Density Tracking by Quadrature for Stochastic Differential Equations

Harish S. Bhat∗† R. W. M. A. Madushani∗

September 25, 2017

Abstract

We develop and analyze a method, density tracking by quadrature (DTQ), to compute the probability density function of the solution of a stochastic differential equation. The derivation of the method begins with the discretization in time of the stochastic differential equation, resulting in a discrete-time Markov chain with continuous state space. At each time step, DTQ applies quadrature to solve the Chapman-Kolmogorov equation for this Markov chain. In this paper, we focus on a particular case of the DTQ method that arises from applying the Euler-Maruyama method in time and the trapezoidal quadrature rule in space. Our main result establishes that the density computed by DTQ converges in $L^1$ to both the exact density of the Markov chain (with exponential convergence rate), and to the exact density of the stochastic differential equation (with first-order convergence rate). We establish a Chernoff bound that implies convergence of a domain-truncated version of DTQ. We carry out numerical tests to show that the empirical performance of DTQ matches theoretical results, and also to demonstrate that DTQ can compute densities several times faster than a Fokker-Planck solver, for the same level of error.

∗Applied Mathematics Unit, University of California, Merced, CA 95343
†email: hbhat@ucmerced.edu
1 Introduction

Consider the stochastic differential equation (SDE) for the scalar process $X_t$ with initial condition $X_0 = C$,

$$dX_t = f(X_t)dt + g(X_t)dW_t,$$

where $W_t$ is the Wiener process. $X_t$ is an Itô diffusion; neither the drift $f$ nor the diffusion $g$ feature explicit time-dependence. Assuming regularity of $f$ and $g$, the process $X_t$ has a probability density function $p(x,t)$ (Rogers 1985). In this paper, we develop a convergent numerical method to solve for $p$. We call our method density tracking by quadrature (DTQ).

To introduce DTQ informally, let us describe the three main steps in its derivation:

1. Discretize the SDE (1) in time using a convergent stochastic time-stepping method.
2. Interpret the time-discretized equation as a discrete-time Markov chain on a continuous state space; let $\tilde{p}$ denote its probability density function. We can then write down a Chapman-Kolmogorov equation that enables us to evolve $\tilde{p}$ forward in time.
3. Discretize both the Chapman-Kolmogorov equation and $\tilde{p}$ in space, e.g., using a spatial grid and numerical quadrature. Let $\hat{p}$ denote the discrete-space approximation of $\tilde{p}$.

We use in step 1 the explicit Euler-Maruyama method and the trapezoidal rule in step 3; unless stated otherwise, this is the DTQ method analyzed in this paper. Please note that the above steps give a blueprint for many possible algorithms; it is entirely possible that by choosing a different time integrator and a different quadrature rule, one could derive a DTQ method that improves upon the default method studied here.

In this paper, we prove that $\hat{p}$ converges to $p$ as the discretization parameters tend to zero. Because there are existing results on the convergence of $\tilde{p}$ to $p$, the main task of this paper is to show that $\hat{p} \to \tilde{p}$.

The foundational work of Bally and Talay (1996) established conditions under which $\tilde{p}$ converges to $p$, in the case where the Euler-Maruyama method is used to discretize the SDE (1) in time. Let $\|f\|_1$ denote the $L^1$ norm of a function $f$. Suppose we seek the density of (1) at time $T > 0$. Let $h > 0$ denote the temporal step size; as we take $h \to 0$, we assume $T = Nh$ stays fixed. Then the results of Bally and Talay (1996) imply that $\|p(\cdot,T) - \tilde{p}(\cdot,T)\|_1 = O(h)$.

Our work builds on this result. The DTQ method analyzed here combines Euler-Maruyama temporal discretization with the trapezoidal rule on an equispaced grid. This results in a fast, simple method to compute an approximation $\hat{p}$ such that $\|\hat{p}(\cdot,T) - \tilde{p}(\cdot,T)\|_1 = O(h^{-1} \exp(-rh^{-k}))$ for positive constants $r, \kappa$. The user can control $\kappa$ by adjusting the relationship between the spatial and temporal grid spacings.

The primary application of this work that we envision is in statistical inference for diffusion processes. DTQ can be used to numerically approximate the likelihood function for a diffusion that is observed at discrete points in time. We have begun to use DTQ to devise both Bayesian and frequentist inference algorithms (Bhat and Madushani 2016; Bhat et al. 2016). The present work lays a theoretical foundation for these statistical applications. Additionally, note that when inference procedures for diffusions have been compared, a method that approximates the likelihood by numerically solving the Fokker-Planck (or Kolmogorov) equation achieves superior accuracy at the cost of excessive computational time (Hurn et al. 2007). The results of the present paper indicate that DTQ achieves
the same accuracy as a Fokker-Planck solver with less computational effort, further motivating the potential use of DTQ in many inference applications.

We now review alternative approaches to compute the density of \( \text{(1)} \), including prior work on DTQ and its relatives.

### 1.1 Alternative Approaches

If the drift \( f \) and diffusion \( g \) are sufficiently smooth, then \( p \) satisfies the forward Kolmogorov (or Fokker-Planck) equation (Rogers 1985):

\[
\frac{\partial}{\partial t} p(x, t) = -\frac{\partial}{\partial x} [f(x, t)p(x, t)] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [g^2(x, t)p(x, t)].
\]  

(2)

Prescribing an initial condition \( p(x, 0) \), we may then solve (2) to obtain the density \( p(x, T) \) at time \( T > 0 \). The solution of (2) must satisfy the normalization condition \( \int_{\mathbb{R}} \! p(\cdot, t) \, dx = 1 \), which implies boundary conditions of the form \( \lim_{|x| \to \infty} p(x, t) = 0 \).

We view DTQ as an alternative to numerical methods for the solution of (2). The primary purpose of the present paper is to demonstrate intrinsic properties—both theoretical and empirical—of DTQ. We compare DTQ with a finite difference method for the solution of (2); this is a logical choice given the particular version of DTQ studied here. In the present version, the density is numerically approximated (i.e., finite"-dimensionalized) by a sequence of values on an equispaced grid, just as in a finite difference method for a partial differential equation (PDE). By instead choosing to represent the unknown density as an expansion in a basis or frame, we can derive different versions of the DTQ method that are akin to finite element, meshless, and Hermite spectral methods for (2) (Paola and Soft 2002; Pichler et al. 2013; Canor and Denoël 2013; Luo and Yau 2013). We will pursue this line of reasoning and resulting comparisons in future work. In the present work, we compare DTQ against a finite difference method that is first-order in time and second-order in space. For a particular test problem at the finest grid resolution we consider, DTQ computes a solution with \( L^1 \) error \( \approx 3 \times 10^{-3} \) more than 100 times faster than the Fokker-Planck method.

Both numerical Fokker-Planck solvers and DTQ are deterministic approaches that avoid random sampling. We also place in this category the closed-form approximation methods of Aït-Sahalia for both univariate and multivariate (Aït-Sahalia 2002, 2008) diffusions. In the univariate case, this method approximates the transition density \( p \) using a Hermite function expansion. Prior work has established convergence and shown that the method works well for models of financial interest, requiring only a small number of terms (Aït-Sahalia 2002). In the present work, we compare likelihood surfaces computed via DTQ and the closed-form Hermite approximation method—see Section 1.3 for a summary of these and other results.

Besides the numerical solution of (2), one might try to estimate the density of (1) by sampling. Specifically, one can employ any convergent numerical method to step (1) forward in time from \( t = 0 \) to \( t = T \), thereby generating one sample of \( X_T \). Repeating this procedure many times, one can obtain enough samples of \( X_T \) to compute a statistical estimate of the density at time \( T \). For instance, one could compute a histogram or a kernel density estimate. Several existing methods can be viewed as special cases and/or extensions of this approach (Hu and Watanabe 1996; Kohatsu-Higa 1997; Milstein et al. 2004; Giles et al. 2015). In such methods, the accuracy of the density will be controlled by two parameters: the temporal step size and the number of sample paths. If there are \( N_S \) samples, then a typical stochastic time-stepping method will contribute an error of \( N_S^{-1/2} \) and kernel density estimation will contribute an error of, e.g., \( N_S^{-4/5} \). In comparison, the DTQ method’s accuracy is also
controlled by two parameters, the temporal step size and the grid spacing. Note that the trapezoidal rule on the real line contributes an error that decays exponentially in the grid spacing (Trefethen and Weideman 2014). For this reason, we believe DTQ will be a strong alternative to a sampling-based method.

Returning to the forward Kolmogorov or Fokker-Planck equation (2), we see that smoothness of \( f \) and \( g \) is required in order to have classical solutions. The implementation of DTQ itself does not utilize derivatives (whether exact or approximate) of \( f \) and \( g \). At the same time, our convergence theory assumes analyticity of \( f \) and \( g \) on a strip in the complex plane that contains the real line. We give two reasons for assuming analyticity. First, many models of scientific interest involve functions \( f \) and \( g \) that do satisfy these hypotheses. Second, in order to apply exponential error estimates for the trapezoidal rule (Trefethen and Weideman 2014), it is essential that our integrand, which depends on \( f \) and \( g \), be analytic on a strip. Ultimately, we expect that the hypotheses in the present convergence proof can be relaxed, both by changing the quadrature scheme/estimates and by making use of improved estimates for the convergence of \( \tilde{p} \) to \( p \) (Gobet and Labart 2008). Still, the present results are sufficient for the inference tasks we have in mind.

1.2 Prior Work

The DTQ method proposed here is an outgrowth of prior work on computing densities for stochastic delay differential equations (Bhat and Kumar 2012; Bhat 2014; Bhat and Madushani 2015). The method from Bhat and Madushani (2015), when adapted to equations with no time delay, is the method in the present paper. Our prior works did not address convergence from a theoretical standpoint, nor did they present empirical results of monotonic convergence that are in strict accordance with theory. The present paper addresses both of these issues.

DTQ has been described previously as numerical path integration. The method has achieved accurate results on a variety of problems in, e.g., nonlinear mechanics and finance—see Wehner and Wolfer (1983); Naess and Johnsen (1993); Linetsky (1997); Yu et al. (1997); Rosa-Clot and Taddei (2002); Skaug and Naess (2007). Recently, numerical path integration has been studied using semigroup methods (Chen et al. 2017); though convergence of \( \tilde{p} \) to \( p \) in \( L^1 \) is established, a fully discrete scheme (i.e., discretized in both time and space) is not analyzed. Interestingly, Chen et al. (2017) do not require that the drift \( f \) or diffusion \( g \) are bounded above, nor do they require more than 4 continuous derivatives for either function. These results complement ours, especially as we seek in future work to relax hypotheses and to improve our error estimates for quantities computed in practice, i.e., \( \hat{p} \) and its truncated domain version, \( \hat{p} \).

