frac{T\sigma^2}{T+4\sigma^2} \right)^{3/2} \right) \bigg/ \left(\sigma - \sqrt{\frac{T\sigma^2}{T+4\sigma^2}} \right) \quad . \end{aligned} \quad (6.48)$$

Therefore the standard deviation of b is given by

$$\sqrt{\frac{\sigma^3 - \left(\frac{T\sigma^2}{T+4\sigma^2} \right)^{3/2}}{\sigma - \sqrt{\frac{T\sigma^2}{T+4\sigma^2}}} - \left(C^{-1} \left(\sigma^2 - \frac{T\sigma^2}{T+4\sigma^2} \right) \right)^2} \quad . \quad (6.49)$$

A less expensive alternative is to sample (a, b) Gaussian and reject with probability $\exp\left(-\frac{(a-b)^2}{T}\right)$.

6.4 Simulation methods for n 1D non-crossing bridges

If we are to claim that a solution to the sign problem has been offered then we must describe how to perform the simulation when there are n particles. In other words, analogously to the situation of the preceding section, we have to demonstrate how to sample a point on the (system) Brownian bridge, when some other points may already have been fixed. As before it turns out that exact sampling is possible, but in general we may need to use a rejection method in order to achieve it.

6.5 Numerical results using non-crossing algorithm for 2 fermions in a 1D Harmonic Oscillator

A program was made to implement the method described above - viz, that we take only bridges that do not cross and estimate formula (6.21). The program estimates $\widetilde{\langle V \rangle}$, ie the approximation to $\langle V \rangle$ obtained by applying Simpson's Rule (3.40) to approximate S , by using direct path sampling (4.8-4.9). The following sequence of results is given, showing that this method is very effective compared to the methods discussed previously; particularly compare Figures 6.9-6.12 with Figures 5.1-5.7. Visually it is clear that a superpolynomial increase in M is required in order to maintain a certain Monte Carlo Standard Error as inverse temperature increases. However, the growth in cost is apparently similar to that found in the case of 1 particle in Subsection 4.3.4, which is because the cost growth is due to a lack of importance sampling, rather than to the sign problem. It seems fair to conjecture, from these results, that we may say that we have succeeded in proving and implementing a solution the fermion sign problem for 1-dimensional systems. The time to perform 10^9 Monte Carlo runs for $T = 10$ was about 82 hours on a 2.0 GHz machine.

Table 6.1 gives an indication of the convergence of the Monte Carlo estimate of $\widetilde{\langle V \rangle}$, when the simulations are performed with $h = 0.1$.

Table 6.1: *Potential energy for 2 noninteracting 1D fermions in a harmonic potential; to 9 d.p.* The results are based on using Simpson's Rule for S ; $h = 0.1$ throughout.

T	M	Realisation of $\widetilde{\langle V \rangle}$	Low $\widetilde{\langle V \rangle}$ (95%)	High $\widetilde{\langle V \rangle}$ (95%)	True value $\langle V \rangle$
2	10^7	1.097576066	1.096432794	1.098721173	1.096916182
4	10^8	1.010107205	1.009327187	1.010888017	1.009664255
6	10^8	1.001711376	0.999801159	1.003626313	1.001248600
8	10^9	1.000137100	0.998525317	1.001752307	1.000167900
10	10^9	1.003130673	0.998769273	1.007514820	1.000022703

It can be seen in Table 6.1 that in every case, the true value of $\langle V \rangle$ is within the interval predicted by the program $\widetilde{\langle V \rangle}$. The program used the Bivariate Gaussian contour method,

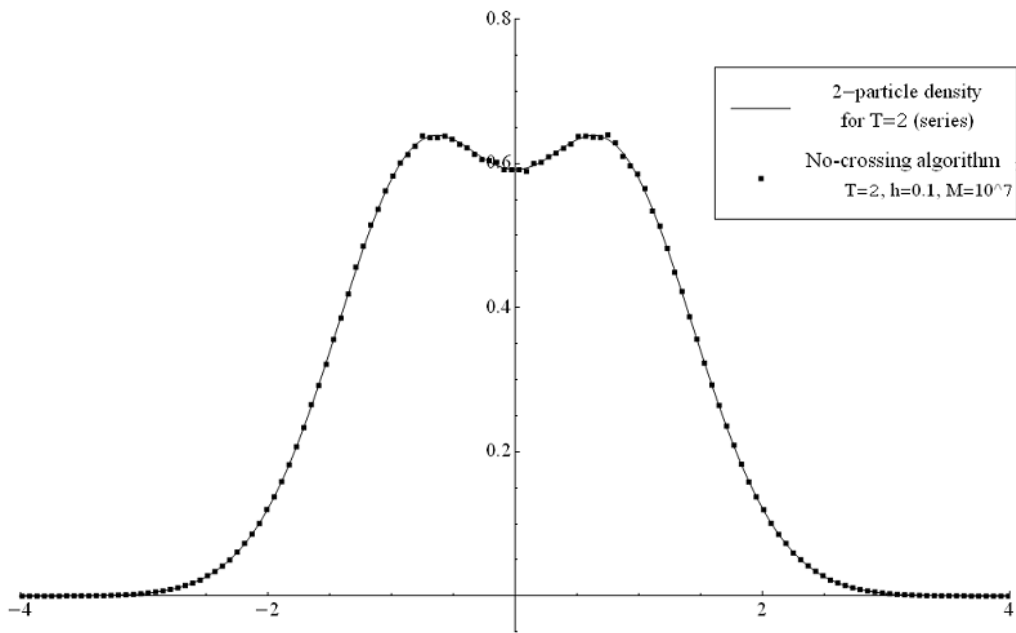


Figure 6.9: Particle density for 2 1D fermions using Simpson's Rule for S and non-crossing algorithm. $T = 2, h = 0.1, M = 10^7$.

described in Chapter 4, to determine these credibility intervals for $\widetilde{\langle V \rangle}$.

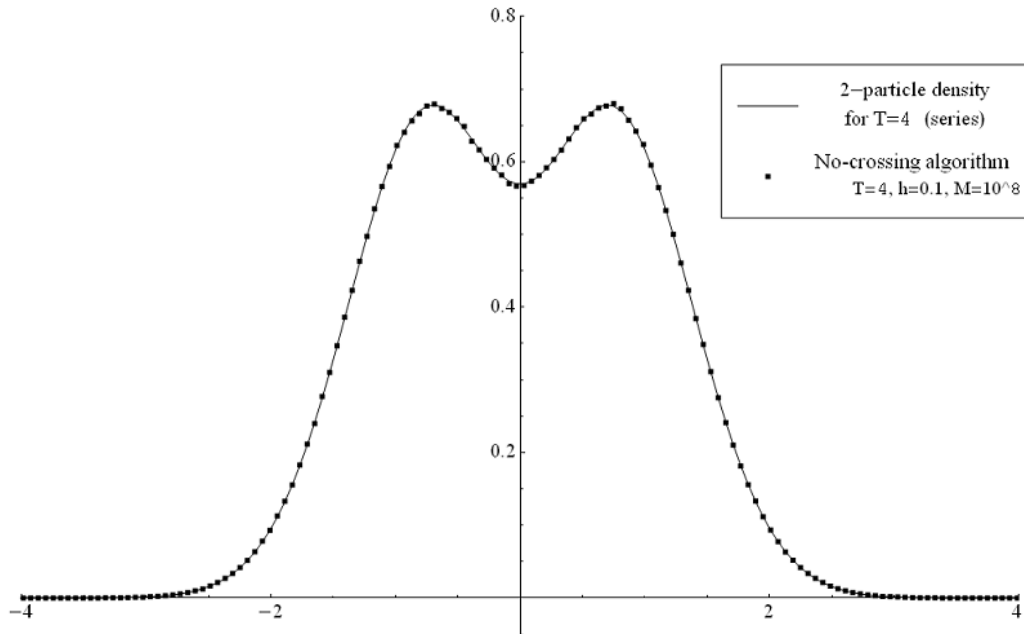


Figure 6.10: Particle density for 2 1D fermions using Simpson's Rule for S and non-crossing algorithm. $T = 4, h = 0.1, M = 10^8$.

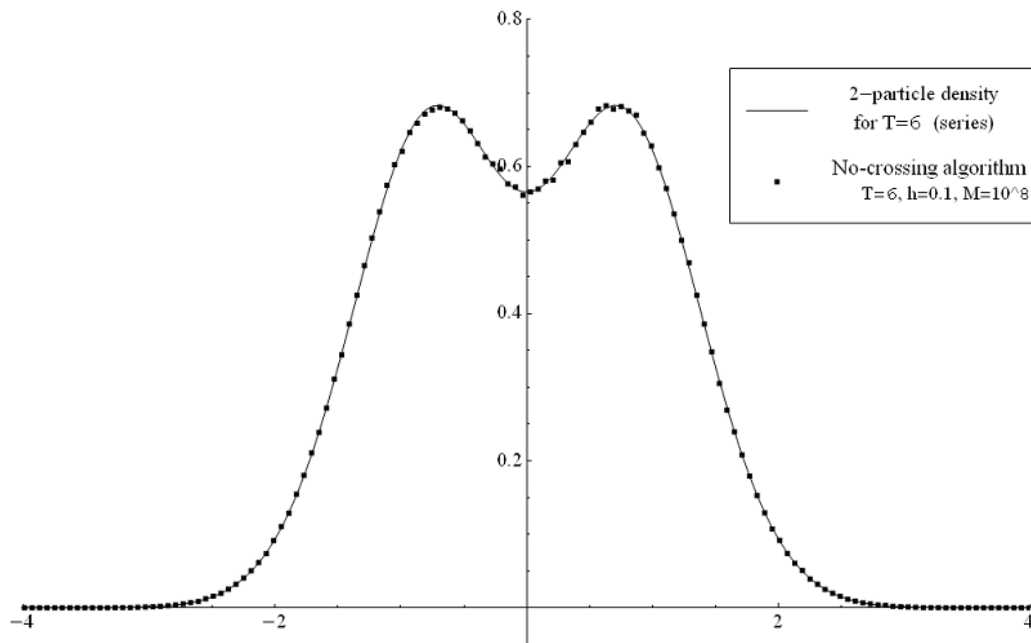


Figure 6.11: Particle density for 2 1D fermions using Simpson's Rule for S and non-crossing algorithm. $T = 6, h = 0.1, M = 10^8$.

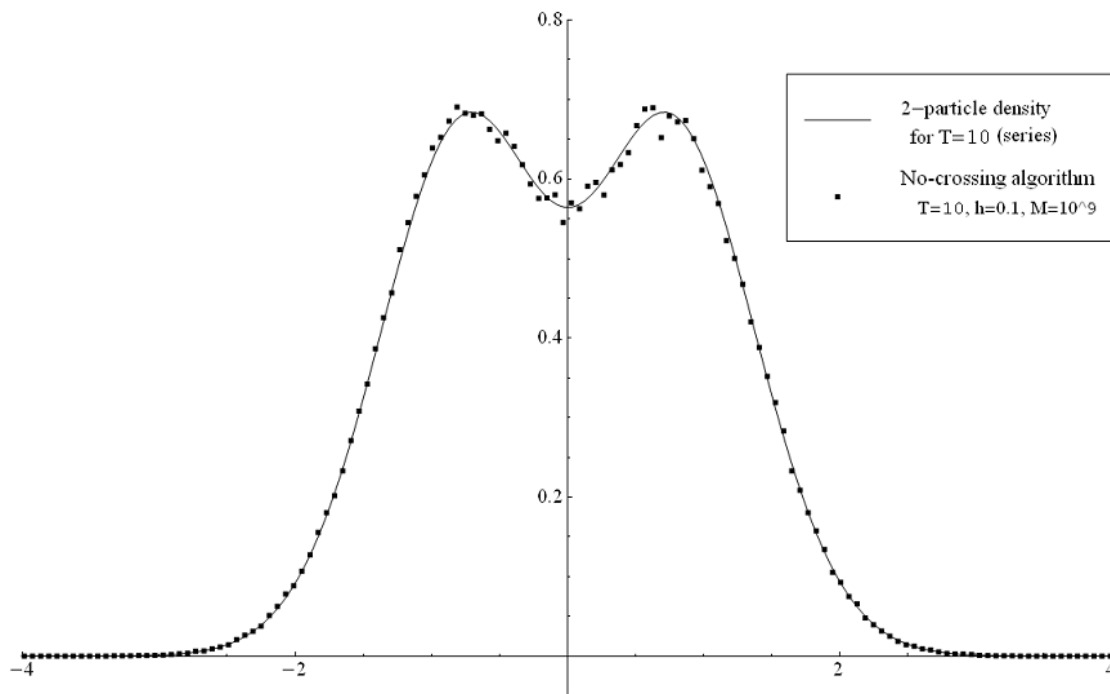


Figure 6.12: Particle density for 2 1D fermions using Simpson's Rule for S and non-crossing algorithm. $T = 10, h = 0.1, M = 10^9$.

Chapter 7

Multidimensional pairs of fermions

In this chapter, we develop and test the Iterated Subdiamonds approach to dealing with the sign problem in the case of a pair of fermions. We begin, in Section 7.1, by discussing what makes our approach different and explaining how we are led to the line of inquiry which is subsequently adopted. In Section 7.2, concentrating on the 2D case, we then build up the Iterated Subdiamonds approach and prove the main result of this chapter, Theorem 7.2.10. We proceed to explain, in Section 7.3, the ramifications of this as regards performing a simulation of a 2D or 3D fermion pair. In Section 7.4, results are presented of numerical experiments to test the average sign being attained by the algorithm and how it changes with T , the length of the time interval over which integration is performed.

7.1 A novel approach to the fermion sign problem

In the previous chapter, we succeeded in effectively solving the fermion sign problem in the 1D case by partitioning the space of paths into subsets where the contributions to the Monte Carlo would be only positive on some, and, as it turned out, the required positive and negative contributions would exhibit perfect cancellation on others. Although our presentation of this result may be novel, its realisation is familiar to practitioners. As mentioned, the physical reasoning behind the 1D result is that particles are prevented from crossing

wavefunction nodes, and the multidimensional generalization of this principle is RPIMC (cf Subsection 5.2.1).

The approach taken here represents a radical departure from the node-avoiding approach. We can instead view the result of the previous chapter in terms of a simultaneous sampling approach, in which the paths sampled together are those which yield a perfect cancellation. Endeavouring to generalise this principle mathematically gives a direction to our research which is different from that of both RPIMC (cf Subsection 5.2.1) and MLB (cf Subsection 5.2.2), although it has commonalities with both. We aim for an exact method which does not depend on external calculations.

As noted previously, it has been shown that the fermion sign problem is insoluble in the sense of producing an exact solution (or one with controlled approximations) which treats 3D fermions and increases in cost polynomially as the number of fermions is increased. The best that it is reasonable to expect is a method for which the cost to achieve a given variance does explode, but with a relatively low exponent. We have already mentioned that although there is a notional equivalence between an increased number of fermions and a reduction in temperature $1/k_B T$, due to the converging relative likelihoods of terminal point permutations, it is not clear that solving the sign problem for T is as hard as solving it for n . In this chapter, it turns out that in fact, we are able to develop a method for the case of just 2 particles where the cost scaling with T is relatively favourable.

7.1.1 Independence and dimensional crossing events

At first sight it might not seem immediately clear what partition of path space might be relevant to use in 2D or 3D. In 1D, the first step was to recognise the events whose probabilities are given by the Gaussian factors on the permutation summands in \mathcal{I}_T (cf (2.62)). In 2D (or 3D), two paths may cross in the x or y (or z) coordinates, so what combinations of crossings are important? A little consideration of the quantities at hand reveals an intuitively likely answer. Letting $a, b \in \mathbb{R}^3$, the factor $\varphi_T((a, b), (b, a))$, which

appears in $\mathcal{I}_T((a, b), (a, b))$ (cf (2.62)), may be seen to be

$$\exp(-\|a - b\|^2 / T) = w_{0,(a,b)}^{T,(a,b)}(E_x) w_{0,(a,b)}^{T,(a,b)}(E_y) w_{0,(a,b)}^{T,(a,b)}(E_z) = w_{0,(a,b)}^{T,(a,b)}(E_x \cap E_y \cap E_z) = w_{0,(a,b)}^{T,(a,b)}(E_0^c), \quad (7.1)$$

where E_x (respectively E_y, E_z) is the set of loop-pairs with a crossing in the x (respectively y, z) coordinate. In other words, this factor is equal to the measure of the set E_0^c of 3-dimensional bridges from (a, b) to (a, b) such that the bridge from a to a and that from b to b meet in every coordinate. This holds because E_x, E_y, E_z are mutually independent events and because from the results of the previous chapter, it follows that $w_{0,(a,b)}^{T,(a,b)}(E_x) = \exp(-(a_x - b_x)^2 / T)$.

Obviously the set of pairs of Brownian bridges that meet in two coordinates simultaneously is a set of measure zero; the paths that are of interest here are those which meet in each coordinate at a different time. For conceptual simplicity we can also interpret the coefficient (7.1) in another way: it is the probability of a longitudinal crossing. That is, if we rotate the coordinate basis so that one "longitudinal" basis vector (corresponding to the x coordinate, for definiteness) has the direction $(b - a) / \|a - b\|$ and the other "transverse" coordinates are orthogonal to this, then we see that clearly $\|a - b\|$ is the distance between the new x positions of a and b and it follows from the independence of coordinates that $\exp(-\|a - b\|^2 / T)$ is the probability of a crossing in x ; clearly there are crossings in the other coordinates simultaneously at time 0. (It also seems clear that for any rotation of coordinates, there is an analogous event of which $\exp(-\|a - b\|^2 / T)$ is the probability.)

The intuition here is that in view of the importance of this rearrangement in the 1D case, regarding path space as partitioned into disjoint events so that on each of these events, we have a set of unweighted summands¹, is a comparatively logical way to proceed. In 1D, the attempts to create a positive covariance between countersigned summands, detailed in Section 5.3, were scuppered by the presence of the Gaussian coefficients such as (7.1), and we solved the problem by rearranging to take account of these coefficients. We can regard

¹Or in general, a set of weighted summands such that for each negative summand there is a positive one with the same weight.

a coefficient on an integral as modifying the measure of integration, and any cancellation between the contributions of different paths must surely rest on comparing paths from terms with equal full measure over the integration domain.

7.2 Subdiamond theory for the case of two fermions in a 2-dimensional space

In this section we begin by developing basic results based on geometrical intuition about how to address the sign problem. We then use these to develop the Iterated Subdiamond method, the justification for which is expressed in Theorem 7.2.10, our main result in this chapter. Although we prove results that demonstrate the required equivalences, the utility of these is only discussed informally.

7.2.1 Diamonds

We shall mostly devote our attention to the unnormalized position density $\mathcal{G}_T((a, b), (a, b))$ (cf (2.62-2.65)) and those quantities which can be derived from it, from here onwards; but it should be borne in mind that the same logic allows us to treat expectations of other exponential-type functionals. In this section, we develop two results which offer a decomposition of $\mathcal{G}_T((a, b), (a, b))$ into a product of integrals conditional on the first and last longitudinal crossing times and points. Although it is not difficult to appreciate how the results of this section generalise to 3D (cf Subsection 7.3.2), we shall focus on the 2D case for ease of exposition.

In the following proposition, we shall use the rotated coordinate basis that makes the x coordinate longitudinal. That is to say, without loss of generality we assume that $a = (a_x, 0)$ and $b = (b_x, 0)$ since this can always be achieved by a rotation and translation of coordinates.

Random variables as measurable functions It is worth recalling at this stage that we regard X as a random variable on the measurable space $(\mathcal{C}_{0,x_0}^T, \mathcal{B}(\mathcal{C}_{0,x_0}^T))$. In this chapter, in fact we shall take X to be the identity: $X(t; \omega) = \omega(t)$ (cf Definition 2.1.4 and the brief discussion following it). A random variable on $(\mathcal{C}_{0,x_0}^T, \mathcal{B}(\mathcal{C}_{0,x_0}^T))$ is a Borel-measurable function on \mathcal{C}_{0,x_0}^T , and expectations of random variables, with any particular conditioning, are integrals of those random variables with respect to measures on $\mathcal{B}(\mathcal{C}_{0,x_0}^T)$. In this chapter it is usually best for clarity, and to avoid longwindedness, to write such integrals explicitly. (Mathematically, however, to constantly redefine the relevant probability measure P to make $(\mathcal{C}_{0,x_0}^T, \mathcal{B}(\mathcal{C}_{0,x_0}^T), P)$ a probability space, and write expectations instead of integrals, would of course be equivalent.)

First crossing time and point We let $\tau(X) : E_x \rightarrow [0, T]$ be given by the first x-crossing time of X (ie, the first crossing time in the longitudinal direction $(b-a)/\|b-a\|$) and let $x_1(X)$ denote $X(\tau(X))$. For $x \in \mathbb{R}^4$ s.t. $x^{(1)1} = x^{(2)1}$, let

$$\begin{aligned} E(t, x) &= \{X \in E_x : \tau(X) = t, x_1(X) = x\} ; \\ E^{\text{SW}}(t, x) &= \left\{ X \in \mathcal{C}_{0;(a,b)}^{T;(b,a)} : \tau(X) = t, x_1(X) = x \right\} \end{aligned}$$

Here ^{SW} stands for swap. We let $\mathcal{C}_{0,(a,b)}^{t,x}|_{E(t,x)}$ denote the set of paths in $\mathcal{C}_{0,(a,b)}^{t,x}$ with no x-crossings before time t , so that x is the first x-crossing point.

The space of all continuous paths of finite time-length We shall let $\mathcal{C}^* = \bigcup_{t \in [0, \infty)} \mathcal{C}_0^t$. From here onwards we generalise our definition of the exponentiated action Y so that for any path in \mathcal{C}^* , Y may be considered to be defined analogously with (1.2). Whenever functions are defined on \mathcal{C}^* it should be assumed that they are also defined for \mathcal{C}_s^t , whenever $s < t$, via translating time by a shift of $-s$.

Concatenation operator We shall use the $\&$ binary operator to indicate the concatenation of two path sections, ie $X_{[0,t]} \& X'_{(t,s]}$ is given by X over $[0, t]$ and by X' over $(t, s]$.

Proposition 7.2.1 (First Diamond Proposition) *Let $f_1 : \mathcal{C}^* \rightarrow \mathcal{C}^*$ be s.t. for any $t \in (0, T)$ and $x \in \mathbb{R}^4$ s.t. $x^{(1)1} = x^{(2)1}$, we have that for $X \in \mathcal{C}_{0,(a,b)}^{t,x}|_{E(t,x)}$, we shall have $f_1(X) \in \mathcal{C}_{0,(b,a)}^{t,x}|_{E(t,x)}$. Moreover let f_1 be s.t. for any event $A \in \mathcal{B}(\mathcal{C}_{0,(a,b)}^{t,x})$,*

$$w_{0,(b,a)}^{t,x}|_{E(t,x)}(f_1(A)) = w_{0,(a,b)}^{t,x}|_{E(t,x)}(A).$$

Let $f_2 : \mathcal{C}^ \rightarrow \mathcal{C}^*$ be s.t. for $X \in \mathcal{C}_{t,x}^{T,(a,b)}$, $f_2(X) \in \mathcal{C}_{t,x}^{T,(b,a)}$ and s.t. for any event $A \in \mathcal{B}(\mathcal{C}_{t,x}^{T,(a,b)})$,*

$$w_{t,x}^{T,(b,a)}(f_2(A)) = w_{t,x}^{T,(a,b)}(A).$$

Then

$$\begin{aligned} \mathcal{G}_T((a, b), (a, b)) &= \int_{E_x^c} Y(X) dw_{0,(a,b)}^{T,(a,b)} + \\ &\frac{1}{2} \int_{E_x} [(Y(X_{[0,\tau(X)]}) - Y(f_1(X_{[0,\tau(X)]}))) (Y(X_{[\tau(X),T]}) - Y(f_2(X_{[\tau(X),T]})))] dw_{0,(a,b)}^{T,(a,b)} \quad , \end{aligned} \quad (7.2)$$

where naturally we regard $f_{1,2}$ as applying in the longitudinal-transverse basis.

