Implied Volatility from Local Volatility: A Path Integral Approach

Article *in* SSRN Electronic Journal · January 2014

DOI: 10.2139/ssrn.2457618

Tai-Ho Wang
Baruch College

Implied volatility from local volatility: A path integral approach

Tai-Ho Wang
Department of Mathematics, Baruch College, CUNY
1 Bernard Baruch Way, New York, NY 10010, USA
e-mail: tai-ho.wang@baruch.cuny.edu

Jim Gatheral
Department of Mathematics, Baruch College, CUNY
1 Bernard Baruch Way, New York, NY 10010, USA
email: jim.gatheral@baruch.cuny.edu

Abstract

Assuming local volatility, we derive an exact Brownian bridge representation for the transition density; an exact expression for the transition density in terms of a path integral then follows. By Taylor-expanding around a certain path, we obtain a generalization of the heat kernel expansion of the density which coincides with the classical one in the time-homogeneous case, but is more accurate and natural in the time inhomogeneous case. As a further application of our path integral representation, we obtain an improved most-likely-path approximation for implied volatility in terms of local volatility.

Keywords and phrases: Small time asymptotic expansion, heat kernels expansion, implied volatility, local volatility model, most likely path, path integral

IMPLIED VOLATILITY FROM LOCAL VOLATILITY: A PATH INTEGRAL APPROACH

TAI-HO WANG AND JIM GATHERAL

Abstract. Assuming local volatility, we derive an exact Brownian bridge representation for the transition density; an exact expression for the transition density in terms of a path integral then follows. By Taylor-expanding around a certain path, we obtain a generalization of the heat kernel expansion of the density which coincides with the classical one in the time-homogeneous case, but is more accurate and natural in the time inhomogeneous case. As a further application of our path integral representation, we obtain an improved most-likely-path approximation for implied volatility in terms of local volatility.

In memory of our long term collaborator and friend, a passionate mathematician, Peter Laurence.

1. Introduction

Because of their consistency with the known prices of European options, and despite their unrealistic dynamical implications, local volatility models continue to be used in practice as powerful tools for risk management of equity derivatives portfolios. Under the forward measure (with no drift), local volatility models take the form

$$\frac{dS_t}{S_t} = \sigma_\ell(S_t, t)\, dB_t, \tag{1.1}$$

where B_t is a Brownian motion and σ_ℓ is a local volatility function that depends only on the underlying level S and the time t.

Assume that prices of European options of all strikes K and expirations T are given or equivalently that the Black-Scholes implied volatility function $\sigma_{\mathrm{BS}}(K,T)$ is known. In that case, it is straightforward to compute the local volatility function σ_ℓ from, for example, equation (1.10) of Gatheral [6]:

$$\sigma_\ell{}^2(K,T) = \frac{\frac{\partial w}{\partial T}}{\left(1 - \frac{k}{2\,w}\,\frac{\partial w}{\partial k}\right)^2 - \frac{1}{4}\,\left(\frac{1}{4} + \frac{1}{w}\right)\,\left(\frac{\partial w}{\partial k}\right)^2 + \frac{1}{2}\frac{\partial^2 w}{\partial k^2}} \tag{1.2}$$

Key words and phrases. Small time asymptotic expansion, heat kernels expansion, implied volatility, local volatility model, most likely path, path integral.

where k denotes the log-strike $k := \log K/S$ and w, the Black-Scholes implied total variance, given by $w(K,T) := \sigma_{\text{BS}}^2\,(K,T)\;T$.

In practice, we observe option prices for only a finite set of strikes and expirations. Moreover (see for example Gatheral and Jacquier [8]), it is very hard if not impossible to find a functional form for implied volatility that both matches observed prices and is free from static arbitrage. One alternative approach is to assume a parameterized functional form for the local volatility function $\sigma_\ell(S,t)$ and price a finite set of European options, tuning the parameters of the function until a satisfactory fit is achieved. Such calibration of local volatility models to given option prices is in practice typically performed using numerical PDE techniques. However, numerical PDE techniques are slow and moreover are not practical in higher dimensions.

Alternatively, to achieve better understanding of the qualitative properties of local volatility models, and potentially faster calibration, both academics and practitioners have exploited asymptotic expansions of implied volatility in terms of local volatility. First, Berestycki, Busca and Florent [2] solved the nonlinear PDE (1.2) for the implied total variance w in the small time to expiration limit, obtaining an exact expression for implied volatility as an integral of local volatility. Subsequently, this asymptotic approximation was extended, to first order in time to expiry $\tau = T - t$ by Henry-Labordère (see the article in this volume and also Henry-Labordère [13]), and then to second order in Gatheral, Hsu, Laurence, Ouyang, and Wang [7] using the heat kernel expansion. Jordan and Tier [14] apply similar methods to derive an asymptotic solution for the SABR and CEV models. In related work, the paper of Cheng, Costanzino, Liechty, Mazzucato, and Nistor [5] derives an operator expansion of the density, which up to first order agrees with prior expansions obtained using the heat kernel expansion. As an earlier example of work in a similar spirit to the most-likely-path approach of our paper, Baldi and Caramellino [1] develop a small-time expansion for the hitting probability of a one-dimensional diffusion.

Our contribution in this paper is to derive an exact Brownian bridge representation for the transition density, from which an exact expression for the transition density in terms of a path integral follows. Indeed, the path integral representation of the density has often been used as a powerful tool for the derivation of improved asymptotic expansions of the transition density. For example, in the foregoing, we apply a technique from the paper of Goovaerts, De Schepper, and Decamps [10]. An earlier paper by Linetsky [16] provides a more general survey of the application of path integral techniques to option pricing.

By replacing all paths that contribute to the path integral by the *most-likely-path*, the unique path that minimizes the action functional in the path integral formulation, we obtain a new approximation to the transition density which is both more accurate and natural than the classical heat kernel version. As an application, we obtain an improved most-likely-path approximation for implied volatility in terms of local volatility.

The most-likely-path (MLP) approach has been used to analyze the asymptotic behavior of implied volatility in stochastic volatility models in Gatheral [6]; this analysis is further elaborated in an article by Keller-Ressel and Teichmann [15] in these proceedings. Guyon and Henry-Labordère [11] and Reghai [17] both explore alternative definitions of the most likely path, achieving improved accuracy by considering fluctuations around the MLP. In particular, Guyon and Henry-Labordère [11] compare and contrast various approximations in a unified setting. Though the approach of Guyon and Henry-Labordère [11] differs from our path integral approach in the current paper, it is worth mentioning that their heat kernel approximation is closely related to ours. Once again however, our path integral approach leads to an unambiguously natural definition of the most-likely-path.

Our paper is organized as follows. In Section 2, we derive Brownian bridge and path integral representations for the transition density of one dimensional diffusions. As an application, in Section 3, we present a novel probabilistic derivation of the heat kernel expansion, also referred to as the WKB method in the physics literature. For time homogeneous diffusions, this new expansion recovers the conventional heat kernel expansion; however, in the time-inhomogeneous case, the two expansions differ a little. In Section 4, we present heuristic derivations of known small time asymptotic expansion of implied volatility to zeroth order. From the path integral perspective, these known approximations are suboptimal in the sense that they correspond to computing the optimal path of an approximate but incomplete action functional. By considering the optimal path of the exact action functional, we show how an optimal approximation may be computed. An interesting feature of the optimal approximation is that it recovers the implied volatility of the time dependent Black-Scholes model exactly, which so far, to the best of our knowledge, none of the existing small time approximations are able to achieve. Finally, in Section 5, we summarize and conclude.

Throughout the text, B_t denotes the standard Brownian motion defined on the filtered probability space $(\Omega, \mathcal{F}_t, \mathbb{P})$ satisfying the usual conditions. X_t denotes the Brownian motion with some drift h. $p^X(T, y|t, x)$ denotes the transition density of X from x at time t to y at time T and similarly $p^S(T, s_T|t, s_t)$ is the transition density from s_t to s_T of the process S_t. Moreover, dot will always refer to the partial derivative with respect to the time variable and prime to the space variables x or s.

2. Path integral representations for transition density

In this section, we derive path integral representations of the transition density and of the call prices under local volatility, which will in turn yield the most-likely-path approximation to implied volatility. The key ingredient in this derivation is a Brownian bridge representation for the transition density, which though straightforward, does not appear to be well-known.

We start with the case of one-dimensional Brownian motion with general but Markovian drift. We reduce the more general diffusion case which concerns us here to this one by applying the well-known Lamperti change of variable.

2.1. **Brownian bridge representations.** Two Brownian bridge representations for the transition density of Brownian motion with general but Markovian, smooth and bounded, drift are derived in Theorem 1. The first expression, (2.1), will be used in the derivation of the path integral representation for transition density in Section 2.2 and the second, (2.2), will be used to derive the heat kernel expansion of transition density in Section 3.