When we derive the DTQ method, we make use of the fact that a time-discretization of (1) can be viewed as a discrete-time Markov chain on a continuous state space. Suppose we were to take a different point of view, that of trying to design a discrete-time Markov chain on a discrete state space whose law or density approximates well that of the original SDE. In this case, there are extensive results starting from the work of Kushner (1974). Like a discrete-time, discrete-time Markov chain, the DTQ algorithm can be written in the form \( \hat{p}(t_{n+1}) = A\hat{p}(t_n) \), where \( A \) is a matrix (possibly with an infinite number of rows and columns) and \( \hat{p}(t_j) \) represents the approximate density at time \( t_j \). However, because of the quadrature-based derivation of the DTQ algorithm, the matrix \( A \) is, in general, not a Markov transition matrix. We find it both mathematically interesting and practically useful that, in spite of this, the DTQ method’s \( \hat{p} \) converges exponentially to \( \tilde{p} \).

The Chapman-Kolmogorov equation that is at the center of this paper—see (6)—has appeared in Pedersen (1995); Santa-Clara (1997). In these works, the right-hand side of the Chapman-Kolmogorov
equation is interpreted as an expected value that can be computed using Monte Carlo methods. In our approach, we use deterministic quadrature to evaluate the right-hand side of the Chapman-Kolmogorov equation. There is one prior paper we found that features this approach, albeit in a different context, that of a nonlinear autoregressive time series model (Cai 2003). The convergence results in Cai (2003) are of a different nature than ours, because they involve taking the continuum limit in space but not in time. In the present work, we are interested in the error made by the DTQ method as both the temporal and spatial grid spacings vanish.

1.3 Summary of Results and Outline

The main result of this paper is a provably convergent method for computing an approximation \( \hat{p} \) of the density \( p \) for the SDE (1). Let \( h > 0 \) and \( k > 0 \) denote, respectively, the temporal and spatial step sizes. Assume that \( k \propto h^{\rho} \) for \( \rho > 1/2 \), and assume that \( f \) and \( g \) are sufficiently regular (more precisely, admissible in the sense of Definition 2). Under these conditions, in Sections 4 and 5, we prove that \( \hat{p} \) converges to \( \tilde{p} \) in \( L^1 \), and that the error decays exponentially in \( h \). Specifically, there exists a constant \( r > 0 \) such that the leading order \( L^1 \) error term is proportional to \( h^{-1} \exp(-rh^{1/2-\rho}) \)—see Theorem 2. As a consequence of this result and the results of Bally and Talay (1996), we conclude that \( \hat{p} \) converges to \( p \) in \( L^1 \), and that the error decays linearly with \( h \)—see Corollary 1.

Up to and including Section 5 our results pertain to an idealized version of the DTQ algorithm in which we track the density \( \hat{p} \) at an infinite number of discrete grid points. In Section 6 we study the effect of boundary truncation. Our main tool in this section is a Chernoff bound on the tail sum of \( \hat{p} \) that we establish through the moment generating function. Let \( \hat{p} \) denote the approximation of \( \hat{p} \) obtained by summing over precisely \( 2M + 1 \) grid points from \(-y_M = -Mk \) to \( y_M = Mk \). The quantity \( \hat{p} \) is what we actually compute when we implement DTQ. In Lemma 9 we show that if \( y_M \to \infty \) at a logarithmic rate, i.e., \( y_M \propto \log h^{-1} \), then the \( L^1 \) error between \( \hat{p} \) and \( \tilde{p} \) is \( O(h) \). Combining this with our earlier results, this establishes \( L^1 \) convergence of \( \hat{p} \) to the true density \( p \)—see Corollary 2.

In Section 7 we study the performance of the DTQ method. For a suite of six test problems for which we have access to the exact solution, our numerical tests confirm \( O(h) \) convergence of \( \hat{p} \) to \( p \). This remains true for drift \( f \) and diffusion functions \( g \) that do not strictly satisfy the hypotheses of our convergence theory. We also present a finite difference method for solving (2); we compare this method against three different implementations of DTQ, and find that DTQ is competitive.

In Section 7 we also explain how to use DTQ to compute likelihoods given an SDE model and data. For two test problems, we compare the likelihood surfaces computed via DTQ against those computed using the closed-form Hermite approximation method (Aıt-Sahalia 2002). For an Ornstein-Uhlenbeck test problem, both methods’ likelihood surfaces are similar and would yield reasonable parameter estimates. For a test problem featuring bistability and bimodality, DTQ outperforms the closed-form method. In this case, maximum likelihood estimates obtained via DTQ are far closer to the ground truth than those obtained from the closed-form method.

Before proceeding, we give a more detailed derivation of the DTQ method in Section 2 and then introduce necessary assumptions and notation in Section 5.
2 Problem Setup

We begin with a more detailed derivation of the DTQ method. First, we discretize (1) in time using the explicit Euler-Maruyama method:

\[ x_{n+1} = x_n + f(x_n)h + g(x_n)\sqrt{h}Z_{n+1}, \]

where \( h > 0 \) is a fixed time step and \( Z_{n+1} \) is a random variable with a standard (mean zero, variance one) Gaussian distribution. We let \( \tilde{p}(x, t_n) \) denote the probability density function of \( x_n \). Note that this differs from \( p(x, t_n) \).

From (3), we observe that the density of \( x_{n+1} \) given \( x_n = y \) is Gaussian with mean \( y + f(y)h \) and variance \( h^2f^2(y) \). Let us denote this conditional density by \( \tilde{p}_{n+1,n}(x|y) \); then

\[ \tilde{p}_{n+1,n}(x|y) = G(x, y) := \frac{1}{\sqrt{2\pi}\sqrt{h}g^2(y)} \exp\left(-\frac{(x - y - f(y)h)^2}{2gh^2(y)}\right). \]

Note that, for any \( y \in \mathbb{R} \),

\[ \int_{\mathbb{R}} G(x, y) \, dx = 1. \]

With these definitions, we obtain

\[ \tilde{p}(x, t_{n+1}) = \int_{\mathbb{R}} \tilde{p}_{n+1,n}(x|y)\tilde{p}(y, t_n) \, dy, \]

the Chapman-Kolmogorov equation for the discrete-time, continuous-space Markov chain given by (3). Similar equations are often employed in the literature on inference for diffusions—see Pedersen (1995), Santa-Clara (1997), Fuchs (2013, Chap 6.3.3); and Kou et al. (2012).

Let us define an equispaced temporal grid by \( t_n = nh \) with \( h = T/N \). In principle, we can now repeatedly apply (6) to determine \( \tilde{p}(x, T) \). This assumes we can perform the integral over the real line.

To compute (6), we use numerical quadrature. Here we employ the trapezoidal rule, enabling the use of exponential error estimates (Trefethen and Weideman 2014; Stenger 2012; Lund and Bowers 1992). To begin with, we apply the trapezoidal rule on the real line. Later, we explain how to incorporate the effects of a finite, truncated integration domain.

Assume the domain \( \mathbb{R} \) is discretized via an equispaced grid \( y_j = jk \) where \( k > 0 \) is fixed. Then our discrete-time, discrete-space evolution equation is

\[ \tilde{p}(x, t_{n+1}) = k \sum_{j=-\infty}^{\infty} G(x, y_j)\tilde{p}(y_j, t_n). \]

Except for the fact that we have not yet truncated the infinite sum, this is the DTQ method.

In what follows, we assume a constant initial condition \( X_0 = C \), which implies \( p(x, 0) = \tilde{p}(x, 0) = \delta(x - C) \). This choice is not essential to either the use or convergence of the DTQ method. In fact, the choice of a point mass initial condition requires special handling, because we cannot discretize \( \tilde{p}(x, 0) \) directly. We insert \( n = 0 \) into (5), use \( \tilde{p}(x, 0) = \delta(x - C) \), and obtain the non-singular initial condition

\[ \tilde{p}(x, t_1) = \tilde{p}(x, t_1) = G(x, C). \]

This enables us to iteratively use (7) for \( n \geq 1 \).

Our main task in Sections 4 and 5 is to estimate \( ||\tilde{p}(\cdot, T) - \tilde{p}(\cdot, T)||_1 \). Before we start the proof of Theorem 2, we introduce necessary notation and assumptions.
3 Notation and Assumptions

We will use the Roman $i$ for the imaginary unit ($i = \sqrt{-1}$) and reserve the Italic $i$ for an index of summation. We denote the $L^1$ norm of a function $f : \mathbb{R} \to \mathbb{R}$ by

$$\|f\|_{L^1} = \int_{\mathbb{R}} |f(x)| \, dx.$$ 

We denote the $\ell^1$ norm of the sequence $\{\omega_j\}_{j=-\infty}^{\infty}$ by

$$\|\omega\|_{\ell^1} = \sum_{j=-\infty}^{\infty} |\omega_j|.$$ 

For a function $f : \mathbb{R} \to \mathbb{R}$, we understand $\|f\|_{\ell^1}$ to be the norm of the sequence obtained by applying $f$ on a spatial grid:

$$\|f\|_{\ell^1} = \sum_{j=-\infty}^{\infty} |f(jk)|,$$

where again $k > 0$ denotes the grid spacing. We use $\lceil x \rceil$ to denote the smallest integer greater than or equal to $x$, and $\lfloor x \rfloor$ to denote the largest integer less than or equal to $x$. The following definition is from the literature (Lund and Bowers 1992).

**Definition 1.** For $a > 0$, let $S_a$ denote the infinite strip of width $2a$ given by

$$S_a = \{z \in \mathbb{C} : |\Re(z)| < a\}.$$ 

Then $B(S_a)$ is the set of functions such that $\varphi \in B(S_a)$ iff $\varphi$ is analytic in $S_a$.

$$\int_{-a}^{a} \varphi(x + iy) \, dy = O(|x|^{\alpha}), \quad x \to \pm \infty, \quad 0 \leq \alpha < 1,$$

and

$$\mathcal{N}(\varphi, S_a) \equiv \lim_{y \to a} \left( \int_{\mathbb{R}} |\varphi(x + iy)| \, dx + \int_{\mathbb{R}} |\varphi(x - iy)| \, dx \right) < \infty. \quad (10)$$

The next definition encapsulates the constraints that the coefficient functions $f$ and $g$ in the original SDE (11) must satisfy in order for us to show exponential convergence of $\hat{p}$ to $\bar{p}$.