Proof. It follows from the definition (2.62-2.65) and the definition of \mathcal{J}_T (2.32) that

$$\mathcal{G}_T((a, b), (a, b)) = \int_{\mathcal{C}_{0,(a,b)}^{T,(a,b)}} Y(X) dw_{0,(a,b)}^{T,(a,b)} - \exp\left(-\frac{\|a-b\|^2}{T}\right) \int_{\mathcal{C}_{0,(a,b)}^{T,(b,a)}} Y(X) dw_{0,(a,b)}^{T,(b,a)} \quad . \quad (7.3)$$

Therefore from (7.1),

$$\begin{aligned} \mathcal{G}_T((a, b), (a, b)) &= \int_{E_x^c} Y(X) dw_{0,(a,b)}^{T,(a,b)} \\ &+ \int_{E_x} Y(X) dw_{0,(a,b)}^{T,(a,b)} - w_{0,(a,b)}^{T,(a,b)}(E_x) \int_{\mathcal{C}_{0,(a,b)}^{T,(b,a)}} Y(X) dw_{0,(a,b)}^{T,(b,a)} \quad . \end{aligned} \quad (7.4)$$

Note that the x-coordinates of X form a pair of 1-dimensional Brownian bridges, independent from the y-coordinates. Therefore, for the x-coordinates, the same logic applies as in the proof of Theorem 6.1.4; viz, the distribution of (τ, x_1) , the first crossing time and point in \mathbf{x} , is the same for $w_{0,(a,b)}^{T,(b,a)}$ as for the concentration of $w_{0,(a,b)}^{T,(a,b)}$ on E_x , so we may condition on a first crossing time and point (t, x) which is identical in law for both summands. Let η

signify the measure on $[0, T] \times \mathbb{R}$ induced by the distribution of (τ, x_1) for the x-coordinate of a bridge in E_x under $w_{0,(a,b)}^{T,(a,b)} \Big|_{E_x}$. Then

$$\begin{aligned}
\mathcal{G}_T((a, b), (a, b)) &= \int_{E_x^c} Y(X) dw_{0,(a,b)}^{T,(a,b)} \\
&\quad + w_{0,(a,b)}^{T,(a,b)}(E_x) \mathcal{J}_T^{E_x}((a, b), (a, b)) - w_{0,(a,b)}^{T,(a,b)}(E_x) \mathcal{J}_T((a, b), (b, a)) \\
&= \int_{E_x^c} Y(X) dw_{0,(a,b)}^{T,(a,b)} \\
&\quad + w_{0,(a,b)}^{T,(a,b)}(E_x) \int_{[0,T] \times \mathbb{R}} \left[\mathcal{J}_T^{E(t,x)}((a, b), (a, b)) - \mathcal{J}_T^{E^{SW}(t,x)}((a, b), (b, a)) \right] d\eta(t, x) \\
&= \int_{E_x^c} Y(X) dw_{0,(a,b)}^{T,(a,b)} \\
&\quad + \int_{E_x} \left[\mathcal{J}_T^{E(\tau(X), x_1(X))}((a, b), (a, b)) - \mathcal{J}_T^{E^{SW}(\tau(X), x_1(X))}((a, b), (b, a)) \right] dw_{0,(a,b)}^{T,(a,b)}. \quad (7.5)
\end{aligned}$$

In other words, intuitively speaking, since the x-crossing in the negative summand is identically distributed with that in the positive summand, we might as well use the same realisation for both. However, notice that under $w_{0,(a,b)}^{T,(a,b)}$,

$$\begin{aligned}
f_1(X_{[0,\tau(X)]}) \& X_{[\tau(X),T]} &\sim w_{0,(b,a)}^{T,(a,b)} \\
X_{[0,\tau(X)]} \& f_2(X_{[\tau(X),T]}) &\sim w_{0,(a,b)}^{T,(b,a)} \\
f_1(X_{[0,\tau(X)]}) \& f_2(X_{[\tau(X),T]}) &\sim w_{0,(b,a)}^{T,(b,a)} \Big|_{E_x^*},
\end{aligned}$$

where $E_x^* \subseteq \mathcal{C}_{0,(b,a)}^{T,(b,a)}$ is the Borel set consisting of those paths from (b, a) to (b, a) with an x-crossing. Therefore it follows from permutational symmetry of the potential that

$$\begin{aligned}
\int_{E_x} \mathcal{J}_T^{E(\tau(X), x_1(X))}((a, b), (a, b)) dw_{0,(a,b)}^{T,(a,b)} &= \int_{E_x} Y(X) dw_{0,(a,b)}^{T,(a,b)} \\
&= \int_{E_x} Y(f_1(X_{[0,\tau(X)]}) \& f_2(X_{[\tau(X),T]})) dw_{0,(a,b)}^{T,(a,b)} \\
\int_{E_x} \mathcal{J}_T^{E^{SW}(\tau(X), x_1(X))}((a, b), (b, a)) dw_{0,(a,b)}^{T,(b,a)} &= \int_{E_x} Y(X_{[0,\tau(X)]} \& f_2(X_{[\tau(X),T]})) dw_{0,(a,b)}^{T,(b,a)} \\
&= \int_{E_x} Y(f_1(X_{[0,\tau(X)]}) \& X_{[\tau(X),T]}) dw_{0,(b,a)}^{T,(a,b)}
\end{aligned}$$

and since Y is multiplicative in the sense that $Y(X) = Y(X_{[0,t]})Y(X_{[t,T]})$, it follows that

$$\begin{aligned}
\mathcal{G}_T((a, b), (a, b)) &= \int_{E_x^c} Y(X) dw_{0,(a,b)}^{T,(a,b)} + \frac{1}{2} \int_{E_x} [Y(X) - Y(f_1(X_{[0,\tau(X)]}) \& X_{[\tau(X),T]}) \\
&\quad - Y(X_{[0,\tau(X)]} \& f_2(X_{[\tau(X),T]})) \\
&\quad + Y(f_1(X_{[0,\tau(X)]}) \& f_2(X_{[\tau(X),T]}))] dw_{0,(a,b)}^{T,(a,b)} \\
&= \int_{E_x^c} Y(X) dw_{0,(a,b)}^{T,(a,b)} \\
&\quad + \frac{1}{2} \int_{E_x} [(Y(X_{[0,\tau(X)]}) - Y(f_1(X_{[0,\tau(X)]}))) \times \\
&\quad (Y(X_{[\tau(X),T]}) - Y(f_2(X_{[\tau(X),T]})))] dw_{0,(a,b)}^{T,(a,b)} .
\end{aligned}$$

■

Basically, f_1 and f_2 are measure-conjugating bijections. Proposition 7.2.1 provides a geometrical reinterpretation of the permutation formula (2.62) for \mathcal{I}_T .

The first summand in the right-hand side of (7.2) is computationally unproblematic, as we discovered in Chapter 6. Essentially we can read the problematic "diamond" part of $\mathcal{I}_T(x, x)$ as $\int_{E_x} [Y(X) - Y(X')] dw_{0,(a,b)}^{T,(a,b)}$ where $X' \in E^{\text{SW}}(\tau(X), x_1(X))$ is constructed from X in a manner which for every $(t, x) \in (0, T) \times \mathbb{R}^4$ conjugates $w_{0,(a,b)}^{T,(a,b)} \Big|_{E(t,x)}$ with $w_{0,(a,b)}^{T,(b,a)} \Big|_{E^{\text{SW}}(t,x)}$. If we conceive of two different ways of doing this so that X' agrees with X either before $\tau(X)$ or after $\tau(X)$, then we can combine them both to create two negative and two positive contributions, as illustrated in Figure 7.1 (see a detailed explanation below).

Subdiamond schematics

Some of the figures in this chapter are described as schematics. This means that rather than illustrating a realisation of the paths relevant to a particular expression for \mathcal{G}_T , or even their expectation positions, for clarity we instead simply draw the set of links (representing path segments) showing the relationships between the paths involved. Any *suitable* sequence of path segments described in a schematic may be thought of as a path which gives rise to either a positive or negative contribution to a functional whose expectation is \mathcal{G}_T . For example, in Figure 7.1, if we follow the blue (negative) segment of length τ from b and then

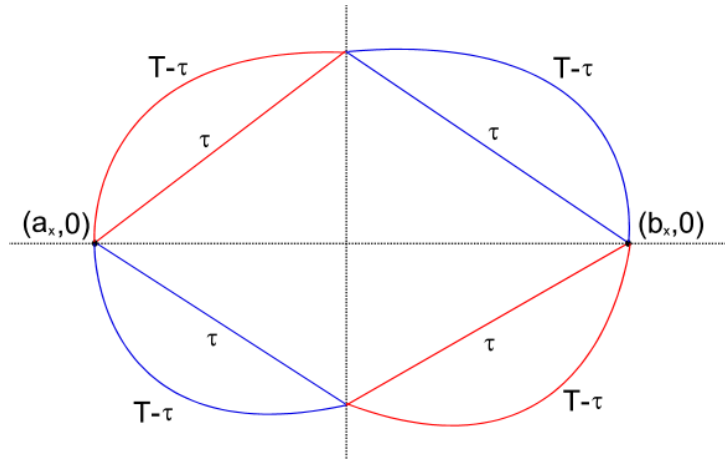


Figure 7.1: Schematic of the contributing path segments for a path in E_x , according to the First Diamond Proposition for finding $\mathcal{G}_T((a_x, 0), (b_x, 0), (a_x, 0), (b_x, 0))$. Red is positive and blue is negative.

follow the blue segment of length $T - \tau$ back to b , this represents a path which would give rise to a positive contribution, since the two negative signs are multiplied. On the other hand, following the blue segment of length τ from b and then following the red segment of length $T - \tau$ to a represents a path with a negative contribution. Following an odd number of blue segments gives rise to a negative-contributing path; following an even number gives rise to a positive-contributing path. More properly we should really say, of course, that we follow two segments at once: if we follow the blue segment of length τ from b then simultaneously we follow the blue segment of length τ from a .

In Figure 7.1, we may say that if X were sampled to be the red path segments, then the blue path segments supply our mental image of what $f_1(X_{[0, \tau(X)]})$ & $f_2(X_{[\tau(X), T]})$ would look like. A similar interpretation applies in the other schematics.

We do not use different colours for links representing path segments which are unconstrained Brownian bridges and for links representing path segments for which crossing constraints are implied in the text.

On relative magnitudes

One direction taken from (7.2) would seem to be to assess the relative magnitudes of $\int_{E_x^c} Y(X) dw_{0,(a,b)}^{T,(a,b)}$ and the remainder "diamond" part, to establish which dominates as $T \rightarrow \infty$. This is especially of interest knowing that, using the techniques of the previous chapter, we can certainly simulate $\int_{E_x^c} Y(X) dw_{0,(a,b)}^{T,(a,b)}$, which is the expectation of a positive functional, and indeed could easily extend this to the case of n 3D particles. However, the undertaking of determining relative magnitudes via partial differential equations leads to significant difficulties. In view of this, an empirical test was performed to see whether using the non-crossing part would produce results converging to the correct value, for 2 noninteracting fermions in a coulombic potential $V(x, y) = 1/\sqrt{x^2 + y^2}$. The conclusion from this was that the results did *not* seem to converge to the correct value. Therefore we shall assume in what follows that we must take an exact approach to (7.2) and not neglect its problematic "diamond" terms.

On vertical symmetry

Proposition 7.2.1 possesses a certain symmetry: since if we condition on $\tau = t$ and consider the distribution of the first crossing point $X(\tau) = ((x_1, y_1), (x_1, y_2))$, we find that (y_2, y_1) is identical in law with (y_1, y_2) , a natural question is, what will happen if we consider adding contributions that arise from conditioning on a first crossing at $((x_1, y_2), (x_1, y_1))$, as well as those that arise from $((x_1, y_1), (x_1, y_2))$ as already considered here? The answer is that the same quantity will be restored, since exchange of (a, b) is the same as exchange of (y_1, y_2) , and thus the conditional expectation of $Y(f_1(X_{[0,\tau(X)]}))Y(f_2(X_{[\tau(X),T]}))$ under crossing with (y_1, y_2) will be equal to the conditional expectation of $Y(X)$ under crossing with (y_2, y_1) .

Remark 7.2.2 *It is also valid to conclude, since we could have selected just one each of the*

positive and negative contributing terms, that we could say

$$\begin{aligned}
 \mathcal{G}_T((a, b), (a, b)) &= \int_{E_x^c} Y(X) dw_{0, (a, b)}^{T, (a, b)} \\
 &\quad + \int_{E_x} [Y(X) - Y(f_1(X_{[0, \tau(X)]}) \& X_{[\tau(X), T]})] dw_{0, (a, b)}^{T, (a, b)} , \\
 \mathcal{G}_T((a, b), (a, b)) &= \int_{E_x^c} Y(X) dw_{0, (a, b)}^{T, (a, b)} \\
 &\quad + \int_{E_x} [Y(X) - Y(X_{[0, \tau(X)]} \& f_2(X_{[\tau(X), T]})] dw_{0, (a, b)}^{T, (a, b)} .
 \end{aligned} \tag{7.6}$$

These may be conceptually helpful, but seem less useful because of the lack of vertical symmetry. In particular, comparing (7.6) with (7.2) shows that in (7.6) we are missing terms that may nearly cancel what we have: if T is large then $Y(f_2(X_{[\tau(X), T]}))$ may usually be relatively close to $Y(X_{[\tau(X), T]})$ for a suitable choice of f_2 . Intuitively speaking, vertical symmetry is important for obtaining a lower variance because without it, we face the danger that we might be equally likely to obtain a "flipped" sample with a close but countersigned contribution. By contrast, an estimator of $\mathcal{I}_{T/2}$ s.t. the samples possess vertical symmetry of this kind, has the property that we may always consider that $y_1 > y_2$; in other words, we may arbitrarily choose that in each diamond, the leftmost point connects to the topmost point to give the positive contributions. This is important.

Examples of measure-preserving bijections

There are a number of alternative practical expressions for \mathcal{I}_T which can be adopted since there are various different apparent choices for f_1 and f_2 . The most obvious are based on coordinate exchange, linear translation, and reflections.

Coordinate exchanges We can use an exchange of coordinates between bridges to define either f_1 or f_2 or both, in applying the Diamond Proposition.

We define the operator $\neg_{[s_1, s_2]}^x$ such that it has the effect of reversing the x-coordinates

on $[s_1, s_2]$. Then

$$\mathcal{G}_T((a, b), (a, b)) = \int_{E_x^c} Y(X) dw_{0,(a,b)}^{T,(a,b)} + \frac{1}{2} \int_{E_x} \left[\left(Y(X_{[0,\tau(X)]}) - Y(\neg_{[0,\tau(X)]}^x X_{[0,\tau(X)]}) \right) \left(Y(X_{[\tau(X),T]}) - Y(\neg_{[\tau(X),T]}^x X_{[\tau(X),T]}) \right) \right] dw_{0,(a,b)}^{T,(a,b)}.$$

Alternatively, considering the equivalence, already noted, of exchanging y_1, y_2 rather than a, b ,

$$\mathcal{G}_T((a, b), (a, b)) = \int_{E_x^c} Y(X) dw_{0,(a,b)}^{T,(a,b)} + \frac{1}{2} \int_{E_x} \left[\left(Y(X_{[0,\tau(X)]}) - Y(\neg_{[0,\tau(X)]}^y X_{[0,\tau(X)]}) \right) \left(Y(X_{[\tau(X),T]}) - Y(\neg_{[\tau(X),T]}^y X_{[\tau(X),T]}) \right) \right] dw_{0,(a,b)}^{T,(a,b)}.$$

These are in fact equivalent: both are equivalent to saying that where $X = (X_1, Y_1, X_2, Y_2)$, the X' sample is found from (X_1, Y_2, X_2, Y_1) . It turns out that this way of generating countersigned contributions, at first glance not advantageous, is by far the most flexible.

Linear translation Given $\tau(X) = t$, the conditional distribution of $X_{[t,T]}$ induces $w_{t,x}^{T,(a,b)}$; there is no constraint. Therefore on this interval we could also construct a section of $X' \in E^{\text{SW}}(t, x)$ via a linear map from the corresponding coordinates of X . (In other words, the trajectory obtained from the corresponding Brownian bridge SDE when using the same realisation of the Wiener process.) Viz, if we define a random variable, whose values are functions $\ell((a, b), \tau(X), T) : [0, T] \rightarrow \mathbb{R}^4$ given by

$$\ell((b - a, 0, \tau(X), T); t) = 1_{t > \tau(X)} \frac{t - \tau(X)}{T - \tau(X)} (b - a, 0, a - b, 0)$$

then

$$\begin{aligned} \mathcal{G}_T((a, b), (a, b)) &= \int_{E_x^c} Y(X) dw_{0,(a,b)}^{T,(a,b)} \\ &\quad + \int_{E_x} [Y(X) - Y(X + \ell((b - a, 0), \tau(X), T))] dw_{0,(a,b)}^{T,(a,b)} \end{aligned}$$

and moreover, allowing f_1 in Proposition 7.2.1 be supplied by coordinate exchange,

$$\begin{aligned} \mathcal{G}_T((a, b), (a, b)) &= \int_{E_x^c} Y(X) dw_{0, (a, b)}^{T, (a, b)} \\ &\quad + \frac{1}{2} \int_{E_x} [(Y(X_{[0, \tau(X)]}) - Y(-\tau_{[0, \tau(X)]}^x X_{[0, \tau(X)]})) \times \\ &\quad (Y(X_{[\tau(X), T]}) - Y(X_{[\tau(X), T]} + \ell((b - a, 0), \tau(X), T)_{[\tau(X), T]}))] dw_{0, (a, b)}^{T, (a, b)}. \end{aligned}$$

The reader should note that when $(\tau(X), x_1(X))$ has been found to be (t, x) , a horizontal linear map does **not** generate a sample from the measure induced on $\mathcal{C}_{0, (b, a)}^{t, x}$ by concentrating on $E^{\text{SW}}(t, x)$. This is immediately evident since it is possible that two bridges without crossings on this interval, when linearly transformed horizontally, will cross. However, it is possible to use a linear map for the y-coordinate, since so far the y coordinate is unconstrained. This is clear because then the x-coordinate of X' is still sampled from the correct measure. Let a random variable $\ell((0, y_2 - y_1), 0, \tau(X)) : [0, T] \rightarrow \mathbb{R}^4$ be given by

$$\ell((0, y_2 - y_1), 0, \tau(X); t) = \frac{t}{\tau(X)}(0, y_2 - y_1, 0, y_1 - y_2) \quad ,$$

and let $\ell((0, y_2 - y_1), T, \tau(X)) : [0, T] \rightarrow \mathbb{R}^4$ be given by

$$\ell((0, y_2 - y_1), T, \tau(X); t) = \frac{T - t}{T - \tau(X)}(0, y_2 - y_1, 0, y_1 - y_2).$$

Then

$$\begin{aligned} \mathcal{G}_T((a, b), (a, b)) &= \int_{E_x^c} Y(X) dw_{0, (a, b)}^{T, (a, b)} \\ &\quad + \frac{1}{2} \int_{E_x} [(Y(X_{[0, \tau(X)]}) - Y(X_{[0, \tau(X)]} + \ell((0, y_2 - y_1), 0, \tau(X)))) \\ &\quad \times (Y(X_{[\tau(X), T]}) - Y(X_{[\tau(X), T]} + \ell((0, y_2 - y_1), T, \tau(X))))] dw_{0, (a, b)}^{T, (a, b)}. \end{aligned}$$

In principle, this gives some scope to proceed with the kind of favourable sampling scheme already discussed in Subsection 5.3.1.

Reflection Another example of a possible choice of f_1 or f_2 is given by reflecting coordinates. For example, if we have $X = (X_1, Y_1, X_2, Y_2)$ with $X(t) = (x, y_1, x, y_2)$ then we may

construct the y-coordinate of X' on $[0, t]$ via $2l_{0, \frac{y_1+y_2}{2}} - Y_1$. The viability of this is evident as follows: suppose $Y_1 = l_{0, y_1} + Z$; then whereas linearly mapping Y_1 by adding $l_{0, y_2 - y_1}$ will give $Y'_1 = l_{0, y_2} + Z$, in this case we instead have $Y'_1 = l_{0, y_2} - Z$. We could similarly reflect the x-coordinate X'_1 from X_2 via $2l_{\frac{x_a+x_b}{2}, x} - X_2$.

Since the average distance between positive and negative contributing paths will usually be larger, it is not intuitively obvious why reflection confers greater advantages than coordinate exchange. Unlike linear translation, however, reflection does map non-crossing bridges to non-crossing bridges.

Generalising the Diamond Proposition

It is expedient to prove a generalisation of the result to the case that we stipulate only that the initial and terminal points have parallel longitudinal vectors, ie where we have $x_0 = (a_0, b_0)$ and $x_T = (a_T, b_T)$, we assume that $(b_0 - a_0) / \|b_0 - a_0\| = (b_T - a_T) / \|b_T - a_T\|$. As before we shall treat the case that the initial and terminal points are both on horizontal axes, since this always can be obtained by a rotation of coordinates. The following result shall be useful in Subsection 7.2.2. The proof is essentially similar to that of Proposition 7.2.1. We introduce the operator $\neg_x : \mathbb{R}^4 \rightarrow \mathbb{R}^4$ given by $\neg_x((a_x, a_y), (b_x, b_y)) = ((b_x, a_y), (a_x, b_y))$.

Proposition 7.2.3 (Generalised Diamond Proposition) *Let*

$$\begin{aligned} x_0 &= (a_0, b_0) = ((a_{0x}, y_0), (b_{0x}, y_0)) \quad , \\ x_T &= (a_T, b_T) = ((a_{Tx}, y_T), (b_{Tx}, y_T)) \quad . \end{aligned}$$

For $x \in R^4$ s.t. $x^{(1)1} = x^{(2)1}$, let $f_1 : \mathcal{C}^ \rightarrow \mathcal{C}^*$ be s.t. for $X \in \mathcal{C}_{0, x_0}^{t, x}|_{E(t, x)}$, $f_1(X) \in \mathcal{C}_{0, \neg_x x_0}^{t, x}|_{E(t, x)}$, and s.t. for any event $A \in \mathcal{B}(\mathcal{C}_{0, x_0}^{t, x})$, $w_{0, \neg_x x_0}^{t, x}|_{E(t, x)}(f_1(A)) = w_{0, x_0}^{t, x}|_{E(t, x)}(A)$. Let $f_2 : \mathcal{C}^* \rightarrow \mathcal{C}^*$ be s.t. for $X \in \mathcal{C}_{t; x}^{T; x_T}$, $f_2(X) \in \mathcal{C}_{t; x}^{T; \neg_x x_T}$ and s.t. for any event $A \in \mathcal{B}(\mathcal{C}_{t; x}^{T; x_T})$,*

$w_{t;x}^{T;\neg_x x_T}(f_2(A)) = w_{t;x}^{T;x_T}(A)$. Then

$$\begin{aligned} \mathcal{G}_T(x_0, x_T) &= \int_{E_x^c} Y(X) dw_{0,x_0}^{T,x_T} \\ &+ \frac{1}{2} \int_{E_x} [(Y(X_{[0,\tau(X)]}) - Y(f_1(X_{[0,\tau(X)]}))) (Y(X_{[\tau(X),T]}) - Y(f_2(X_{[\tau(X),T]})))] dw_{0,x_0}^{T,x_T}. \end{aligned} \quad (7.7)$$

Proof. By applying the definition (2.62-2.65) and (7.1), we get, analogously to (7.4):

$$\mathcal{G}_T(x_0, x_T) = \int_{E_x^c} Y(X) dw_{0,x_0}^{T,x_T} + \int_{E_x} Y(X) dw_{0,x_0}^{T,x_T} - w_{0,x_0}^{T,x_T}(E_x) \int_{E_x^{\text{sw}}} Y(X) dw_{0,x_0}^{T,\neg_x x_T}, \quad (7.8)$$

where we should recognise that quite possibly, $w_{0,x_0}^{T,x_T}(E_x) = 1$ and where $E_x^{\text{sw}} \in \mathcal{B}(\mathcal{C}_{0,(a,b)}^{T,(b,a)})$ is the Borel set consisting of paths with an x-crossing.

Note that the x-coordinates of X form a pair of 1-dimensional Brownian bridges, independent from the y-coordinates. Therefore, for the x-coordinates, the same logic applies as in the proof of Theorem 6.1.4; viz, the distribution of (τ, x_1) , the first crossing time and point in the x direction, is the same for $w_{0,x_0}^{T,\neg_x x_T}$ as for the concentration of w_{0,x_0}^{T,x_T} on E_x , so we may condition on a first crossing time and point (t, x) which is identical in law for both summands. Let η signify the measure on $[0, T] \times \mathbb{R}$ induced by the distribution of (τ, x_1) for the x-coordinate of a bridge in E_x under $w_{0,x_0}^{T,x_T} \Big|_{E_x}$. Then

$$\begin{aligned} \mathcal{G}_T(x_0, x_T) &= \int_{E_x^c} Y(X) dw_{0,x_0}^{T,x_T} + w_{0,x_0}^{T,x_T}(E_x) \mathcal{J}_T^{E_x}(x_0, x_T) - w_{0,x_0}^{T,x_T}(E_x) \mathcal{J}_T^{E_x^{\text{sw}}}(x_0, \neg_x x_T) \\ &= \int_{E_x^c} Y(X) dw_{0,x_0}^{T,x_T} \\ &+ w_{0,x_0}^{T,x_T}(E_x) \int_{[0,T] \times \mathbb{R}} [\mathcal{J}_T^{E(t,x)}(x_0, x_T) - \mathcal{J}_T^{E^{\text{sw}}(t,x)}(x_0, \neg_x x_T)] d\eta(t, x) \\ &= \int_{E_x^c} Y(X) dw_{0,x_0}^{T,x_T} \\ &+ \int_{E_x} [\mathcal{J}_T^{E(\tau(X), x_1(X))}(x_0, x_T) - \mathcal{J}_T^{E^{\text{sw}}(\tau(X), x_1(X))}(x_0, \neg_x x_T)] dw_{0,x_0}^{T,x_T}. \end{aligned} \quad (7.9)$$

In other words, since the x-crossing in the negative summand is identically distributed with that in the positive summand, we might as well use the same realisation (t, x) for both.

However, notice that

$$\begin{aligned} f_1(X_{[0,\tau(X)]}) \& X_{[\tau(X),T]} &\sim w_{0,\neg_x x_0}^{T,x_T} \\ X_{[0,\tau(X)]} \& f_2(X_{[\tau(X),T]}) &\sim w_{0,x_0}^{T,\neg_x x_T} \\ f_1(X_{[0,\tau(X)]}) \& f_2(X_{[\tau(X),T]}) &\sim w_{0,\neg_x x_0}^{T,\neg_x x_T} \Big|_{E_x^*} \end{aligned}$$

where $E_x^* \subseteq \mathcal{C}_{0,\neg_x x_0}^{T,\neg_x x_T}$ is the Borel set consisting of those paths from $\neg_x x_0$ to $\neg_x x_T$ with an x -crossing. Therefore it follows from permutational symmetry of the potential that

$$\begin{aligned} \int_{E_x} \mathcal{J}_T^{E(\tau(X),x_1(X))}(x_0, x_T) dw_{0,x_0}^{T,x_T} &= \int_{E_x} Y(X) dw_{0,x_0}^{T,x_T} \\ &= \int_{E_x} Y(f_1(X_{[0,\tau(X)]}) \& f_2(X_{[\tau(X),T]})) dw_{0,x_0}^{T,x_T} \\ \int_{E_x} \mathcal{J}_T^{E^{SW}(\tau(X),x_1(X))}(x_0, \neg_x x_T) dw_{0,x_0}^{T,x_T} &= \int_{E_x^{SW}} Y(X_{[0,\tau(X)]} \& f_2(X_{[\tau(X),T]})) dw_{0,x_0}^{T,\neg_x x_T} \\ &= \int_{E_x^{SW}} Y(f_1(X_{[0,\tau(X)]}) \& X_{[\tau(X),T]}) dw_{0,\neg_x x_0}^{T,x_T} \end{aligned}$$

and since Y is multiplicative in the sense that $Y(X) = Y(X_{[0,t]})Y(X_{[t,T]})$, it follows that

$$\begin{aligned} \mathcal{G}_T(x_0, x_T) &= \int_{E_x^c} Y(X) dw_{0,x_0}^{T,x_T} + \frac{1}{2} \int_{E_x} [Y(X) - Y(f_1(X_{[0,\tau(X)]}) \& X_{[\tau(X),T]}) \\ &\quad - Y(X_{[0,\tau(X)]} \& f_2(X_{[\tau(X),T]})) + Y(f_1(X_{[0,\tau(X)]}) \& f_2(X_{[\tau(X),T]}))] dw_{0,x_0}^{T,x_T} \\ &= \int_{E_x^c} Y(X) dw_{0,x_0}^{T,x_T} \\ &\quad + \frac{1}{2} \int_{E_x} [(Y(X_{[0,\tau(X)]}) - Y(f_1(X_{[0,\tau(X)]}))) \\ &\quad \times (Y(X_{[\tau(X),T]}) - Y(f_2(X_{[\tau(X),T]})))] dw_{0,x_0}^{T,x_T}. \end{aligned}$$

■

Remark 7.2.4 Naturally if $a_{Tx} > b_{Tx}$ and $a_{0x} < b_{0x}$ or vice versa then $w_{0,x_0}^{T,x_T}(E_x^c) = 0$.