Theorem 1. *Let X_t be a Brownian motion with drift driven by*

$$dX_t = dB_t + h(X_t, t)dt,$$

where the drift h is assumed smooth and bounded. Let H be an antiderivative of h with respect to x, i.e., $\frac{\partial}{\partial x}H(x,t) = h(x,t)$, for all x and t. The transition density p^X of X_t has the following two equivalent Brownian bridge representations:

$$p^X(T, y|t, x) = \phi(T-t, y-x)\,\tilde{\mathbb{E}}_{x,y}\left[e^{\int_t^T h(X_s,s)dX_s - \frac{1}{2}\int_t^T h^2(X_s,s)ds}\right] \tag{2.1}$$

and

$$p^X(T,y|t,x) = \phi(T-t,y-x)\, e^{H(y,T)-H(x,t)} \times \tilde{\mathbb{E}}_{x,y}\left[e^{-\frac{1}{2}\int_t^T h^2(X_s,s)+h_x(X_s,s)+2H_t(X_s,s)ds}\right], \tag{2.2}$$

where ϕ is the Gaussian density $\phi(t,\xi) = \frac{1}{\sqrt{2\pi t}}\, e^{-\frac{\xi^2}{2t}}$. The notation $\tilde{\mathbb{E}}_{x,y}[\cdot]$ denotes the expectation under the Brownian bridge measure from x to y.

Proof. Note that X_t under the original measure $\mathbb{P}$ is a Brownian motion with drift h. Define a new probability measure $\tilde{\mathbb{P}}$ through the Radon-Nikydom derivative

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} = e^{-\int_t^T h(X_s,s)dB_s - \frac{1}{2}\int_t^T h^2(X_s,s)ds}.$$

By the Girsanov theorem, X_t is a Brownian motion under $\tilde{\mathbb{P}}$. Given any bounded measurable function f, we have, since $dB_t = dX_t - h(X_t,t)dt$,

$$\mathbb{E}_{t,x}[f(X_T)] = \tilde{\mathbb{E}}_{t,x}\left[f(X_T)\frac{d\mathbb{P}}{d\tilde{\mathbb{P}}}\right] = \tilde{\mathbb{E}}_{t,x}\left[f(X_T)e^{\int_t^T h(X_s,s)dX_s - \frac{1}{2}\int_t^T h^2(X_s,s)ds}\right],$$

where, for notational simplicity, $\mathbb{E}_{t,x}[\cdot]$ denotes the conditional expectation $\mathbb{E}[\cdot|X_t = x]$, and similarly for $\tilde{\mathbb{E}}_{t,x}[\cdot]$. It follows that, for any bounded measurable function f,

$$\int f(y)p^X(T,y|t,x)dy = \int f(y)\tilde{\mathbb{E}}_{x,y}\left[e^{\int_t^T h(X_s,s)dX_s - \frac{1}{2}\int_t^T h^2(X_s,s)ds}\right]\phi(T-t,y-x)dy,$$

where $\tilde{\mathbb{E}}_{x,y}[\cdot] = \tilde{\mathbb{E}}[\cdot|X_t = x, X_T = y]$. Consequently, (2.1) follows, i.e.,

$$p^X(T,y|t,x) = \phi(T-t,y-x)\tilde{\mathbb{E}}_{x,y}\left[e^{\int_t^T h(X_s,s)dX_s - \frac{1}{2}\int_t^T h^2(X_s,s)ds}\right].$$

Furthermore, Ito's formula implies that

$$\int_t^T h(X_s,s)dX_s = H(X_T,T) - H(X_t,t) - \int_t^T \left[H_t(X_s,s) + \frac{h_x(X_s,s)}{2}\right]ds,$$

where we recall that H is an antiderivative of h with respect to x. Thus,

$$e^{\int_t^T h(X_s,s)dX_s} = e^{H(X_T,T)-H(X_t,t)-\int_t^T\left[H_t(X_s,s)+\frac{h_x(X_s,s)}{2}\right]ds}.$$

We further rewrite the transition density as

$$p^X(T,y|t,x) = \phi(T-t,y-x)e^{H(y,T)-H(x,t)}\tilde{\mathbb{E}}_{x,y}\left[e^{-\frac{1}{2}\int_t^T h^2(X_s,s)+h_x(X_s,s)+2H_t(X_s,s)ds}\right].$$

This completes the proof of (2.2). □

Remark 1. We remark that the conditional expectations in both (2.1) and (2.2) are under the Brownian bridge measure since X_t is a Brownian motion under $\tilde{\mathbb{P}}$. One intriguing feature of the representation (2.2) is that, if we Taylor expand the conditional expectation for small $T-t$ around the straight line connecting the initial and terminal points, we recover the heat kernel expansion in the time-homogeneous case and probably do better than the heat kernel expansion in the time-inhomogeneous case. See Section 3 for more detailed discussions on the heat kernel expansion.

Now for the general diffusion case, consider the process S_t driven by the stochastic differential equation (SDE)

$$dS_t = \mu(S_t, t)dt + a(S_t, t)dB_t, \quad S_0 = s_0,$$

where for simplicity, we assume the coefficients μ and a are Lipschitz and of linear growth; a is further assumed strictly away from zero. By applying the Lamperti transformation $x = \int_{s_0}^{s} \frac{d\xi}{a(\xi,t)}$, the process S_t is transformed into a Brownian motion with drift. Specifically, denote the transformation from s to x by $x = \varphi(s,t) = \int_{s_0}^{s} \frac{d\xi}{a(\xi,t)}$. Applying Ito's formula to $X_t = \varphi(S_t, t)$ yields

$$\begin{aligned} & dX_t = d\varphi(S_t, t) \\ = & \left[\dot{\varphi}(S_t, t) + \mu(S_t, t)\varphi_s(S_t, t) + \frac{a^2(S_t, t)}{2}\varphi_{ss}(S_t, t)\right] dt + \varphi_s(S_t, t)a(S_t, t)dB_t \\ = & \left[\dot{\varphi}(S_t, t) + \frac{\mu(S_t, t)}{a(S_t, t)} - \frac{a_s(S_t, t)}{2}\right] dt + dB_t \\ = & \; dB_t + h(X_t, t)dt, \end{aligned}$$

where subindices of φ and a refer to partial derivatives. The function h is defined as $h(x,t) = \dot{\varphi}(s,t) + \frac{\mu(s,t)}{a(s,t)} - \frac{a_s(s,t)}{2}$, with $s = \varphi^{-1}(x,t)$. The transition densities p^S for S_t and p^X for X_t are then related as

$$p^S(T, s_T|t, s_t) = \frac{1}{a(s_T, T)} p^X(T, x_T|t, x_t),$$

with $x_T = \varphi(s_T, T)$ and $x_t = \varphi(s_t, t)$. Thus, the transition from the Brownian bridge representation for p^X to a similar representation for p^S is straightforward by applying Theorem 1. Theorem 2 below formalizes this result.

Theorem 2. *Let S_t be the diffusion process driven by the stochastic differential equation*

$$dS_t = \mu(S_t, t)dt + a(S_t, t)dB_t, \quad S_0 = s_0.$$

Denote the Lamperti transformation from s to x by $x = \varphi(s,t) = \int_{s_0}^{s} \frac{d\xi}{a(\xi,t)}$. Define the function h by $h(x,t) = \dot{\varphi}(s,t) + \frac{\mu(s,t)}{a(s,t)} - \frac{a_s(s,t)}{2}$, with $s = \varphi^{-1}(x,t)$, where subindices refer to corresponding partial derivatives. Let H be an antiderivative of h with respect to x, namely, $\frac{\partial}{\partial x}H(x,t) = h(x,t)$, for all x and t. Then the transition density p^S of

S_t from (t, s_t) to (T, s_T) has the following Brownian bridge representations:

$$p^S(T, s_T|t, s_t) = \frac{\phi(T-t, \varphi(s_T,T) - \varphi(s_t,t))}{a(s_T,T)} \tilde{\mathbb{E}}_{\varphi(s_t,t),\varphi(s_T,T)} \left[e^{\int_t^T h(X_s,s)dX_s - \frac{1}{2}\int_t^T h^2(X_s,s)ds} \right] \tag{2.3}$$

and

$$\begin{aligned} p^S(T, s_T|t, s_t) &= \frac{\phi(T-t, \varphi(s_T,T) - \varphi(s_t,t))}{a(s_T,T)} e^{H(\varphi(s_T,T),T) - H(\varphi(s_t,t),t)} \times \\ &\quad \tilde{\mathbb{E}}_{\varphi(s_t,t),\varphi(s_T,T)} \left[e^{-\frac{1}{2}\int_t^T h^2(X_s,s) + h_x(X_s,s) + 2H_t(X_s,s)ds} \right], \end{aligned} \tag{2.4}$$

where again ϕ denote the Gaussian density $\phi(t,\xi) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{\xi^2}{2t}}$. As before, the notation $\tilde{\mathbb{E}}_{x,y}[\cdot]$ denotes the expectation under the Brownian bridge measure from x to y.