**Definition 2.** In this paper, we say that $f$ and $g$ are admissible if they satisfy the following properties. First, there exists $d > 0$ such that $f$ and $g$ are analytic on the strip $S_d$. Additionally, there exist positive, finite, real constants $M_1$, $M_2$, $M_3$, and $M_4$ such that for all $z \in S_d$,

$$|f'(z)| \leq M_1 \quad (11a)$$

$$M_2 \leq |g(z)| \leq M_3 \quad (11b)$$

$$\Re(g(z)) \neq 0 \quad (11c)$$

$$|g'(z)| \leq M_4. \quad (11d)$$

We now state a theorem that gives an exponential error estimate for the trapezoidal rule (Lund and Bowers 1992), one that we shall use to bound the error made in one step of the DTQ method. Other error estimates, with different hypotheses, can be found in the literature (Stenger 2012; Trefethen and Weideman 2014).
Theorem 1. Suppose $\varphi \in B(S_a)$ and $k > 0$. Let

$$\eta = \int_\mathbb{R} \varphi(x) \, dx - k \sum_{j=-\infty}^{\infty} \varphi(jk).$$

Then

$$|\eta| \leq \frac{N(\varphi, S_a)}{2 \sinh(\pi a/k)} \exp(-\pi a/k).$$

Proof. See Lund and Bowers (1992, Theorem 2.20). □

4 Preliminary Theory

In this section, we prove several lemmas that are essential ingredients for the convergence theorem in Section 5. The overall goal of these lemmas is to show that the integrand

$$\varphi(x, y, t_n) = G(x, y) \hat{p}(y, t_n),$$

considered as a function of $y$ for the purposes of quadrature, satisfies the hypotheses of Theorem 1.

The first lemma enables us to pass from an estimate of the error made in one time step to an estimate of the error made across a non-zero interval of time, even as the number of time steps becomes infinite.

Lemma 1. Suppose for the function $\xi : \mathbb{R}^+ \to \mathbb{R}^+$ there exist $\gamma > 1$, $\epsilon > 0$ and $h_0 > 0$ such that $\xi(h) \leq \epsilon h^\gamma$ for all $h < h_0$. Fix $T > 0$ and define $h = T/N$ where $N \in \mathbb{N}^+$. Then

$$\lim_{N \to \infty} \left[ h \sum_{j=1}^{N-1} (1 + \xi(h))^{j-1} \right] = T.$$

Proof. Take $N$ sufficiently large so that $h < 1$ and $h < h_0$. Then we calculate

$$\sum_{j=1}^{N-1} (1 + \xi(h))^{j-1} = \xi(h)^{-1} \left[ (1 + \xi(h))^{N-1} - 1 \right]$$

$$= \sum_{j=1}^{N-1} \binom{N-1}{j} \xi(h)^{j-1}$$

$$\leq \frac{T}{h} + \sum_{j=2}^{N-1} \frac{T^j \epsilon^{j-1}}{j!} h^{\gamma(j-1)} \exp(T \epsilon).$$

Using $h < 1$, we have

$$h \sum_{j=1}^{N-1} (1 + \xi(h))^{j-1} \leq T + \sum_{j=2}^{N-1} \frac{T^j \epsilon^{j-1}}{j!} h^{\gamma(j-1)} \exp(T \epsilon).$$

We have shown that the limit is $T$, and that the correction term to the limit is $O(h^{\gamma-1})$. □
The following lemma specializes an \( \ell^1 \)-norm estimate of a discrete Gaussian to the case of our function \( G \).

**Lemma 2.** Suppose \( g \) is admissible and \( h, k > 0 \) satisfy

\[
k \leq 2\pi (\log 2)^{-1/2} M_2 h^{1/2}
\]

Then for all \( y \in \mathbb{R} \), we have

\[
\left| 1 - k \| G(\cdot, y) \|_{\ell^1} \right| \leq 4 \exp \left( -\frac{2\pi^2 g^2(y) h}{k^2} \right).
\]

**Proof.** Let

\[
\phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(x - \mu)^2}{2\sigma^2} \right).
\]

Note that \( G \) and \( \phi \) coincide when \( \mu = y + f(y) h \) and \( \sigma^2 = g^2(y) h \). For any \( d > 0 \), on the strip \( S_d \), \( \phi \) satisfies the hypotheses of Theorem \([1]\). We restrict attention to those \( d \) satisfying \( d > (2\pi)^{-1} k \log 2 \), so that \( (\sinh(\pi d / k))^{-1} \leq 4 \exp (-\pi d / k) \). Then

\[
\int_{x \in \mathbb{R}} \left| \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(x + id - \mu)^2}{2\sigma^2} \right) \right| dx = e^{d^2/(2\sigma^2)}.
\]

As the right-hand side does not change when we replace \( d \) by \(-d\), we have \( N(\phi, S_d) = 2 \exp(d^2/(2\sigma^2)) \). Using Theorem \([1]\) and \( \int_{\mathbb{R}} \phi(x) dx = 1 \),

\[
\left| 1 - k \sum_{j=-\infty}^{\infty} \phi(jk) \right| \leq \exp \left( d^2/(2\sigma^2) \right) \frac{\exp(-\pi d / k)}{\sinh(\pi d / k)} \leq 4 \exp \left( \frac{d^2}{2\sigma^2} - \frac{2\pi d}{k} \right).
\]

When \( \sigma^2 = g(y) h \), we know by \([11b]\) and \([13]\) that \( d_* = 2\pi\sigma^2 / k \geq (2\pi)^{-1} k \log 2 \), so we can choose \( d = d_* \), the minimizer of \([16]\) with respect to \( d \), and maintain consistency. Making this substitution and setting \( \sigma^2 = g^2(y) h \), we have \([14]\). \( \square \)

For each \( t_n \), we think of \( \{\hat{p}(x, t_n)\}_{j=-\infty}^{\infty} \) as an infinite sequence. It is important to estimate the \( \ell^1 \) norm of this sequence.

**Lemma 3.** Suppose \( g \) is admissible and \( h, k > 0 \) satisfy \([13]\). Then for \( n \geq 1 \),

\[
(1 - 4 \exp(-2\pi^2 M_2^2 h / k^2))^{n-1} \leq \| \hat{p}(\cdot, t_n) \|_{\ell^1} / \| \hat{p}(\cdot, t_1) \|_{\ell^1} \leq (1 + 4 \exp(-2\pi^2 M_2^2 h / k^2))^{n-1},
\]

and the series defined by \ ([7] \) converges uniformly.

**Proof.** We prove this by induction with the base case of \( n = 1 \), for which \([17]\) is trivial. Consider the infinite series on the right-hand side of \ ([7] \) for \( n = 1 \) and fixed \( h \) and \( k \). Using \([11b]\), we have the elementary bound \( 0 \leq G(x, y) \leq (2\pi M_2^2 h)^{-1/2} \). Note that \([8]\) and \([14]\) together give us an \( \ell^1 \) bound on \( \{\hat{p}(j, t_1)\}_{j=-\infty}^{\infty} \). Combining these two bounds, it is clear that \([7]\) converges uniformly for \( n = 1 \), i.e., \( \hat{p}(y, t_2) \) converges uniformly.
Now assume for a particular \( n \geq 1 \) that (17) holds, \( \hat{p}(y, t_n) \geq 0, \| \hat{p}(\cdot, t_n) \|_{\ell^1} < \infty \), and \( \hat{p}(y, t_{n+1}) \) converges uniformly. We now show that these properties hold with \( n \) incremented by 1. By the induction hypotheses, we see that all terms of the convergent series on the right-hand side of (7) are nonnegative. Hence \( \hat{p}(y, t_{n+1}) \geq 0 \). We evaluate (7) at \( x = x_i \):

\[
\hat{p}(x_i, t_{n+1}) = k \sum_{j=-\infty}^{\infty} G(x_i, y_j) \hat{p}(y_j, t_n).
\]  

We take absolute values, sum over all \( i \), and interchange the order of summation—this is all justified because all terms are nonnegative. We obtain

\[
\| \hat{p}(:, t_{n+1}) \|_{\ell^1} = \sum_{j=-\infty}^{\infty} \left[ k \sum_{i=-\infty}^{\infty} G(x_i, y_j) \right] \hat{p}(y_j, t_n).
\]

Applying (14) and (11b), we have

\[
(1 - 4 \exp(-2\pi^2 M_2^2 h/k^2)) \| \hat{p}(\cdot, t_n) \|_{\ell^1} \leq \| \hat{p}(\cdot, t_{n+1}) \|_{\ell^1} \leq (1 + 4 \exp(-2\pi^2 M_2^2 h/k^2)) \| \hat{p}(\cdot, t_n) \|_{\ell^1}.
\]  

This shows that \( \| \hat{p}(\cdot, t_{n+1}) \|_{\ell^1} < \infty \). Now we return to the right-hand side of (7) with \( n \) replaced by \( n+1 \). Combining our elementary bound on \( G \) with the \( \ell^1 \) bound on \( \hat{p}(\cdot, t_{n+1}) \), it is clear that the series converges uniformly. Finally, from (19) we obtain upper and lower bounds for \( \| \hat{p}(\cdot, t_{n+1}) \|_{\ell^1} / \| \hat{p}(\cdot, t_n) \|_{\ell^1} \). Multiplying appropriately by (17), we advance \( n \) by 1. \( \square \)

One consequence of Lemma 3 is that it enables us to give asymptotic conditions on \( h \) and \( k \) such that \( \hat{p} \) is normalized correctly.