Using first and last crossing times

In addition to being able to choose various ways of constructing X' from X , we can also create similar formulae using other x -crossing times than the first crossing time $\tau(X)$, as

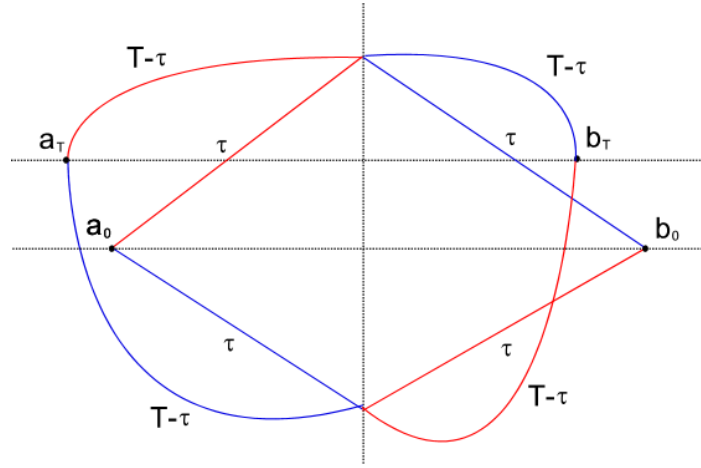


Figure 7.2: Schematic of the contributing path segments for a path in E_x , according to the Generalised Diamond Proposition.

long as we choose a crossing time that is well-defined. For example, we might use the last crossing time, or the crossing time which is nearest to $T/2$, as long as we choose bijections which conjugate the corresponding measures. We shall use τ_1 to denote the first crossing time and τ'_1 to denote the last crossing time.

In particular, it is notable that (7.2) is not symmetric in time, even in the narrow sense of looking at $[0, T]$ forwards or backwards. This seems counter-intuitive in the solution to a problem where such a time symmetry is very evident. We can further derive an expression which does conceptually have this symmetry, by combining the diamond for the first crossing with that for the last.

Overloaded notation for crossing events To achieve greater simplicity, for the remainder of this chapter we shall adopt a streamlined (ie overloaded) notation about events. From now on, we shall let $E(s, x)$ indicate the event that the *first* x -crossing takes place at time s and system position x , regardless of the path space relevant to the integral (for us the meaning shall always be clear). We shall let $E'(s, x)$ indicate the event that the *last* x -crossing takes place at time s and system position x .

Conditional independence when a last crossing time and point is specified Consider a pair of 1-dimensional Brownian bridges, $X^{(1)}$ and $X^{(2)}$. The last crossing time τ'_1

is not a Markov time and so we cannot simply invoke the strong Markov property to infer conditional independence of $X_{[0,\tau'_1]}$ and $X_{[\tau'_1,T]}$. However, it shall be important to recognise that a closely related conditional independence does apply: given $t \in (0, T)$ and $x \in \mathbb{R}$, we need to be able to say that given the event that $\tau'_1 = t$ and given that $X(t) = (x, x)$, $X_{[0,t]}$ and $X_{(t,T]}$ are independent. It is sufficient to show that $X(s_1)$ and $X(s_2)$ are independent when conditioning that $\tau'_1 = t$, for some $t \in (s_1, s_2)$, and that $X(t) = (x, x)$. We now advance three different perspectives to explain why this is so.

The most direct derivation is as follows. If we recall our definition of conditioning (2.17), it is clear that the conditional pdf for $X(s_1), X(s_2)$, at some point (z_1, z_2) , given $\tau'_1 = t$ and the last crossing point $x_2 = (x, x)$, is proportional to the conditional pdf for τ'_1, x_2 given $(X(s_1), X(s_2)) = (z_1, z_2)$, evaluated at $t, (x, x)$, multiplied with the unconditional pdf for $X(s_1), X(s_2)$. So our task is to show that this yields a product of functions of z_1 and z_2 . The joint conditional pdf of τ'_1, x_2 may be found by taking the product of the conditional pdf for τ'_1 , with the pdf for x_2 conditional on all of $X(s_1) = z_1, X(s_2) = z_2$ and $\tau'_1 = t$. Recalling the pdf for τ_1 , that is, $\frac{d\eta}{d\Lambda}$ in (6.13), it is clear from symmetry that the pdf for τ'_1 is given by evaluating this $\frac{d\eta}{d\Lambda}$ at $s_2 - t + s_1$ rather than at $t - s_1$. This forms a product of functions of z_1 and z_2 multiplied by $\exp\left(-\frac{(z_1^{(2)} - z_1^{(1)})(z_2^{(2)} - z_2^{(1)})}{2(s_2 - s_1)}\right)$. It follows that the conditional pdf for τ'_1 multiplied by the unconditional pdf for $(X(s_1), X(s_2))$ at (z_1, z_2) is a product of functions of z_1 and z_2 .

Meanwhile, conditioning on $\tau'_1 = t$ can only affect the bridge component $\frac{X^{(2)} - X^{(1)}}{\sqrt{2}}$, as previously explained; this means that conditioning on $X(s_1) = z_1, X(s_2) = z_2, \tau'_1 = t$ yields the same conditional pdf for $\frac{X^{(2)}(t) + X^{(1)}(t)}{\sqrt{2}}$ as just conditioning on $X(s_1) = z_1, X(s_2) = z_2$. This is of course a product of functions of $\frac{z_1^{(2)} + z_1^{(1)}}{\sqrt{2}}$ and $\frac{z_2^{(2)} + z_2^{(1)}}{\sqrt{2}}$. But given $\tau'_1 = t$, $x_2 = \frac{1}{\sqrt{2}}\left(\frac{X^{(2)}(t) + X^{(1)}(t)}{\sqrt{2}}, \frac{X^{(2)}(t) + X^{(1)}(t)}{\sqrt{2}}\right)$. It follows that the conditional pdf for τ'_1, x_2 given $X(s_1) = z_1, X(s_2) = z_2$ is a product of functions of z_1 and z_2 , as required.

For a less explicit but more intuitive point of view, recognise that due to continuity of paths, the event $X(t) = (x, x)$ is $\sigma(X_{[s_1, t]})$ -measurable whereas the event $\tau'_1 = t$ is $\sigma(X_{(t, s_2]})$ -

measurable. Therefore (cf (2.20)) joint conditioning is repeated conditioning. So consider what happens if we condition first on $X(t) = (x, x)$ and then on the fact that this is actually the last crossing in (s_1, s_2) . The former conditioning leads to independence of $X(s_1), X(s_2)$; the latter involves conditioning on an event which is measurable with respect to $\sigma(X_{(t, s_2)})$ and so cannot introduce any dependence.

Yet another line of reasoning is based on the symmetry involved in reversing time. Recall that where $X \sim w_{0, x_0}^{T, x_T}$, $\circlearrowleft X$ given by $\circlearrowleft X(t) = X(T - t)$ has $\circlearrowleft X \sim w_{0, x_T}^{T, x_0}$. Then for $\circlearrowleft X$ we have conditional independence, given a first crossing time τ_1° and point x_1° . But it is clear that $\tau_1' = T - \tau_1^\circ$ and $x_2 = x_1^\circ$. Consequently, $(\circlearrowleft X)_{[0, t]}$ and $(\circlearrowleft X)_{(t, T]}$ have independence conditional on $\tau_1' = T - t$ and $x_2 = x$; but this is the same thing as saying that conditionally, $X_{[0, T-t]}$ and $X_{(T-t, T]}$ are independent.

Conditional independence with first and last crossings The fact that this conditional independence applies in the 1-dimensional case implies that we may use it for each component in the multidimensional case. Moreover, since joint conditioning is repeated conditioning, we may conclude that when conditioning on both the first and last crossing times and points, the resulting conditional measure is a product of the measures relevant to the three sections $[0, \tau_1), (\tau_1, \tau_1')$ and $(\tau_1', T]$.

Proposition 7.2.5 (Second Diamond Proposition) *Let $f_1 : \mathcal{C}^* \rightarrow \mathcal{C}^*$ be s.t. for any $x \in \mathbb{R}^4$ with $x^{(1)1} = x^{(2)1}$, for $X \in \mathcal{C}_{0, (a, b)}^{t, x} \Big|_{E(t, x)}$, we shall have $f_1(X) \in \mathcal{C}_{0, (b, a)}^{t, x} \Big|_{E(t, x)}$, and for any event $A \in \mathcal{B} \left(\mathcal{C}_{0, (a, b)}^{t, x} \right)$, $w_{0, (b, a)}^{t, x} \Big|_{E(t, x)}(f_1(A)) = w_{0, (a, b)}^{t, x} \Big|_{E(t, x)}(A)$. Meanwhile let $f_2 : \mathcal{C}^* \rightarrow \mathcal{C}^*$ be s.t. for any $x' \in \mathbb{R}^4$ with $x'^{(1)1} = x'^{(2)1}$, for $X \in \mathcal{C}_{t', x'}^{T, (a, b)} \Big|_{E'(t', x')}$, we shall have $f_2(X) \in \mathcal{C}_{t', x'}^{T, (b, a)} \Big|_{E'(t', x')}$, and for any event $A \in \mathcal{B} \left(\mathcal{C}_{0, (a, b)}^{t, x} \right)$, $w_{t', x'}^{T, (b, a)} \Big|_{E'(t', x')}(f_2(A)) = w_{t', x'}^{T, (a, b)} \Big|_{E'(t', x')}(A)$. Then where τ_1 (dependence on X unannotated) is the first x -crossing*

time of X and where τ_2 is the last x -crossing time of X ,

$$\begin{aligned} \mathcal{G}_T((a, b), (a, b)) &= \int_{E_x^c} Y(X) dw_{0,(a,b)}^{T,(a,b)} \\ &+ \frac{1}{2} \int_{E_x} \left[(Y(X_{[0,\tau_1]}) - Y(f_1(X_{[0,\tau_1]}))) Y(X_{[\tau_1,\tau_2]}) \right. \\ &\quad \left. \times (Y(X_{[\tau'_1,T]}) - Y(f_2(X_{[\tau'_1,T]}))) \right] dw_{0,(a,b)}^{T,(a,b)}. \end{aligned} \quad (7.10)$$

Proof. The proof follows from the same logic as the Diamond Proposition. Let $E_x^* \in \mathcal{B}(\mathcal{C}_{0,(b,a)}^{T,(b,a)})$ be the set of paths with a crossing in x . Consider as before that when $X \sim w_{0,(a,b)}^{T,(a,b)}$ and τ_1 is the first x -crossing time and τ_2 is the last, then because $w_{0,(a,b)}^{T,(a,b)} \Big|_{E_x}, w_{0,(a,b)}^{T,(b,a)}, w_{0,(b,a)}^{T,(a,b)}, w_{0,(b,a)}^{T,(b,a)} \Big|_{E_x^*}$ all give rise to the same joint measure on first x -crossing times and last x -crossing times, and because of conditional independence,

$$\begin{aligned} f_1(X_{[0,\tau_1]}) \& X_{[\tau_1,T]} &\sim w_{0,(b,a)}^{T,(a,b)} \\ X_{[0,\tau'_1]} \& f_2(X_{[\tau'_1,T]}) &\sim w_{0,(a,b)}^{T,(b,a)} \\ f_1(X_{[0,\tau_1]}) \& X_{[\tau_1,\tau'_1]} \& f_2(X_{[\tau'_1,T]}) &\sim w_{0,(b,a)}^{T,(b,a)} \Big|_{E_x^*}. \end{aligned}$$

Therefore

$$\begin{aligned} \mathcal{G}_T((a, b), (a, b)) &= \int_{E_x^c} Y(X) dw_{0,(a,b)}^{T,(a,b)} + \\ &\frac{1}{2} \int_{E_x} \left[Y(X) - Y(f_1(X_{[0,\tau_1]})) Y(X_{[\tau_1,T]}) \right. \\ &\quad \left. - Y(X_{[0,\tau'_1]}) Y(f_2(X_{[\tau'_1,T]})) + Y(f_1(X_{[0,\tau_1]})) Y(X_{[\tau_1,\tau'_1]}) Y(f_2(X_{[\tau'_1,T]})) \right] dw_{0,(a,b)}^{T,(a,b)} \end{aligned}$$

and the result follows. ■

This result informs us that it is possible to decompose further the contributions to \mathcal{I}_T that arise from E_x : rather than splitting out only a diamond up to the first x -crossing time, we can simultaneously do the same thing for time from the last crossing onwards. This is a logical step because there are strong reasons to expect a sign problem to occur because of the longer part of time (τ_1, T) : for since this is a diamond consisting of Brownian bridges, for long times we shall find that a given value of the action Y is almost equilikely under the

positive or negative-contributing paths. Consequently, it is logical to apply the principle of Proposition 7.2.1 again to this section of time in order to ameliorate matters. Figure 7.3 illustrates this geometric meaning of Proposition 7.2.5.

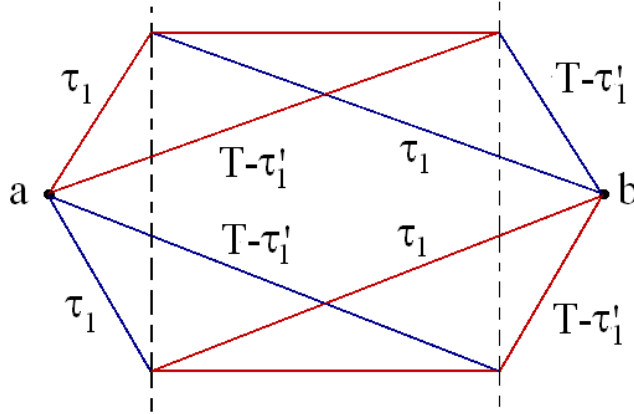


Figure 7.3: A schematic of the decomposition of $\mathcal{G}_T((a, b), (a, b))$ made available by the Second Diamond Proposition. Blue segments are negative and red segments are positive.

Proposition 7.2.5 is the basis of the Subdiamond method described in what follows.

Informal discussion of how the fermion sign problem may resurrect after the Second Diamond Proposition is used

We may consider the variance that will be obtained by using an algorithm based on (7.2) or (7.10) by conditioning on the values of X at τ_1 and τ'_1 . Again it is the integral over E_x that is of concern. According to the usual conditional variance formula (e.g. [Wil01]), we must add the variance of the conditional expectation of \mathcal{G}_T to the conditional variance of our collected functional. Naturally the law of $X(\tau_1)$ and $X(\tau'_1)$ is best considered as conditional on τ_1, τ'_1 . For T large, the distribution of first and last crossing times is converging to that seen under Brownian motion. This is a long-tailed distribution with no expectation (it is evident from Figure 7.4 that $E\tau_1 \rightarrow \infty$ as $T \rightarrow \infty$). However, for any moderate T the distribution of τ_1 is such that the bulk of probability can be thought of as close to 0. If we tentatively form the mental picture that for long times, τ_1 and $(T - \tau'_1)$ are small proportions

of T , then we may think of an increase in T as basically extending the middle section of time, $[\tau_1, \tau'_1]$ (note that this is merely an intuition - we have provided no rigorous argument that the growth of the diamonds should not be important also).

We consider first the conditional variance of \mathcal{G}_T given $\tau_1, \tau'_1, X(\tau_1), X(\tau'_1)$ and shall find that the situation looks reasonably encouraging. In integrating the functional in the E_x summand of (7.10), viz

$$F(X) = (Y(X_{[0, \tau_1]}) - Y(f_1(X_{[0, \tau_1]}))) Y(X_{[\tau_1, \tau'_1]}) (Y(X_{[\tau'_1, T]}) - Y(f_2(X_{[\tau'_1, T]}))) ,$$

we are trying to find the expectation of a product of 3 independent random variables, since the only dependence of $X_{[0, \tau_1]}$ and $X_{[\tau_1, \tau'_1]}$ is through $X(\tau_1)$, etc). Moreover, where these are denoted F_1, F_2, F_3 , we can see that $|F_1|, |F_2|, |F_3|$ are also clearly independent. Therefore we may recognise that conditional on first and last crossings,

$$\frac{EF}{E|F|} = \frac{EF_1 F_2 F_3}{E|F_1 F_2 F_3|} = \frac{EF_1}{E|F_1|} \frac{EF_2}{E|F_2|} \frac{EF_3}{E|F_3|},$$

so that in fact there is a sign problem for EF if and only if there is a sign problem for at least one of the components F_1, F_2, F_3 . If the variances of the components are thought of as small relative to expectations - then there should be no sign problem for the product. On the middle section, clearly the variance is simply the variance of a positive functional under a Brownian bridge; we have already mentioned that this grows with T polynomially relative to the expectation, under the action measure. Meanwhile, we might hope that the variance on the diamond sections $[0, \tau_1], [\tau'_1, T]$ is considerably improved from a more naïve formulation, for the following reasoning. The expected time length grows quite slowly with T (see Figure 7.4). Thinking of τ_1 and τ'_1 as fixed, we know that the x-coordinate paths are not crossing within each diamond, and so very roughly, we might therefore visualise the situation by pretending to ourselves that both paths lie entirely on their own side of the meeting point x . Then before we take y-crossings within the diamond into account, it seems that positive and negative contributions are drawn from paths which are (roughly speaking) located in different areas; intuitively then a sign problem should not result. In reality of

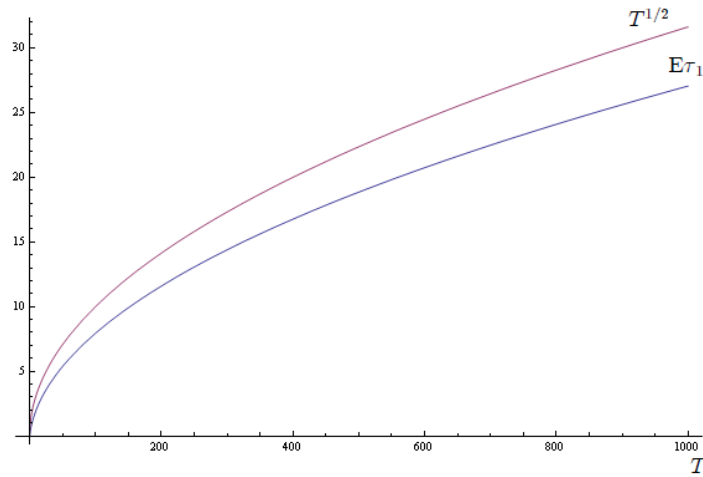


Figure 7.4: The growth of the expected first crossing time with T , taking the initial and terminal distances to be 1, compared with $T^{1/2}$.

course, the paths may cross $X_1 = x$ and $X_2 = x$ as long as they do not cross each other, and if time is projected out then we will usually see some x-crossings in the projection, which is all that our functionals depend upon. Worse, y-crossings will create a sign problem within the diamond: on an intermediate section of time, the distributions of the positive and negative paths are weakly converging; this means that almost equilikely contributions will be close in magnitude and countersigned, but may be encountered on different occasions. (ie, on one occasion, positive paths basically occupy position 1 and negative position 2; on another occasion negative may occupy position 1 and positive may occupy position 2.)

The natural conclusion is that it would be better to deal only with sections of time where either there are no crossings of x or there are no crossings of y . This is possible and is a natural development of the Subdiamond method which we shall call the *Last-to-First Subdiamond method*.

Now let us consider the variance of the conditional expectation of \mathcal{G}_T , using the functional from (7.10), given values for $X(\tau_1), X(\tau'_1)$; we shall immediately find it to be problematic.

Consider an inversion of one pair of y -values (for the sake of argument, the y -values at τ'_1). The only thing that makes inverted coordinates less than equilikely with uninverted is covariance between the y -values at τ_1 and the y -values at τ'_1 , ie when the values at τ_1 are

(y_1, y_2) with $y_1 > y_2$, if $y_3 > y_4$ then covariance makes it more likely to encounter (y_3, y_4) than (y_4, y_3) at τ'_1 . If we think of the middle section as growing with T then this covariance is fading linearly and (y_4, y_3) is becoming equilikely with (y_3, y_4) . But at the same time, the expectation of $Y(X_{(\tau_1, \tau'_1)})$ for X from (x_1, y_1, x_1, y_2) to (x_2, y_3, x_2, y_4) is becoming very close to that for X from (x_1, y_1, x_1, y_2) to (x_2, y_4, x_2, y_3) because when the time interval is long, the unconditional distribution of the paths' location on an intermediate section of time is very similar. Thus we expect to encounter a high variance of our estimate of \mathcal{I}_T because the conditional expectation of \mathcal{I}_T given almost equilikely values of $y_{1,2,3,4}$ is almost the same but countersigned. This is addressed by the Iterated Subdiamond method, described in the following subsection: *recognising that the fermion sign problem on the middle section of time is similar to the fermion sign problem found when estimating \mathcal{I}_T naïvely*, we basically proceed by iterating the Second Diamond Proposition, alternating coordinates each time.

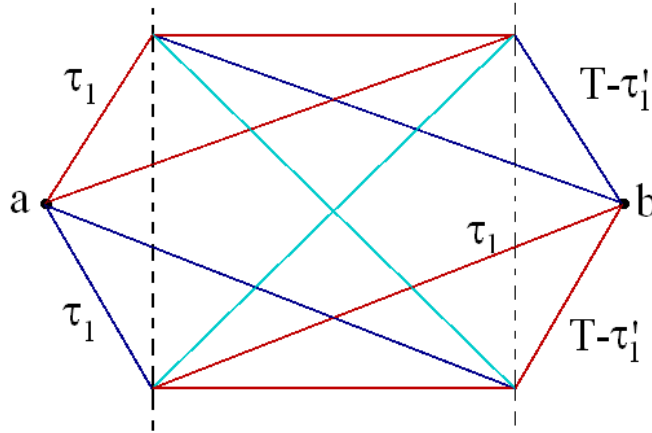


Figure 7.5: How the sign problem may reoccur after 1 iteration of the subdiamond method.

Conceptually, the Iterated Subdiamond method does not have time symmetry in the sense that if we look backwards from $T/2$ then the structure of diamonds is different than if we look forwards from 0. Thus, having proven the Diamond Proposition and then shown how to obtain greater time symmetry with the Second Diamond Proposition, we shall in the next subsection prove the validity of the Iterated Subdiamond method and then show how to

obtain greater time symmetry using the Last-to-First method. To conclude this subsection, we shall also need the following generalisation of the Second Diamond Proposition.

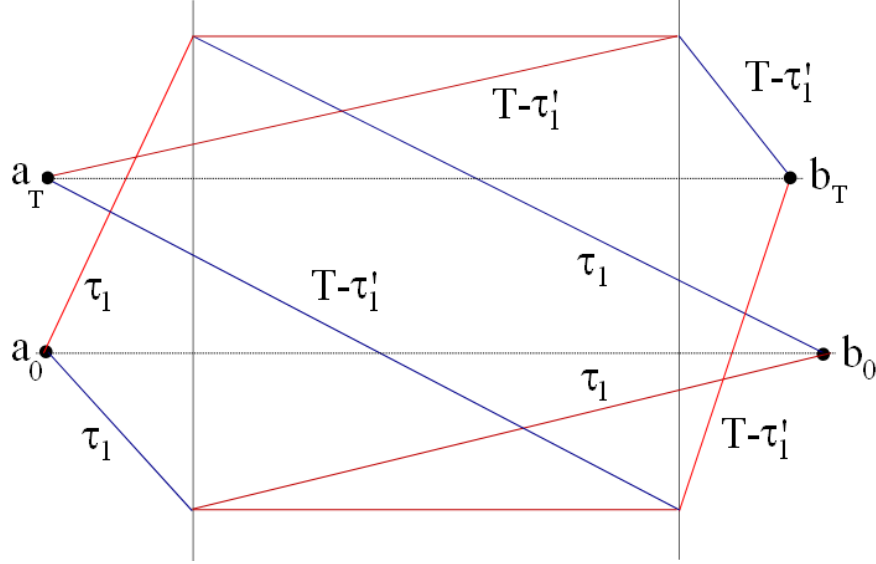


Figure 7.6: Schematic of path segments in Generalised Second Diamond Proposition.

Proposition 7.2.6 (Generalised Second Diamond Proposition) *Let*

$$x_0 = (a_0, b_0) = ((a_{0x}, y_0), (b_{0x}, y_0)) \quad ,$$

$$x_T = (a_T, b_T) = ((a_{Tx}, y_T), (b_{Tx}, y_T)) \quad .$$

Let $f_1 : \mathcal{C}^ \rightarrow \mathcal{C}^*$ be s.t. for any $x \in \mathbb{R}^4$ with $x^{(1)1} = x^{(1)2}$, for $X \in \mathcal{C}_{0,x_0}^{t,x}|_{E(t,x)}$, we shall have*

$f_1(X) \in \mathcal{C}_{0,\neg x x_0}^{t,x}|_{E(t,x)}$, and for any event $A \in \mathcal{B}(\mathcal{C}_{0,x_0}^{t,x})$, $w_{0,\neg x x_0}^{t,x}|_{E(t,x)}(f_1(A)) = w_{0,x_0}^{t,x}|_{E(t,x)}(A)$.

Let $f_2 : \mathcal{C}^ \rightarrow \mathcal{C}^*$ be s.t. for any $x' \in \mathbb{R}^4$ with $x'^{(1)1} = x'^{(1)2}$, for $X \in \mathcal{C}_{t',x'}^{T,x_T}|_{E'(t',x')}$, we shall*

have $f_2(X) \in \mathcal{C}_{t',x'}^{T,\neg x x_T}|_{E'(t',x')}$ and for any event $A \in \mathcal{B}(\mathcal{C}_{t',x'}^{T,x_T})$, $w_{t',x'}^{T,\neg x x_T}|_{E'(t',x')}(f_2(A)) = w_{t',x'}^{T,x_T}|_{E'(t',x')}(A)$. Then

$$\begin{aligned} \mathcal{G}_T(x_0, x_T) &= \int_{E_x^c} Y(X) dw_{0,x_0}^{T,x_T} \\ &+ \frac{1}{2} \int_{E_x} \left[\begin{aligned} &(Y(X_{[0,\tau_1]}) - Y(f_1(X_{[0,\tau_1]}))) Y(X_{[\tau_1,\tau_2]}) \times \\ &(Y(X_{[\tau_2,T]}) - Y(f_2(X_{[\tau_2,T]}))) \end{aligned} \right] dw_{0,x_0}^{T,x_T} . \quad (7.11) \end{aligned}$$

The proof is omitted, since it is not substantially different from the proof of the Second Diamond Proposition.

7.2.2 The iterated subdiamond approach

We have explained, in the preceding discussion, why a sign problem arises in the case of applying an estimator based on (7.10), due to the possibility of an inversion of y-coordinates at only one of τ_1, τ'_1 . We now come to show that the same procedure for reducing variance can be iterated: we can decompose the problematic middle section of (7.10) in the same way as (7.10) decomposes (7.4). We shall need to introduce the following notation.