Note that the X_t process in both expressions (2.3) and (2.4) is a Brownian bridge from $x_T = \varphi(s_T,T)$ to $x_t = \varphi(s_t,t)$. One application of such Brownian bridge representations of transition densities is to devise more efficient simulation schemes. For example, for some given function f, we may compute numerically the expectation $\mathbb{E}_{t,s_t}[f(S_T)]$ in x-space as

$$\begin{aligned} \mathbb{E}_{t,s_t}[f(S_T)] = \mathbb{E}_{t,x_t}[f(\varphi^{-1}(X_T))] &= \int f \circ \varphi^{-1}(x_T,T)\, p^X(T, x_T|t, x_t)dx_T \\ = \quad & \int f \circ \varphi^{-1}(x_T,T)\, \phi(T-t, x_T - x_t) e^{H(x_T,T) - H(x_t,t)} \\ & \times \tilde{\mathbb{E}}_{x_t,x_T} \left[e^{-\frac{1}{2}\int_t^T h^2(X_\tau,\tau) + h_x(X_\tau,\tau) + 2H_t(X_\tau,\tau)d\tau} \right] dx_T \\ = \quad & \mathbb{E}\left\{ f \circ \varphi^{-1}(Y,T) e^{H(Y,T) - H(x_t,t)}\, \tilde{\mathbb{E}}_{x_t,Y} \left[e^{-\frac{1}{2}\int_t^T h^2(X_\tau,\tau) + h_x(X_\tau,\tau) + 2H_t(X_\tau,\tau)d\tau} \right] \right\} \end{aligned}$$

where Y is a normal random variable with mean x_t and variance $T-t$. Therefore, if there is an efficient method to calculate or approximate the conditional expectation $\tilde{\mathbb{E}}_{x_t,Y}\left[e^{-\frac{1}{2}\int_t^T h^2(X_\tau,\tau) + h_x(X_\tau,\tau) + 2H_t(X_\tau,\tau)d\tau} \right]$ in the Brownian bridge measure, the expectation $\mathbb{E}_{t,s_t}[f(S_T)]$ could potentially be computed more efficiently. Since X_t is a Brownian bridge from x_t to Y, one obvious approximation is to simply replace the integral in the exponent of $\tilde{\mathbb{E}}_{x_t,Y}\left[e^{-\frac{1}{2}\int_t^T h^2(X_\tau,\tau) + h_x(X_\tau,\tau) + 2H_t(X_\tau,\tau)d\tau} \right]$ with the integrand evaluated along the straight line $x_\tau = \frac{T-\tau}{T-t} x_t + \frac{\tau - t}{T-t} Y$ for $\tau \in [t,T]$. In other words,

$$\begin{aligned} & \tilde{\mathbb{E}}_{x_t,Y} \left[e^{-\frac{1}{2}\int_t^T h^2(X_t,t) + h_x(X_t,t) + 2H_t(X_t,t)dt} \right] \\ \approx \quad & e^{-\frac{1}{2}\int_t^T h^2(x_\tau,\tau) + h_x(x_\tau,\tau) + 2H_t(x_\tau,\tau)d\tau}. \end{aligned} \tag{2.5}$$

Hence,

$$\mathbb{E}_{t,s_t}[f(S_T)] \approx e^{-\frac{1}{2}\int_t^T h^2(x_\tau,\tau) + h_x(x_\tau,\tau) + 2H_t(x_\tau,\tau)d\tau}\, \mathbb{E}\left[f \circ \varphi^{-1}(Y,T)\, e^{H(Y,T) - H(x_t,t)} \right],$$

after which we need only simulate the normal random variable Y.

The straight-line approximation in (2.5) seems somewhat *ad hoc*. Would we do better with another path? Why not add two extra paths to take some account of the variability of the random paths in the full integral? More generally, is there an optimal or systematic way of picking these paths? The path integral representation in Section 2.2 below may provide a partial answer to this question. The Brownian bridge representations (2.1) and (2.3) play a key role in the derivation of this path integral representation.

2.2. **The path integral representation of the density.** In this section, we provide a formal derivation of the path integral representation exploiting the Brownian bridge representations of Section 2.1 and the Chapman-Kolmogorov equation. As in Section 2.1, we will use the notations $\varphi(s_t, t)$ and x_t interchangeably.

Let $\{t = t_0 < t_1 < \cdots < t_n = T\}$ be a partition of the time interval $[t, T]$ with $\Delta t_i = t_i - t_{i-1} = \frac{T}{n}$, for $i = 1, \cdots, n$. By iteratively applying the Chapman-Kolmogorov equation, the transition density $p^S(T, s_T|t, s_t)$ can be written as

$$p^S(T, s_T|t, s_t) = \int \cdots \int \prod_{i=1}^{n} p^S(t_i, s_i|t_{i-1}, s_{i-1})\, ds_1 \cdots ds_{n-1}, \tag{2.6}$$

where we set $s_0 = s_t$ and $s_n = s_T$. Recall from (2.3) that the transition density p^S of S_t from (t_{i-1}, s_{i-1}) to (t_i, s_i) has the Brownian bridge representation

$$\begin{aligned} & p^S(t_i, s_i|t_{i-1}, s_{i-1}) \\ = \; & \frac{\phi(\Delta t, \varphi(s_i, t_i) - \varphi(s_{i-1}, t_{i-1}))}{a(s_i, t_i)}\, \tilde{\mathbb{E}}_{\varphi(s_{i-1}, t_{i-1}), \varphi(s_i, t_i)} \left[e^{\int_{t_{i-1}}^{t_i} h(X_\tau, \tau) dX_\tau - \int_{t_{i-1}}^{t_i} \frac{1}{2} h^2(X_\tau, \tau) d\tau} \right], \end{aligned}$$

where X_τ is a Brownian bridge from $\varphi(s_{i-1}, t_{i-i})$ to $\varphi(s_i, t_i)$. We next compute the limit of (2.6), as $\Delta t \to 0^+$ (or equivalently $n \to \infty$), assuming that, for $i = 1, \cdots, n$, the s_i's form a discretization of a differentiable curve s_τ, for $\tau \in [t, T]$.

We have

$$\begin{aligned} & \lim_{n\to\infty} \prod_{i=1}^{n} \tilde{\mathbb{E}}_{\varphi(s_{i-1}, t_{i-1}), \varphi(s_i, t_i)} \left[e^{\int_{t_{i-1}}^{t_i} h(X_\tau, \tau) dX_\tau - \frac{1}{2} \int_{t_{i-1}}^{t_i} h^2(X_\tau, \tau) d\tau} \right] \\ & = e^{\int_t^T h(\varphi(s_\tau, \tau), \tau) \dot{x}_\tau d\tau - \frac{1}{2} \int_t^T h^2(\varphi(s_\tau, \tau), \tau) d\tau} \end{aligned}$$

and

$$\begin{aligned} & \lim_{n\to\infty}\prod_{i=1}^{n} e^{-\frac{1}{2\Delta t}[\varphi(s_i,t_i)-\varphi(s_{i-1},t_{i-1})]^2} \\ = & \lim_{n\to\infty} e^{-\frac{1}{2\Delta t}\sum_{i=1}^{n}[\varphi(s_i,t_i)-\varphi(s_{i-1},t_{i-1})]^2} \\ = & \lim_{n\to\infty} e^{-\frac{1}{2}\sum_{i=1}^{n}\left[\varphi'_{i-1}\frac{\Delta s_i}{\Delta t}+\dot\varphi_{i-1}\right]^2\Delta t+\mathcal{O}(\Delta s_i{}^2+\Delta t)^2} \\ = & \; e^{-\frac{1}{2}\int_t^T[\varphi'(s_\tau,\tau)\dot s_\tau+\dot\varphi(s_\tau,\tau)]^2 d\tau}. \end{aligned}$$

Substitution into (2.6) and taking the limit $n\to\infty$ yields the following path integral representation for the transition density p^S

$$p^S(T,s_T|t,s_t)=\int_{\mathcal{C}_s} e^{-\frac{1}{2}\int_t^T[\varphi'(s_\tau,\tau)\dot s_\tau+\dot\varphi(s_\tau,\tau)-h(\varphi(s_\tau,\tau),\tau)]^2 d\tau}\mathcal{D}[s], \tag{2.7}$$

where

$$\mathcal{D}[s]=\lim_{n\to\infty}\frac{1}{\sqrt{2\pi\Delta t}\,a(s_T,T)}\prod_{i=1}^{n-1}\frac{1}{\sqrt{2\pi\Delta t}}\frac{ds_i}{a(s_i,t_i)},$$

and $\mathcal{C}_s$ denotes the collection of all differentiable curves from (t,s_t) to (T,s_T). Equivalently, because $dx_i=\frac{ds_i}{a(s_i,t_i)}$, we may rewrite the path integral representation (2.7) more neatly and simply in x-space as

$$p^S(T,s_T|t,s_t)=\frac{1}{a(S_T,T)}\int_{\mathcal{C}_x} e^{-\frac{1}{2}\int_t^T[\dot x_\tau-h(x_\tau,\tau)]^2 d\tau}\mathcal{D}[x] \tag{2.8}$$

where

$$\mathcal{D}[x]=\lim_{n\to\infty}\frac{1}{\sqrt{2\pi\Delta t}}\prod_{i=1}^{n-1}\frac{dx_i}{\sqrt{2\pi\Delta t}}$$

and $\mathcal{C}_x$ denotes the collection of all differentiable curves from (t,x_t) to (T,x_T). We shall henceforth deal mostly with the simpler expression (2.8). Heuristically, one could think of the path integral representation (2.8) of the density as an exponentially-weighted average over all possible differentiable curves connecting x_t to x_T. $\mathcal{D}[x]$ could then be regarded as the "Lebesgue" measure on the space of differentiable curves connecting x_t to x_T, though mathematically such a measure does not really exist.