**Lemma 4.** Suppose, in addition to the hypotheses of Lemmas 2 and 3 that \( k = r_1 h^\rho \) for constants \( r_1 > 0 \) and \( \rho > 1/2 \). Assume that \( N = T/h \) for some fixed \( T > 0 \). Then for \( 1 \leq n \leq N + 1 \),

\[
\lim_{h \to 0} k \| \hat{p}(\cdot, t_n) \|_{\ell^1} = 1.
\]  

**Proof.** Applying the hypotheses to the exponential terms in (17) with \( n = N = T/h \), we have

\[
\lim_{h \to 0} \left( 1 + 4 \exp(-2\pi^2 M_2^2 h^2 r_1^{-2\rho+1}) \right)^{T/h} = 1.
\]  

Consequently, for any \( n \in \{0, 1, \ldots, N\} \), we have

\[
\lim_{h \to 0} \frac{\| \hat{p}(\cdot, t_{n+1}) \|_{\ell^1}}{\| \hat{p}(\cdot, t_n) \|_{\ell^1}} = 1.
\]  

From (8) and (11), we conclude that \( k \| \hat{p}(\cdot, t_1) \|_{\ell^1} \to 1 \) as \( k \to 0 \). Then (20) follows immediately from (22). \( \square \)

**Lemma 5.** Suppose \( f \) and \( g \) are admissible and that \( a < \min\{d, M_2^2/(2 M_3 M_4)\} \). Then for any \( x, y \in \mathbb{R} \), there exist \( A_2 > 0 \) and \( A_1, A_0 \in \mathbb{R} \) such that

\[
|G(x, y + ia)| = \frac{1}{\sqrt{2\pi h |g(y + ia)|^2}} \exp \left( -\frac{A_2 x^2 + A_1 x + A_0}{4g(y + ia)^2 h} \right),
\]  

and there exists \( \gamma_0 \in (0, 2) \) such that

\[
|G(x, y + ia)| \leq \frac{1}{\sqrt{2\pi h M_2^2}} \exp \left( \frac{a^2(1 + h M_4)^2}{h\gamma_0 M_2^2} \right).
\]
We suppose that a defined by ξ. Note that By “c.c.” we mean the complex conjugate of all preceding terms. We have used the fact that because f and g are analytic on S_d, and because they are real-valued when restricted to the real axis, both f and g commute with complex conjugation. That is, f(y + ia) = f(y − ia) and similarly for g and g^2. The upshot is that A_2, A_1, and A_0 are all real.

Let us now prove that A_2 > 0. Define the function

θ(y, ε) = g^2(y − ie) + g^2(y + ie),

for ε ∈ [0, d). For each fixed y, by the mean-value theorem, there exists ξ such that

θ(y, ε) − θ(y, 0) = ε \frac{∂θ}{∂ε}(y, ξ).

Note that ξ may depend on ε and y. Now we use (11) to compute

\sup_{y ∈ S_d, ε ∈ (−d, d)} \left| \frac{∂θ}{∂ε} \right| = 4 \sup_{y ∈ S_d, ε ∈ (−d, d)} \left| S(g(y + iε)g'(y + iε)) \right| ≤ 4M_3M_4. \tag{25}

Then using the previous two equations together with (11b), we have

θ(y, ε) ≥ θ(y, 0) − 4εM_3M_4 ≥ 2M_2^2 − 4εM_3M_4. \tag{26}

The right-hand side will be positive as long as ε < min{d, M_2^2/(2M_3M_4)}. Given the hypothesis on a in the statement of the lemma, θ(y, a) = A_2 will be positive. Because A_2 > 0, we can maximize the right-hand side of (23) as a function of x; the global maximum occurs at x = −A_1/(2A_2). Then we have

|G(x, y + ia)| ≤ \frac{1}{\sqrt{2πhM_2^2}} \exp \left( \frac{(2a + ih(f(y − ia) − f(y + ia)))^2}{4h(g^2(y + ia) + g^2(y − ia))} \right).

We suppose that a = bM_2^2/(2M_3M_4) for some b ∈ (0, 1) such that a < d. Then the lower bound (26) implies θ(y, a) ≥ 2M_2^2(1 − b). We define γ_0 = 2(1 − b) ∈ (0, 2) and write

|G(x, y + ia)| ≤ \frac{1}{\sqrt{2πhM_2^2}} \exp \left( \frac{(2a + ih(f(y − ia) − f(y + ia)))^2}{hγ_0M_2^2} \right).

Let Γ be the segment connecting y − ia to y + ia. Note that a < d implies that Γ is completely contained
in the strip $S_d$ where $f$ is analytic. Using (11a), we have
\[
|2a + ih(f(y - ia) - f(y + ia))| 
\leq 2|a| + h|f(y + ia) - f(y - ia)| 
\leq 2|a| + h \int f'(z) \, dz 
\leq 2|a| + h \int |f'(z)| \, |dz|
\]
\[
\leq 2|a|(1 + hM_1)
\]
Using this estimate in (4) finishes the proof.

\[\square\]

**Lemma 6.** Suppose that $f$ and $g$ are admissible, that $h, k > 0$ satisfy (13), and that $a < \min(d, M_2^2/(2M_3M_4))$. Then the integrand (12), considered as a function of $y$, is a member of $B(S_d)$, i.e., $\varphi(x, \cdot, t_n) \in B(S_d)$.

**Proof.** There are three conditions for membership in $B(S_d)$, which we verify in turn. First, we check that $\varphi$ is analytic on $S_d$. At time step $t_1$, we have $\hat{p}(y, t_1) = G(y, C)$, the analyticity of which follows from (11c) and the lower bound in (11b). The arguments made earlier regarding the convergence of (18) hold equally well with $x_t$ replaced by any $x$. This implies that for $n \geq 1$, $\hat{p}(y, t_{n+1})$ is analytic in $y$ on $S_d$, so the integrand $\varphi$ is analytic on $S_a \subset S_d$. Next, we consider
\[
\Phi(x, y, t_n) = \int_{b=-a}^{a} |\varphi(x, y + ib, t_n)| \, db.
\]
Let $z_j = jk$. Since
\[
\hat{p}(y + ia, t_{n+1}) = k \sum_{j=-\infty}^{\infty} G(y + ia, z_j) \hat{p}(z_j, t_n),
\]
we have
\[
\Phi(x, y, t_{n+1}) \leq k \sum_{j=-\infty}^{\infty} \hat{p}(z_j, t_n) \int_{b=-a}^{a} |G(y + ib, z_j)||G(x, y + ib)| \, db 
\leq k \sum_{j=-\infty}^{\infty} \hat{p}(z_j, t_n) G(y, z_j) \int_{b=-a}^{a} \exp \left( \frac{b^2}{2g^2(z_j)h} \right) |G(x, y + ib)| \, db 
\leq \frac{1}{\sqrt{2\pi hM_2^2}} \int_{b=-a}^{a} \exp \left( \frac{b^2}{2M_2^2h} \right) \exp \left( \frac{h^2(1 + hM_1)^2}{2h\gamma hM_2^2} \right) \, db \sum_{j=-\infty}^{\infty} \hat{p}(z_j, t_n) G(y, z_j).
\]
To derive the last inequality, we have applied Lemma 5 and (11b). By Lemma 3 we know $\hat{p}(y, t_{n+1})$ converges uniformly. We integrate with respect to $y$, bring the integral into the sum, and use (5). In this way, we derive $||\hat{p}(\cdot, t_{n+1})||_1 = k||\hat{p}(\cdot, t_0)||_1 < \infty$. Therefore, $\hat{p}(y, t_{n+1}) \to 0$ as $|y| \to \infty$; in the same limit, we have $\Phi(x, y, t_{n+1}) = O(|y|^\alpha)$ for $\alpha = 0$, satisfying (5).

Next, we establish a bounded, real function $L_n$ such that for each $x \in \mathbb{R},$
\[
N := \int_{\mathbb{R}} |G(x, y + ia)\hat{p}(y + ia, t_n)| \, dy + \int_{\mathbb{R}} |G(x, y - ia)\hat{p}(y - ia, t_n)| \, dy \leq L_n(x).
\]

11
We need this estimate in order to apply Theorem 1. For this purpose, we seek an upper bound on $N$ that does not depend essentially on the spatial discretization parameter $k$. Starting again from (28), we have

$$
\int_{y \in \mathbb{R}} |G(x, y + ia)| \hat{p}(y + ia, t_{n+1}) |dy| \leq k \sum_{j=-\infty}^{\infty} \hat{p}(z_j, t_{n}) \int_{\mathbb{R}} |G(y + ia, z_j)| |G(x, y + ia)| |dy|
$$

$$
= k \sum_{j=-\infty}^{\infty} \hat{p}(z_j, t_{n}) \int_{\mathbb{R}} \exp \left( \frac{a^2}{2g^2(z_j)} \right) G(y, z_j) G(x, y + ia) |dy|
$$

$$
\leq k \exp \left( \frac{a^2}{2M_2^2 h} \right)
\times \sum_{j=-\infty}^{\infty} \hat{p}(z_j, t_{n}) \int_{\mathbb{R}} G(y, z_j) G(x, y + ia) |dy|
$$

$$
\leq k \exp \left( \frac{a^2}{2M_2^2 h} \right) \| \hat{p}(\cdot, t_{n}) \|_{L^1} \psi(x, a),
$$

where

$$
\psi(x, a) = \sup_{z \in \mathbb{R}} \left[ \int_{y \in \mathbb{R}} G(y, z) G(x, y + ia) |dy| \right].
$$

Examining (23), we see that the right-hand side of (31) is invariant under the reflection $a \mapsto (-a)$. We define the real-valued function

$$
L_{n+1}(x) = 2k \exp \left( \frac{a^2}{2M_2^2 h} \right) \| \hat{p}(\cdot, t_n) \|_{L^1} \psi(x, a),
$$

and note that (31) implies $N \leq L_n(x)$, as required by (29). Our task now is to demonstrate that $L_n$ is finite. By Lemma 5 and (5), we have

$$
\psi(x, a) \leq \frac{1}{\sqrt{2\pi hM_2^2}} \exp \left( \frac{a^2(1 + hM_1)^2}{h\gamma_0 M_2^2} \right).
$$

Using this estimate in (31), we obtain

$$
L_{n+1}(x) \leq 2k \exp \left( \frac{a^2}{2M_2^2 h} \right) \| \hat{p}(\cdot, t_n) \|_{L^1} \frac{1}{\sqrt{2\pi hM_2^2}} \exp \left( \frac{a^2(1 + hM_1)^2}{h\gamma_0 M_2^2} \right).
$$

Note that the bound on the right-hand side does not depend on $x$ at all. The dependence on $k$ is confined to the terms $k\| \hat{p}(\cdot, t_n) \|_{L^1}$. By Lemmas 2 and 3 together with (14),

$$
k\| \hat{p}(\cdot, t_n) \| \leq \left( 1 + 4 \exp(-2\pi^2 g^2(C)/k^2) \right) \left( 1 + 4 \exp(-2\pi^2 M_2^2 h/k^2) \right)^{n-1} \leq 5^n < \infty
$$

for all $k \geq 0$. In sum, we have shown that for fixed $h > 0$, fixed $n \geq 1$, and $a < \min\{d, M_2^2/(2M_3 M_1)\}$, $L_n(x)$ is bounded uniformly in $x$ and $k$. We have demonstrated that (29) holds. Thus $\varphi(x, \cdot, t_n) \in B(S_a)$.