The sequence of crossing times Let $\tau_0 = 0, \tau'_0 = T$. As mentioned, let τ_1 be the first x-crossing time, and let τ'_1 be the last x-crossing time. Let τ_2 be the first y-crossing time in (τ_1, τ'_1) , and let τ'_2 be the last y-crossing time in (τ_1, τ'_1) . For odd $i \geq 1$, let τ_i, τ'_i be the first and last x-crossing times within $(\tau_{i-1}, \tau'_{i-1})$, should there be any; and for even $i \geq 1$, let τ_i, τ'_i be the first and last y-crossing times within $(\tau_{i-1}, \tau'_{i-1})$, should there be any. If this sequence is finite, call the maximum index χ . Thus, χ is s.t. within (τ_χ, τ'_χ) there is no crossing of the coordinate corresponding to $\chi + 1$. (Thus χ is a random variable; its dependence on X is unannotated.) We shall use θ_i to represent x for i odd, and y for i even. Moreover, we use \neg_i for \neg_{θ_i} and \neg^i for \neg^{θ_i} .

We shall let $E_i(s, x)$ indicate the event that the *first* θ_i crossing takes place at time s and system position x , regardless of the path space relevant to the integral (again, the meaning shall always be clear). We shall let $E'_i(s, x)$ indicate the event that the *last* θ_i crossing takes place at time s and system position x . We let $E_i(s)$ indicate the event that the first θ_i crossing takes place at time s , and $E'_i(s)$ the event that the last θ_i crossing takes place at time s .

Measure-preserving bijections We shall let $f_i : \mathcal{C}^* \rightarrow \mathcal{C}^*$ be a measure-preserving bijection in the following sense. Let $0 \leq t_{i-1} < t_i \leq T$, $s \in [t_{i-1}, t_i]$ and let $x_{i-1}, x_i, x \in \mathbb{R}^4$, with x assigning equal θ_i coordinates to both particles. Then f shall act as a bijection from

$\mathcal{C}_{t_{i-1}, x_{i-1}}^{t_i, x_i} \Big|_{E_i(s, x)}$ to $\mathcal{C}_{t_{i-1}, x_{i-1}}^{t_i, \neg_{i-1} x_i} \Big|_{E_i(s, x)}$, and for any event $A \in \mathcal{B}(\mathcal{C}_{t_{i-1}, x_{i-1}}^{t_i, x_i})$, we require that

$$w_{t_{i-1}, x_{i-1}}^{t_i, \neg_{i-1} x_i} \Big|_{E_i(s, x)} (f_i(A)) = w_{t_{i-1}, x_{i-1}}^{t_i, x_i} \Big|_{E_i(s, x)} (A).$$

(Of course the main the interest shall be in conditioning on $E_i(t_i, x_i)$.) Meanwhile, we shall likewise let $f'_i : \mathcal{C}^* \rightarrow \mathcal{C}^*$ be a measure-preserving bijection in the following sense. Let $0 \leq t'_i < t'_{i-1} \leq T$, $s' \in [t'_i, t'_{i-1}]$ and let $x'_i, x'_{i-1}, x' \in \mathbb{R}^4$, with x' assigning equal θ_i coordinates to both particles. Then f'_i shall act as a bijection from $\mathcal{C}_{t'_i, x'_i}^{t'_{i-1}, x'_{i-1}} \Big|_{E'_i(s', x')}$ to $\mathcal{C}_{t'_i, \neg_{i-1} x'_i}^{t'_{i-1}, x'_{i-1}} \Big|_{E'_i(s', x')}$, and for any event $A \in \mathcal{B}(\mathcal{C}_{t'_i, x'_i}^{t'_{i-1}, x'_{i-1}})$, we require that

$$w_{t'_i, \neg_{i-1} x'_i}^{t'_{i-1}, x'_{i-1}} \Big|_{E'_i(s', x')} (f'_i(A)) = w_{t'_i, x'_i}^{t'_{i-1}, x'_{i-1}} \Big|_{E'_i(s', x')} (A).$$

(The main interest shall be in conditioning on $E'_i(t'_i, x'_i)$.)

Diamond notation Furthermore, in order to avoid longwindedness we shall introduce the notation

$$\begin{aligned} \diamond_i(X_{[\tau_{i-1}, \tau_i]}) &= Y(X_{[\tau_{i-1}, \tau_i]}) - Y(f_i(X_{[\tau_{i-1}, \tau_i]})) , \\ \diamond'_i(X_{[\tau'_i, \tau'_{i-1}]}) &= Y(X_{[\tau'_i, \tau'_{i-1}]}) - Y(f'_i(X_{[\tau'_i, \tau'_{i-1}]})) \end{aligned} \quad (7.12)$$

or use \diamond_i, \diamond'_i for short.

Lemma 7.2.7 (Iterated subdiamonds) Where $x_0 = ((x_{a0}, 0), (x_{b0}, 0))$,

$$\begin{aligned} \mathcal{G}_T(x_0, x_0) &= \int_{\{\chi=0\}} Y(X) dw_{0, x_0}^{T, x_0} \\ &+ \frac{1}{2} \int_{\{\chi \geq 1\}} \left[\frac{1}{4^{\chi-1}} \prod_{i=1}^{\chi} [\diamond_i(X_{[\tau_{i-1}, \tau_i]}) \diamond'_i(X_{[\tau'_i, \tau'_{i-1}]})] Y(X_{[\tau_\chi, \tau'_\chi]}) \right] dw_{0, x_0}^{T, x_0} , \end{aligned} \quad (7.13)$$

or equivalently,

$$\mathcal{G}_T(x_0, x_0) = \int_{\mathcal{C}_{0, x_0}^{T, x_0}} \left[4^{-\max(\chi - \frac{1}{2}, 0)} \prod_{i=1}^{\chi} [\diamond_i \diamond'_i] Y(X_{[\tau_\chi, \tau'_\chi]}) \right] dw_{0, x_0}^{T, x_0} .$$

Proof. Consider the statement, for a given $k \geq 1$, that

$$\begin{aligned} \mathcal{G}_T(x_0, x_0) &= \int_{\{\chi=0\}} Y(X) dw_{0,x_0}^{T,x_0} \\ &+ \frac{1}{2} \int_{\{\chi \geq 1, \chi < k\}} \left[\frac{1}{4^{\chi-1}} \prod_{i=1}^{\chi} [\diamond_i(X_{[\tau_{i-1}, \tau_i]}) \diamond'_i(X_{[\tau'_i, \tau'_{i-1}])}] Y(X_{[\tau_\chi, \tau'_\chi]}) \right] dw_{0,x_0}^{T,x_0} \\ &+ \frac{1}{2} \int_{\{\chi \geq k\}} \left[\frac{1}{4^{k-1}} \prod_{i=1}^k [\diamond_i(X_{[\tau_{i-1}, \tau_i]}) \diamond'_i(X_{[\tau'_i, \tau'_{i-1}])}] Y(X_{[\tau_k, \tau'_k]}) \right] dw_{0,x_0}^{T,x_0}. \end{aligned} \quad (7.14)$$

It follows from the Second Diamond Proposition that the statement (7.14) holds for $k = 1$. The conclusion (7.13) will hold by induction if we can establish that for $k \geq 1$, the statement (7.14) for k implies the statement (7.14) for $k + 1$. So let us consider the quantity

$$\begin{aligned} &\int_{\{\chi \geq k\}} \left[\frac{1}{4^{k-1}} \prod_{i=1}^k [\diamond_i(X_{[\tau_{i-1}, \tau_i]}) \diamond'_i(X_{[\tau'_i, \tau'_{i-1}])}] Y(X_{[\tau_k, \tau'_k]}) \right] dw_{0,x_0}^{T,x_0} \\ &= \int_{\{\chi \geq k\} \cap \boxtimes_k^c} \left[\frac{1}{4^{k-1}} \prod_{i=1}^k [\diamond_i(X_{[\tau_{i-1}, \tau_i]}) \diamond'_i(X_{[\tau'_i, \tau'_{i-1}])}] Y(X_{[\tau_k, \tau'_k]}) \right] dw_{0,x_0}^{T,x_0} \\ &+ \int_{\{\chi \geq k\} \cap \boxtimes_k} \left[\frac{1}{4^{k-1}} \prod_{i=1}^k [\diamond_i(X_{[\tau_{i-1}, \tau_i]}) \diamond'_i(X_{[\tau'_i, \tau'_{i-1}])}] Y(X_{[\tau_k, \tau'_k]}) \right] dw_{0,x_0}^{T,x_0}, \end{aligned}$$

where \boxtimes_k is defined as the event of a reverse of coordinate order: for k even, that where $X(\tau_k) = \left((x_k, y_k^{(1)}), (x_k, y_k^{(2)}) \right)$ and $X(\tau'_k) = \left((x'_k, y_k'^{(1)}), (x'_k, y_k'^{(2)}) \right)$, that $y_k^{(1)} - y_k^{(2)}$ is of different sign to $y_k'^{(1)} - y_k'^{(2)}$; and likewise for the x-coordinate in the case that k is odd. However, given an obverse (ie s.t. $X \in \boxtimes_k^c$) pair of θ_k, θ'_k , it is clear that the relative likelihood of obtaining $(\theta_k, \neg_{k-1}\theta'_k)$ is $\exp \left(-\frac{(\theta_k^{(1)} - \theta_k^{(2)})(\theta_k'^{(1)} - \theta_k'^{(2)})}{\tau'_k - \tau_k} \right)$. Therefore, by the definition of \mathcal{G}_T , it is clear then that (using conditional independence),

$$\begin{aligned} &\int_{\{\chi \geq k\}} \left[\frac{1}{4^{k-1}} \prod_{i=1}^k [\diamond_i \diamond'_i] Y(X_{[\tau_k, \tau'_k]}) \right] dw_{0,x_0}^{T,x_0} \\ &= \int_{\{\chi \geq k\} \cap \boxtimes_k^c} \left[\frac{1}{4^{k-1}} \prod_{i=1}^k [\diamond_i \diamond'_i] \mathcal{G}_{\tau'_k - \tau_k}(X(\tau_k), X(\tau'_k)) \right] dw_{0,x_0}^{T,x_0} \\ &= \int_{\{\chi = k\}} \left[\frac{1}{4^{k-1}} \prod_{i=1}^k [\diamond_i \diamond'_i] \mathcal{J}_{\tau'_k - \tau_k}^{E_0}(X(\tau_k), X(\tau'_k)) \right] dw_{0,x_0}^{T,x_0} \\ &+ \frac{1}{2} \int_{\{\chi > k\} \cap \boxtimes_k^c} \left[\frac{1}{4^{k-1}} \prod_{i=1}^{k+1} [\diamond_i \diamond'_i] \mathcal{J}_{\tau'_{k+1} - \tau_{k+1}}(X(\tau_{k+1}), X(\tau'_{k+1})) \right] dw_{0,x_0}^{T,x_0} \\ &= \int_{\{\chi = k\}} \left[\frac{1}{4^{k-1}} \prod_{i=1}^k [\diamond_i \diamond'_i] Y(X_{[\tau_k, \tau'_k]}) \right] dw_{0,x_0}^{T,x_0} \\ &+ \frac{1}{2} \int_{\{\chi > k\} \cap \boxtimes_k^c} \left[\frac{1}{4^{k-1}} \prod_{i=1}^{k+1} [\diamond_i \diamond'_i] Y(X_{[\tau_{k+1}, \tau'_{k+1}])} \right] dw_{0,x_0}^{T,x_0}, \end{aligned}$$

where the second equality holds by applying the Generalized Second Diamond Proposition to $\mathcal{G}_{\tau'_k - \tau_k}(X(\tau_k), X(\tau'_k))$. However,

$$\begin{aligned} \int_{\{\chi > k\} \cap \boxtimes_k} \left[\frac{1}{4^{k-1}} \prod_{i=1}^{k+1} [\diamond_i \diamond'_i] Y \left(X_{[\tau_{k+1}, \tau'_{k+1}]} \right) \right] dw_{0,x_0}^{T,x_0} \\ = \int_{\{\chi > k\} \cap \boxtimes_k^c} \left[\frac{1}{4^{k-1}} \prod_{i=1}^{k+1} [\diamond_i \diamond'_i] Y \left(X_{[\tau_{k+1}, \tau'_{k+1}]} \right) \right] dw_{0,x_0}^{T,x_0} \end{aligned}$$

since conditional on $X(\tau_k), X(\tau'_k)$, the distribution of $X(\tau_{k+1}), X(\tau'_{k+1})$ is the same under θ_k, θ'_k as under $\theta_k, \neg_{k-1} \theta'_k$, but \diamond'_k is antisymmetric for θ'_k and so is \diamond'_{k+1} , making the integrand symmetric. Consequently,

$$\begin{aligned} \int_{\{\chi \geq k\}} \left[\frac{1}{4^{k-1}} \prod_{i=1}^k [\diamond_i \diamond'_i] Y \left(X_{[\tau_k, \tau'_k]} \right) \right] dw_{0,x_0}^{T,x_0} \\ = \int_{\{\chi = k\}} \left[\frac{1}{4^{k-1}} \prod_{i=1}^k [\diamond_i \diamond'_i] Y \left(X_{[\tau_k, \tau'_k]} \right) \right] dw_{0,x_0}^{T,x_0} + \\ \frac{1}{4} \int_{\{\chi > k\}} \left[\frac{1}{4^{k-1}} \prod_{i=1}^{k+1} [\diamond_i \diamond'_i] Y \left(X_{[\tau_{k+1}, \tau'_{k+1}]} \right) \right] dw_{0,x_0}^{T,x_0} . \end{aligned}$$

The result follows. ■

Figures 7.7-7.9 illustrate the principle of applying the Iterated Subdiamond Lemma to avoid the sign problem.

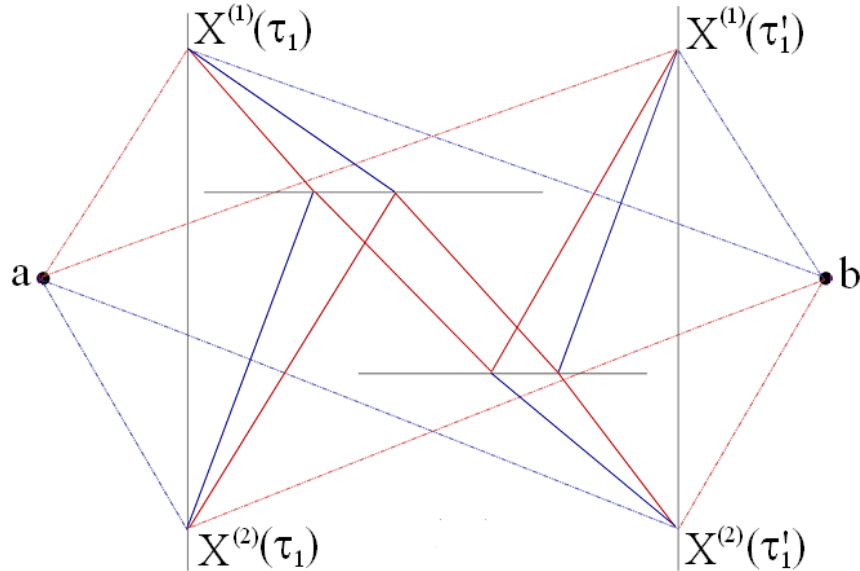


Figure 7.7: Schematic of two iterations of the subdiamond method.

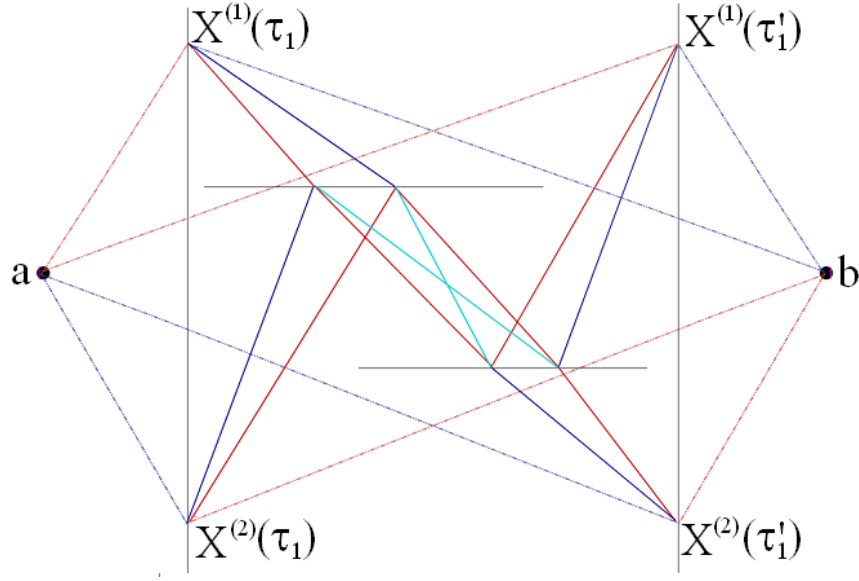


Figure 7.8: How the fermion sign problem may recur if two iterations are performed.

Remark 7.2.8 Due to continuity, the subset of $\mathcal{C}_{0,x_0}^{T,x_0}$ for which the sequence (τ_i, τ'_i) may be continued indefinitely is the set of paths with a simultaneous crossing in the x and y directions. The net contribution to $\mathcal{G}_T(x_0, x_0)$ from this set is zero because of cancellation: paths giving positive and negative contributions share the same law under this constraint. (However, it is well-known that this set is of measure zero anyway.)

Remark 7.2.9 It is valid to choose $f_i(X_{[\tau_{i-1}, \tau_i]}) = \neg^{i-1} X_{[\tau_{i-1}, \tau_i]}$, $f'_i(X_{[\tau'_i, \tau'_{i-1}]}) = \neg^{i-1} X_{[\tau'_i, \tau'_{i-1}]}$ in every case.

Based on the foregoing discussion about the problems that will be encountered on just one diamond iteration, the benefits of partitioning path space in this way are clear. Effectively, path space is partitioned here into parts enumerated by the different sequences of crossing times. The problem described before for the middle section $[\tau_1, \tau'_1]$ does not arise here for $[\tau_\chi, \tau'_\chi]$. If we take the point of view here that the τ_i, τ'_i are drawn first and then we take a product of expectations, the appropriate measure for $X_{[\tau_n, \tau'_n]}$ is now concentrated on paths which do not cross in the coordinate corresponding to $\chi + 1$. Under this condition, there is clearly no need to be concerned about a sign problem arising through a similar likelihood being attributed to $X(\tau_\chi), X(\tau'_\chi) = (x_\chi^{(1)}, x_\chi^{(2)}), (x_\chi'^{(1)}, x_\chi'^{(2)})$ as to $X(\tau_\chi), X(\tau'_\chi) =$

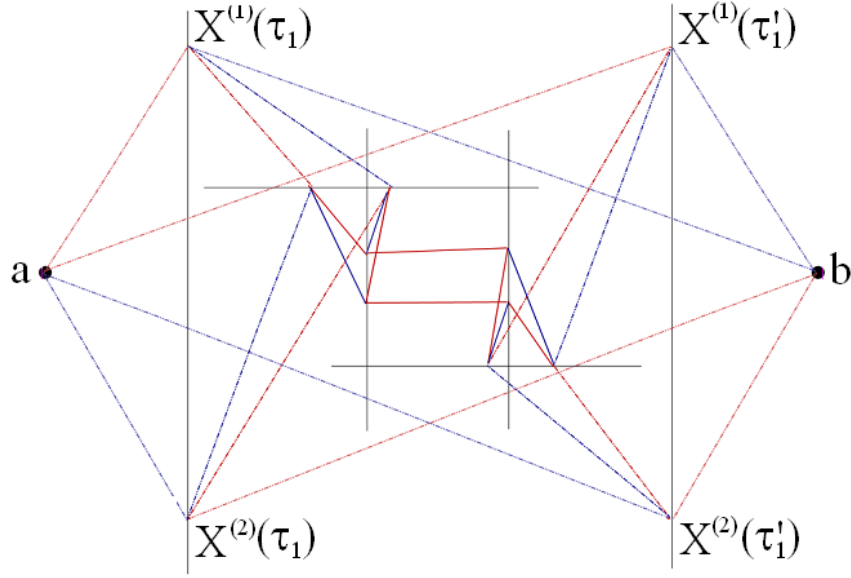


Figure 7.9: Schematic of three iterations of the subdiamond method.

$\left(x_{\chi}^{(1)}, x_{\chi}^{(2)}\right), \left(x_{\chi}^{\prime(2)}, x_{\chi}^{\prime(1)}\right)$. This is in contrast to the former case of the Second Diamond Proposition, where the corresponding measure for $X_{[\tau_1, \tau'_1]}$ is simply conditional Wiener.

The Last-to-first method

There is something different that we can do in order to reduce the variance, which becomes especially apparent once one considers the case of more than 2 particles. This is to only start diamonds from the last crossing of X^1 (ie, the x-coordinates) to the first crossing of X^2 (ie, the y-coordinates), and vice versa. This is advantageous because diamonds then take less time and promote variance less. (Moreover, this also means that diamonds may be more likely to fall in between two integration points, as discussed in the following section - thus, efficiently, cancelling contributions from a larger part of path space than before.)

More notation for crossing times Where τ_i indicates a crossing of coordinate θ_i , for $0 \leq i < \chi$, we let ς_i represent the last crossing in θ_i before τ_{i+1} . Likewise, we let ς'_i represent the first crossing in θ_i after τ'_{i+1} .

We shall now require f_i, f'_i to satisfy further conditions: we stipulate that as well as

preserving the first crossing of θ_i at the end of the interval, f_i will now also preserve the last crossing time of θ_{i-1} and the system position. That is, in our new notation, when $X \in \mathcal{C}_{t_{i-1}, x_{i-1}}^{t_i, x_i}$ is in both $E_i(s, x)$ and $E'_{i-1}(s_{i-1}, x')$, for some $s_{i-1} \in [t_{i-1}, t_i)$, $f_i(X) \in \mathcal{C}_{t_{i-1}, x_{i-1}}^{t_i, \neg_i x_i}$ shall be in both $E_i(s, x)$ and $E'_{i-1}(s_{i-1}, x')$; and moreover, we assume that for any $A \in \mathcal{B}(\mathcal{C}_{t_{i-1}, x_{i-1}}^{t_i, x_i})$,

$$w_{t_{i-1}, x_{i-1}}^{t_i, \neg_i x_i} \Big|_{E_i(s, x) \cap E'_{i-1}(s_{i-1}, x')} (f_i(A)) = w_{t_{i-1}, x_{i-1}}^{t_i, x_i} \Big|_{E_i(s, x) \cap E'_{i-1}(s_{i-1}, x')} (A) .$$

Subject to assuming that in fact t_{i-1} is a crossing of θ_{i-1} , one example of such an f_i is, once again, that we apply \neg^{i-1} ; that is, we simply exchange the θ_{i-1} coordinate paths of the particles. This shall be the only situation which we shall need to worry about. Meanwhile, f'_i is assumed to satisfy similar properties with regard to $E_{i-1}(s'_{i-1}, x') \cap E'_i(s, x)$.

The reader should note that *the meaning of \diamond_i (cf (7.12)) is enhanced accordingly*.

We are now in a position to prove the following theorem, which forms the basis of the Last-to-first Subdiamonds algorithm.

Theorem 7.2.10 (Last-to-first Subdiamonds Theorem) *Where*

$$x_0 = ((x_{a0}, 0), (x_{b0}, 0)),$$

$$\begin{aligned} \mathcal{G}_T(x_0, x_0) &= \int_{\{\chi=0\}} Y(X) dw_{0, x_0}^{T, x_0} \\ &+ \frac{1}{2} \int_{\{\chi \geq 1\}} \left[\frac{1}{4^{\chi-1}} \prod_{i=1}^{\chi} [Y(X_{[\tau_{i-1}, s_{i-1}]}) \diamond_i (X_{[s_{i-1}, \tau_i])} \right. \\ &\quad \left. \diamond'_i (X_{[\tau'_i, s'_{i-1}])} Y(X_{[s'_{i-1}, \tau'_{i-1}]) \right] Y(X_{[\tau_\chi, \tau'_\chi]}) dw_{0, x_0}^{T, x_0} . \quad (7.15) \end{aligned}$$

Proof. Due to conditional independence, it follows from Lemma 7.2.7 that

$$\begin{aligned}
\mathcal{G}_T(x_0, x_0) &= \int_{\{\chi=0\}} Y(X) dw_{0,x_0}^{T,x_0} \\
&+ \frac{1}{2} \int_{\{\chi \geq 1\}} \left[\frac{1}{4^{\chi-1}} \prod_{i=1}^{\chi} \left[\left(\mathcal{J}_{\tau_i - \tau_{i-1}}^{E'_{i-1}(\varsigma_{i-1}, X(\varsigma_{i-1})) \cap E_i(\tau_i, X(\tau_i))} (X(\tau_{i-1}), X(\tau_i)) \right. \right. \right. \\
&\quad \left. \left. - \mathcal{J}_{\tau_i - \tau_{i-1}}^{E'_{i-1}(\varsigma_{i-1}, X(\varsigma_{i-1})) \cap E_i(\tau_i, \neg_{i-1} X(\tau_i))} (X(\tau_{i-1}), \neg_{i-1} X(\tau_i)) \right) \right. \\
&\quad \left. \times \left(\mathcal{J}_{\tau'_i - \tau'_{i-1}}^{E_{i-1}(\varsigma'_{i-1}, X(\varsigma'_{i-1})) \cap E'_i(\tau'_i, X(\tau'_i))} (X(\tau'_i), X(\tau'_{i-1})) \right. \right. \\
&\quad \left. \left. - \mathcal{J}_{\tau'_i - \tau'_{i-1}}^{E_{i-1}(\varsigma'_{i-1}, X(\varsigma'_{i-1})) \cap E'_i(\tau'_i, \neg_{i-1} X(\tau'_i))} (\neg_{i-1} X(\tau'_i), X(\tau'_{i-1})) \right) \right] Y(X_{[\tau_\chi, \tau'_\chi]}) \right] dw_{0,x_0}^{T,x_0}. \quad (7.16)
\end{aligned}$$

However, in view of the new definition of f_i , it is evident that where

$$X_{[t_{i-1}, t_i]}^\heartsuit \sim w_{t_{i-1}, x_{i-1}}^{t_i, x_i} \Big|_{E'_{i-1}(s_{i-1}, x') \cap E_i(t_i, x_i)} \quad ,$$

we have

$$X_{[t_{i-1}, s_{i-1}]}^\heartsuit \ \& \ f_i \left(X_{[s_{i-1}, t_i]}^\heartsuit \right) \sim w_{t_{i-1}, x_{i-1}}^{t_i, \neg_{i-1} x_i} \Big|_{E'_{i-1}(s_{i-1}, x') \cap E_i(t_i, x_i)} \quad ;$$

and likewise where

$$X_{[t'_i, t'_{i-1}]}^\heartsuit \sim w_{t'_i, x'_i}^{t'_{i-1}, x'_{i-1}} \Big|_{E_{i-1}(s'_{i-1}, x') \cap E'_i(t'_i, x'_i)} \quad ,$$

we have

$$f'_i \left(X_{[t'_i, s'_{i-1}]}^\heartsuit \right) \ \& \ X_{[s'_{i-1}, t'_{i-1}]}^\heartsuit \sim w_{t'_i, \neg_{i-1} x'_i}^{t'_{i-1}, x'_{i-1}} \Big|_{E_{i-1}(s'_{i-1}, x') \cap E'_i(t'_i, \neg_{i-1} x'_i)} \quad .$$

Consequently, for $x, x_{i-1}, x_i \in \mathbb{R}^4$, and $t_i, t_{i-1}, s_{i-1} \in (0, T]$,

$$\begin{aligned}
&\mathcal{J}_{t_i - t_{i-1}}^{E'_{i-1}(s_{i-1}, x') \cap E_i(t_i, x_i)}(x_{i-1}, x_i) - \mathcal{J}_{t_i - t_{i-1}}^{E'_{i-1}(s_{i-1}, x') \cap E_i(t_i, \neg_{i-1} x_i)}(x_{i-1}, \neg_{i-1} x_i) \\
&= \mathcal{J}_{s_{i-1} - t_{i-1}}^{E_i^c}(x_{i-1}, x') \left(\mathcal{J}_{t_i - s_{i-1}}^{E'_{i-1}(s_{i-1}, x') \cap E_i(t_i, x_i)}(x', x_i) \right. \\
&\quad \left. - \mathcal{J}_{t_i - s_{i-1}}^{E'_{i-1}(s_{i-1}, x') \cap E_i(t_i, \neg_{i-1} x_i)}(x', \neg_{i-1} x_i) \right) \quad , \quad (7.17)
\end{aligned}$$

and likewise

$$\begin{aligned}
&\mathcal{J}_{t'_{i-1} - t'_i}^{E_{i-1}(s'_{i-1}, x') \cap E'_i(t'_i, \neg_{i-1} x'_i)}(x'_i, x'_{i-1}) - \mathcal{J}_{t'_{i-1} - t'_i}^{E_{i-1}(s'_{i-1}, x') \cap E'_i(t'_i, \neg_{i-1} x'_i)}(\neg_{i-1} x'_i, x'_{i-1}) \\
&= \mathcal{J}_{s'_{i-1} - t'_i}^{E_i^c}(x'_i, x') \left(\mathcal{J}_{t'_{i-1} - s'_{i-1}}^{E_{i-1}(s'_{i-1}, x') \cap E'_i(t'_i, \neg_{i-1} x'_i)}(x, x_i) \right. \\
&\quad \left. - \mathcal{J}_{t'_{i-1} - s'_{i-1}}^{E_{i-1}(s'_{i-1}, x') \cap E'_i(t'_i, \neg_{i-1} x'_i)}(\neg_{i-1} x'_i, x'_{i-1}) \right) \quad . \quad (7.18)
\end{aligned}$$

By applying the relations (7.17), (7.18) in (7.16), the result follows. \blacksquare

Remark 7.2.11 *Despite the route that we have taken towards this result, we naturally may now recognise that it has a geometrical meaning all its own, if we instead look at the sequence of crossings incrementally. Suppose we continue to call the first crossing of x by τ_1 , and continue to use τ_2 for the first subsequent crossing of y , but carry on with this labelling until $\tau_{2\chi}$, which is the first crossing of y that is not followed by a crossing in x . Likewise we may define ς_i as before for $0 \leq i < 2\chi$ as the last crossing of coordinate θ_i before τ_{i+1} and define $\varsigma_{2\chi} = T$. Then it follows immediately from (7.15) that*

$$\mathcal{G}_T(x_0, x_0) = \int_{\mathcal{C}_{0,x_0}^{T,x_0}} \prod_{i=0}^{2\chi-1} \left[\frac{1}{2} Y(X_{[\tau_i, \varsigma_i]}) \diamond_i (X_{[\varsigma_i, \tau_{i+1}]}) \right] Y_{[\tau_{2\chi}, \varsigma_{2\chi}]} dw_{0,x_0}^{T,x_0} . \quad (7.19)$$

7.3 The Last-to-first Subdiamond algorithm and its implementation

In this section, we shall discuss the practical implications of the theory that has been developed in the previous section. In particular, we detail exactly how the expression (7.15) is to be utilised in performing a Monte Carlo simulation to give information about the expectation of an observable.