Assume now that under the pricing measure (assuming zero interest rate and dividend yield), the price S_t of the underlying is driven by the SDE of local volatility type

$$dS_t=a(S_t,t)dB_t.$$

The path integral representation (2.7) of the transition density p^S in this case has the following simpler form

$$p^S(T, s_T|t, s_t) = \int_{\mathcal{C}_s} e^{-\frac{1}{2}\int_t^T \left[\frac{\dot{s}_\tau}{a(s_\tau,\tau)} + \frac{a_s(s_\tau,\tau)}{2}\right]^2 d\tau} \mathcal{D}[s].$$

Integrating the payoff function over the transition density, the path integral representation for call price is immediate:

$$C(t, s_t, K, T) = \int_K^\infty (s_T - K) \int_{\mathcal{C}_s} e^{-\frac{1}{2}\int_t^T \left[\frac{\dot{s}_\tau}{a(s_\tau,\tau)} + \frac{a_s(s_\tau,\tau)}{2}\right]^2 d\tau} \mathcal{D}[s] ds_T,$$

or equivalently in x-space,

$$C(t, s_t, K, T) = \int_K^\infty \frac{s_T - K}{a(s_T, T)} \int_{\mathcal{C}_x} e^{-\frac{1}{2}\int_t^T |\dot{x}_\tau - h(x_\tau,\tau)|^2 d\tau} \mathcal{D}[x] ds_T, \tag{2.9}$$

where $h(x, t) = \dot{\varphi}(s, t) - \frac{a_s(s,t)}{2}$.

3. Probabilistic derivation of the heat kernel expansion

The heat kernel expansion is a small time asymptotic expansion of the fundamental solution of the heat equation over a Riemannian manifold. Reexpressing the transition density of a diffusion process in terms of this fundamental solution leads naturally to a small time asymptotic expansion of the transition density. This topic is well-studied in the Riemannian geometry literature, see Chavel [4] for a geometric analytical approach and Hsu [12] for a probabilistic approach. In the physics literature, the heat kernel approach to deriving small time asymptotic expansions is also known as the WKB method or the ray solution, see Jordan and Tier [14]. Deriving such expansions in one dimension is much simpler than in higher dimensions where no analogue of the Lamperti transformation exists.

Though the heat kernel expansion is very well-known, the Brownian bridge representation (2.4) of Theorem 2 leads to a novel probabilistic derivation which we will now present. To fix ideas and illustrate the methodology employed, we start with the case of Brownian motion with drift; as before, the general diffusion case follows via the Lamperti transformation. To minimize mathematical technicalities, we shall assume (at least in this section) that all functions are bounded with bounded derivatives.

3.1. **Heat kernel expansion for Brownian motion with drift.**

Theorem 3. *Let X_t be the Brownian motion with drift h, i.e., X_t satisfies the SDE $dX_t = dB_t + h(X_t, t)dt$. Denote by H an antiderivative of h with respect to x, namely,*

$\frac{\partial}{\partial x}H(x,t) = h(x,t)$, for all x and t. The transition density p^X of X_t has, as $t \to T^-$, the following small time asymptotic expansion:

$$\begin{aligned} p^X(T,y|t,x) &= \phi(T-t,y-x)\,e^{H(y,T)-H(x,t)} \times \qquad (3.1)\\ &\left\{1-\frac{1}{2}\int_t^T \left[h^2(x_s^*,s)+h_x(x_s^*,s)+2H_t(x_s^*,s)\right]ds + \mathcal{O}(T-t)^2\right\} \end{aligned}$$

where ϕ is the Gaussian density $\phi(t,\xi) = \frac{1}{\sqrt{2\pi t}}e^{-\frac{\xi^2}{2t}}$. x_s^* denotes the straight line from (t,x) to (T,y), i.e., $x_s^* = x + \frac{s-t}{T-t}(y-x)$ for $s \in [t,T]$.

Notice that in the time-inhomogeneous case $h = h(x,t)$, the approximation (3.1) is different from the heat kernel expansion (see, for example, (3.3), (3.6), and (3.7) on page 603 of Gatheral et al. [7]) in that the approximation in (3.1) involves an integration from t to T whereas, in the classical heat kernel expansion, all quantities are evaluated at the fixed initial time t. Of course, in the time homogeneous case where the drift $h = h(x)$ has no explicit dependence on t, the expansion (3.1) coincides with the classical heat kernel expansion as formalized in the following corollary.

Corollary 1. *(Heat kernel expansion for Brownian motion with drift)*
For Brownian motion with time homogeneous drift $h = h(x)$, the transition density p^X of X_t from (t,x) to (T,y) has the asymptotic expansion up to first order as

$$p(T,y|t,x) = \phi(T-t,y-x)e^{H(y)-H(x)}\left\{1-\frac{T-t}{2(y-x)}\int_x^y \left[h^2(\xi)+h'(\xi)\right]d\xi + \mathcal{O}(T-t)^2\right\},$$

which coincides with the classical heat kernel expansion up first order (see, for instance, Gatheral et al. [7]).

Proof. In this case, $H_t = 0$ because $h_t = 0$. The integral in (3.1) can be evaluated as

$$\begin{aligned} &\int_t^T \left[h^2(x_s^*)+h'(x_s^*)\right]ds \\ =& \int_t^T \left[h^2\left(x+\frac{s-t}{T-t}(y-x)\right)+h'\left(x+\frac{s-t}{T-t}(y-x)\right)\right]ds \\ =& \frac{T-t}{y-x}\int_x^y \left[h^2(\xi)+h'(\xi)\right]d\xi, \end{aligned}$$

where in the last equation we used the change of variable $\xi = x + \frac{s-t}{T-t}(y-x)$. ☐

Let Y_t denote the Brownian bridge from x at time t to y at time T and $\tilde{\mathbb{E}}_{x,y}[\cdot]$ be the expectation under the Brownian bridge measure. The proof of the asymptotic expansion (3.1) requires the following two lemmas.

Lemma 1. *For a bounded function $g = g(x,s)$, $|g| \le M$ say, we have the following estimate*

$$\tilde{\mathbb{E}}_{x,y}\left[e^{\int_t^T g(Y_s,s)ds}\right] = 1 + \int_t^T \tilde{\mathbb{E}}_{x,y}\left[g(Y_s,s)\right] ds + \mathcal{O}(T-t)^2.$$

Proof. The proof is based on a clever application of the convex order for random variables first observed, to our knowledge, in the paper by Goovaerts et al. [10] (see Proposition 6.2 on page 348). Denote by $Q_{g(Y_s,s)}(q)$ the qth quantile of the random variable $g(Y_s,s)$. Since exponential functions are convex, it follows from Proposition 6.2 of Goovaerts et al. [10] that

$$\tilde{\mathbb{E}}_{x,y}\left[e^{\int_t^T g(Y_s,s)ds}\right] \le \int_0^1 e^{\int_t^T Q_{g(Y_s,s)}(q)ds} dq.$$

We establish an upper bound for the right hand side. First we Taylor expand the integrand and rewrite the integral as

$$\int_0^1 e^{\int_t^T Q_{g(Y_s,s)}(q)ds} dq = \sum_{k=0}^\infty \frac{1}{k!} \int_0^1 \left(\int_t^T Q_{g(Y_s,s)}(q)ds\right)^k dq.$$

An upper bound for $\int_0^1 \left(\int_t^T Q_{g(Y_s,s)}(q)ds\right)^k dq$ is then determined as

$$\begin{aligned}
& \int_0^1 \left(\int_t^T Q_{g(Y_s,s)}(q)ds\right)^k dq \\
\le\ & \int_0^1 (T-t)^{k-1} \int_t^T \left|Q_{g(Y_s,s)}(q)\right|^k ds\, dq \quad \text{(by Hölder's inequality)} \\
=\ & (T-t)^{k-1} \int_t^T \tilde{\mathbb{E}}_{x,y}|g(Y_s,s)|^k ds \\
\le\ & M^k (T-t)^k \quad (\text{since } |g| \le M).
\end{aligned}$$

Thus,

$$\begin{aligned}
& \int_0^1 e^{\int_t^T Q_{g(Y_s,s)}(q)\, ds}\, dq \\
=\ & 1 + \int_0^1 \int_t^T Q_{g(Y_s,s)}(q)\, ds\, dq + \sum_{k=2}^\infty \frac{1}{k!} \int_0^1 \left(\int_t^T Q_{g(Y_s,s)}(q)ds\right)^k dq \\
\le\ & 1 + \int_t^T \tilde{\mathbb{E}}_{x,y}\left[g(Y_s,s)\right]\, ds + \sum_{k=2}^\infty \frac{1}{k!} M^k (T-t)^k \\
\le\ & 1 + \int_t^T \tilde{\mathbb{E}}_{x,y}\left[g(Y_s,s)\right]\, ds + M^2 (T-t)^2 e^{M(T-t)},
\end{aligned}$$

which completes the proof. □

Lemma 2 asserts that the time integral of the conditional expectation in Lemma 1 is approximately, up to order $(T-t)^2$, equal to the integral along a straight line connecting x at time t to y at time T.