□
5 Convergence Theorem

Let

\[ E(y, t_n) = \tilde{p}(y, t_n) - \hat{p}(y, t_n). \]  

In this section, we establish conditions under which \( \|E(\cdot, T)\|_1 \) goes to zero at an exponential rate.

**Theorem 2.** Suppose that \( f \) and \( g \) are admissible in the sense of Definition 2. Assume that

\[ k = r_1 h^\rho \]  

for constants \( r_1 > 0 \) and \( \rho > 1/2 \). Choose

\[ a < \min\{d, M_2^2/(2M_3M_4)\} \]

such that

\[ a = r_2 h^{1/2} \]  

for some \( r_2 > 0 \). Assume that \( h, k \) satisfy (34) and that \( k < 2\pi a/\log 2 \). For fixed \( T > 0 \), choose

\[ h \in \left(0, \min\{T, (M_2^2/(4M_3M_4r_2)^2)\}\right) \]

such that \( N = T/h \in \mathbb{N}^+ \). To be clear, \( r_1 \) and \( r_2 \) are constants that do not depend on \( h \). Then

\[ \|E(\cdot, T)\|_1 \leq c_* h^{-1} \exp(-2\pi r_2^{-1} h^{1/2-r})(1 + o(h) + o(k)) \]  

where \( o(h) \) and \( o(k) \) stand for terms that vanish as \( h \to 0 \) and \( k \to 0 \), and \( c_* > 0 \) is a constant that does not depend on \( h \).

**Proof.** We begin with

\[ \tilde{p}(x, t_{n+1}) = \int_{y \in \mathbb{R}} G(x, y) \tilde{p}(y, t_n) \, dy \]

\[ = \int_{y \in \mathbb{R}} G(x, y) \hat{p}(y, t_n) \, dy + \int_{y \in \mathbb{R}} G(x, y) E(y, t_n) \, dy. \]

We now apply the trapezoidal rule to the first integral. For each \( x \) and \( t_n \), we let \( \tau(x, t_n) \) denote the quadrature error incurred, i.e.,

\[ \int_{y \in \mathbb{R}} G(x, y) \hat{p}(y, t_n) \, dy = k \sum_{j=-\infty}^{\infty} G(x, y_j) \hat{p}(y_j, t_n) + \tau(x, t_n) \]

\[ = \hat{p}(x, t_{n+1}) + \tau(x, t_n). \]  

We use this in the previous equation to derive

\[ E(x, t_{n+1}) = \int_{y \in \mathbb{R}} G(x, y) E(y, t_n) \, dy + \tau(x, t_n). \]

Taking absolute values, we apply the triangle inequality together with \( G \geq 0 \) to obtain

\[ |E(x, t_{n+1})| \leq \int_{y \in \mathbb{R}} G(x, y)|E(y, t_n)| \, dy + |\tau(x, t_n)|. \]
Integrating over \( x \) and using Fubini’s theorem and (5),
\[
\|E(\cdot, t_{n+1})\|_1 - \|E(\cdot, t_n)\|_1 \leq \|\tau(\cdot, t_n)\|_1.
\] (39)

Summing both sides from \( n = 1 \) to \( n = N - 1 \) and using (5),
\[
\|E(\cdot, T)\|_1 \leq \sum_{n=1}^{N-1} \|\tau(\cdot, t_n)\|_1.
\] (40)

We apply Lemma 6 and Theorem 1 to produce
\[
|\tau(x, t_n)| \leq \frac{N}{2 \sinh(\pi a/k)} \exp(-\pi a/k),
\] (41)

where \( \tau \) and \( N \) are defined by (38) and (29), respectively. Combining (39) with (23), we have
\[
\int_{y \in \mathbb{R}} |G(x, y + ia)\hat{p}(y + ia, t_{n+1})| \, dy \leq \exp\left(\frac{a^2}{2M^2 h}\right) k \sum_{j=-\infty}^{\infty} \hat{p}(z_j, t_n)
\times \int_{y \in \mathbb{R}} \frac{G(y, z_j)}{\sqrt{2\pi h|g(y + ia)|^2}} \exp\left(-\frac{A_2 x^2 + A_1 x + A_0}{4|g(y + ia)|^4 h}\right) \, dy,
\]

where again \( A_2, A_1, \) and \( A_0 \) are defined by (24). We see that the right-hand side of this inequality is invariant under \( a \mapsto -a \), and so we write
\[
N \leq 2 \exp\left(\frac{a^2}{2M^2 h}\right) k \sum_{j=-\infty}^{\infty} \hat{p}(z_j, t_n) \int_{y \in \mathbb{R}} \frac{G(y, z_j)|g(y + ia)|}{\sqrt{g^2(y + ia) + g^2(y - ia)}} \times \exp\left(\frac{(2a + ih(f(y - ia) - f(y + ia)))^2}{4h(g^2(y + ia) + g^2(y - ia))}\right) \, dy.
\]

For \( a < \min\{d, M^2/(2M_3 M_4)\} \), we have shown that the coefficient \( A_2 \) is positive on \( S_a \). This enables us to integrate both sides with respect to \( x \):
\[
\int_{x \in \mathbb{R}} N \, dx \leq 2 \sqrt{\pi} \exp\left(\frac{a^2}{2M^2 h}\right) k \sum_{j=-\infty}^{\infty} \hat{p}(z_j, t_n)
\times \int_{y \in \mathbb{R}} \frac{G(y, z_j)|g(y + ia)|}{\sqrt{g^2(y + ia) + g^2(y - ia)}} \times \exp\left(\frac{(2a + ih(f(y - ia) - f(y + ia)))^2}{4h(g^2(y + ia) + g^2(y - ia))}\right) \, dy.
\]

On the right-hand side, we have carried out the \( x \) integral first; the changing of the order of summation and integration is justified by the nonnegativity of every term. Next, we apply estimates established in the proof of Lemma 5. We obtain
\[
\int_{x \in \mathbb{R}} N \, dx \leq 2 \sqrt{\pi} \exp\left(\frac{a^2}{2M^2 h}\right) \frac{M_3}{\gamma_0^{1/2} M_2} \exp\left(\frac{a^2(1 + hM_1)^2}{h\gamma_0 M^2_2}\right) k \sum_{j=-\infty}^{\infty} \hat{p}(z_j, t_n)
\]

Combining this with (41) and the estimates from the proof of Lemma 2, we have
\[
\int_{x \in \mathbb{R}} |\tau(x, t_n)| \, dx \leq 4 \sqrt{\pi} \exp\left(\frac{a^2}{2M^2 h}\right) \frac{M_3}{\gamma_0^{1/2} M_2} \exp\left(\frac{a^2(1 + hM_1)^2}{h\gamma_0 M^2_2}\right) \exp(-2\pi a/k) \|\tau(\cdot, t_n)\|_{\ell^1}.
\]
Using (17), we obtain
\[
\|\tau(\cdot, t_n)\|_1 \leq 4 \sqrt{2} M_3 \gamma_0^{-1/2} M_2^{-1} \times \exp \left( \frac{a^2}{2 M_2^2 h} \right) \exp \left( \frac{a^2(1 + hM_1)^2}{h \gamma_0 M_2^2} \right) \exp(-2\pi a/k) \times k \|\hat{\theta}(\cdot, t_1)\|_{\ell^1} (1 + 4 \exp(-2\pi^2 M_2^2 h/k^2))^{n-1}.
\]

We sum both sides from \( n = 1 \) to \( n = N - 1 \):
\[
\sum_{n=1}^{N-1} \|\tau(\cdot, t_n)\|_1 \leq \sqrt{2} M_3 \gamma_0^{-1/2} M_2^{-1} \times \exp \left( \frac{r_2^2}{2 M_2^2} \right) \exp \left( \frac{r_2^2(1 + hM_1)^2}{\gamma_0 M_2^2} \right) T \exp(-2\pi r_2 h^{-1/2} \gamma_{1/2}) \exp(-2\pi h^{-1} k \|\hat{\theta}(\cdot, t_1)\|_{\ell^1}) \times \left[ h \sum_{n=1}^{N-1} (1 + 4 \exp(-2\pi^2 M_2^2 h/k^2))^{n-1} \right].
\]
We now use (40) and hypotheses (34) and (35):
\[
\|E(\cdot, T)\|_1 \leq \sqrt{2} M_3 \gamma_0^{-1/2} M_2^{-1} \times \exp \left( \frac{r_2^2}{2 M_2^2} \right) \exp \left( \frac{r_2^2(1 + hM_1)^2}{\gamma_0 M_2^2} \right) T \exp(-2\pi r_2 h^{-1/2} \gamma_{1/2}) \exp(-2\pi h^{-1} k \|\hat{\theta}(\cdot, t_1)\|_{\ell^1}) \times \left[ h \sum_{n=1}^{N-1} (1 + 4 \exp(-2\pi^2 M_2^2 h/k^2))^{n-1} \right].
\]

By (36), we have \( h \leq T \). By the definition of \( \gamma_0 \) in Lemma 5, we have that \( \gamma_0 = 2(1 - b) \) where
\[
b = 2M_3 M_4 a/M_2^2 = 2M_3 M_4 r_2 h^{1/2}/M_2^2.
\]
Assumption (36) now implies that \( b \leq 1/2 \) and \( \gamma_0^{-1} \leq 1 \). We write
\[
c_* = \sqrt{2} M_3 M_2^{-1} \exp \left( \frac{r_2^2}{2 M_2^2} \right) \exp \left( \frac{r_2^2(1 + T M_1)^2}{M_2^2} \right)\]

Let \( \xi(h) = 4 \exp(-c_1 h^{-c_2}) \), where \( c_1 \) and \( c_2 \) are positive constants with no dependence on \( h \). We check that \( \xi \) satisfies the hypotheses of Lemma 11 \( h^{-\gamma} \xi(h) \) has a global maximum at \( h_* = (c_1 c_2 / \gamma)^{1/c_2} \), and so we have \( \xi(h) \leq \epsilon h^\gamma \) for \( \epsilon = h_*^{\gamma} \xi(h_*) \), any choice of \( \gamma > 1 \), and all \( h > 0 \). With \( c_1 = 2\pi^2 \gamma_2 \) and \( c_2 = 2\rho - 1 \), we apply Lemma 11 to the term in square brackets on the right-hand side of (43). We conclude that
\[
\frac{h}{T} \sum_{n=1}^{N-1} (1 + 4 \exp(-2\pi^2 M_2^2 r_2^{-1} h^{1-2\rho}))^{n-1} = 1 + o(h)
\]
as \( h \to 0 \) with \( N = T/h \). By Lemma 2, \( k \|\hat{\theta}(\cdot, t_1)\|_{\ell^1} = 1 + o(k) \) as \( k \to 0 \). Putting everything together, we are left with (57). \( \square \)
We are now in a position to combine our result with an earlier result from the literature to establish the convergence of \( \hat{p} \) to \( p \).