7.3.1 Notes on implementation

The most simple way to construct the algorithm is based on (7.19). We first sample a set of discretisation points X_k according to $\tilde{w}_{0,x_0}^{T,x_0}$. Supposing that we intend to use Simpson's Rule (cf (3.40)) for $S(X)$, we are going to need to use the values of $V(X(t_k))$ at $t_k = \frac{kh}{2}, k = 0, \dots, 2N$, so all these points need to be included in our sampled discretisation. Conceptually, we might say that conditional on $\{X_k\}_{k=0}^{2N}$, we then sample the sequence of crossing times τ_i and ς_i , where $0 \leq i < 2\chi$. (We do not need to draw the crossing points; indeed, we shall see shortly that we do not really even have to draw the times).

In essence, the idea is then to collect the approximate functionals $\tilde{Y}(X_{[\tau_i, \varsigma_i]})$ and $\tilde{\diamond}_i(X_{[\varsigma_i, \tau_{i+1}]})$, and multiply. We use \neg^{i-1} as f_i in every case. However, it should be noted that applying

Simpson's Rule means that a technical point is being elided here. More accurately, we should say that (7.19) may be regarded as giving us a linear combination of terms of the form $\tilde{Y}(X^\neg)$ for some X^\neg which is generated from X by some application of coordinate exchanges at each timeslice. For each of these terms, the sequence of quadrature coefficients is the same: $(\alpha_k)_{k=0}^{2N} = (\frac{h}{6}, \frac{2h}{3}, \frac{h}{3}, \frac{2h}{3}, \frac{h}{3}, \dots, \frac{h}{3}, \frac{2h}{3}, \frac{h}{6})$. This means that we *can* proceed by collecting multiplicand approximations $\tilde{Y}(X_{[\tau_i, \varsigma_i]}^\neg)$, $\tilde{\diamond}_i(X_{[\varsigma_i, \tau_{i+1}]}^\neg)$ and multiplying, but need to recognise that in generating these approximations, the coefficients that apply are the α_k corresponding to the timeslices involved. Corresponding to each of the component paths X^\neg , this means that we are collecting a product of approximate exponentiated actions,

$$\begin{aligned} \tilde{Y}(X^\neg) &= \prod_{i=0}^{2\chi-1} \left[\exp\left(-\tilde{S}(X_{[\tau_i, \varsigma_i]}^\neg)\right) \exp\left(-\tilde{S}(X_{[\varsigma_i, \tau_{i+1}]}^\neg)\right) \right] \exp\left(-\tilde{S}(X_{[\tau_{2\chi}, \varsigma_{2\chi}]}^\neg)\right) \\ &= \exp\left(-\sum_{k=0}^{2N} \alpha_k V(X_k^\neg)\right). \end{aligned}$$

(Of course, $X_{[\tau_i, \varsigma_i]}^\neg = X_{[\tau_i, \varsigma_i]}$ for all the paths we are designating as X^\neg .)

However, it should now be clear that there is no reason to actually sample the sequence of ς_i and τ_i , even though this is possible, and indeed a version of the algorithm using the approach of drawing ς_i and τ_i was programmed, for verification purposes. (Because it is very useful, we explain how to draw the crossing time sequence, in Subsection 7.3.4.) In order to collect $\tilde{Y}(X_{[\tau_i, \varsigma_i]}^\neg)$, $\tilde{\diamond}_i(X_{[\varsigma_i, \tau_{i+1}]}^\neg)$, we need only to know what discretisation timeslices lie within $[\tau_i, \varsigma_i]$ and $[\varsigma_i, \tau_{i+1}]$. For this purpose, it is sufficient to determine which intervals (t_k, t_{k+1}) contain x-crossings and/or y-crossings. Moreover, in view of the cancellation property which we shall discuss next, we do not even have to worry at all about what happens if many x and y crossings occur within the same (t_k, t_{k+1}) , making the task very straightforward.

Pseudocode for the algorithm shall be provided in Subsection 7.3.3.

Advantageous cancellation property

Using Simpson's Rule, or any other integrator based on quadrature of V , we may recognise that when both x and y cross between two discretisation times (for these purposes, by a "discretisation time", we mean any point in time at which V is evaluated), the total

contribution to our Monte Carlo for \mathcal{G}_T shall be zero. This is because this means that for some i and k we have $[\varsigma_i, \tau_{i+1}] \subset (t_{k-1}, t_k)$, so that $\tilde{\Diamond}_i(X_{[\varsigma_i, \tau_{i+1}]})$ is zero: $[\varsigma_i, \tau_{i+1}]$ contains no discretisation timeslices, so $\tilde{\Diamond}_i(X_{[\varsigma_i, \tau_{i+1}]}) = 1 - 1 = 0$.

The advantage of this interaction between the integration method and the manner of path sampling is thus considerable: on a substantial part of path space, we are able to get a functional contribution of zero. By using the Last-to-first Subdiamonds method, we identify where there are equally likely paths which would give exactly equal countersigned contributions under the approximation used to produce the estimate of the functional. The advantage of this cancellation grows as T increases: whereas as the situation would ordinarily be getting worse because of the propensity for more crossings to occur, for us this can only improve matters!

Remark 7.3.1 *We note that in the program, we have no need to sample from the finite-dimensional distribution for $w_{0,(a_0,b_0)}^{t,((x,y_1),(x,y_2))} \Big|_{E(t,x)}$ directly. According to (2.17) this would be possible to do. For $0 < s_1 < s < s_2 < T$, when $X(s_1)$ and $X(s_2)$ are known, the conditional density of $X(s)$ may be found by conditioning on the event that no x -crossing happens within (s_1, s_2) , by the same kind of reasoning as for $w_{0,a_0}^{T,a_T} \Big|_{E_0}$ in the 1D case. For the case that no extra information about points between s and t is known, however, we would have to condition on (t, x) being the first crossing subsequent to s , and this is a slightly more awkward conditioning.*

7.3.2 Further developments that give rise to the full algorithm

It now behoves us to introduce certain straightforward extensions to the algorithm, to complete our discussion of this topic.

Using native coordinates

For exposition we have adopted the perspective, in the previous section, that, by using longitudinal-transverse coordinates, it is sufficient to treat the case $x_0 = ((x_{a0}, 0), (x_{b0}, 0))$.

However, we can also apply the same kind of thinking without changing the coordinate basis. This is important since as the number of particles is increased and dimension is increased much beyond 2, it is impossible to find coordinate bases which have the same special properties. In order to avoid looking at the problem with $n = 2, d = 2$ in the light of tricks which apply only to this relatively simple case, it seems preferable to perform simulations with diamonds in native coordinates. The only difference is that now we do not have an automatic y -crossing at time 0 and time T . Therefore we search for the first crossing in either x or y , and this we may call our τ_0 ; we proceed to use diamonds on the intervals which lie between a crossing of x and a crossing of y , as before. The resulting disposition of diamonds is illustrated in Figure 7.10; here \diamond indicates where, given this sequence of crossing points, a diamond is contributed to the product of expectations for the corresponding interval, and bold lines indicate where instead, a simple positive contribution is collected.

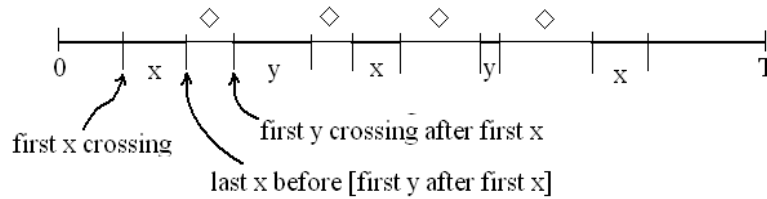


Figure 7.10: Diagram of how diamonds occur in the last-to-first method, using the native coordinate basis.

Extending to the case of two fermions in a 3-dimensional space

In the case of a pair of fermions in a 3-dimensional space, similar logic to that of Section 7.2 leads to a method where contributions are collected from paths which are still based on exchanging coordinates at crossing times. Now, however, we need to apply dimensional coordinate exchange over "diamond" intervals in which we have a crossing in each coordinate: x, y, z .

However, this means that one cannot necessarily use every possible "diamond" when

there are 3 dimensions. This is illustrated by the following example. Suppose the sequence of crossings is as shown in Figure 7.11. The connections marked (i), (ii) and (iii) illustrate possible 3-way diamonds which we might choose to use. While (iii) appears inconsistent with our methodology thus far (why would we wait until the last y-crossing to end the diamond?), both (i) and (ii) are consistent with it. Intuitively, (i) seems the best in this

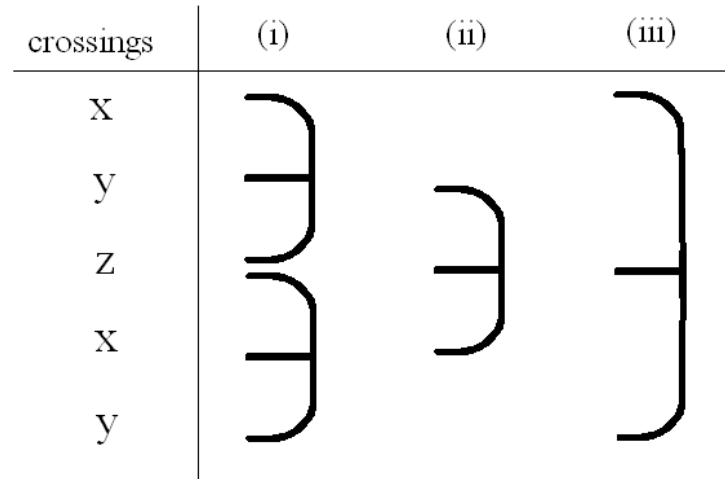


Figure 7.11: Illustration of some ideas for placing 3-way diamonds given a sequence of crossing dimensions

case, given no information about the time distances involved. It seems clear that the more time is left diamond-free and contains crossings, intuitively the more likely it is that a countersigned sample of similar magnitude could exist. However, in general, when there is a long sequence of crossings, it would be troublesome to perform a sorting to find out which arrangement gave the greatest number of diamonds. Consequently, in experiments (see Section 7.4) an algorithm was used which simply started and ended diamonds as soon as possible, incrementally. (In the case of Figure 7.11, this means that we would be using alternative (i).) Note that in each diamond there is still only one negative and one positive contribution: the negative from eventually exchanging coordinates in all dimensions, the positive from not exchanging them in any of the dimensions.

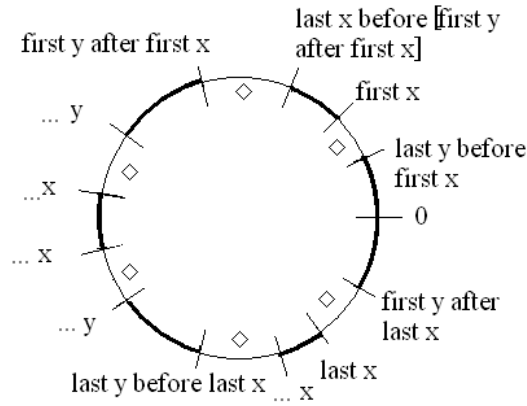


Figure 7.12: Diagram of how diamonds occur in the Last-to-first method when time is a circle.

Identifying the endpoints of time

As before it is desirable to perform a Markov Chain Monte Carlo in order to apply importance sampling for X . Now that we have replaced integration against Y with integration against the functional being integrated in the right-hand-side of (7.19), which we may call Y^\diamond , it would make sense to set the Radon-Nikodym derivative of the sampling measure with respect to w^* equal to $|Y^\diamond(X)|$. In order to apply the Multilevel Metropolis method, we need to regard time 0 and time T as identified. This does not stop us from still recognising the ordering of crossing times. Figure 7.12 displays how the situation of Figure 7.10 may be understood in this case.

However, Figure 7.12 also prompts us to consider whether we should be distinguishing time 0 by not placing a diamond over it, given that we no longer work in a coordinate basis with transverse crossings at time 0. The answer is that we can, without difficulty, place a diamond here, and contributions then also accrue to the alternative initial point. This gives rise to an approach in which all timeslices are treated the same way, which is then a perfect situation for performing a Markov Chain Monte Carlo simulation.

7.3.3 The Last-to-first algorithm in pseudocode

For understanding, the following pseudocode describes a direct path sampling algorithm based on (7.15), to find the ensemble expectation of some function of system position (e.g. potential energy). Simpson's Rule is used for Y , although naturally this is an inessential detail. We apply the diamonds using native coordinates rather than longitudinal-transverse coordinates.

```
void main procedure
```

```
{
```

```
    Set up coefficient array  $\alpha$ , with  $T/\text{halfh}+1$  elements, as 1,4,2,4,2,...,4,2,4,1 times
    halfh/3
```

```
    Loop  $M$  times:
```

```
    {
```

```
        Draw initial point  $x_0$ , according to some predetermined measure  $\mu$ 
```

```
         $Y_{\text{mc}} = \text{Ysamples}(x_0) / \text{Radon-Nikodym derivative of } \mu \text{ at } x_0.$ 
```

```
         $F_{\text{mc}} = A(x_0)Y_{\text{mc}}$ 
```

```
        Store contributions to the sample averages of  $Y_{\text{mc}}, F_{\text{mc}}, Y_{\text{mc}}^2, F_{\text{mc}}Y_{\text{mc}}, F_{\text{mc}}^2$ 
```

```
    };
```

```
    Compute sample estimate of Monte Carlo covariance matrix
```

```
    (e.g. sample  $\text{cov}(Y_{\text{mc}}, F_{\text{mc}})$  is based on sample averages of  $F_{\text{mc}}, Y_{\text{mc}},$  and  $F_{\text{mc}}Y_{\text{mc}}$ ).
```

```
    Write output to file
```

```
}
```

```
subroutine: double Ysamples (initial point  $x_0$ )
```

```
{
```

Step 1. Sample a sequence of T/halfh system positions, ie points in \mathbb{R}^4 , sampled as per the finite-dimensional distributions of a Brownian bridge: from x_0 back to x_0 .

We shall store the x and y components in separate arrays, $Xx[][]$ and $Xy[][]$, where the first subscript is for which particle and the second is for which timeslice.

Step 2a. Populate two arrays, to record which time intervals are considered to contain crossings, in x and in y respectively.

Loop through timeslices with index k ,

{

Set $p_x = \exp(-(Xx[0][k]-Xx[1][k])*(Xx[0][k+1]-Xx[1][k+1])/halfh)$;

Set $p_y = \exp(-(Xy[0][k]-Xy[1][k])*(Xy[0][k+1]-Xy[1][k+1])/halfh)$;

Generate a random number $u \sim U[0,1]$. Iff $u < p_x$ then record that this interval (ie between k and $k + 1$) has a crossing in x.

Generate a random number $u \sim U[0,1]$. Iff $u < p_y$ then record that this interval (ie between k and $k + 1$) has a crossing in y.

If the interval between k and $k + 1$ now has both x and y crossings, **return**
 $Y^\diamond = 0$.

};

Step 2b. (If either there are no crossings in x, or no crossings in y, then we can skip this step.) We now populate arrays to index which timeslices are at the beginning and end of each diamond. One array stores 1 for timeslices which are the first within a diamond, 0 otherwise. Another array stores 1 for timeslices which are the last within a diamond, 0 otherwise. To begin with, we seek the timeslice before the first crossing of either x or y (call it θ); then we seek the first crossing in the other coordinate (θ'); the timeslice previous to this is the last in a diamond. We then seek backwards to find

the timeslice after the last crossing of θ ; this is the first in a diamond. We then seek the first crossing of θ following the first crossing of θ' and so on.

Step 3. Collect estimate of Y^\diamond .

$$Y^\diamond = 1$$

Counting through timeslices (index k):

{

If $k = 0$ or k is between diamonds,

{

$$S_{\text{section}} = 0$$

Count on with k until k is at the start of a diamond or $k = T/\text{halfh}$:

{

Add to S_{section} this: $\alpha[k]*V(X_x[0][k], X_y[0][k], X_x[1][k], X_y[1][k])$

};

Multiply Y^\diamond by $\exp(-S_{\text{section}})$

};

If k is now at the start of a diamond,

{

$$S_{\text{positive}} = 0$$

$$S_{\text{negative}} = 0$$

Count through timeslices k until k reaches the end of the diamond:

{

Add to S_{positive} this: $\alpha[k]*V(X_x[0][k], X_y[0][k], X_x[1][k], X_y[1][k])$

Add to S_{negative} this: $\alpha[k]*V(X_x[0][k], X_y[1][k], X_x[1][k], X_y[0][k])$

}

Multiply Y^\diamond by $(\exp(-S_{\text{positive}}) - \exp(-S_{\text{negative}}))/2$.

}

}

If any diamonds were encountered, now multiply Y^\diamond by 2, to compensate for fact that we do not need 1/2 factor on first diamond.

return Y^\diamond

7.3.4 Drawing from the distributions of first and last crossing times

Drawing first and last crossing times for a pair of fermions

We earlier made some discussion of exact methods of sampling draws of a random variable when the pdf is known. We described rejection (this includes adaptive rejection, ziggurat methods and so on), volume decomposition, and the transformation method.

In order to sample the sequence of crossing times ς_i, τ_i , we need to sample crossing times in two distinct situations. One is where the termini are different and the other is where the termini are equal. For an interval of length T_0 , once can obtain that the pdfs, respectively, are as follows. In the case that the initial distance is Δ_1 and the terminal distance is Δ_2 :

$$f(s) = \frac{\Delta_1}{(4\pi)^{1/2}} \frac{T_0^{1/2}}{s^{3/2}(T_0 - s)^{1/2}} \exp\left(-\frac{\Delta_1^2}{4s} - \frac{\Delta_2^2}{4(T_0 - s)} + \frac{(\Delta_1 + \Delta_2)^2}{4T_0}\right)$$

In the case that the termini are equal (ie $\Delta_2 = 0$):

$$\begin{aligned} f(s) &= \frac{\Delta_1}{(4\pi)^{1/2}} \frac{T_0^{1/2}}{s^{3/2}(T_0 - s)^{1/2}} \exp\left(-\frac{\Delta_1^2}{4s} + \frac{\Delta_1^2}{4T_0}\right) \\ &= \frac{\Delta_1}{(4\pi)^{1/2}} \frac{T_0^{1/2}}{s^{3/2}(T_0 - s)^{1/2}} \exp\left(-\frac{\Delta_1^2(T_0 - s)}{4sT_0}\right) \end{aligned}$$

However, it turns out that especially the former case presents substantial difficulties if one attempts to use either rejection or volume decomposition. For volume decomposition, information about the number of modes is needed; to form a candidate for rejection, at least this information is needed, and usually one also expects to know the global maximum of the pdf, and for this function it is not trivial to compute analytically. (It is possible to find it for the other case by solving a cubic equation.) Thus it is highly fortuitous that in the more difficult case, it is possible to find a superior method entirely: it can be obtained via transformation from a Gaussian draw.

A transformation method for drawing from the first crossing time pdf when the points are constrained to meet at a time T_0 . Now where τ is the first crossing time, let $\Upsilon = \left(\frac{\Delta_1^2}{4\tau} - \frac{\Delta_1^2}{4T_0} \right)^{1/2}$. We can show that the pdf of S , for $S > 0$, is given by

$$\frac{d}{dx} P(\Upsilon < x) = \frac{2}{\sqrt{\pi}} \exp(-x^2) \quad .$$

For if we assume this then we shall find that where $x \equiv x(s) = \left(\frac{\Delta_1^2(T_0-s)}{4sT_0} \right)^{1/2} = \left(\frac{\Delta_1^2}{4s} - \frac{\Delta_1^2}{4T_0} \right)^{1/2}$, a monotonic function of s ,

$$\begin{aligned} \frac{d}{ds} P(\tau < s) &= \frac{dx}{ds} \frac{d}{dx} P(\Upsilon > x(s)) = -\frac{dx}{ds} \frac{d}{dx} P(\Upsilon < x) \\ &= \frac{1}{2x} \left(\frac{\Delta_1^2}{4s^2} \right) \frac{2}{\sqrt{\pi}} \exp(-x^2) \\ &= \frac{\Delta_1}{4s^2} \frac{2s^{1/2}T_0^{1/2}}{(T_0-s)^{1/2}} \frac{1}{\sqrt{\pi}} \exp\left(-\frac{\Delta_1^2(T_0-s)}{4sT_0}\right) \end{aligned}$$

as required. Therefore to draw samples of τ , we may draw samples of Υ , which is easy, and then apply $s = \frac{\Delta_1^2}{4} \left(x^2 + \frac{\Delta_1^2}{4T_0} \right)^{-1}$. In fact it is then clear that to draw Υ as standard Gaussian instead would be sufficient.

It was also found that this creates an extremely expedient candidate for a rejection method in the case that the termini are not equal.

In order to produce programs based on sampling the crossing time sequence, it is sometimes also necessary to be able to draw crossing points, such as $X(\tau_i)$, in order to iteratively

sample further crossing times, but it is not difficult to sample $X(\tau_i)$ conditional on the sampled value of τ_i .

Drawing first and last crossing times in the case $n > 2$

Knowing how to sample first and last crossing times in an interval (in fact, the sequence of distinct crossings) becomes more important when there are more than two particles, if we use the method suggested in Section 7.5. In order to illustrate how to proceed, we now consider the case that $n = 3$. Again we are concerned with crossings between the coordinate paths corresponding to one dimension, as we may find the x-crossing, y-crossing, z-crossing sequences separately.

We shall use E_3 to denote the event that $X^{(1)} > X^{(2)} > X^{(3)}$ over the whole interval $[t_{k-1}, t_k]$.

Writing

$$\theta_1 = \frac{X^{(1)} + X^{(2)} + X^{(3)}}{3}; \theta_2 = X^{(2)} - \frac{X^{(1)} + X^{(3)}}{2}; \theta_3 = \frac{X^{(1)} - X^{(3)}}{2}$$

then we can see that for any t , $\theta_{1,2,3}(t)$ are pairwise independent random variables. Moreover, $X^{(1)}(t) > X^{(2)}(t) > X^{(3)}(t)$ is equivalent to

$$|\theta_2(t)| < \theta_3(t) .$$

Let us set

$$\xi_1 = \frac{X^{(1)} + X^{(2)} + X^{(3)}}{\sqrt{3}}; \xi_2 = \frac{X^{(2)} - \frac{X^{(1)} + X^{(3)}}{2}}{\sqrt{3/2}}; \xi_3 = \frac{X^{(1)} - X^{(3)}}{\sqrt{2}} .$$

Then $X^{(1)}(t) > X^{(2)}(t) > X^{(3)}(t)$ is equivalent to

$$\xi_2(t) < \sqrt{3} |\xi_3(t)| .$$

The path X is determined by the path ξ , which also consists of three independent Brownian bridges with the same variance as the originals. The constraint for ξ represented by E_3 is, however, simpler than that for X . In fact its locus at any time instant is an infinite sector,

of angle 60 degrees. The question therefore arises of how we shall determine, given values for $\xi_{2,3}(t_{k-1})$ and $\xi_{2,3}(t_k)$, the probability that ξ has exited this sector during $[t_{k-1}, t_k]$, and if so at what point it exited. This can be done using an iterative method, as follows:

1. Call $\xi_{2,3}(t_{k-1})$ by z_0 .
2. Let A_{1i} be the locus boundary line nearer to z_i and let A_{2i} be that which is further from z_i . Find the point $r_i \in A_{2i}$ which minimizes the distance to z_i . Find the point $r'_i \in A_{1i}$ which minimizes the distance to r_i . Then consider R_i to be an infinite rectangle with corners r_i, r'_i , extending away from the origin with one of its boundaries a subset of A_{1i} (these are blue rectangles in the Figure).
3. There are three possibilities:
 - (a) $\xi_{2,3}$ leaves R_i at $t < t_k$ via A_{1i} . This means the path sampled was not in E_3 . In this case, t is the first exit time of X from the locus $x^{(1)}(t) > x^{(2)}(t) > x^{(3)}(t)$ and we can also sample the exit point.
 - (b) $\xi_{2,3}$ reaches $\xi_{2,3}(t_k)$ without leaving R_i . This means the sampled path was in E_3 .
 - (c) $\xi_{2,3}$ leaves R_i at $t < t_k$ via another boundary than A_{1i} . In this case, sample the exit point $\xi_{2,3}(t)$ and return to step 2 to continue, calling $\xi_{2,3}(t) = z_{i+1}$.