Lemma 2. *For a bounded function $g = g(x,s)$ with bounded second partial derivative with respect to x, the following asymptotic holds.*

$$\int_t^T \tilde{\mathbb{E}}_{x,y}\left[g(Y_s,s)\right]\, ds = \int_t^T g(x_s,s)ds + \mathcal{O}(T-t)^2,$$

where x_s denotes the straight line $x_s = x + \frac{s-t}{T-t}(y-x)$ from (t,x) to (T,y).

Proof. Taylor's theorem implies that

$$g(Y_s,s) = g(x_s,s) + g_x(x_s,s)(Y_s - x_s) + \frac{g_{xx}(\xi_s,s)}{2}(Y_s-x_s)^2,$$

for some ξ_s between Y_s and x_s. Since Y_s is a Brownian bridge from (t,x) to (T,y), Y_s is normally distributed: $Y_s \sim N\left(x_s, \frac{(s-t)(T-s)}{T-t}\right)$. Therefore,

$$\begin{aligned}\tilde{\mathbb{E}}_{x,y}\left[g(Y_s,s)\right] &= g(x_s,s) + g_x(x_s,s)\,\tilde{\mathbb{E}}_{x,y}\left[Y_s - x_s\right] + \frac{g_{xx}(\xi_s,s)}{2}\tilde{\mathbb{E}}_{x,y}\left[(Y_s-x_s)^2\right] \\ &= g(x_s,s) + \frac{g_{xx}(\xi_s,s)}{2}\frac{(s-t)(T-s)}{T-t}.\end{aligned}$$

Hence, by the assumption that $|g_{xx}| \leq K$,

$$\begin{aligned}\int_t^T \tilde{\mathbb{E}}_{x,y}\left[g(Y_s,s)\right]\, ds &= \int_s^T g(x_s,s)ds + \int_t^T \frac{g_{xx}(\xi_s,s)}{2}\frac{(s-t)(T-s)}{T-t}\,ds \\ &\leq \int_t^T g(x_s,s)ds + \frac{K}{2}\int_t^T \frac{(s-t)(T-s)}{T-t}ds \\ &= \int_t^T g(x_s,s)ds + \frac{K}{12}(T-t)^2.\end{aligned}$$

□

The proof of Theorem 3 is now straightforward.

Proof. (Proof of Theorem 3)
By combining the two asymptotics in Lemma 1 and Lemma 2 with $g(x,s) = h^2(x,s) + h_x(x,s) + 2H_t(x,s)$, under the assumption that g is bounded with bounded second partial derivative with respect to x, we obtain

$$\begin{aligned}&\tilde{\mathbb{E}}_{x,y}\left[e^{-\frac{1}{2}\int_t^T h^2(X_s,s)+h_x(X_s,s)+2H_t(X_s,s)ds}\right] \\ =\ & 1 - \frac{1}{2}\int_t^T \left[h^2(x_s,s) + h_x(x_s,s) + 2H_t(x_s,s)\right] ds + \mathcal{O}(T-t)^2.\end{aligned}$$

Recall expression (2.2) for the transition density:

$$\begin{aligned} p^X(T,y|t,x) &= \phi(T-t,y-x)\, e^{H(y,T)-H(x,t)} \times \\ &\qquad \tilde{\mathbb{E}}_{x,y}\left[e^{-\frac{1}{2}\int_t^T h^2(X_s,s)+h_x(X_s,s)+2H_t(X_s,s)ds}\right]. \end{aligned}$$

Substituting the approximation of the conditional expectation above, we obtain

$$\begin{aligned} p^X(T,y|t,x) &= \phi(T-t,y-x)\, e^{H(y,T)-H(x,t)} \times \\ &\quad \left\{1-\frac{1}{2}\int_t^T \left[h^2(x_s,s)+h_x(x_s,s)+H_t(x_s,s)\right]ds+\mathcal{O}(T-t)^2\right\}. \end{aligned}$$

□

3.2. **Heat kernel expansion for nondegenerate diffusions.** For general nondegenerate diffusions, consider the process S_t driven by the SDE:

$$dS_t = a(S_t,t)dB_t + \mu(S_t,t)dt. \tag{3.2}$$

Again the Lamperti transformation allows us to carry over the small time asymptotic expansion (3.1) in x-space to s-space. Specifically, recall that the Lamperti transformation $x_t = \varphi(s_t,t) = \int_{s_0}^{s_t}\frac{d\xi}{a(\xi,t)}$ transforms the SDE (3.2) into a Brownian motion with drift $dX_t = dB_t + h(X_t,t)dt$, where $h(x_t,t) = \dot{\varphi}(s_t,t) + \frac{\mu(s_t,t)}{a(s_t,t)} - \frac{a_s(s_t,t)}{2}$ and that the transition densities p^S for S_t and p^X for X_t are related by

$$p^S(T,s_T|t,s_t) = \frac{1}{a(s_T,T)}p^X(T,x_T|t,x_t),$$

with $x_T = \varphi(s_T,T)$ and $x_t = \varphi(s_t,t)$. Hence, a small time asymptotic expansion as $t\to T^-$ for p^S can be obtained by simply applying the expansion (3.1). This argument is formalized in Theorem 4.

Theorem 4. *The transition density p^S of the process S_t driven by the SDE*

$$dS_t = a(S_t,t)dB_t + \mu(S_t,t)dt$$

has the small time asymptotic expansion as $t\to T^-$

$$\begin{aligned} p^S(T,s_T|t,s_t) &= \frac{\phi(T-t,\varphi(s_T,T)-\varphi(s_t,t))}{a(s_T,T)}\, e^{H(\varphi(s_T,T),T)-H(\varphi(s_t,t),t)} \qquad (3.3)\\ &\times \left\{1-\frac{1}{2}\int_t^T h^2(\varphi_\tau,\tau)+h_x(\varphi_\tau,\tau)+2H_t(\varphi_\tau,\tau)d\tau + O(T-t)^2\right\}, \end{aligned}$$

where $\varphi_\tau = \frac{T-\tau}{T-t}\varphi(s_t,t)+\frac{\tau-t}{T-t}\varphi(s_T,T)$.

We stress once again that in the time-inhomogeneous case, $a = a(s,t)$, the expansion in (3.3) is not identical to the classical heat kernel expansion as it involves an integral along the path φ_τ. On the other hand, in the time-homogeneous case

$a = a(s)$, (3.3) does recover the classical heat kernel expansion. In this sense therefore, we have derived a natural generalization of the classical heat kernel expansion.

Corollary 2. *(Heat kernel expansion for time-homogeneous diffusions)* *The transition density p^S of the process S_t driven by the time-homogeneous SDE*

$$dS_t = a(S_t)dB_t + \mu(S_t)dt$$

has the small time asymptotic expansion as $t \to T^-$ up to first order

$$\begin{aligned} p(T, s_T|t, s_t) = {} & \frac{\phi(T-t, \varphi(s_T) - \varphi(s_t))}{a(s_T)} \, e^{H\circ\varphi(s_T) - H\circ\varphi(s_t)} \\ & \times \left\{ 1 - \frac{T-t}{2(\varphi(s_T)-\varphi(s_t))} \int_{s_t}^{s_T} \left[h^2(\varphi(s)) + h' \circ \varphi(s) \right] \frac{ds}{a(s)} + O(T-t)^2 \right\}, \end{aligned} \tag{3.4}$$

where $\varphi(s) = \int_{s_0}^{s} \frac{d\xi}{a(\xi)}$, $h \circ \varphi(s) = \frac{\mu(s)}{a(s)} - \frac{a'(s)}{2}$, and H is an antiderivative of h. ϕ denotes the Gaussian density $\phi(t,\xi) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{\xi^2}{2t}}$. The small time asymptotic expansion coincides with the classical heat kernel expansion up to first order.