**Corollary 1.** In addition to the hypotheses of Theorem 2, suppose there exist constants \( F_k, G_k > 0 \) such that \( \sup_{x \in \mathbb{R}} |f(x)| \leq F_k \) and \( \sup_{x \in \mathbb{R}} |g(x)| \leq G_k \) for all \( k \geq 0 \). Note that for \( k = 1 \), the first condition is redundant with (11a); for \( k = 0 \) and \( k = 1 \), the second condition is redundant with (11b) and (11d). Then we have

\[
\| p(\cdot, T) - \hat{p}(\cdot, T) \|_1 = O(h)
\]

**Proof.** We have

\[
\| p(\cdot, T) - \hat{p}(\cdot, T) \|_1 \leq \| p(\cdot, T) - \tilde{p}(\cdot, T) \|_1 + \| \tilde{p}(\cdot, T) - \hat{p}(\cdot, T) \|_1
\]

To handle the first term, we appeal to Corollary 2.1 from Bally and Talay (1996). Our lower bound on \( g \) in (11b) corresponds to Bally and Talay’s uniform ellipticity hypothesis “H1”; we may then apply Equations (27-28) from Bally and Talay (1996) to derive

\[
| p(x, T) - \hat{p}(x, T) | \leq h \mathcal{H}_1 \exp \left( - \mathcal{H}_2 x^2 / T \right)
\]

for constants \( \mathcal{H}_1, \mathcal{H}_2 > 0 \) that do not depend on \( h \). Therefore,

\[
\| p(\cdot, T) - \tilde{p}(\cdot, T) \|_1 \leq h \mathcal{H}_1 \left( \frac{\pi T}{\mathcal{H}_2} \right)^{1/2} = O(h).
\]

Returning to (44), by Theorem 2, the second term on the right-hand side goes to zero much faster than \( h \), finishing the proof. \( \square \)

### 6 Boundary Truncation

In practice, in lieu of the infinite sum (7), we compute approximate densities using the following truncation:

\[
\hat{p}(x, t_{n+1}) = k \sum_{j=-M}^{M} G(x, y_j) \hat{p}(y_j, t_n)
\]

(45)

As in (8), we take \( \hat{p}(x, t_1) = G(x, C) \) and use (45) starting with \( n = 1 \). Let us denote the error due to truncation by

\[
r(x, t_{n+1}) = \hat{p}(x, t_{n+1}) - \hat{p}(x, t_{n+1})
\]

(46)

By (8), we have \( r(x, t_1) \equiv 0 \). For \( n \geq 1 \), we have

\[
r(x, t_{n+1}) = k \left( \sum_{|j|>M} G(x, y_j) \hat{p}(y_j, t_n) + \sum_{|j|\leq M} G(x, y_j) r(y_j, t_n) \right).
\]

(47)

Based on the right-hand side, we see that it will be important to estimate the tail sum \( \sum_{|j|>M} \hat{p}(x_j, t_n) \). We accomplish this using a Chernoff bound. To arrive at this bound, we construct a sequence of random variables \( \{ Q_n \}_{n \geq 1} \). We first define a normalization constant at time \( n \):

\[
K_n = \| \hat{p}(\cdot, t_n) \|_1 = \sum_{i} \hat{p}(x_i, t_n).
\]

(48)
By (17), we know that \(K_n < \infty\) for \(k > 0\) and \(h > 0\). Let

\[
q(x_i, t_n) = \frac{\hat{p}(x_i, t_n)}{K_n},
\]

so that \(\sum_i q(x_i, t_n) = 1\). For each \(n\), we postulate a random variable \(Q_n\) with state space \(k\mathbb{Z}\) and probability mass function \(q(\cdot, t_n)\). In order to apply a Chernoff bound to \(Q_n\), we must estimate its moment generating function.

**Lemma 7.** Suppose \(f\) and \(g\) are admissible. Suppose \(k = h^p\) for some \(p > 1/2\), and that \(h, k > 0\) satisfy (13). Then there exists \(h_s\) such that for all \(h \in [0, h_s]\), all \(s \in \mathbb{R}\), and all \(n\) satisfying \(0 \leq n \leq (N - 1)\),

\[
k\mathbb{E}[e^{tQ_{n+1}}] < \frac{3}{2} \exp \left( T \left( \frac{M_{2,1}^2 s^2}{2} + f(0)s \right) \right) \left( \frac{1}{2} + \exp(Cse^{M_1T}) \right) < \infty.
\]

**Proof.** We begin with our estimate of the moment generating function of \(Q_{n+1}\). The calculation proceeds in two phases. The first phase is exact; note that in what follows we use the notation \(y_j = jk\), \(w_j = y_j + f(y_j)h\), and \(\sigma^2 = g^2(y_j)\):

\[
\mathbb{E}[e^{tQ_{n+1}}] = \sum_{i=-\infty}^{\infty} e^{tx_i} q(x_i, t_{n+1})
\]

\[
= k \frac{1}{K_{n+1}} \sum_i e^{tx_i} \sum_j \frac{1}{\sqrt{2\pi g^2 h}} \exp \left( -\frac{(x_i - w_j)^2}{2 g^2 h} \right) \hat{p}(y_j, t_n)
\]

\[
= k \frac{1}{K_{n+1}} \sum_j \sum_i \frac{1}{\sqrt{2\pi g^2 h}} \exp \left( -\frac{x_i^2 - 2x_i w_j + w_j^2 - 2g^2 h s x_i}{2 g^2 h} \right) \hat{p}(y_j, t_n)
\]

\[
= k \frac{1}{K_{n+1}} \sum_j \xi_s(j) \exp \left( -\frac{w_j^2 - (w_j + g^2 h s)^2}{2 g^2 h} \right) \hat{p}(y_j, t_n),
\]

where

\[
\xi_s(j) = k \sum_i \frac{1}{\sqrt{2\pi g^2 h}} \exp \left( -\frac{(x_i - (w_j + g^2 h s))^2}{2 g^2 h} \right).
\]

It is at this point that we begin to estimate. Note that the summand is in fact a discrete Gaussian \(\phi(x_i)\), as in (15), with \(\mu = w_j + g^2(y_j)h s\) and \(\sigma^2 = g^2(y_j)h\). Hence we may apply the same reasoning from the proof of Lemma 2 to obtain

\[
\xi_s(j) \leq 1 + 4 \exp \left( -\frac{2\pi^2 g^2(y_j)^2 h}{k^2} \right) \leq 1 + 4 \exp \left( -\frac{2\pi^2 M_{2,1}^2 h}{k^2} \right).
\]

Next, we turn our attention to the remaining exponential in (50). We use (11b), the mean value theorem, and the definition of \(w_j\) to obtain:

\[
\exp \left( -\frac{w_j^2 - (w_j + g^2 h s)^2}{2 g^2 h} \right) = \exp \left( w_j s + \frac{1}{2} g^2(y_j)h s^2 \right)
\]

\[
\leq e^{M_{2,1}^2 h s^2/2} \exp(y_j s + f(y_j)h s)
\]

\[
\leq e^{M_{2,1}^2 h s^2/2} \exp(y_j s + f(0)h s + \beta y_j h s)
\]

\[
\leq e^{M_{2,1}^2 h s^2/2 + f(0)h s} \exp(y_j s(1 + \beta h))
\]
Here $\beta = f'(y)$ for some $y \in (0, y_j)$. Now we combine (50), (51), and (52). The result is

$$E[e^{\beta Q_{n+1}^+}] \leq \frac{K_n}{K_{n+1}} \left( 1 + 4 \exp(-2\pi^2 M_2^2 h/k^2) \right) e^{M_2^2 h^2 + f(0)h_s} \frac{1}{K_n} \sum_j \exp(y_j s(1 + \beta h)) \tilde{p}(y_j, t_n) \quad (53)$$

We recognize the expression on the second line as the moment generating function of $Q_n$ evaluated at $s' = s(1 + \beta h)$. Therefore,

$$kE[e^{\beta Q_{n+1}^+}] \leq \frac{K_n}{K_{n+1}} \left( 1 + 4 \exp(-2\pi^2 M_2^2 h/k^2) \right) e^{M_2^2 h^2 + f(0)h_s} \frac{1}{K_n} \sum_j \exp(y_j s(1 + \beta h)) \tilde{p}(y_j, t_n) \quad (53)$$

The main question now is what happens as $h \to 0$ and $N \to \infty$ such that $hN = T$. We assume that $0 \leq n \leq (N - 1)$. Because $k = r_1 h^\rho$ for $\rho > 1/2$, we know by Lemma $[5]$ that $\zeta_1(h) \to 1$ as $h \to 0$. Hence there exists $h_1^*$ such that $h \in [0, h_1^*)$ ensures that $|\zeta_1(h) - 1| < 1/2$, i.e., $\zeta_1(h) < 3/2$. Next, consider

$$\zeta_2(h) = kE[e^{(1 + \beta h)^\nu Q_n}]$$

$$= k \sum_{i=-\infty}^\infty e^{(1 + \beta h)^\nu x_i} \tilde{p}(x_i, t_1)$$

$$= k \sum_{i=-\infty}^\infty e^{(1 + \beta h)^\nu x_i} G(x_i, C)$$

$$= \exp \left( (C + f(C)h)s(1 + \beta h)^n + \frac{hg^2(C)}{2}(1 + \beta h)^{2n} \right) k \sum_{i=-\infty}^\infty \phi(x_i),$$

where $\phi(x)$ is the Gaussian density defined in (15) with

$$\mu = C + f(C)h + hg^2(C)s(1 + \beta h)^n$$

$$\sigma^2 = hg^2(C)$$

Now we apply Lemma $[2] n \leq (N - 1)$, and (11a) to obtain

$$\zeta_2(h) \leq \exp \left( (C + f(C)h)s(1 + M_1 h)^N + \frac{hg^2(C)}{2}(1 + M_1 h)^{2N} \right) \times (1 + 4 \exp(-2\pi^2 g^2(C)h/k^2)).$$

As before, $hk^{-2} = r_1^{-2} h^{1-2\rho} \to +\infty$ as $h \to 0$, and the term on the second line goes to 1 as $h \to 0$. Since $\lim_{h \to 0^+} (1 + M_1 h)^N = e^{M_1 T}$, we have

$$\lim_{h \to 0^+} \zeta_2(h) \leq \exp \left( Cse^{M_1 T} \right).$$

Thus there exists $h_2^*$ such that $h \in [0, h_2^*)$ implies

$$\left| \zeta_2(h) - \exp \left( Cse^{M_1 T} \right) \right| \leq \frac{1}{2}.$$

Taking $h_* = \min \{ h_1^*, h_2^* \}$ finishes the proof.
Lemma 8. Suppose \( f \) and \( g \) are admissible in the sense of Definition 2. Suppose \( k = h^\rho \) for \( \rho > 1/2 \), and that \( h, k > 0 \) satisfy (13). For \( \varepsilon \geq 1 \), let
\[
M = [(\varepsilon + \rho + 1)(-\log h)/k].
\]
Let \( h_* \) be defined as in Lemma 7. Then for \( h < h_* \), we have \( k \sum_{\|i\| \leq M} |r(x_i, T)| = O(h) \).