It is possible for us to make the determination between (a),(b) and (c) because the problem of first exit from a rectangle is soluble in the case of independent Brownian bridges. Viz, we may consider each coordinate separately and taking the minimum of interval exit times (sampling these is discussed in [MT99]) shall give us the first exit time from the rectangle; we may also sample the other coordinates at the exit time, if we are careful about it (cf the approach taken in [MT04, MT99] for the Brownian motion case).

No claim is made that this algorithm is the optimal one of its type, and one direction of further research in this area would be to try to ascertain an approximately optimal determination of the rectangle coordinates r_i . It would also be helpful to try to discover an elegant

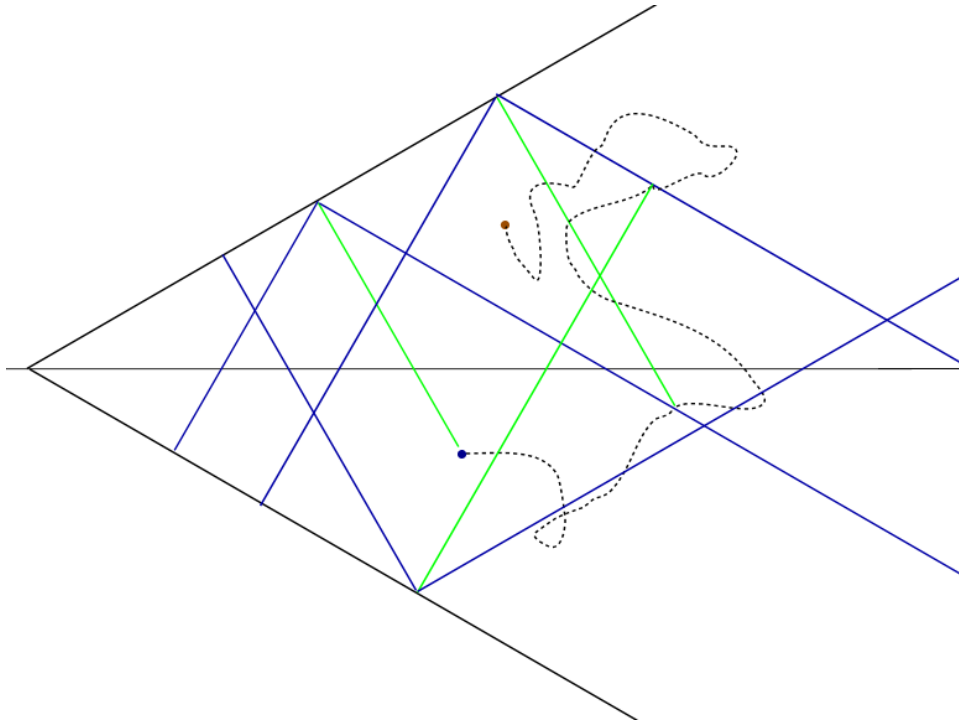


Figure 7.13: A schematic of how to sample whether a path lies in E_3 , in a case where 3 iterations are required.

way to solve the problem when more than 3 particles are considered together, because a change of coordinate basis apparently only reduces the dimensionality by 1 in general. In practice, in many simulations particles will be sufficiently far apart that it is safe to neglect crossing events between most of them; consequently, the problem breaks down into performing the above procedures for clusters of 2 or 3 particles.

7.4 Numerical results using the Last-to-first Subdiamond algorithm for 2 fermions in a multidimensional space

In this section, we describe the results of testing the Last-to-first Subdiamond algorithm on several simple examples. The results are, broadly speaking, encouraging.

7.4.1 Additive potentials and the Correlated Oscillator

In order to be sure that the programming is correct, it is preferable to have a system for which the action integral $\mathcal{G}_T(x_0, x_T)$ is analytically soluble.

The analytical solution for the harmonic oscillator potential It turns out that because the harmonic oscillator potential is dimensionally additive, the solution to the action integral for the 1-dimensional case (4.38) allows us to easily find the solution for the 2-dimensional case (and indeed the d -dimensional case). Recall that we use x_0^1 and x_0^2 , respectively, for the x-coordinates and y-coordinates of $x_0 = (x_0^{1(1)}, x_0^{2(1)}, x_0^{1(2)}, x_0^{2(2)}) \in \mathbb{R}^4$, and $x_0^{(1)}, x_0^{(2)}$ for the particle coordinates; likewise for x_T and other points. We observe that, due to the independence of Brownian bridge components,

$$\mathcal{J}_T(x_0, x_T) = \mathcal{J}_T(x_0^1, x_T^1) \mathcal{J}_T(x_0^2, x_T^2) .$$

Where $x_T = (x_T^{(1)}, x_T^{(2)})$, let $\neg x_T = (x_T^{(2)}, x_T^{(1)})$. When values for $\mathcal{J}_T(x_0^1, x_T^1)$, $\mathcal{J}_T(x_0^1, \neg x_T^1)$, $\mathcal{J}_T(x_0^2, x_T^2)$, $\mathcal{J}_T(x_0^2, \neg x_T^2)$ are known exactly, as by using (4.38) in the harmonic oscillator case, we can of course use them to give an exact value for $\mathcal{G}_T(x_0, x_T)$:

$$\begin{aligned} \mathcal{G}_T(x_0, x_T) &= \mathcal{J}_T(x_0, x_T) - \exp\left(-\frac{\|x_T^{(2)} - x_T^{(1)}\|^2}{T}\right) \mathcal{J}_T(x_0, \neg x_T) \\ &= \mathcal{J}_T(x_0^1, x_T^1) \mathcal{J}_T(x_0^2, x_T^2) - \exp\left(-\frac{\|x_T^{(2)} - x_T^{(1)}\|^2}{T}\right) \mathcal{J}_T(x_0^1, \neg x_T^1) \mathcal{J}_T(x_0^2, \neg x_T^2) . \end{aligned} \quad (7.20)$$

The solution to the sign problem for dimensionally additive potentials However, we can go even further than this: in the case of any dimensionally additive potential we can avoid the fermion sign problem entirely. Recognise that based on (7.20),

$$\begin{aligned} \mathcal{G}_T(x_0, x_T) &= \\ &\mathcal{J}_T(x_0^2, x_T^2) \mathcal{G}_T(x_0^1, x_T^1) + \exp\left(-\frac{(x_T^{(2)1} - x_T^{(1)1})^2}{T}\right) \mathcal{J}_T(x_0^1, \neg x_T^1) \mathcal{G}_T(x_0^2, x_T^2) . \end{aligned}$$

Since in the 1-dimensional case, we may apply the very effective simulation method for \mathcal{G}_T that was described in Chapter 6, it is then possible to find $\mathcal{G}_T(x_0, x_T)$ without further difficulties. So there is a way to perform simulations for d -dimensional systems which does not suffer from an insuperable fermion sign problem, whenever the potential is dimensionally additive.

However, it will be apparent that this means that such potentials do not provide good examples for testing, precisely because they exhibit this degeneracy. If we wish to test the efficacy of the Last-to-first Subdiamond algorithm then it seems preferable to choose an example in which the sign problem is not so easy to remove.

The correlated oscillator

We shall now use the notation (x_0, y_0) for the initial point, and so on. If we stretch one of the coordinate functions in the harmonic potential then there is only one coordinate basis in which it remains degenerate. So that this is not the basis in which the program works, we shall consider the following potential:

$$V((x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)})) = V((x^{(1)}, y^{(1)})) + V((x^{(2)}, y^{(2)})) \quad (7.21)$$

$$V((x, y)) = \frac{x^2 + y^2 - xy}{3/2} \text{ for } x, y \in \mathbb{R} \quad (7.22)$$

This is degenerate only when the coordinate basis is $\left(\frac{x+y}{\sqrt{2}}, \frac{x-y}{\sqrt{2}}\right)$. This allows us to find the solution for u and still achieve a sign problem in methods which would give no sign problem for a degenerate potential. Writing $(x', y') = \left(\frac{x+y}{\sqrt{2}}, \frac{x-y}{\sqrt{2}}\right)$,

$$V(x, y) = \frac{1}{3}x'^2 + y'^2 \quad .$$

Therefore for (7.22) we have the following single-particle action integral:

$$\begin{aligned}
\mathcal{J}_T((x_0, y_0), (x_T, y_T)) &= \int_{\mathcal{C}_{0, (x_0, y_0)}^{T, (x_T, y_T)}} \exp(-S(X)) dw_{0, (x_0, y_0)}^{T, (x_T, y_T)}(X) \\
&= \int_{\mathcal{C}_{0, x'_0}^{T, x'_T}} \exp\left(-\int_0^T \frac{1}{3} X(t)^2 dt\right) dw_{0, x'_0}^{T, x'_T}(X) \int_{\mathcal{C}_{0, y'_0}^{T, y'_T}} \exp\left(-\int_0^T X(t)^2 dt\right) dw_{0, y'_0}^{T, y'_T}(X) \\
&= f\left(x'_0, x'_T, \sqrt{\frac{2}{3}}\right) f\left(y'_0, y'_T, \sqrt{2}\right) \\
&= f\left(\frac{x_0 + y_0}{\sqrt{2}}, \frac{x_T + y_T}{\sqrt{2}}, \sqrt{\frac{2}{3}}\right) f\left(\frac{x_0 - y_0}{\sqrt{2}}, \frac{x_T - y_T}{\sqrt{2}}, \sqrt{2}\right),
\end{aligned}$$

where from (4.38),

$$\begin{aligned}
f(x_0, x_T, \gamma) \\
&= (\gamma T \operatorname{csch}(\gamma T))^{1/2} \exp\left(\frac{(x_T - x_0)^2}{2T} - \frac{1}{2}\gamma(x_0^2 + x_T^2) \coth(\gamma T) + \gamma x_0 x_T \operatorname{csch}(\gamma T)\right).
\end{aligned}$$

Since the particles are noninteracting in (7.22), for 2 particles $\mathcal{G}_T : \mathbb{R}^4 \times \mathbb{R}^4 \longrightarrow \mathbb{R}$ may be found as

$$\begin{aligned}
\mathcal{G}_T((x_0, y_0), (x_T, y_T)) &= \mathcal{J}_T\left(\left(x_0^{(1)}, y_0^{(1)}\right), \left(x_T^{(1)}, y_T^{(1)}\right)\right) \mathcal{J}_T\left(\left(x_0^{(2)}, y_0^{(2)}\right), \left(x_T^{(2)}, y_T^{(2)}\right)\right) \\
&\quad - \exp\left(-\frac{\left(x_0^{(2)} - x_0^{(1)}\right)\left(x_T^{(2)} - x_T^{(1)}\right) + \left(y_0^{(2)} - y_0^{(1)}\right)\left(y_T^{(2)} - y_T^{(1)}\right)}{T}\right) \\
&\quad \times \mathcal{J}_T\left(\left(x_0^{(1)}, y_0^{(1)}\right), \left(x_T^{(1)}, y_T^{(1)}\right)\right) \mathcal{J}_T\left(\left(x_0^{(2)}, y_0^{(2)}\right), \left(x_T^{(2)}, y_T^{(2)}\right)\right). \quad (7.23)
\end{aligned}$$

Figure 7.14 was obtained by taking $\int_{\mathbb{R}^2} \mathcal{G}_6((x, y), (x, y)) d(x^{(2)}, y^{(2)})$ to give a function of $x^{(1)}, y^{(1)}$.

The expression resulting from (7.23) is useful for providing exact solutions for specific bridge endpoints in order to perform intermediate analyses. It is difficult to perform numerical integrations of \mathcal{G} based on (7.23), but by working in the degenerate coordinates, it is

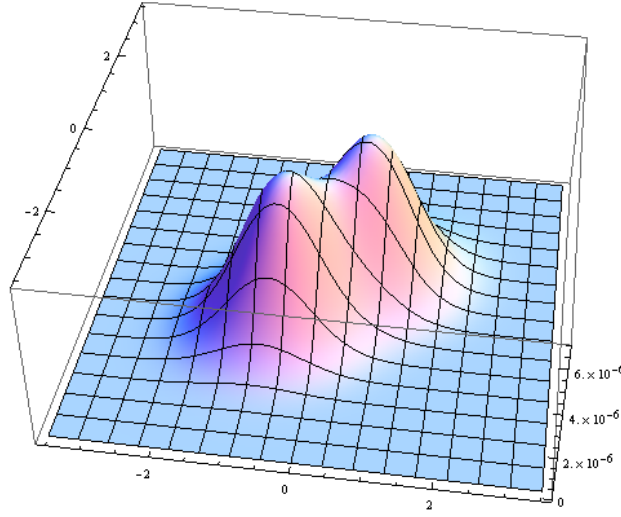


Figure 7.14: Particle density for 2 fermions in noninteracting correlated oscillator, $T=6$

possible to find \mathcal{Z} by decomposing the integrand:

$$\begin{aligned}
 \mathcal{Z} &= \int_{\mathbb{R}^4} \left[\frac{f\left(x^{(1)}, x^{(1)}, \sqrt{\frac{2}{3}}\right) f\left(y^{(1)}, y^{(1)}, \sqrt{2}\right) f\left(x^{(2)}, x^{(2)}, \sqrt{\frac{2}{3}}\right) f\left(y^{(2)}, y^{(2)}, \sqrt{2}\right)}{-\exp\left(-\frac{(x^{(2)}-x^{(1)})^2 + (y^{(2)}-y^{(1)})^2}{T}\right)} f\left(x^{(1)}, x^{(2)}, \sqrt{\frac{2}{3}}\right)^2 f\left(y^{(1)}, y^{(2)}, \sqrt{2}\right)^2} \right] d(x_0, y_0) \\
 &= \left(\int_{\mathbb{R}} f\left(x, x, \sqrt{\frac{2}{3}}\right) dx \right)^2 \left(\int_{\mathbb{R}} f\left(y, y, \sqrt{2}\right) dy \right)^2 \\
 &\quad - \int_{\mathbb{R}^2} \left[\exp\left(-\frac{(x^{(2)}-x^{(1)})^2}{T}\right) f\left(x^{(1)}, x^{(2)}, \sqrt{\frac{2}{3}}\right)^2 \right] dx \\
 &\quad \times \int_{\mathbb{R}^2} \left[\exp\left(-\frac{(y^{(2)}-y^{(1)})^2}{T}\right) f\left(y^{(1)}, y^{(2)}, \sqrt{2}\right)^2 \right] dy
 \end{aligned}$$

and the components of this expression are amenable to numerical integration in a standard software package. This allowed it to be verified that the program gave correct results.

The correlated oscillator in 3D

In 3 dimensions the potential used as the 3D correlated oscillator was

$$V(x, y, z) = \frac{13}{18}x^2 - \frac{5}{9}xy + \frac{13}{18}y^2 - \frac{2}{9}xz - \frac{2}{9}yz + \frac{5}{9}z^2.$$

It is possible to find \mathcal{Z} for a pair of noninteracting fermions using the same method as above, and in this way it was verified that the program for 2 3D fermions gave correct results.

7.4.2 Last-to-first experiments: a pair of 2D fermions

Experiments were performed with the intention of measuring the efficiency of the simulation, by collecting the average sign. The results for 2 noninteracting 2D fermions using the correlated oscillator potential are given in Figure 7.15. The initial point sampling measure was Gaussian with standard deviation 1. The credibility interval for $\langle \widetilde{\pm} \rangle$ was approximated by using Bivariate Gaussian Monte Carlo. Clearly our familiar assumption that prior information can be neglected only holds valid as long as the interval reported is relatively small; the results for $T = 16$ should be ignored completely, but are included to demonstrate that further progress with direct path sampling is infeasible. In all of these experiments, $M = 10^{10}$. The results seem to suggest that the asymptotic value of $\langle \widetilde{\pm} \rangle$ is above 0.85, unless there is a change in behaviour after $T = 10$ which we are unable to observe. Using $h = 0.2$, we observe a possible increase of the average sign between $T = 12$ and $T = 14$; this is probably due to an increased proportion of paths for which cancellation applies because of encountering an x and a y crossing within the same discretisation interval. As T increases, the variance is observed to increase exponentially due to the lack of importance sampling. For $T = 10$, the time to run the experiment on a 2.0 GHz machine was about 14.25 days.

For comparison, a simulation using the Coulombic potential was performed. Specifically, we let $V : \mathbb{R}^4 \rightarrow \mathbb{R}$ be given by

$$V(x^{(1)}, x^{(2)}) = -\min \left\{ \frac{1}{\|x^{(1)}\|}, 12 \right\} - \min \left\{ \frac{1}{\|x^{(2)}\|}, 12 \right\} + \frac{1}{2\|x^{(1)} - x^{(2)}\|} + R(x^{(1)}, x^{(2)}) + 10, \quad (7.24)$$

where $R : \mathbb{R}^4 \rightarrow \mathbb{R}$ applies harmonic walls to enforce a discrete spectrum of energy values (recall that all of our results are predicated on the assumption that a discrete spectrum exists; cf Theorem 2.2.1):

$$R(x^{(1)}, x^{(2)}) = \frac{1}{2} \left(1_{\|x^{(1)}\| > 10} (\|x^{(1)}\| - 10)^2 + 1_{\|x^{(2)}\| > 10} (\|x^{(2)}\| - 10)^2 \right).$$

The cutoff on the attractive terms in V avoids any problems with stability without substantially affecting the results. For $T = 6$, the posterior expectation of $\langle \widetilde{\pm} \rangle$ was 0.937015 (to

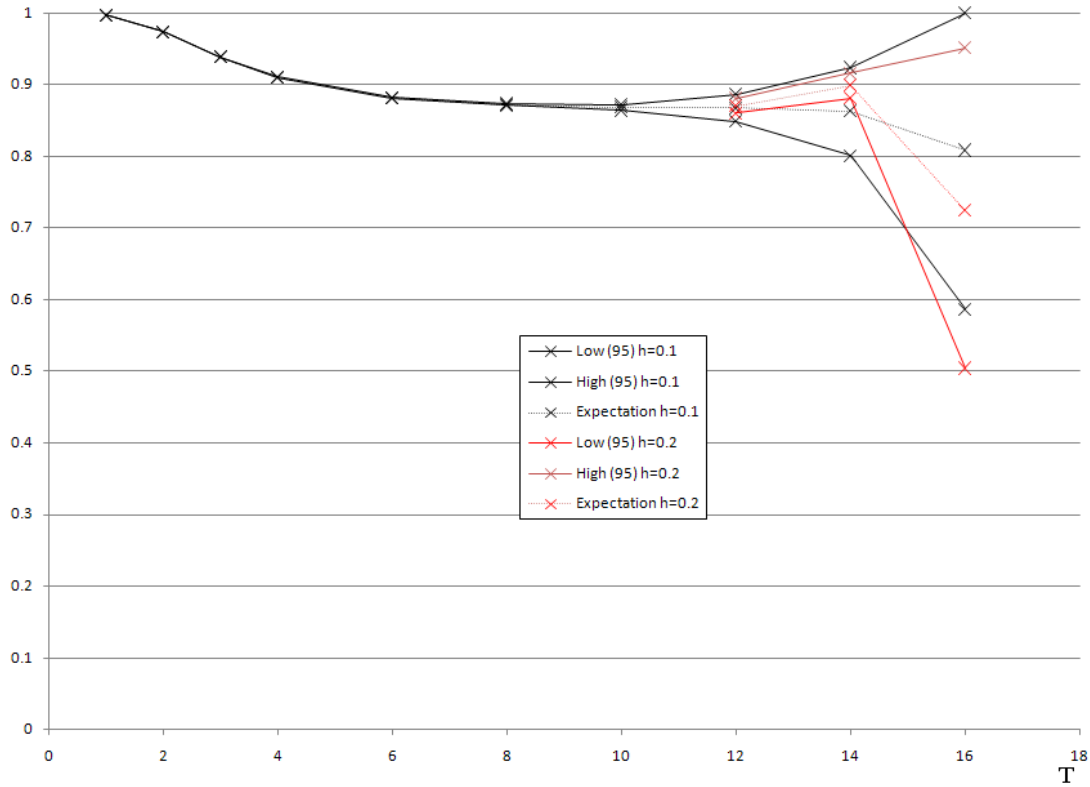


Figure 7.15: Average sign using the Last-to-first algorithm for a pair of 2D fermions, noninteracting correlated oscillator.

6 d.p.) and the 95% credibility interval was $(0.926909, 0.946587)$. In other words, a higher average sign was observed than for the noninteracting correlated oscillator. This is almost certainly because repulsion in the potential decreases the contribution of paths that bring the particles together simultaneously, and therefore increases the relative magnitude of the contribution from the subsets of $\mathcal{C}_{0,x_0}^{T,x_0}$ with lower values of χ . It is encouraging to note that this does actually translate into a higher average sign using our algorithm.

For the correlated oscillator in 2D, an interesting effect emerges: whenever the number of diamonds is even (so, the last variable to cross is also the first to cross), only positive contributions are collected. This is shown in Table 7.1, for the $T = 10$ result. Here the overall average sign was between 0.864511 and 0.872479, with posterior expectation 0.868506 (to 6 d.p.). For brevity the results with 22 or more diamonds are omitted. The standard error on the average sign from each subset was calculated only using the rough approximation (4.7) and as such, is only displayed for those results where the reported value was low enough for

us to regard this method as even giving an indicative figure. The table row marked "Kills" indicates the number of paths on which perfect cancellation occurred due to encountering both x and y crossings within the same discretisation interval.

The fact that the average sign is 1 if we condition on there being an even number of diamonds is a peculiar and unexplained artifact of the potential, which does not hold for the Coulombic potential (7.24), and in particular implies that if longitudinal-transverse coordinates were used, making the number of diamonds always even, then all contributions (ie, all samples of $Y^\diamond(X)$) would be positive. Moreover it demonstrates incontrovertibly that in native coordinates, the average sign cannot tend to zero, or fall much below $1/2$, since for T large, the proportion of paths with an even number of diamonds clearly will not fall much below $1/2$, and it is evident that as one might expect, the magnitude of EY^\diamond is decreasing as the number of diamonds increases, not remaining high only when the number of diamonds is odd! It should be noted that the number of diamonds, and anything conditional upon it, is meaningless here in intrinsic terms: in fact, it is dependent on the initial point sampling measure.

By way of comparison, the average sign for a completely naïve simulation at $T = 10$ (which we can find by taking the ratio of the boson partition function to the fermion partition function) would be 0.000285 (*to 3 s.f.*). Thus, the average sign under our algorithm is at least 3033 times greater, and the cost saving, when using importance sampling, should therefore be about 7 orders of magnitude.

7.4.3 Last-to-first experiments: a pair of 3D fermions

Figure 7.16 shows the results obtained using the Last-to-first algorithm for the 3D non-interacting correlated oscillator. The results are much less conclusive than those of Figure 7.15; in particular there is nothing to immediately dispel the concern that $\langle \widetilde{\pm} \rangle$ may decrease more or less linearly towards $T = 12$, and therefore be carrying on towards 0 thereafter.

Table 7.2, which breaks down the average sign according to number of diamonds in the

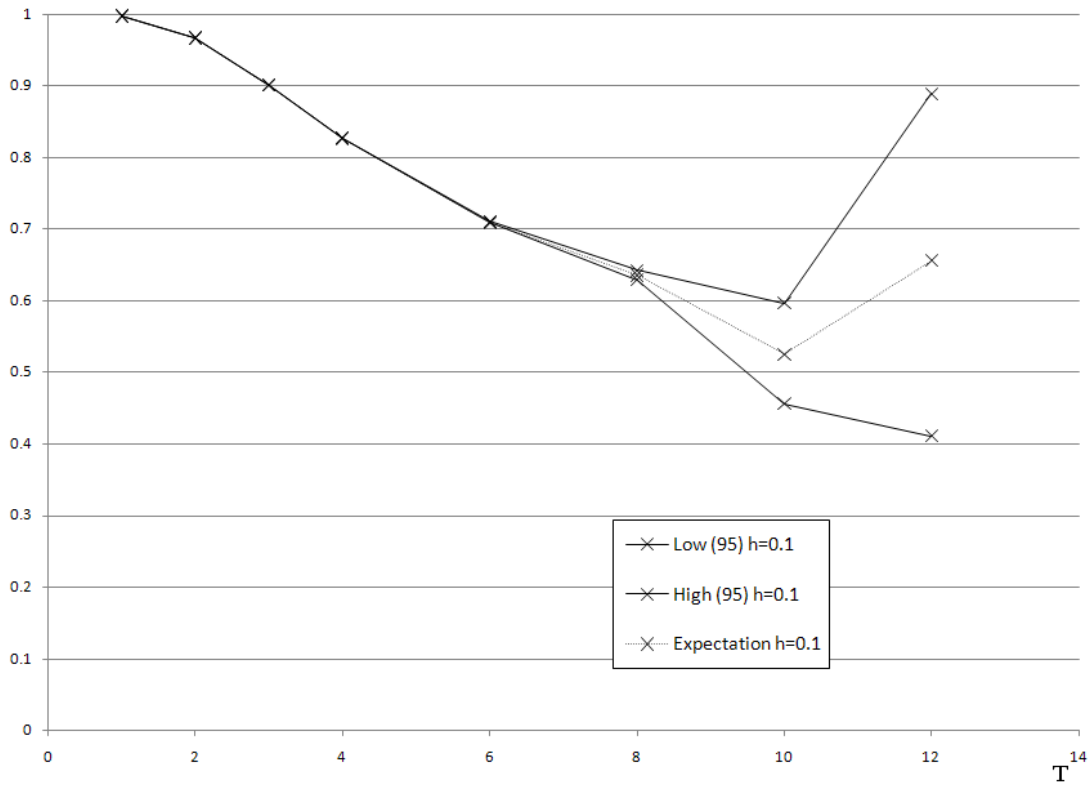


Figure 7.16: Average sign using the Last-to-first algorithm for a pair of 3D fermions, noninteracting correlated oscillator

same manner as Table 7.1. Again $T = 10$ and in this case the overall average sign was found to be between 0.456 and 0.596. Although the estimates of average signs may well be quite poor in terms of Monte Carlo variance, through most of the table, and there is a somewhat decreasing trend up to 6 diamonds, it seems that the results are nonetheless *broadly inconsistent* with a trend that increasing the number of diamonds would bring the average sign towards 0. This latter would have to be the case in order for our simulation to have a sign problem. Further experiments are required in order to determine whether or not this is the case, however, using MCMC to simulate effectively at higher values of T .