Proof. We verify that the expansion (3.4) is indeed the classical heat kernel expansion. The classical heat kernel expansion up to first order (see, for instance, Gatheral et al. [7]) reads in our notation

$$\begin{aligned} p(T, s_T|t, s_t) \;\; \approx \;\; & \frac{\phi(T-t, \varphi(s_T) - \varphi(s_t))}{a(s_T)} \, u(s_t, s_T) \\ & \times \left\{ 1 + \frac{T-t}{\varphi(s_T)-\varphi(s_t)} \int_{s_t}^{s_T} \frac{\mathcal{L}u(s, s_T)}{u(s, s_T)} \frac{ds}{a(s)} \right\}, \end{aligned}$$

where $u(s, s_T) = e^{\int_s^{s_T} \frac{\mu(\eta)}{a^2(\eta)} d\eta} \sqrt{\frac{a(s)}{a(s_T)}}$ and $\mathcal{L} = \frac{a^2(s)}{2}\partial_s^2 + \mu(s)\partial_s$ is the infinitesimal generator associated with the process S_t. In this case, the asymptotic expansion (3.3) reduces to

$$\frac{\phi(T-t, \varphi(s_T) - \varphi(s_t))}{a(s_T)} \, e^{H\circ\varphi(s_T) - H\circ\varphi(s_t)} \times \left\{ 1 - \frac{1}{2} \int_t^T h^2(\varphi_\tau) + h'(\varphi_\tau) d\tau \right\},$$

where $\varphi_\tau = \frac{T-\tau}{T-t}\varphi(s_t) + \frac{\tau - t}{T-t}\varphi(s_T)$. Therefore, it suffices to show that

$$e^{H\circ\varphi(s_T) - H\circ\varphi(s_t)} = u(s_t, s_T) \tag{3.5}$$

and

$$-\frac{1}{2} \int_t^T h^2(\varphi_\tau) + h'(\varphi_\tau) d\tau = \frac{T-t}{\varphi(s_T)-\varphi(s_t)} \int_{s_t}^{s_T} \frac{\mathcal{L}u(s, s_T)}{u(s, s_T)} \frac{ds}{a(s)}. \tag{3.6}$$

For (3.5), since $h\circ\varphi(s)=\frac{\mu(s)}{a(s)}-\frac{a'(s)}{2}$, $\varphi'(s)=\frac{1}{a(s)}$, and H is an antiderivative of h, we have

$$H\circ\varphi(s_T)-H\circ\varphi(s_t)=\int_{\varphi(s_t)}^{\varphi(s_T)}h(\xi)d\xi=\int_{s_t}^{s_T}h\circ\varphi(s)d\varphi(s)$$
$$=\int_{s_t}^{s_T}\left[\frac{\mu(s)}{a(s)}-\frac{a'(s)}{2}\right]\frac{ds}{a(s)}=\int_{s_t}^{s_T}\frac{\mu(s)}{a^2(s)}ds-\frac{1}{2}\log\left[\frac{a(s_T)}{a(s_t)}\right].$$

Therefore,

$$e^{H\circ\varphi(s_T)-H\circ\varphi(s_t)}=e^{\int_{s_t}^{s_T}\frac{\mu(s)}{a^2(s)}ds}\sqrt{\frac{a(s_t)}{a(s_T)}}=u(s_t,s_T).$$

As for (3.6), since $\varphi_\tau=\frac{T-\tau}{T-t}\varphi(s_t)+\frac{\tau-t}{T-t}\varphi(s_T)$, we have

$$\int_t^T h^2(\varphi_\tau)+h'(\varphi_\tau)d\tau$$
$$=\int_t^T h^2\left(\frac{T-\tau}{T-t}\varphi(s_t)+\frac{\tau-t}{T-t}\varphi(s_T)\right)+h'\left(\frac{T-\tau}{T-t}\varphi(s_t)+\frac{\tau-t}{T-t}\varphi(s_T)\right)d\tau$$
$$=\frac{T-t}{\varphi(s_T)-\varphi(s_t)}\int_{s_t}^{s_T}h^2(\varphi(s))+h'(\varphi(s))d\varphi(s).$$

Note that $d\varphi(s)=\frac{ds}{a(s)}$ and

$$h'\circ\varphi(s)=\frac{1}{\varphi'(s)}\frac{d}{ds}\left[h\circ\varphi(s)\right]=a(s)\times\frac{d}{ds}\left[\frac{\mu(s)}{a(s)}-\frac{a'(s)}{2}\right],$$

consequently,

$$\int_{s_t}^{s_T}\left[h^2(\varphi(s))+h'(\varphi(s))\right]d\varphi(s)=\int_{s_t}^{s_T}\left[\left(\frac{\mu}{a}-\frac{a'}{2}\right)^2+a\left(\frac{\mu}{a}-\frac{a'}{2}\right)'\right]\frac{ds}{a(s)},$$

where we suppressed the dependence on s for notational simplicity. On the other hand, for the right hand side of (3.6), by straightforward calculation we have

$$\mathcal{L}u(s,s_T)=\frac{a^2(s)}{2}\partial_s^2u(s,s_T)+\mu(s)\partial_su(s,s_T)$$
$$=\left[-\frac{\mu^2}{2a^2}-\frac{(a')^2}{8}-\frac{a}{2}\left(\frac{\mu}{a}-\frac{a'}{2}\right)'+\frac{a'\mu}{2a}\right]u(s,s_T)$$
$$=-\frac{1}{2}\left[\left(\frac{\mu}{a}-\frac{a'}{2}\right)^2+a\left(\frac{\mu}{a}-\frac{a'}{2}\right)'\right]u(s,s_T).$$

It follows that

$$\int_{s_t}^{s_T}\frac{\mathcal{L}u(s,s_T)}{u(s,s_T)}\frac{ds}{a(s)}=-\frac{1}{2}\int_{s_t}^{s_T}\left[\left(\frac{\mu}{a}-\frac{a'}{2}\right)^2+a\left(\frac{\mu}{a}-\frac{a'}{2}\right)'\right]\frac{ds}{a(s)}$$
$$=-\frac{1}{2}\int_{s_t}^{s_T}\left[h^2(\varphi(s))+h'(\varphi(s))\right]d\varphi(s),$$

which completes the proof of (3.6). □

4. Implied volatility approximation

The implied volatility $\sigma_{\mathrm{BS}} = \sigma_{\mathrm{BS}}(K, T)$ is defined implicitly by solving the nonlinear equation

$$C(s, t, K, T) = C_{\mathrm{BS}}(s, t, K, T, \sigma_{\mathrm{BS}}(K, T)), \tag{4.1}$$

where the function C_{BS} on the right hand side is the celebrated Black-Scholes pricing formula for call options (assuming zero interest rate and dividend yield):

$$C_{\mathrm{BS}}(s, t, K, T, \sigma_{\mathrm{BS}}) = sN(d_1) - KN(d_2)$$

with $d_1 = \frac{\log s - \log K}{\sigma_{\mathrm{BS}}\sqrt{T-t}} + \frac{\sigma_{\mathrm{BS}}\sqrt{T-t}}{2}$, $d_2 = d_1 - \sigma_{\mathrm{BS}}\sqrt{T-t}$, and $N(\cdot)$ is the cumulative normal distribution function. The Black-Scholes formula is monotonic increasing in the volatility parameter σ_{BS}, and for this reason amongst others, it is often market practice to quote options in terms of Black-Scholes implied volatility. Moreover, practitioners often calibrate their option pricing models to implied volatilities rather than price quotes. In this regard, efficient and accurate approximations of implied volatility not only permit faster calibration of option pricing models but also help build intuition.

Conventionally, asymptotic expansions of implied volatility for small time to expiry (to lowest order) are generated by matching exponents in respectively, an asymptotic approximation for a far out-of-the-money (OTM) option under Black-Scholes, and an asymptotic approximation to the option price from direct integration over the (approximated) density. For such far out-of-the-money (OTM) options, as time approaches expiry, the event that the underlying will end up in-the-money at expiry is a rare event. According to the theory of large deviations, such a rare event has exponentially small probability, so the option price is of the form

$$\int_K^\infty e^{-\frac{d(x)}{T-t}} f(x)\, dx. \tag{4.2}$$

As $t \to T^-$, the main contribution to the integral comes from the minimum point of d, which in this case is the boundary point of the support of f because, in the OTM case, $d(x)$ is strictly increasing in x, and $f(x)$ has the payoff function as a factor (see (4.4) below). To zeroth order, the Laplace asymptotic formula (for example, see (5.2.23) on page 193 of Bleistein and Handelsman [3]) then reads

$$\int_K^\infty e^{-\frac{d(x)}{T-t}} f(x) dx \approx (T-t)^2 e^{-\frac{d(K)}{T-t}} \frac{f'(K)}{|d'(K)|^2} \tag{4.3}$$

as $t \to T^-$, provided $f'(K)$ and $d'(K)$ are nonzero. Thus, the small time asymptotic expansion of the implied volatility is obtained by applying the Laplace asymptotic formula (4.3) to both sides of (4.1) then matching the corresponding coefficients. As one might expect, the dominating term of such expansions is typically the zeroth order term. Our objective in this section is to demonstrate how to implement this matching procedure from the path integral perspective.