**Proof.** We start with
\[
|r(x_i, t_{n+1}| \leq k \sum_{\|j\| > M} G(x_i, y_j) \hat{p}(y_j, t_n) + k \sum_{\|j\| \leq M} G(x_i, y_j)\rho(y_j, t_n)|.
\]
Summing over \( i \), we obtain
\[
\sum_{\|i\| \leq M} |r(x_i, t_{n+1})| \leq k \sum_{\|j\| > M} \sum_{\|i\| \leq M} G(x_i, y_j) \hat{p}(y_j, t_n) + k \sum_{\|j\| \leq M} \sum_{\|i\| \leq M} G(x_i, y_j)\rho(y_j, t_n)|.
\]
Using (14) together with (11b), we have
\[
\sum_{\|i\| \leq M} |r(x_i, t_{n+1})| \leq (1 + 4 \exp(-2\pi^2 M^2_2 h/k^2)) \sum_{\|j\| > M} \hat{p}(y_j, t_n)
+ (1 + 4 \exp(-2\pi^2 M^2_2 h/k^2)) \sum_{\|j\| \leq M} |\rho(y_j, t_n)|.
\]
This is of the form
\[
R_{n+1} \leq \alpha \pi_n + \alpha R_n.
\]
We derive from this the sequence of inequalities \( \alpha R_n \leq \alpha^2 \pi_{n-1} + \alpha^2 R_{n-1}, \ldots, \alpha^n R_2 \leq \alpha^n \pi_1 + \alpha^n R_1 \).

Summing these together with (56), we derive
\[
R_{n+1} \leq \sum_{i=1}^n \alpha^i \pi_{n-i+1} + \alpha^n R_1.
\]
Applying this to (55) and using \( r(\cdot, t_1) \equiv 0 \), we have
\[
\sum_{\|i\| \leq M} |r(x_i, t_{n+1})| \leq \sum_{i=1}^n (1 + 4 \exp(-2\pi^2 M^2_2 h/k^2))^i \sum_{\|j\| > M} \hat{p}(y_j, t_{n-i+1}).
\]
Now we use [49] and the Chernoff bound to derive:
\[
\sum_{\|j\| > M} \hat{p}(y_j, t_{n-i+1}) = K_{n-i+1} \sum_{\|j\| > M} q(y_j, t_{n-i+1}) \\
\quad \leq K_{n-i+1} \left[ P(Q_{n-i+1} \geq y_M) + P(Q_{n-i+1} \leq -y_M) \right]
\quad \leq K_{n-i+1} e^{-\gamma_M}(E[e^{\gamma Q_{n-i+1}}] + E[e^{-\gamma Q_{n-i+1}}])
\]
We apply Lemma 7 to obtain
\[
k \cdot \sum_{\|j\| > M} \hat{p}(y_j, t_{n-i+1}) \leq 3/2 K_{n-i+1} e^{-\gamma_M} \exp \left[T\left(M^2_2 s^2 + |f(0)s|\right)(1 + 2 \cosh(C e^M_1 T))\right]
\]
19
Applying this result to (57), we have

\[ k \sum_{|l| \leq M} |r(x_i, t_{n+1})| \leq \frac{3}{2} e^{-g y_M} \exp \left[ T \left( \frac{M_3^2 s^2}{2} + |f(0)s| \right) \right] \left( 1 + 2 \cosh(C s e^{M_1 T}) \right) \times \sum_{i=1}^n (1 + 4 \exp(-2\pi^2 M_2^2 h/k^2))^i K_{n-i+1}. \]

By (48) and (17), we have

\[ K_{n-i+1} \leq \| \hat{p}(\cdot, t_1) \|_{\ell^1} (1 + 4 \exp(-2\pi^2 M_2^2 h/k^2))^{n-i} \]

Using this and \( n \leq N = T/h \),

\[ k \sum_{|l| \leq M} |r(x_i, t_{n+1})| \leq \frac{3}{2} e^{-g y_M} \exp \left[ T \left( \frac{M_3^2 s^2}{2} + |f(0)s| \right) \right] \left( 1 + 2 \cosh(C s e^{M_1 T}) \right) \times \| \hat{p}(\cdot, t_1) \|_{\ell^1} \frac{T}{h} (1 + 4 \exp(-2\pi^2 M_2^2 h/k^2))^{T/h}. \]  \hspace{1cm} (59)

Let \( s = 1 \). Note that

\[ \lim_{h \to 0} (1 + 4 \exp(-2\pi^2 M_2^2 h/k^2))^{T/h} = 1 \]

and \( \lim_{h \to 0} k \| \hat{p}(\cdot, t_1) \|_{\ell^1} = 1 \). Thanks to (54), we know that \( y_M \geq (\varepsilon + \rho + 1)(-\log h) \). Putting things together, the right-hand side of (59) behaves like

\[ h^{\varepsilon + \rho + 1} k^{-1} h^{-1} = h^\varepsilon = O(h) \]

as desired. \( \square \)

So long as \( M \) remains a positive integer, we can add a constant to (54) and still prove Lemma 8. What is important is how \( M \) scales as a function of \( h \); the logarithmic rate given in (54) is the rate at which we have to push \( M \) to \( +\infty \) so that we obtain \( O(h) \) convergence. If we push \( M \) to \( +\infty \) at a faster rate, e.g., by replacing \((-\log h)\) with \( h^{-1} \), then \( r \) will converge at a rate that is exponential in \( h \).

Thus far we have considered convergence of \( r \) in a truncated and scaled version of the \( \ell^1 \) norm. Convergence in \( L^1 \) is an easy consequence.

**Lemma 9.** Suppose \( f \) and \( g \) are admissible in the sense of Definition 2. Suppose \( k = h^\rho \) for \( \rho > 1/2 \), and that \( h, k > 0 \) satisfy (13). For \( \varepsilon \geq 1 \), let \( M \) be defined as in (54). Let \( h_* \) be defined as in Lemma 7. Then for \( h < h_* \), we have \( \| r(\cdot, T) \|_{\ell^1} = O(h) \).

**Proof.** Note that

\[ |r(x, T)| \leq k \sum_{|l| > M} G(x, y_j) \hat{p}(y_j, t_{N-1}) + k \sum_{|l| \leq M} G(x, y_j) |r(y_j, t_{N-1})|. \]

This is similar to what we wrote above, except that the discrete variable \( x_i \) has been replaced by the continuous variable \( x \). We now integrate both sides with respect to \( x \) to obtain

\[ \| r(\cdot, T) \|_{\ell^1} \leq k \sum_{|l| > M} \hat{p}(y_j, t_{N-1}) + k \sum_{|l| \leq M} |r(y_j, t_{N-1})|. \]
The second term is $O(h)$ by Lemma 8. For the first term, we use (58) to write

$$k \sum_{|j| > M} \hat{p}(y_j, t_{N-1}) \leq \frac{3}{2} K_{N-1} e^{-y_M} \exp \left[ T \left( \frac{M^2}{2} + |f(0)| \right) \right] \left( 1 + 2 \cosh(C e^{M_1 T}) \right).$$

(60)

Since $\lim_{k \to 0} kK_{N-1} = 1$ and $e^{-y_M} = O(h^{\epsilon+1})$, the right-hand side of (60) behaves like $h^{\epsilon+1} = O(h^2)$. □

It is now immediately clear that, under certain conditions, we have established $O(h)$ convergence of $\hat{p}$ to the true density $p$ in the $L^1$ norm.

**Corollary 2.** Suppose that all hypotheses of Corollary 2 and Lemma 9 are satisfied. Then, combining these results, we have $\|p(\cdot, T) - \hat{p}(\cdot, T)\|_1 = O(h)$.

### 7 Numerical Experiments

In this section, we use R and C++ implementations of DTQ to study its empirical convergence behavior, and also to compare against a numerical solver for (2), the Fokker-Planck or Kolmogorov equation. All codes described in this section, together with instructions on how to reproduce Figures 12, 3, and 4 are available online.1 We caution the reader that, in the present work, we do not deal with all important implementation issues. Here we are primarily concerned with demonstrating properties of DTQ. This can be done quite well even with the assumptions on the initial condition and domain sizes given below. Relaxing these assumptions poses no conceptual difficulties, but may require changes to technical details in the codes.

#### 7.1 Convergence

First, we conduct an empirical study of DTQ convergence. We verify that under the conditions given by Theorem 2, we do observe convergence in practice. We also show numerical evidence that such convergence takes place when one or more of the hypotheses do not hold. Each SDE we consider is an equation for a scalar unknown $X_t$.

Let us describe the way in which we conduct numerical tests for each SDE. We begin with the initial condition $X_0 = 0$ and solve forward in time until $T = 1$. That is, we apply DTQ (45) to compute $\hat{p}(x, 1)$. We use the following values of the temporal step $h$:

$$\{0.5, 0.2, 0.1, 0.05, 0.02, 0.01, 0.005, 0.002, 0.001\}. \quad (61)$$

For $h \geq 0.01$, we find that an implementation of DTQ written completely in R is able to run in a reasonable amount of time. For $h = 0.005$ and below, we use an implementation where computationally intensive parts of the code are written in C++; this code is glued to our R code using the Repp and Repp/Armadillo packages (Eddelbuettel and Françoix 2011; Eddelbuettel 2013; Eddelbuettel and Sanderson 2014; Sanderson and Curtin 2016).