By way of comparison, the average sign for a completely naïve simulation at $T = 10$, found as the ratio of the fermion partition function to the boson partition function, would be 0.000294 (*to 3 s.f.*). The average sign under our algorithm is thus at least 1551 times greater and therefore the cost saving, when using importance sampling, should be about 6 orders of magnitude. In conclusion, the empirical evidence that our algorithm is successful

in 2D is fairly positive. The evidence as to whether it is successful in 3D is inconclusive, but it is at least moderately effective.

7.5 Extending the Subdiamond approach to n particles

In this section we provide some comments on the challenges inherent in extending our method to $n > 2$ particles, and some speculative remarks about how these challenges might be addressed.

7.5.1 Towards a general method

Having found a logical approach to dealing with the sign problem in the case of just a pair of fermions with $d > 1$, it is natural to search for a generalisation to $n > 2$ fermions; in other words, a method which uses the same principles to avoid the sign problem, and which features our Last-to-first 2-particle method as a special case.

To extend the principle of forming sampling blocks of paths based on coordinate exchange should be possible if one adopts the perspective that \mathcal{G}_T is found by integrating with respect to $w_{0,x_0}^{T,\Pi_n x_0}$ (compare the proof of Theorem 6.2.1). Considering the 2 particle case in this light, if we condition upon a set of crossings involving crossings in both x and y , then it is easy to recognise that the pair exchange (12) is equally likely with the identity. (Moreover, if we condition on there being either no crossings in x or no crossings in y , then it is obvious that the identity is the only possible permutation.) This remains true if we look at only a section of time (using conditional independence under $w_{0,x_0}^{T,\Pi_n x_0}$ and invoking multiplication of permutations in the same way as discussed in Subsection 4.2.3; see also Subsection 2.1.4) and so it is possible to re-develop the results of Chapter 7 using this perspective. Similarly, it is not difficult to see that in general, if we can identify a set of crossings then one legitimate approach is to collect functionals for a set of paths which are equilikely under the given crossing set².

²Or, more generally, weighted according to their probability conditional on the given crossing set.

It is important to note that from the definition of w_{0,x_0}^T , we know that each binary choice sequence would be equiprobable under w_{0,x_0}^T given a well-defined set of crossings, and therefore it follows that under $w_{0,x_0}^{T,\Pi_n x_0}$, those choices which give rise to $X(T) \in \Pi_n x_0$ correspond to equilikely paths. (This also applies when some of the choices give rise to the same permutation: such permutations receive a commensurately higher weighting.)

The issue of combining flips

However, the problem is deeper than this in two ways. Firstly, we do not have to always collect a path for all the possible binary choices (ie choices of which crossings should be used for particle exchange) corresponding to the sampled crossing sequence. We might choose instead to partition this set into equivalence classes $\mathcal{P}_1, \dots, \mathcal{P}_q$. If our initially sampled path belongs to \mathcal{P}_i then we collect functionals for each of the binary choices in \mathcal{P}_i . This concept is quite important because, due to the fact that a permutation may have multiple different representations as pair swaps, it could be quite difficult to access certain permutations by matching particle swaps between coordinates.

Secondly, if we choose to recognise crossings between our sampled particle loops (from here on denoted AB_x, AB_y etc) then when we have $n > 2$ particles involved, the issue arises that accepting or not accepting certain exchanges of coordinates may change the meaning of other crossings. For example, while A and B are exchanged, say in the x coordinate, BC_x becomes AC_x . We could avoid this by instead recognising crossings between paths indexed not by their sampled loop but by their ordering in the relevant dimension (these crossings being written $12_x, 12_y$). These crossings are well-defined however we choose to map the initially sampled paths to paths for which Y is to be evaluated. However, the same problem immediately returns in that we need to map them to particle crossings in order to perform any simulations. A way to avoid this difficulty would be available if we could always find a decomposition into \mathcal{P}_i so that the elements of \mathcal{P}_i were all accessible from each other without a reinterpretation of which particles should be involved in any crossing that is used. So

apparently, labelling crossings ordinally confers no advantage.

7.5.2 Illustrative examples of crossing sequences

Let us consider some examples of crossing sequences and how we might respond to them. For simplicity we shall consider examples where we assume that the initial points A,B,C are arranged diagonally so that both the x and the y coordinates are ordered A,B,C.

Example 1: no complications

Suppose we simply encounter AB_x and then AB_y . This appears straightforward: it is the same as in the $n = 2$ case. The same thing applies if we encounter this sequence preceding another one; e.g. $AB_x AB_y BC_x BC_y$. We may summarize the possibilities, where 1 indicates that a crossing is used and 0 indicates that it is not, thus:

AB_x	AB_y	
0	0	+
1	1	—

Example 2: mixing the coordinate sequences together

Suppose we encounter a sequence where the same particle exchange sequence applies in both (or all) coordinates, but without necessarily being split into separate segments of time.

This *should not* present significant difficulties. The viable choices, which are all equilikely, are tabulated as follows:

AB_x	BC_x	AB_y	BC_y	
0	0	0	0	+
0	1	0	1	—
1	0	1	0	—
1	1	1	1	+

We do not yet consider flips whose time interval spans zero. Figure 7.17 displays the links that are followed out. The initially sampled paths are loops and they are represented by black curves. The blue curves represent alternative choices under exchange. Green represents more negative curves, so that following blue and then green gives us a positive contribution, $(AB)(BC)$. The numbers 1,...,5 represent segments of time.

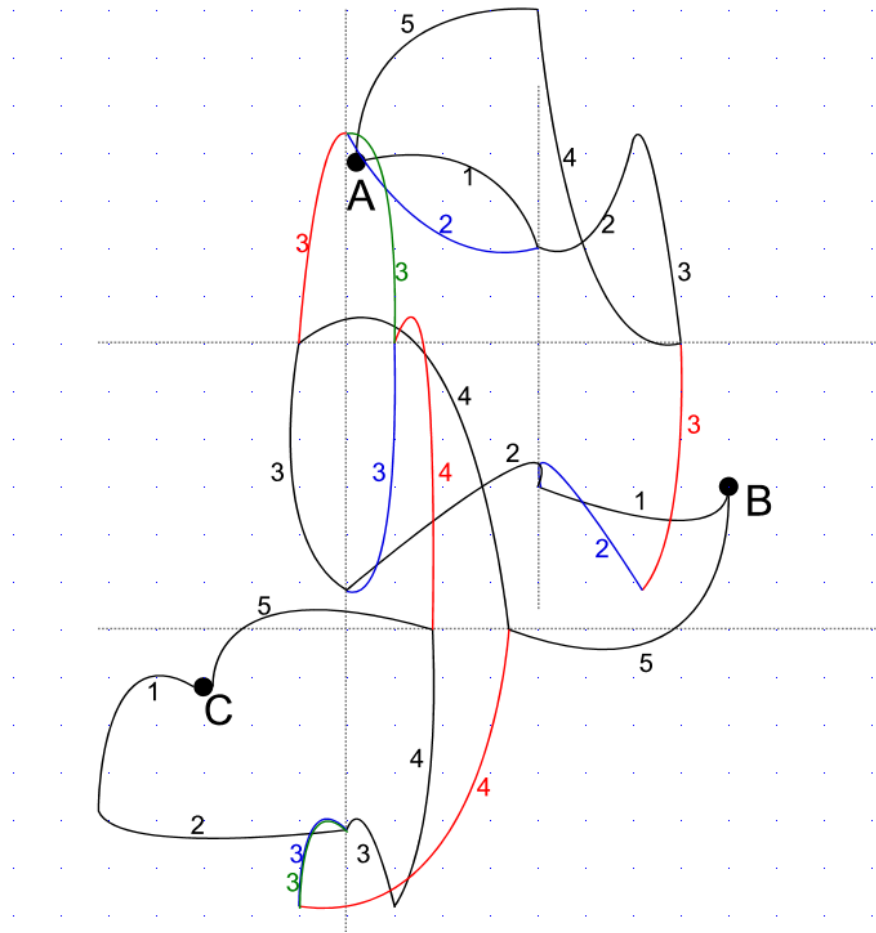


Figure 7.17: Schematic showing a crossing sequence $AB_x BC_x AB_y BC_y$, with permutations (id), (AB), (BC), (AB)(BC).

It is natural to ask how we might actually build an algorithm able to collect these simultaneous samples efficiently. We no longer simply have a product of differences and therefore it seems that some extra computational cost may be unavoidable.

At first it might seem that one solution, to the problem of identifying viable equivalence classes (in effect, sampling blocks), is to include paths in the same equivalence class if and only if they are accessible from each other via flips of coordinates (ie, exchanges over a time interval between two crossings) which do not involve changing the meaning of any other observed crossings. However, this turns out to be an unfortunate choice. Consider the crossing sequence in this example. Superficially we see that we could use $\{AB_x, AB_y\}$ to exchange A and B, or could use $\{BC_x, BC_y\}$ to exchange B and C; in fact these flips are also compatible with each other, and we could use all four switches to get (AB)(BC). But

according to the rule that one crossing should not affect another, we would have to rule out using any of them!

Example 3: incongenial ordering of swaps

On the other hand, if the crossing sequence is $AB_x BC_y AB_y BC_x$ (and we shall assume that there are many other crossings before and after this, so that it is not a simple matter of rotating time) then $\{AB_x, AB_y\}$ and $\{BC_x, BC_y\}$ are incompatible: using all switches leads to $(AB)(BC)$ for x but $(BC)(AB)$ for y . This is problematic because we apparently have more negative than positive contributions of which to take account.

AB_x	BC_y	AB_y	BC_x	
0	0	0	0	+
0	1	0	1	—
1	0	1	0	—

It would certainly be preferable not to pick up two negative contributions and only one positive. We know that under $w_{0,x_0}^{T,\Pi_n x_0}$, the weights of the viable choices are equal. But if we choose to neglect, say, $\{BC_x, BC_y\}$ then since we are using particle labels rather than ordinal ones, we know that on another occasion we might have sampled a path X which did feature the B,C exchange. On the face of it, this leads to a sign problem however; so let us consider our options.

Possible algorithms for determining equivalent paths

Reflecting on these examples, there are a number of possible strategies for constructing the equivalence class of an observed path X , such as:

1. Reject flips which are not compatible with those already encountered (moving clockwise through time from 0).
2. Use all directly accessible flips. ie, our algorithm is:
 - (a) Detect all flips which are directly accessible from X ;
 - (b) Determine a matrix of values for which flips are pairwise compatible.

- (c) Collect those binary choices which do not include any incompatible pairs of flips.
3. Use an algorithm to search for all possible binary choices which are viable. (This could be very expensive.)

In each of these cases we have to bear in mind that we are also interested in using flips (diamonds) whose time interval spans the zero time. It is conjectured that for this reason, one never in fact obtains a net negative number of contributions using strategy (2). However, this still remains to be investigated.

There is also potentially an alternative approach, if it were possible to prove that one might legitimately exchange opposite coordinates from the ones encountered. So for $\{AB_x, AB_y\}$ we might choose to exchange the y-coordinates of A and B at the time of an AB_x crossing and the x-coordinates at the time of AB_y . Intuition suggests this is consistent with the particles being indistinguishable, but mathematically it is not clear. If this approach should be seen to be valid, then one has a fourth algorithm: choose to make a flip of coordinates only if it is possible to do so without affecting an intermediate crossing. So in the case of AB_x, BC_x, AB_y, BC_y we allow exchange of the y-coordinates of A and B between AB_x and AB_y and this is then compatible with exchanging the x-coordinates of B and C between BC_x and BC_y . In the case of AB_x, BC_y, AB_y, BC_x we would use x for A,B and x for B,C. The fact that flips never affect other crossings means that they are all always mutually compatible.

It remains to be seen how much improvement in average sign would be obtained by using these algorithms.

Example 4: flips contained within one another

Now consider what might happen if the order of these crossings were different, so that we have $AB_x BC_x BC_y AB_y$. Note that this is still a different sequence when time is regarded as a circle: it is qualitatively different because the x exchanges appear next to each other.

We now have some difficulty in combining (AB) and (BC) because using AB_x means that we see BC_x then as AC_x .

AB_x	BC_x	BC_y	AB_y	
0	0	0	0	+
0	1	1	0	-
1	0	0	1	-

The answer is supplied by considering time symmetry. If we regard time as a circle, it is immediately seen that this crossing sequence is the same as the innocuous $AB_y AB_x BC_x BC_y$ (see Figure 7.18). Therefore assuming that this kind of time symmetry is to be invoked, we must treat this crossing sequence the same way regardless of our superimposed labelling of time.

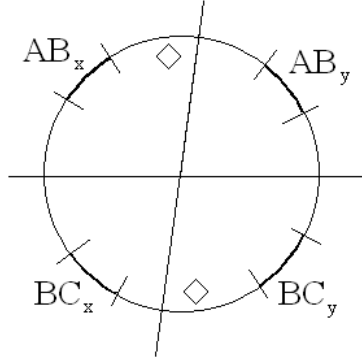


Figure 7.18: $AB_x BC_x BC_y AB_y$ can be regarded as $AB_y AB_x BC_x BC_y$ by moving the 0 label when 0 and T are identified to make $[0, T]$ a circle.

The most obvious way, as illustrated, is to place a diamond between AB_y and AB_x and between BC_x and BC_y . If we think of the 0 label as lying within the first of these diamonds, then we are collecting contributions both for $x_0 = (x^{1(1)}, x^{1(2)}, x^{1(3)}, x^{2(1)}, x^{2(2)}, x^{2(3)})$ and for $x'_0 = (x^{1(2)}, x^{1(1)}, x^{1(3)}, x^{2(1)}, x^{2(2)}, x^{2(3)})$. While 2 negative choices and 1 positive choice correspond to x_0 , 2 positive choices and 1 negative choice correspond to x'_0 .

However, this choice of diamond placement does beg the question, of why we do not use the "long" diamonds, from AB_x to AB_y or BC_x to BC_y . Intuitively this certainly seems less desirable: by collecting paths from diamonds which take place over a shorter time interval, we collect countersigned contributions which have a greater covariance with those from our

originally sampled path and the other paths being collected. Whether neglecting these diamonds, in favour of the shorter ones, leads to a severe sign problem, is a question which requires further investigation.

Example 5: Permutations that are inaccessible by using matching swaps

It is relevant to consider an example where the fact that a permutation may have two different representations as pair swaps becomes an issue. Taking $AB_x BC_x AB_x BC_y AB_y BC_y$, if we seek to match permutations between x and y for a choice to be viable, we have the following table:

AB_x	BC_x	AB_x	BC_y	AB_y	BC_y		Name
0	0	0	0	0	0	+	id1
0	0	0	1	0	1	+	id2
0	0	1	0	1	0	−	(AB)1
0	1	0	1	0	0	−	(BC)1
0	1	0	0	0	1	−	(BC)2
0	1	1	1	1	0	+	(BC)(AB)
1	0	0	0	1	0	−	(AB)2
1	0	1	0	0	0	+	id3
1	0	1	1	0	1	+	id4
1	1	0	0	1	1	+	(AB)(BC)
1	1	1	1	1	1	−	(AC)

This seems to indicate that there are 6 positive choices but only 5 negative choices. But how are these to be accessed? If we could access them all from each other, this would be fine, but to do this, without checking every combination of exchanges and holding a large number of paths (strategy (3) above), seems to be very problematic.

Under any of our algorithms described above, from id1 we must be able to access (BC)1, (AB)1, (BC)(AB):

AB_x	BC_x	AB_x	BC_y	AB_y	BC_y		Name
0	0	0	0	0	0	+	id1
0	0	1	0	1	0	−	(AB)1
0	1	0	1	0	0	−	(BC)1
0	1	1	1	1	0	+	(BC)(AB)

All of the other binary choices are arrived at by using either the positive flip AB_x - AB_x and/or the positive flip BC_y - BC_y . Under strategy (3) we would include these, but otherwise

we exclude them completely. Because we are using particle labels rather than ordinal labels, this is perfectly valid, although again it remains to be seen whether it leads to a severe sign problem or not.

This example also suggests another intuitively appealing concept. For any algorithm to determine path equivalence, if the juxtaposition of the x and y crossing sequences is altered, it is preferable that the permutations sampled, ie the equivalence classes, should remain the same, at least until diamonds across zero are taken into account.

Summary

In conclusion, further study is needed to establish how to efficiently extend the Subdiamonds approach to the case $n > 2$. We have explored, through simple examples, some of the features which we might expect to characterise such a generalised approach. The examples are suggestive that a viable method may well exist, but also indicative that as n is increased we would expect a considerable increase in complexity. There clearly is a lot of potential for further research into this rich subject, as we shall discuss further in Section 8.2.

Table 7.1: *Observed efficiency conditional on number of diamonds, in simulating 2 noninteracting 2D fermions in correlated oscillator.* The results are based on using Simpson's Rule. The statistics are only indicative of what happens with a standard Gaussian as initial point sampling measure. The timestep $h = 0.1$ throughout.

No. of Diamonds	Frequency	EY^\diamond (to 3 s.f.)	$E Y^\diamond $ (to 3 s.f.)	Average sign (to 3 d.p.)	Average sign SE (approx)
0	7152189314	1.72E-10	1.72E-10	1	0
1	684659866	1.18E-10	1.76E-10	0.671	0.00629
2	1018766626	1.56E-10	1.56E-10	1	0
3	295824297	1.19E-11	2.43E-11	0.49	0.0083
4	400287703	7.67E-12	7.67E-12	1	0
5	102143305	1.10E-13	3.08E-13	0.359	0.0169
6	128843315	4.77E-14	4.77E-14	1	0
7	29578995	1.80E-16	6.45E-16	0.28	0.0479
8	35268650	5.98E-17	5.98E-17	1	0
9	7427158	7.47E-20	3.69E-19	0.202	?
10	8459481	4.33E-20	4.33E-20	1	0
11	1657773	2.49E-23	5.54E-23	0.45	?
12	1811859	1.11E-24	1.11E-24	1	0
13	332489	7.94E-28	9.66E-28	0.822	0.119
14	350603	2.30E-28	2.30E-28	1	0
15	60757	5.76E-33	3.19E-32	0.18	?
16	61935	8.07E-35	8.07E-35	1	0
17	10124	1.16E-37	1.18E-37	0.982	0.0191
18	10067	1.12E-39	1.12E-39	1	0
19	1489	1.92E-43	1.92E-43	0.9994	0.000618
20	1449	1.05E-46	1.05E-46	1	0
21	202	-5.09E-52	5.19E-52	-0.98	0.0265
22+	283				
Kills:	132252260	0.00E+00	0.00E+00		

Table 7.2: *Observed efficiency conditional on number of diamonds, in simulating 2 noninteracting 3D fermions in correlated oscillator.* The results are based on using Simpson's Rule. The statistics are only indicative of what happens with a standard Gaussian as initial point sampling measure. The timestep $h = 0.1$ throughout.

No. of Diamonds:	Frequency	EY^\diamond (to 3 s.f.)	$E Y^\diamond $ (to 3 s.f.)	Average sign (to 3 d.p.)	Average sign SE (approx)
0	3962852300	1.02E-13	1.02E-13	1.000	0.0000
1	1326223018	4.50E-14	1.85E-13	0.243	0.0809
2	1669750130	1.05E-13	1.77E-13	0.591	0.0226
3	778183026	1.85E-14	4.93E-14	0.375	0.0408
4	582623313	4.57E-15	9.57E-15	0.478	0.0346
5	304201081	1.63E-16	6.37E-16	0.256	0.0426
6	185801119	1.08E-17	4.25E-17	0.254	0.114
7	97985958	1.73E-19	9.62E-19	0.180	?
8	53655084	-7.34E-22	1.80E-20	-0.041	?
9	27481788	6.02E-23	2.31E-22	0.261	?
10	14054441	5.53E-26	1.12E-24	0.049	?
11	6928690	4.00E-27	7.96E-27	0.503	?
12	3366472	-6.29E-30	4.03E-29	-0.156	?
13	1597575	-3.56E-32	8.20E-32	-0.435	?
14	744415	-7.72E-35	9.80E-35	-0.787	?
15	340581	7.03E-38	9.95E-38	0.706	?
16	153821	9.70E-41	1.20E-40	0.806	?
17	67383	-2.14E-44	2.87E-44	-0.743	?
18	29640	-3.31E-45	3.35E-45	-0.988	0.0162
19	12476	-4.07E-51	6.37E-51	-0.640	?
20+	9025				
Kills:	983938664	0.00E+00	0.00E+00		

Chapter 8

Conclusions and outlook

8.1 Conclusions

In this section we summarize the main conclusions from the preceding chapters.

The background to our endeavours

By developing the viewpoint of works such as [Kac57, Kac56], it is possible to explain the relationship between conditional Wiener integrals and quantum statistical mechanics, via the probabilistic representation of the solutions to a certain parabolic partial differential equation (e.g. [Kac51, Fre85]). This allows us to find the expectation of practically any observable quantity, in principle, for a quantum system in thermal equilibrium at a nonzero temperature. We are able to make a rigorous justification for Path Integral Monte Carlo based on Conditional Wiener integrals, as opposed to needing to employ Feynman path integrals under a Wick rotation.

Regarding simulation of excited system states

Under suitable conditions, such as the Hamiltonian operator having a discrete spectrum, we have established a functional integral expression whose zero-temperature limit yields a density corresponding to the sum of the first k eigenstates of a quantum system. This means

that if a suitable method of zero-temperature simulation is adopted, and the multiplicity of eigenstates is known, then it is possible to simulate statistics corresponding to the average of the states at each energy level.

Using the well-known Boltzmann expansion, simulation of excited states provides an approximate simulation method for systems at sufficiently low nonzero temperatures.

Regarding stochastic numerical integration and Path Integral Monte Carlo

In this thesis (see also [DT10]) it has been demonstrated that a piecewise constant numerical method is of second order in the time-step for a relatively general class of functionals, when performing integration with respect to any conditional Wiener measure. In order to prove this, a novel technique was used, applying Taylor's theorem for functionals (our Theorem 3.1.1) to locally expand our approximate functional about the functional being integrated.

The equivalent result relating to Wiener integrals was already known [GM84], and our result was already known in the special case of exponential-type functionals [MT04ii] (see also [MT04]) and in particular had long been known in the case of the action functional [Suz86]. However, the knowledge that the result holds for a broad class of functionals is very useful, because when an observable is 'diagonal' in the momentum representation, it is often possible to express it in the position representation using a functional that is *not* of exponential type; the example of kinetic energy was treated in this respect. The result particularly has implications for the simulation of bosonic systems, for which there is no sign problem. Amongst other experiments offering empirical confirmation of the result, a simulation of the kinetic energy of four bosons in a 1-dimensional space under a harmonic potential was performed, and this illustrated the difference between the second-order piecewise constant method and the first-order Euler method.

In order to estimate expectations of observables it is usually necessary to find information about ratios of functional integrals. This can be accomplished probabilistically using Path

Integral Monte Carlo procedures, and it is then possible to estimate a credibility interval for the ratio that is of interest. Importance sampling is essential to any serious simulation, and we were able to determine the optimal path sampling measure. (Nonetheless, the most popular choice of sampling measure is the action measure, which simplifies the analysis by avoiding the need to simulate a partition function.) In order to sample according to such measures, Markov Chain Monte Carlo is used; however, we have explained that our results concerning numerical integration error still hold true regardless of how sampling is performed, and of the sampling measure used. This was demonstrated by performing a simulation of the potential energy of a system of 64 boltzmannons under a Lennard-Jones potential.

It was also seen empirically that in the case of an exponential-type functional, applying Simpson's Rule to the exponent (a method already known to be second-order for this class of functionals) sometimes incurs a bias that is almost 10 times smaller than that incurred by using the Trapezoidal Rule for the exponent (which coincides with our piecewise constant method).

On the fermion sign problem in 1-dimensional systems

The fermion sign problem, defined in terms of finding an exact fermion simulation algorithm with polynomial cost scaling in both the inverse temperature T and the number of particles n , is almost certainly insoluble [TW05]. When a sampling measure is used that removes the denominator integral, the relative cost of a simulation (to obtain the same variance) due to the presence of signs is governed by the square of the average sign, $\langle \pm \rangle^2 = \mathcal{E}$, and we have shown that it is always more than \mathcal{E} times greater (cf [Cep96]). This creates significant obstacles to fermionic simulations since without a special scheme being constructed to avoid it, the rate of decrease of the average sign with T or n is exponential. It was seen that using a method based on linear translation to create covariances between countersigned functional contributions achieves only a limited amelioration of cost. It was proven that for a system of noninteracting fermions (not necessarily in a 1-dimensional space), it is possible

to replace the sum over permutations with a sum over loops of different time lengths. A translational method based on this was tested; it was seen that the benefits of this method, on its own, are also limited.

For 1-dimensional systems, the fermion sign problem is soluble and in this thesis, a rigorous proof - based on mathematics rather than on physics - has been advanced for the solution. It is possible to prove that constraining Brownian bridges not to cross is sufficient to allow only positive contributions to be sampled, avoiding any sign problem. Moreover, it has been demonstrated that it is possible to implement exact sampling of path discretisations subject to this constraint. In fact, we were able to calculate the pdf for one intermediate point on a non-crossing Brownian bridge (a result closely related to the classical Karlin-McGregor theorem [KM59]), and then we established a relatively efficient method for sampling according to this pdf. A simulation was carried out for the particle density of the 1-dimensional harmonic oscillator, and seen to give sound results.

On a geometrical approach to the general fermion sign problem in Path Integral Monte Carlo

One existing way to extend this approach to the multidimensional case is to regard the one-dimensional solution as meaning that paths are prevented from crossing the nodes of the wavefunction. In this thesis, we have instead taken the point of view that it arises because in the case of crossing paths, we are able to find paths whose contributions exactly cancel, and sample them simultaneously. It has been proven that at least for the case of just two fermions, there is a way to generalise this principle, forming sampling blocks based on the crossings encountered for each coordinate, to achieve a powerful algorithm for carrying out simulations at relatively low temperatures. Specifically, we have proved that for the action functional, we can substitute a functional involving a product of diamonds (differences of functionals over sections of path between coordinate crossings) and positive functionals. It was seen empirically that the average sign does not tend to zero as T tends to infinity in the

case of a pair of fermions in a 2-dimensional space, and there was some limited evidence to suggest that this may also hold true for a pair of fermions in a 3-dimensional space.