Recasting equation (4.1) for implied volatility using our path integral representation of the density, and using our earlier representation (2.9) of the call price, we obtain

$$\begin{aligned}
& \int_K^\infty \frac{S_T - K}{a(S_T, T)} \int_{\mathcal{C}_x} e^{-\frac{1}{2}\int_t^T |\dot{x}_\tau - h(x_\tau,\tau)|^2 d\tau}\, \mathcal{D}[x]\, dS_T, \\
= & \int_K^\infty \frac{S_T - K}{\sigma_{\mathrm{BS}} s_T} e^{-\frac{\sigma_{\mathrm{BS}}}{2}(x_T - x_t) - \frac{\sigma_{\mathrm{BS}}^2}{8}(T-t)} \int_{\mathcal{C}_x} e^{-\frac{1}{2}\int_t^T |\dot{x}_\tau|^2 d\tau} \mathcal{D}[x] dS_T \\
= & \int_K^\infty \frac{S_T - K}{\sqrt{2\pi(T-t)}\sigma_{\mathrm{BS}} S_T} e^{-\frac{1}{2}\left(\frac{\log s_T - \log s_t}{\sigma_{\mathrm{BS}}\sqrt{T-t}} + \frac{\sigma_{\mathrm{BS}}\sqrt{T-t}}{2}\right)^2} dS_T. \qquad (4.4)
\end{aligned}$$

Equation (4.4) provides an implicit expression for Black-Scholes implied volatility in terms of local volatility. In the foregoing, we first show how to recover from (4.4) the heat kernel approximations of Gatheral et al. [7] and the most-likely-path approximation of Gatheral and Wang [9]. Finally, in Section 4.3, we show how to improve on these approximations by adopting the path integral perspective.

4.1. **Recovery of the Berestycki-Busca-Florent (BBF) formula.** To rederive the results in Berestycki et al. [2] and Gatheral et al. [7] from (4.4), we approximate both sides of (4.4) as Laplace type integrals as in (4.2). The path integral on the left hand side of (4.4) is approximated as follows:

$$\begin{aligned}
& \int_{\mathcal{C}_x} e^{-\frac{1}{2}\int_t^T |\dot{x}_\tau - h(x_\tau,\tau)|^2 d\tau} \mathcal{D}[x] \\
= & \int_{\mathcal{C}_x} e^{-\frac{1}{2}\int_t^T |\dot{x}_\tau|^2 - 2\int_t^T h(x_\tau,\tau) dx_\tau + \int_t^T h^2(x_\tau,\tau) d\tau} \mathcal{D}[x] \\
\approx & \int_{\mathcal{C}_x} e^{-\frac{1}{2}\int_t^T |\dot{x}_\tau|^2 d\tau} \left[1 - 2\int_t^T h(x_\tau,\tau) dx_\tau + \int_t^T h^2(x_\tau,\tau) d\tau\right] \mathcal{D}[x] \\
\approx & \; e^{-\frac{(x_T - x_t)^2}{2(T-t)}} \left[1 + \mathcal{O}(T-t)\right],
\end{aligned}$$

where in the last step we approximated the path integral by evaluating the integral in the exponent along a single path: the straight line connecting x_t and x_T. Recall that $x_t = \varphi(s_t, t) = \int_{s_0}^{s_t} \frac{d\xi}{a(\xi,t)}$. Substitution back into the left hand side of (4.4) gives

$$\begin{aligned} &\int_K^\infty \frac{S_T - K}{a(S_T,T)} \int_{\mathcal{C}_x} e^{-\frac{1}{2}\int_t^T |\dot{x}_\tau - h(x_\tau,\tau)|^2 d\tau} \mathcal{D}[x]\, dS_T \\ \approx\ &\int_K^\infty e^{-\frac{|\varphi(s_T,T)-\varphi(s_t,t)|^2}{2(T-t)}} \frac{S_T - K}{a(S_T,T)} \left[1 + \mathcal{O}(T-t)\right] dS_T, \end{aligned}$$

which is of Laplace type as in (4.2). Applying the Laplace asymptotic formula (4.3), we obtain that, up to a factor,

$$C(s,t,K,T) \approx e^{-\frac{|\varphi(K,T)-\varphi(s,t)|^2}{2(T-t)}}. \tag{4.5}$$

Likewise, the Black-Scholes price on the right hand side of (4.4) is given, up to a factor, by

$$C_{\rm BS}(s,t,K,T) \approx e^{-\frac{|\log K - \log s|^2}{2\sigma_{\rm BS}^2(T-t)}}. \tag{4.6}$$

Finally, by matching the exponents in (4.5) and (4.6), we obtain the zeroth order approximation of the implied volatility as

$$\sigma_{\rm BS} \approx \frac{\log K - \log s}{\varphi(K,T) - \varphi(s,t)}.$$

In the time homogeneous case,

$$\varphi(K) - \varphi(s) = \int_s^K \frac{d\xi}{a(\xi)}$$

and we recover the BBF formula as in Berestycki et al. [2] and Gatheral et al.[7].

4.2. **Recovery of the variational-most-likely-path (vMLP) approximation of Gatheral and Wang** [9]**.** The path integral term in (4.4) is in x-space. Alternatively, in s-space it reads

$$\int_{\mathcal{C}_s} e^{-\frac{1}{2}\int_t^T \left[\frac{\dot{s}_\tau}{a(s_\tau,\tau)} + \frac{a_s(s_\tau,\tau)}{2}\right]^2 d\tau} \mathcal{D}[s],$$

where

$$\mathcal{D}[s] = \lim_{n\to\infty} \frac{1}{\sqrt{2\pi\Delta t}\, a(s_T,T)} \prod_{i=1}^{n-1} \frac{1}{\sqrt{2\pi\Delta t}} \frac{ds_i}{a(s_i,t_i)}.$$

Hence, we can rewrite the left hand side of (4.4) in s-space as

$$C(t,s_t,K,T) = \int_K^\infty (s_T - K) \int_{\mathcal{C}_s} e^{-\frac{1}{2}\int_t^T \left[\frac{\dot{s}_\tau}{a(s_\tau,\tau)} + \frac{a_s(s_\tau,\tau)}{2}\right]^2 d\tau} \mathcal{D}[s] ds_T.$$

The variational most-likely-path approximation of implied volatility developed in Gatheral and Wang [9] is obtained by dropping the second term $\frac{a_s(s_\tau,\tau)}{2}$ in the path integral and evaluating the resulting path integral along the path that minimizes the functional

$$e^{-\frac{1}{2}\int_t^T \left|\frac{\dot{s}_\tau}{a(s_\tau,\tau)}\right|^2 d\tau}.$$

In other words,

$$C(s,t,K,T) \approx \int_K^\infty (s_T - K) e^{-\frac{1}{2}\int_t^T \left|\frac{\dot{s}^*_\tau}{a(s^*_\tau,\tau)}\right|^2 d\tau} ds_T,$$

where s^*_τ is the optimal path that maximizes the action functional $\int_t^T \left|\frac{\dot{s}_\tau}{a(s_\tau,\tau)}\right|^2 d\tau$ subject to the constraints that initial and terminal points are fixed at s_t and s_T respectively. Moreover, since the resulting integral is of Laplace type, the call price is given asymptotically, up to a factor, by

$$C(s,t,K,T) \approx e^{-\frac{1}{2}\int_t^T \left|\frac{\dot{s}^*_\tau}{a(s^*_\tau,\tau)}\right|^2 d\tau},$$

where the optimal path s^*_τ has initial and terminal points s and K respectively. Finally, by matching the exponent with the Black-Scholes asymptotic as in (4.6), the zeroth order approximation of implied volatility is given by

$$\sigma_{\mathrm{BS}} \approx \frac{|\log K - \log s|}{\sqrt{T-t}} \left[\int_t^T \left|\frac{\dot{s}^*_\tau}{a(s^*_\tau,\tau)}\right|^2 d\tau\right]^{-\frac{1}{2}}$$

which recovers the variational most-likely-path approximation of the implied volatility presented in Gatheral and Wang [9].

4.3. **New and improved most-likely-path (MLP) approximation.** As is obvious from our presentation, the approximations obtained in Gatheral et al. [7] and in Gatheral and Wang [9] are suboptimal from the perspective of our path integral representation (4.4) in the sense that they both drop terms. This suggests that we should define the path-integral-most-likely-path to be the path that maximizes the full action functional

$$\frac{1}{2}\int_t^T |\dot{x}_\tau - h(x_\tau,\tau)|^2 \, d\tau \tag{4.7}$$

or equivalently in s-space the functional

$$\frac{1}{2}\int_t^T \left[\frac{\dot{s}_\tau}{a(s_\tau,\tau)} + \frac{a_s(s_\tau,\tau)}{2}\right]^2 d\tau \tag{4.8}$$

without dropping terms. The Euler-Lagrange equation associated with the functional in (4.7) is

$$\ddot{x}_\tau = h\, h_x + h_t \tag{4.9}$$

with boundary conditions x_t and x_T at times t and T respectively. Matching exponents as before gives

$$\left|\frac{\log K - \log s}{\sigma_{\mathrm{BS}}\sqrt{T-t}} + \frac{\sigma_{\mathrm{BS}}\sqrt{T-t}}{2}\right|^2 = \int_t^T [\dot{x^*_\tau} - h(x^*_\tau,\tau)]^2 d\tau, \tag{4.10}$$

where x_τ^* is the optimal path which maximizes the functional (4.7) (or equivalently solves (4.9)) with initial and terminal points given by $\varphi(s,t)$ and $\varphi(K,T)$ respectively. Solving (4.10) for σ_{BS} yields our new-and-improved zeroth order approximation for implied volatility.

To illustrate the accuracy of our new approximation (4.10), consider the case of time dependent Black-Scholes, where rather pleasingly, (4.10) gives the exact solution. Note in passing that, to the best of our knowledge, none of the existing small time approximations is able to recover this very simple case.

Example 1. *(Implied volatility in the time dependent Black-Scholes model)*
Assume the price S_t of the underlying satisfies the following under the pricing measure:

$$dS_\tau = \sigma(\tau)\, S_\tau\, dB_\tau, \quad S_t = s_t.$$

In order to apply (4.10), we proceed as follows:

a) Transform the model into x-space.
b) Solve the Euler-Lagrange equation (4.9) for the optimal path.
c) Evaluate the the action functional (4.9) along the optimal path, substitute into (4.10) and solve for the implied volatility.

a) *Transform into x-space*: In this case, $x = \varphi(s,t) = \int_{s_0}^{s} \frac{1}{\sigma(t)\xi} d\xi = \frac{\log s - \log s_0}{\sigma(t)}$. Dropping the explicit dependence on t for ease of notation, and applying Ito's formula to $X_t = \varphi(S_t,t)$ we obtain

$$\begin{aligned} dX_t &= \dot{\varphi}(S_t,t)dt + \varphi_s(S_t,t)dS_t + \frac{1}{2}\varphi_{ss}(S_t,t)d[S]_t \\ &= dB_t - \left(\frac{\sigma'}{\sigma}X_t + \frac{\sigma}{2}\right) dt. \end{aligned}$$

Thus $h(x,t) = -\frac{\sigma}{2} - \frac{\sigma'}{\sigma}x$.

b) *Solve the Euler-Lagrange equation*: The associated Euler-Lagrange equation (4.9) in this case reads

$$\ddot{x} = h\, h_x + h_t = \left[\left(\frac{\sigma'}{\sigma}\right)^2 - \left(\frac{\sigma'}{\sigma}\right)'\right] x.$$

With the change of variable $x = \frac{z}{\sigma}$, the above ODE for x is transformed into the following ODE for z

$$\ddot{z} - \frac{2\sigma'}{\sigma}\dot{z} = 0 \quad \Longrightarrow \quad \frac{d}{d\tau}\left(\frac{\dot{z}}{\sigma^2}\right) = 0.$$

With boundary conditions $z_t = \sigma_t\, x_t$ and $z_T = \sigma_T\, x_T$, the solution to the Euler-Lagrange equation is given by

$$\sigma_\tau\, x_\tau = z_\tau = \sigma_t\, x_t + \frac{\sigma_T\, x_T - \sigma_t\, x_t}{\int_t^T \sigma^2(s)ds} \int_t^\tau \sigma^2(s)ds.$$

(c) *Solve for implied volatility*: It follows that the functional (4.7) evaluated along the optimal path, taking into account that $\frac{\dot z}{\sigma^2} = \frac{\sigma_T x_T - \sigma_t x_t}{\int_t^T \sigma^2(s)ds}$ is a constant, is given by

$$\begin{aligned} &\int_t^T |\dot x_\tau - h(x_\tau,\tau)|^2 d\tau = \int_t^T \left| \dot x_\tau + \frac{\sigma}{2} + \frac{\sigma'}{\sigma} x \right|^2 d\tau \\ =\ & \int_t^T \left| \frac{\partial_\tau(\sigma\, x)}{\sigma} + \frac{\sigma}{2} \right|^2 d\tau = \int_t^T \left| \frac{\dot z}{\sigma} + \frac{\sigma}{2} \right|^2 d\tau \\ =\ & \left(\frac{\sigma_T x_T - \sigma_t x_t}{\int_t^T \sigma^2(s)ds} + \frac{1}{2} \right)^2 \int_t^T \sigma_\tau^2 d\tau \\ =\ & \left(\frac{\log s_T - \log s_t}{\sqrt{\int_t^T \sigma^2(s)ds}} + \frac{1}{2} \sqrt{\int_t^T \sigma^2(s)ds} \right)^2. \end{aligned}$$

Finally, substituting this last expression into (4.10) gives the well-known result

$$\sigma_{\mathrm{BS}}^2 = \frac{1}{T-t} \int_t^T \sigma^2(s)ds,$$

which is exact.

5. Conclusion

We have shown, up to first order in $\tau = T - t$, that the classical heat kernel expansion can be derived using a novel probabilistic approach. This new probabilistic derivation of the heat kernel expansion inspires a path integral representation of the transition density; natural definitions of the most-likely-path approximation of the transition density, the call price, and the implied volatility then follow. In the time homogeneous case, we recover well-known classical results. However, in the time inhomogeneous case, we obtain a new asymptotic expansion that generalizes the classical one. We showed how the lowest order approximation of Berestycki, Busca and Florent as well as the higher order approximations of Gatheral et al. [7] and Gatheral and Wang [9] correspond to dropping terms in our lowest order path integral representation. We further showed that by restoring the dropped terms, our new representation recovers the exact expression for Black-Scholes implied volatility in the time-dependent Black-Scholes model, which no existing asymptotic expansion technique has so far been able to achieve, to the best of our knowledge. Further applications of this promising approach to the important practical problem of accurately approximating implied volatility under local volatility is left for future research.

Acknowledgement

We thank the anonymous reviewer for his helpful and constructive comments. We are also grateful for helpful discussions with the participants of the following seminars: Math Finance and PDE Seminar at Rutgers University, Probability Seminar at TU Berlin, Probability Seminar at Academia Sinica, Mathematics Colloquium at Ritsumeikan University, Mathematical Finance Seminar at Osaka University. All errors are our own responsibility.

References

[1] Baldi P. and Caramellino L., Asymptotics of hitting probabilities for general one-dimensional diffusions, *Annals of Applied Probability* **12** pp.1071–1095, 2002.

[2] Berestycki, H., Busca, J., and Florent, I., Asymptotics and calibration of local volatility models, *Quantitative Finance*, **2**, pp.61–69, 2002.

[3] Bleistein, N. and Handelsman, R.A., Asymptotic expansions of integrals, *Dover Publications*, 1986.

[4] Chavel, I., Eigenvalues in Riemannian geometry, *Pure and Applied Mathematics*, Book 115, *Academic Press*, 1984.

[5] Cheng, W., Costanzino, N., Liechty, J., Mazzucato, A.L., and Nistor, V., Closed-form asymptotics and numerical approximations of 1D parabolic equations with applications to option pricing, *SIAM Journal on Financial Mathematics*, **2**(1), pp.901–934, 2011.

[6] Gatheral, J., The Volatility Surface: A Practitioner's Guide, *Wiley Finance*, 2006.

[7] Gatheral, J., Hsu, E.P., Laurence, P., Ouyang, C., and Wang, T.-H., Asymptotics of implied volatility in local volatility models, *Mathematical Finance*, **22**(4), pp.591–620, 2012.

[8] Gatheral, J., and Jacquier, A., Arbitrage-free SVI volatility surfaces, *Quantitative Finance*, **14**(1), pp.59–71, 2014.

[9] Gatheral, J. and Wang, T.-H., The heat kernel most-likely-path approximation, *International Journal of Theoretical and Applied Finance*, **15**(1), 1250001, 2012.

[10] Goovaerts, M., De Schepper, A., and Decamps, M., Closed-form approximations for diffusion densities: a path integral approach, *Journal of Computational and Applied Mathematics*, **164-165**, pp.337–364, 2004.

[11] Julien Guyon and Henry-Labordère, P., From spot volatilities to implied volatilities, *Risk Magazine*, pp.79–84, June 2011.

[12] Hsu, E.P., Stochastic analysis on manifolds, *Graduate Studies in Mathematics*, *American Mathematical Society*, 2002.

[13] Henry-Labordère, P., Analysis, geometry, and modeling in finance, *Chapman& Hall/CRC*, Financial Mathematics Series, 2008.

[14] Jordan, R. and Tier, C., Asymptotic approximations to deterministic and stochastic volatility models, *SIAM Journal on Financial Mathematics*, **2**(1), pp.935–964, 2011.

[15] Keller-Ressel, M. and Teichmann, J., A remark on Gatheral's 'most-likely path approximation' of implied volatility, *in these Proceedings*, 2014.

[16] Linetsky, V., The path integral approach to financial modeling and options pricing, *Computational Economics*, **11**(1–2), pp.129–163, 1997.
[17] Reghai A., The hybrid most likely path, *Risk Magazine*, pp.34–35, April 2006.

Tai-Ho Wang
Department of Mathematics
Baruch College, The City University of New York
1 Bernard Baruch Way, New York, NY10010

E-mail address: tai-ho.wang@baruch.cuny.edu

Jim Gatheral
Department of Mathematics
Baruch College, The City University of New York
1 Bernard Baruch Way, New York, NY10010,

E-mail address: jim.gatheral@baruch.cuny.edu