We also have offered some indications of how the method may generalise to the case of a system with $n > 2$ fermions in a multidimensional space.

8.2 Outlook for research

In this section, we shall discuss some possible directions of future research to develop the work of this thesis further. The most obvious is to rigorously develop a generalisation of the Subdiamonds approach introduced in Chapter 7. However, we discuss the topics in the order that the relevant work appears in the thesis.

Towards a zero-temperature method based on path integrals

So far the focus of our discussion has been mostly on the case of a system at a nonzero finite temperature. There is probably a larger body of literature treating the zero-temperature case; there exist multiple approaches, but probably the most popular is to use the Diffusion Monte Carlo method already mentioned in Subsection 5.2.1. This method is based on modelling the ground state wavefunction as a positive probability density; this fact leads to certain inherent limitations [KFS96]. Rather than simply taking a sufficiently large value of T in Path Integral Monte Carlo, is it possible that an elegant method of zero-temperature simulation using conditional Wiener integrals could be constructed?

For simplicity, we consider a system of distinguishable particles. We recall the notation of Section 2.2 and let q_t denote the unnormalized thermal density matrix at inverse temperature t ; ie $q_t(x_1, x_2) = \mathcal{J}_t(x_1, x_2)\varphi_t(x_1, x_2)$. Furthermore, let

$$\rho_T(x_1, x_2) = \frac{q_T(x_1, x_2)}{\int q_T(x, x)dx} \quad ; \quad \rho_\infty(x_1, x_2) = \lim_{T \rightarrow \infty} \rho_T(x_1, x_2) \quad .$$

By using asymptotic properties of Brownian bridges, it is possible to obtain that

$$\rho_\infty(x_1, x_2) = \frac{\int_{\mathbb{R}^{2nd}} [q_t(x_1, z_1) q_t(x_2, z_2) \rho_\infty(z_1, z_2)] dz_1 dz_2}{\int_{\mathbb{R}^{3nd}} [q_t(x, z_1) q_t(x, z_2) \rho_\infty(z_1, z_2)] dz_1 dz_2 dx} \quad . \quad (8.1)$$

If it is possible to let each of the real-valued ground state eigenfunctions r_i solve

$$r_i(x) \propto \int_{\mathbb{R}^{nd}} [q_t(x, z) r_i(z)] dz , \quad (8.2)$$

along with the usual orthonormality conditions, then this provides a solution to (8.1). If this integral equation (8.2) can be solved approximately (using, for example, an iterative method) by using a Path Integral Monte Carlo method to collect information about q_t , then we may fairly say that it is possible to approach zero-temperature simulations via conditional Wiener integrals.

If this were to lead to a method which is able to find $\lim_{T \rightarrow \infty} I_{T,k}$ (see Section 2.3) then clearly, simulation of excited states as well as ground states is facilitated (albeit with sign problems arising, as is the usual case). Being able to simulate excited states provides a controlled approximation to low (non-zero) temperature systems, since at sufficiently low temperatures a good approximation will come from the first few terms of the Boltzmann distribution (2.51). Bounding the error from remainder terms is not difficult.

Investigation of possible higher-order ‘random series’-based numerical method for a broad class of functionals

As mentioned in Subsection 3.2.1, it is apparently the case that methods involving functional evaluations away from the sampled path discretisation are able to attain higher orders of convergence for action integrals [PD03, Pre04]. For the same reasons that it was desirable to prove that our piecewise constant method is second-order on the broad class of functionals described in Subsection 3.1.1, it would be worthwhile to find out whether a numerical method analogous to that of [Pre04] could be rigorously proven to attain higher orders of convergence on this class.

It seems that to employ the same method of proof as used for Theorem 3.2.1 might be an expedient way to attempt this.

The extension of the Subdiamonds approach

As discussed in Section 7.5, it seems quite reasonable to think that it would be possible to extend the Subdiamonds approach to systems of $n > 2$ fermions.

It seems unlikely, given the results of [TW05], that it is possible to create a method which has polynomial cost scaling in both n and T . However, it might be possible, for instance, for there to be a method involving a subroutine of nonpolynomial cost which does not involve potential evaluations. Since these are usually thought of as representing a heavy computational expense, in simulations of practical interest, this might allow some progress to be made with performing exact simulations at increased T and n . There is some reason to believe that our approach might lead to such a method, since identifying the paths belonging to the same sampling block as the sampled path X could well require an algorithm with nonpolynomial cost scaling in n . In default of this, it may nonetheless be possible to seek a generalised algorithm where the cost growth is nonpolynomial in n but sufficiently slow that it is computationally feasible to work with moderate n , e.g. $n = 12$. This would enable simple atomic simulations to be carried out.

The first step in investigating the possibilities is to rigorously prove a generalisation of Theorem 7.2.10, corresponding to the intuitive direction sketched out in Section 7.5. It shall probably then be apparent that nonpolynomial cost growth in some part of the algorithm is inevitable, but experiments would be needed to determine the actual rate of cost growth.

It is also notable that further research is needed to study the properties of the Last-to-first method as it applies to just a pair of 2D or 3D fermions, under different potentials. It is clear that MCMC simulations are needed, to avoid the explosion of variance that was witnessed in the experiments of Subsection 7.4.3 when T was increased even moderately.

Other extensions of the Subdiamonds approach are also of interest:

Combining the Subdiamonds approach with zero-temperature simulations It is notable that an analogous fermion sign problem is encountered when performing zero-

temperature simulations of fermion systems, using DMC or any other method. Therefore, given a zero-temperature simulation method, it would be of interest to know whether a coordinate crossing-based blocking strategy for avoiding the sign problem exists, by analogy with our Subdiamond approach in the nonzero temperature case.

Besides the fixed-node and release-node methods, other methods to address the fermion sign problem in DMC exist; for example see [Mis06] or [And95]. It is of interest to try to recognise whether these exhibit commonalities with the Subdiamond approach and to investigate the relative performance under different conditions, should a generalisation of the Subdiamond approach to zero-temperature be possible.

Nonequilibrium problems / quantum dynamics Application of path integrals to nonequilibrium problems is an active area of research (e.g. [BSKF03, MR07, Mak09]) in which a sign problem is apparently encountered in nearly all cases, that may be even more severe than that of fermion PIMC [Cep96]. It seems an exciting possibility that parallels to the work of Chapter 7 may exist that relate to such simulations. It is also notable that some dynamical quantities can be accessed via the off-diagonal values of the thermal density matrix [Cep96]; our method is perfectly adequate for finding these values.

Bibliography

- [AF97] P. W. Atkins and R. S. Friedman. *Molecular Quantum Mechanics*. OUP, 1997.
- [And76] J. B. Anderson. Quantum chemistry by random walk. *J. Chem. Phys.* **65** (1976), 4121.
- [And95] J. B. Anderson. Exact Quantum Chemistry by Monte Carlo Methods. in *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy (Understanding Chemical Reactivity)*, ed. S. Langhoff. Springer, 1995.
- [AT05] A. Alavi and A. J. W. Thom. A combinatorial approach to the electron correlation problem. *J. Chem. Phys.* **123** (2005), 204106.
- [AT06] A. Alavi and A. J. W. Thom. Path resummations and the Fermion sign problem, *Lect. Notes Phys.* **703** (2006), 685-704.
- [Ber06] J. Berger. The case for objective Bayesian analysis. *Bayesian Analysis* **1** (2006), 385-402.
- [BM99] G. Birkhoff and S. Mac Lane. *Algebra*. Chelsea, 1999.
- [BMNR01] K. Binder, M. H. Müser, P. Nielaba and Chr. Rickwardt, Path integral Monte Carlo simulations of silicates. *Phys. Rev. B* **63**, 045204 (2001).
- [BS99] R. M. Bowley and M. Sánchez. *Introductory Statistical Mechanics*. (2nd edn) OUP, 1999.

- [BS02] A. N. Borodin and P. Salminen. *Handbook of Brownian Motion - Facts and Formulae. (2nd edn.)* Birkhäuser, 2002.
- [BSKF03] M. Bonitz, D. Semkat, D. Kremp and V. S. Filinov. Combination of Quantum Kinetic Theory and First Principle Simulations for Strongly Correlated Quantum Plasmas, in *Progress in Nonequilibrium Green's Functions II*, eds. M. Bonitz and D. Semkat. World Scientific, 2003.
- [BPS06] M. Boninsegni, N. V. Prokofiev, and B. V. Svistunov, Worm algorithm and diagrammatic Monte Carlo: A new approach to continuous-space path integral Monte Carlo simulations. *Phys. Rev. E* **74**, 036701 (2006).
- [CA80] D. M. Ceperley and B. J. Alder. Ground State of the Electron Gas by a Stochastic Method. *Phys Rev Lett* **45** (1980), 566-569.
- [CA84] D. M. Ceperley and B. J. Alder. Quantum Monte Carlo for molecules: Green's function and nodal release. *J. Chem. Phys.* **81** (1984), 5833.
- [CB88] D. M. Ceperley and B. Bernu. The Calculation of Excited State Properties with Quantum Monte Carlo, *J. Chem. Phys.* **89** (1988), 6316.
- [CD04] J.F. Corney and P.D. Drummond. Gaussian Quantum Monte Carlo Methods for Fermions and Bosons. *Phys. Rev. Lett.* **93**, 260401 (2004).
- [Cep91] D. M. Ceperley. Fermion nodes. *J. Stat. Phys.* **63** (1991), 1237-1267.
- [Cep92] D. M. Ceperley. Path-integral calculations of normal liquid ^3He . *Phys. Rev. Lett.* **69** (1992), 331-334.
- [Cep95] D. M. Ceperley. Path integrals in the theory of condensed helium. *Rev. Mod. Phys.* **67** (1995), 279-355.

- [Cep95ii] D. M. Ceperley. The Simulation of Quantum Systems, in *Recent Progress in Many-Body Theories, Vol 4*; eds. E. Schachinger, H. Mitter, and M. Sormann. Plenum Press, 1995.
- [Cep96] D. M. Ceperley. Path Integral Monte Carlo Methods for Fermions. in *Monte Carlo and Molecular Dynamics of Condensed Matter Systems*. eds. K. Binder and G. Ciccotti. Editrice Compositori (Bologna, Italy), 1996.
- [Cep96ii] D. M. Ceperley. Solving Quantum Many-Body Problems with Random Walks. in *Computational Physics, Proc. Ninth Physics Summer School, Australian National University*. eds. H. J. Gardner and C. M. Savage. World Scientific, 1997.
- [Cep00] D. M. Ceperley. Lecture notes from QMC Summer School 2007, Materials Computation Center, University of Illinois at Urbana-Champaign. Found on <http://cms.mcc.uiuc.edu/>
- [Cep03] D. M. Ceperley. Metropolis Methods for Quantum Monte Carlo Simulations, in *The Monte Carlo Method in the Physical Sciences*, edited by J. E. Gubernatis (AIP Conference Proceedings, 2003), p.85-98. arXiv:physics/0306182v1 [physics.comp-ph]
- [CJL06] E. Cancès, B. Jourdain and T. Lelièvre. Quantum Monte Carlo simulations of fermions: A mathematical analysis of the fixed-node approximation. *Math. Model. Method. Appl. Sci.* **16** (2006), 1403-1449.
- [CM00] D. M. Ceperley and B. Militzer, Path Integral Monte Carlo Calculation of the Deuterium Hugoniot. *Phys Rev. Lett.* **85** (2000), 1890-1893.
- [Coh80] D. L. Cohn. *Measure Theory*. Birkhäuser, 1980.
- [Cor08] P. R. Corboz. Simulations of strongly correlated fermions and bosons (PhD thesis), Eidgenössische Technische Hochschule Zürich, 2008.

- [CP84] D. M. Ceperley and E. L. Pollock, Simulation of quantum many-body systems by path-integral methods. *Phys. Rev. B.* **30**, 2555-2568 (1984).
- [Dir33] P. A. M. Dirac. The Lagrangian in quantum mechanics. *Physikalische Zeitschrift der Sowjetunion* **3** (1933), 64.
- [DM01] M. V. Dikovsky and C. H. Mak. Analysis of the Multilevel Blocking approach to the fermion sign problem: Accuracy, errors and practice. *Phys. Rev. B.* **63** (2001), 235105.
- [DT82] I. M. Davies and A. Truman. Laplace asymptotic expansions of conditional Wiener Integrals and generalised Mehler kernel formulas. *J. Math. Phys.* **23**, 2059 (1982)
- [DT82ii] I. M. Davies and A. Truman. Laplace asymptotic expansions of conditional Wiener Integrals and applications to quantum physics. *Springer Lecture Notes in Physics* **173**, 40 (1982)
- [DT10] W. M. Dumas and M. V. Tretyakov. Computing conditional Wiener integrals of functionals of a general form. *IMA Journal of Numerical Analysis* (accepted; to appear)
- [Dyn65] E. B. Dynkin. *Markov Processes*. Academic Press, 1965.
- [FC81] M. Creutz and B. A. Freedman. A statistical approach to quantum mechanics. *Ann. Phys.* **132** (1981), 427-462.
- [Ein26] A. Einstein, *Investigations on the Theory of the Brownian Movement*. Methuen, 1926.
- [EMWG98] C. H. Mak, R. Egger and H. Weber-Gottschick. Multilevel Blocking Approach to the Fermion Sign Problem in Path-Integral Monte Carlo Simulations. *Phys. Rev. Lett.* **81** (1998), 4533-4536.

- [ESY93] A. D. Egorov, P. I. Sobolevsky and L. A. Yanovich. *Functional Integrals: approximate evaluation and applications*. Kluwer, 1993.
- [Fey48] R. P. Feynman, Space-Time Approach to Non-Relativistic Quantum Mechanics. *Rev. Mod. Phys.* **20**, 367-387 (1948).
- [Fey53] R. P. Feynman, Atomic Theory of the λ Transition in Helium. *Phys. Rev.* **91**, 1291-1301 (1953).
- [Fey72] R. P. Feynman. *Statistical Mechanics*. Westview Press, 1972.
- [Fey83] R. P. Feynman. Simulating physics with computers. *Int. J. of Theo. Phys.* **21** (1982) 467-488.
- [FH65] R. P. Feynman and A. R. Hibbs. *Quantum Mechanics and Path Integrals*. McGraw-Hill, 1965.
- [Fre85] M. I. Freidlin, *Functional integration and partial differential equations*. Princeton University Press, 1985.
- [Fri64] A. Friedman. *Partial Differential Equations of Parabolic Type*. Prentice-Hall, 1964.
- [Fri73] K. O. Friedrichs. *Spectral Theory of Operators in Hilbert Space*. Springer-Verlag, 1973.
- [Gen98] J. E. Gentle. *Random Number Generation and Monte Carlo Methods*. Springer, 1998.
- [GIV98] P. N. Vorontsov-Velyaminov, S. D. Ivanov and R. I. Gorbunov. Quantum gas in an external field: Exact grand canonical expressions and numerical treatment. *Phys. Rev. E* **59** (1999), 168-176.
- [GY56] I. M. Gelfand and A. M. Yaglom. Integration in functional spaces and its application in quantum physics. *J. Math. Phys.* **1** (1960), 48-69.

- [GM84] S. A. Gladyshev and G. N. Milstein. Runge-Kutta method for evaluation of Wiener integrals of exponential-type functionals. *Zh. Vychisl. Mat. i Mat. Fiz.* **24** (1984), 1136-1149.
- [GM98] A. Gelman and X. L. Meng. Simulating Normalizing Constants: From Importance Sampling to Bridge Sampling to Path Sampling. *Statistical Science* **13** (1998), 163-185.
- [GRS96] W. R. Gilks, S. Richardson and D. J. Spiegelhalter (eds.) *Markov Chain Monte Carlo in practice*. Chapman & Hall, 1996.
- [GS72] I. I. Gikhman and A. V. Skorokhod. *Stochastic Differential Equations*. Springer, 1972.
- [GS89] J. Goodman and A. D. Sokal. Multigrid Monte Carlo method: Conceptual foundations. *Phys. Rev. D* **40** (1989), 2035-2071.
- [HF04] A. O'Hagan and J. Forster. *Bayesian Inference*. Arnold, 2004.
- [Hin69] D. V. Hinkley. On the ratio of two correlated normal random variables. *Biometrika* **56** (1969), 635.
- [HR05] C. P. Herrero and R. Ramírez, Rare-gas solids under pressure: A path-integral Monte Carlo simulation, *Phys. Rev. B* **71**, 174111 (2005)
- [IM74] K. Itô and H. P. McKean. *Diffusion Processes and their Sample Paths*. Springer-Verlag, 1974.
- [Iva05] S. D. Ivanov. Path Integral studies of quantum systems at finite temperatures (PhD thesis), Stockholm University, 2005.
- [IT01] T. Ichinose and H. Tamura. The Norm Convergence of the Trotter-Kato Product Formula with Error Bound. *Commun. Math. Phys.* **217** (2001), 489-502.

- [IW81] N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes*. North-Holland, 1981.
- [Kac51] M. Kac, On some connections between probability theory and differential and integral equations, in *Proc. Second Berkeley Symp. Math. Statist.Prob.* edited by J. Neyman. Univ. of California Press, 1951. p.189-215.
- [Kac56] M. Kac. *Some Stochastic Problems in Physics and Mathematics*. Socony Mobil Oil Co., Inc. Field Research Laboratory, 1956.
- [Kac57] M. Kac. *Probability and Related Topics in Physical Sciences*. Interscience, 1957.
- [KFS96] I. Kosztin, B. Faber and K. Schultzen. Introduction to the Diffusion Monte Carlo Method. *Am. J. Phys.* **64** (1996), 633. arXiv:physics/9702023v1 [physics.comp-ph]
- [KS98] I. Karatzas, S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, 1998.
- [Kle95] H. Kleinert. *Path Integrals in Quantum Mechanics, Statistics and Polymer Physics*. World Scientific, 1995.
- [KF57] A. N. Kolmogorov and S. V. Fomin. *Elements of the Theory of Functions and Functional Analysis*. Dover, 1957.
- [KM59] S. Karlin and J. McGregor. Coincidence Probabilities. *Pac. J. Math.* **9** (1959), 1141-1164.
- [Kry99] N. V. Krylov. On Kolmogorov's equations for finite dimensional diffusions, in *Stochastic PDE's and Kolmogorov Equations in Infinite Dimensions* edited by G. Da Prato. Springer, 1999. p.1-63.
- [LL70] L. D. Landau and E. M. Lifschitz. *Statistical Physics*. Addison-Wesley, 1970.
- [LMR97] P. V. Landshoff, A. Metherell and W. G. Rees. *Essential Quantum Physics*. CUP, 1997.

- [Lev39] P. Lévy, Sur certains processus stochastiques homogenes. *Compos. Math.* **7**, 283-339 (1939).
- [Lob96] Yu.Yu. Lobanov. Functional integrals for nuclear many-particle systems. *J. Phys. A: Math. Gen.* **29** (1996), 6653-6669.
- [LN04] D. P. Landau and K. Nho. Bose-Einstein condensation temperature of a homogeneous weakly interacting Bose gas: Path integral Monte Carlo study. *Phys. Rev. A* **70** (2004), 053614.
- [Lyu05] A. P. Lyubartsev. Simulation of excited states and the sign problem in the path integral Monte Carlo method. *J. Phys. A: Math. Gen.* **38** (2005) 6659-6674.
- [Mac97] V. Mackevicius. Convergence rate of Euler scheme for stochastic equations: functionals of solutions. *Math. Comp. Simul.* **44** (1997), 109-121.
- [Mak92] C. H. Mak. Stochastic Method for Real-Time Path Integrations. *Phys. Rev. Lett.* **68** (1992), 899-902.
- [Mak09] C. H. Mak. The sign problem in real-time path integral simulations: Using the cumulant action to implement multilevel blocking. *J. Chem. Phys.* **131** (2009), 044125.
- [Met53] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller. Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21** (1953), 1087-92.
- [MH89] B. A. Mason and K. Hess. Quantum Monte Carlo calculations of electron dynamics in dissipative solid-state systems using real-time path integrals. *Phys. Rev. B* **39** (1989), 5051-5069.
- [Mil00] B. Militzer. Path Integral Monte Carlo simulations of hot dense Hydrogen (PhD thesis), University of Illinois at Urbana-Champaign, 2000.

- [Mis06] Y. Mishchenko. Remedy for the fermion sign problem in the Diffusion Monte Carlo method for few fermions with antisymmetric diffusion process. *Phys. Rev. E* **73** (2006), 026706.
- [MR07] L. Muhlbacher and E. Rabani. Real-Time Path Integral Approach to Nonequilibrium Many-Body Quantum Systems. *Phys. Rev. Lett.* **100** (2008), 176403.
- [MT99] G. N. Milstein and M. V. Tretyakov. Simulation of a space-time bounded diffusion. *Ann. Appl. Probab.* **9** (1999), 732-779.
- [MT04] G. N. Milstein and M. V. Tretyakov. *Stochastic Numerics for Mathematical Physics*. Springer, 2004.
- [MT04ii] G. N. Milstein, M.V. Tretyakov. Evaluation of conditional Wiener integrals by numerical integration of stochastic differential equations. *J. Comp. Phys.* **197** (2004), 275-298.
- [MT05] G. N. Milstein and M.V. Tretyakov. Numerical integration of stochastic differential equations with nonglobally Lipschitz coefficients. *SIAM J. Numer. Anal.* **43** (2005), 1139-1154.
- [MT09] G. N. Milstein and M.V. Tretyakov. Monte Carlo algorithms for backward equations in nonlinear filtering. *Advances in Applied Probability* **41** (2009), No 1.
- [Mur07] I. Murray. Advances in Markov Chain Monte Carlo methods (PhD thesis), University College London, 2007.
- [Nea93] R. M. Neal. Probabilistic inference using Markov Chain Monte Carlo methods. Technical Report CRG-TR-93-1, Dept. Comp. Sci., Toronto University, 1993.
- [NK92] W. Newman and A. Kuki. Improved methods for Path Integral Monte Carlo in fermionic systems. *J. Chem. Phys.* **96** (1992), 1409-1417.

- [NKPTR02] R. J. Needs, P. R. C. Kent, A. R. Porter, M. D. Towler, G. Rajagopal. Quantum Monte Carlo Calculations for Ground and Excited States. *Int. J. Quantum. Chem.* **86** (2002), 218-225.
- [Neg86] J. W. Negele, Monte Carlo Studies of Nuclear Many-Particle Systems. *J. Stat. Phys.* **43** (1986), 991-1015 .
- [NO88] J. W. Negele and H. Orland, *Quantum Many-Particle Systems* (Westview Press, 1988).
- [PD03] C. Predescu and J. D. Doll. Random series and discrete path integral methods: The Lévy-Ciesielski implementation. *Phys. Rev. E* **67** (2003), 026124.
- [PM04] E. L. Pollock and B. Miltzer. Dense plasma effects on nuclear reaction rates. *Phys. Rev. Lett.* **92** (2004), 021101.
- [Pre04] C. Predescu. Existence of short-time approximations of any polynomial order for the computation of density matrices by path integral methods. *Phys. Rev. E* **69** (2004), 056701
- [Roe94] G. Roepstorff. *Path Integral Approach to Quantum Physics*. Springer, 1994.
- [RW94] L. C. G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales. Vol 2. Itô Calculus*. Wiley, 1994. CUP, 2000.
- [RY99] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*. Springer, 1999.
- [San03] R. R. dos Santos. Introduction to Quantum Monte Carlo simulations for fermionic systems. *Braz. J. Phys.* **33** (2003), 36. arXiv:cond-mat/0303551v1
- [She05] S. V. Shevkunov. Calculation of the Equation of State of a Dense Hydrogen Plasma by the Feynman Path Integral Method. *Zh. Eksp. Teor. Fiz.* **127**, 696-716 (2005)
[JETP **100**, 617-634 (2005)]

- [Shi89] A. N. Shiryaev. *Probability*. Springer, 1996 (2nd edition, first published in Russian: Nauka, 1989.)
- [Sim05] B. Simon. *Functional Integration and Quantum Physics*. AMS Chelsea Publishing, 2005.
- [Sip92] M. Sipser. The history and status of the P versus NP question. *Proceedings of the twenty-fourth annual ACM symposium on Theory of computing*, 603-618 (1992).
- [Ste99] M. L. Stedman. Ground-state and finite-temperature quantum Monte Carlo simulations of many-particle systems (PhD thesis), Imperial College, London (1999).
- [Suz86] M. Suzuki. Quantum Statistical Monte Carlo Methods and Applications to Spin Systems. *J. Stat. Phys.* **43** (1986), 883-909.
- [Suz91] M. Suzuki. General Theory of fractal path integrals with applications to many-body theories and statistical physics. *J. Math. Phys.* **32** (1991), 400-407.
- [Suz94] M. Suzuki. Convergence of General Decompositions of Exponential Operators. *Commun. Math. Phys.* **163** (1994), 491-508.
- [TI84] M. Takahashi and M. Imada. Monte Carlo calculation of quantum systems. *J. Phys. Soc. Japan* **53** (1984), 963-974.
- [Tit58] E. C. Titchmarsh. *Eigenfunction Expansions Associated with Second-Order Differential Equations, Part II*. OUP, 1958.
- [Tol02] L. Tolmatz. On the distribution of the square integral of the Brownian bridge. *Ann. Appl. Prob.* **30** (2002), 253-269.
- [Tow06] M. D. Towler. The quantum Monte Carlo method. *Phys. Stat. Sol. (b)* **243** (2006), 2573–2598.
- [Tro59] H. F. Trotter. On the product of semi-groups of operators. *Proc. Am. Math. Soc.* **10**, 545-551 (1959).

- [TW05] M. Troyer and U-J. Wiese. Computational Complexity and Fundamental Limitations to Fermionic Quantum Monte Carlo Simulations. *Phys. Rev. Lett.* **94** (2005), 170201.
- [VGM84] A. D. Ventzel, S. A. Gladyshev, G. N. Milstein. Piecewise constant approximation for Monte Carlo calculation of Wiener integrals. *Theor. Prob. Appl.* **29** (1984), 744-752.
- [Wie24] N. Wiener. The Average Value of a Functional. *Proc. London. Math. Soc. Ser. 2*, **22** (1924), 454-467.
- [Wil01] D. Williams. *Weighing the Odds, a Course in Probability and Statistics*. CUP, 2001.
- [Wil91] D. Williams. *Probability with Martingales*. CUP, 1991.
- [Wag88] W. Wagner. Unbiased multi-step estimators for the Monte Carlo evaluation of certain functional integrals. *J. Comp. Phys.* **79** (1988), 336-352.
- [ZNF77] V. M. Zamalin, G. E. Norman, and V. S. Filinov. *The Monte-Carlo Method in Statistical Thermodynamics*. Nauka, 1977.