1 https://github.com/hbh4000/sdeinference/tree/master/DTQpaper

21
The remaining algorithm parameters are set in the following way:

\[
\begin{align*}
  k &= h^{3/4} \\
  \left\{ \begin{array}{l}
  \text{All except Ex. (64c)} & M = \lfloor \pi/k^2 \rfloor \\
  \text{Ex. (64c)} & M = \lfloor \pi/(2k) - 2 \rfloor.
  \end{array} \right.
\end{align*}
\]

(62a) (62b)

\[
  x_j = jk, \quad \text{for} \quad -M \leq j \leq M.
\]

(62c)

For each value of \( h \), we compare \( \hat{p}(x, T) \) computed using DTQ against the exact solution \( p(x, T) \). Let \( F(y, T) = \int_{x=-y}^{x=y} p(x, T) \, dx \) denote the cumulative distribution function associated with the density \( p \). Each comparison is carried out using the following three norms:

\[
\begin{align*}
  \|p(\cdot, T) - \hat{p}(\cdot, T)\|_1 &\approx k \sum_{j=-M}^{j=M} |p(jk, T) - \hat{p}(jk, T)| \quad \text{(63a)} \\
  \|p(\cdot, T) - \hat{p}(\cdot, T)\|_\infty &\approx \sup_{|j| \leq M} |p(jk, T) - \hat{p}(jk, T)| \quad \text{(63b)} \\
  \|F(\cdot, T) - \hat{F}(\cdot, T)\|_\infty &\approx \sup_{|j| \leq M} |F(jk, T) - \hat{F}(jk, T)| \quad \text{(63c)}
\end{align*}
\]

For our tests, we consider six SDE examples, all for a scalar unknown \( X_t \):

\[
\begin{align*}
  dX_t &= -X_t \, dt + dW_t & p(x, t) &= \frac{\exp(-x^2/(1 - \exp(-2t)))}{\sqrt{\pi(1 - \exp(-2t))}} \\
  (64a) \\
  dX_t &= -\frac{1}{2} \tanh X_t \, \text{sech}^2 X_t \, dt + \text{sech} X_t \, dW_t & p(x, t) &= (2\pi)^{-1/2}(\cosh x) \exp(-\sinh^2 x/(2t)) \\
  (64b) \\
  dX_t &= -(\sin X_t \cos^3 X_t) \, dt + (\cos^2 X_t) \, dW_t & p(x, t) &= (2\pi)^{-1/2}(\sec^2 x) \exp(-\tan^2 x/(2t)) \\
  (64c) \\
  dX_t &= \left( \frac{1}{2} X_t + \sqrt{1 + X_t^2} \right) \, dt + \sqrt{1 + X_t^2} \, dW_t & p(x, t) &= (2\pi(1 + x^2))^{-1/2} \\
  \times \exp(-\sinh^{-1} x - t^2/2) \\
  (64d) \\
  dX_t &= \frac{1}{2} X_t \, dt + \sqrt{1 + X_t^2} \, dW_t & p(x, t) &= (2\pi t(1 + x^2))^{-1/2} \\
  \times \exp(-\sinh^{-1} x)^2/(2t)) \\
  (64e) \\
  dX_t &= \left( -\sqrt{1 + X_t^2} \sinh^{-1} X_t + \frac{1}{2} X_t \right) \, dt \\
  &\quad + \sqrt{1 + X_t^2} \, dW_t & p(x, t) &= \frac{\exp(-\sinh^{-1} x)^2/(1 - \exp(-2t))}{\sqrt{\pi(1 - \exp(-2t))(1 + x^2)}} \\
  (64f)
\end{align*}
\]

Note that for each example, we have supplied an exact solution in the form of a probability density function \( p(x, t) \). For each example, we compare the DTQ density with \( p(x, T = 1) \).
Figure 1 shows the convergence results for all six examples. The overall impression we gain from the plots is that the practical $L^1$ error between the DTQ and exact density functions scales like $h$. As we now explain, this first-order convergence is displayed under a variety of conditions.

Example (64a) features drift and diffusion coefficients that clearly satisfy the hypotheses of our convergence theory. In this case, the computational results confirm the theory.

In Example (64b), the drift and diffusion coefficients satisfy all but one of the hypotheses. Specifically, because $\text{sech} \ x \to 0$ as $|x| \to \infty$, the diffusion coefficient is not bounded away from zero. However, as a matter of numerical practice, on any truncated domain of the form (62), the diffusion coefficient never equals zero. We can say, then, that on the computational domain, the diffusion coefficient does have a global lower bound that is greater than zero. The computational results display first-order convergence.

Example (64c) is similar to Example (64b) in that all but one of the hypotheses are satisfied. Again, it is the diffusion coefficient $\cos^2 y$ that is not bounded away from zero. However, either an analysis of the original SDE or inspection of the exact solution reveals that the density will only be supported on the interval $(-\pi/2, \pi/2)$. For this SDE, we set $M = \lceil \pi/(2k) - 1 \rceil$ as in (62b), retaining (62a) and (62c). This way, the spatial grid covers the interior of $(-\pi/2, \pi/2)$ and the diffusion coefficient never reaches zero. Again, the computational results show that the $L^1$ error scales like $h$.

Moving to Examples (64d) and (64e), the diffusion coefficient is now bounded from below by 1 but unbounded above. All other hypotheses of our convergence theory are satisfied. The empirical convergence rates for both examples match what we expect from theory.

Reexamining the situation with slightly more depth, what we find from our proofs is that (25) is the only place where the upper bound on the diffusion coefficient is used. However, for the particular case of the diffusion coefficient $g(x) = (1 + x^2)^{1/2}$ used in Examples 4 and 5, we have that

$$|\Im (g(y + i\epsilon)g'(y + i\epsilon))| = |\Im (y + i\epsilon)| \leq d,$$

meaning that we can substitute $d$ for $M_3M_4$ and the convergence proof follows. This is an example of how, for specific SDE that do not satisfy the hypotheses of the general theorem, we may yet be able to prove convergence of the DTQ method.

Finally, we come to Example (64f). Now we have that the derivative of the drift coefficient is unbounded and that the diffusion coefficient is unbounded above. Though the hypotheses of the convergence theory are not satisfied, we still observe first-order convergence.

For the SDE in Example (64f), even if we are able to patch our proof to prove that $\hat{p}$ converges to $\bar{p}$, we can no longer apply the result of Bally and Talay (1996) to guarantee convergence of $\hat{p}$ to $p$. Overall, we take the numerical results for Example (64f) as evidence that $\hat{p}$ must converge to $p$ under more general conditions than have been established in the literature.

### 7.2 Comparison with Fokker-Planck

Now we compare DTQ against a classical approach, that of numerically solving the Fokker-Planck or Kolmogorov PDE (2). In what follows, we use subscripts to denote partial derivatives, so that (2) is written

$$p_t + (f(x)p(x,t))_x = \frac{1}{2} \left( g^2(x)p(x,t) ight)_{xx}. \quad (66)$$

To solve this equation, we employ a standard finite difference method. To resolve the singular initial condition $p(x, 0) = \delta(x)$, we use a subtraction technique: we set $p = u + v$, where $u$ solves

$$u_t = \frac{1}{2} \kappa u_{xx}, \quad u(x, 0) = \delta(x), \quad (67)$$

and $v$ solves

$$v_t + (f(x)v(x,t))_x = \frac{1}{2} \left( g^2(x)v(x,t) ight)_{xx}, \quad (68)$$

subject to $v(0, t) = 0$. We analyze the error $w = u + v - p$.

We first show that $w$ satisfies

$$w_t + (f(x)w(x,t))_x = \frac{1}{2} \left( g^2(x)w(x,t) ight)_{xx}, \quad (69)$$

subject to $w(0, t) = 0$. Then

$$w(x, t) = \max(0, w(x,t)) = \max(0, \delta(x)) = 0, \quad (70)$$

and

$$\max(0, w(0,t)) = 0, \quad (71)$$

The error $w$ is then

$$w(x,t) = 0.$$
while \( v \) solves

\[
v_t + (f(x)v(x,t))_x = \frac{1}{2} \left( g^2(x)v(x,t) \right)_x + \frac{1}{2} \left[ \left( g^2(x) - \kappa \right) u(x,t) \right]_x - \left[ f(x) u(x,t) \right]_x \tag{68a}
\]

\[v(x,0) = 0.\tag{68b}\]

The point is that (67) can be solved analytically, i.e., for \( t > 0 \),

\[u(x,t) = \frac{1}{\sqrt{2\pi\kappa t}} \exp \left( -\frac{x^2}{2\kappa t} \right).\tag{69}\]

Here \( \kappa > 0 \) is a parameter that we are free to set. In our own tests, we use \( \kappa = 1 \). Since (69) is known, we substitute it into the final two terms on the right-hand side of (68a)—this yields a known forcing term \( F(x,t) \). We then employ the following numerical scheme to solve (68) for \( v(x,t) \):

- We discretize \( v(x,t) \) on fixed spatial and temporal grids with respective spacings \( k \) and \( h \). Let \( V^n_j \) denote our numerical approximation to \( v(jk, nh) \). Here \( 0 \leq n \leq N \) with \( Nh = T > 0 \), the final time. We also have that \( -M \leq j \leq M \). Implicitly, we assume that \( v(x,t) = 0 \) for \( |x| > Mk \).
- We use a first-order approximation to \( v_t \): \( v_t(x,t) \approx (V^{n+1}_j - V^n_j)/h \).
- We treat the drift term explicitly:

\[
(f(x)v(x,t))_x \approx \left( f((j + 1)k)V^n_{j+1} - f((j - 1)k)V^n_{j-1} \right)/(2k).
\]

- We treat the diffusion term implicitly:

\[
\frac{1}{2} \left( g^2(x)v(x,t) \right)_x \approx \frac{1}{2k^2} \left[ g^2((j - 1)k)V^{n+1}_{j-1} - 2g^2(jk)V^{n+1}_j + g^2((j + 1)k)V^{n+1}_{j+1} \right].
\]

Let \( V^n \) be a vector of length \( 2M + 1 \) whose \( j \)-th entry is \( V^n_j \). Then, combining approximations, we obtain the matrix-vector system

\[AV^{n+1} = BV^n + F^n\tag{70}\]

with tridiagonal matrices \( A \) and \( B \) given by (71) and (72) in Table 1. We also define \( F^n \) in (70) by discretizing \( F(x,t) \) in (68a). That is, for \( -M \leq j \leq M \), we define the \( j \)-th component of \( F^n \) by

\[
F^n_j = \frac{h}{2k^2} \left[ g^2((j - 1)k)u((j - 1)k, nh) - 2g^2(jk)u(jk, nh) + g^2((j + 1)k)u((j + 1)k, nh) \right]
- \frac{h}{2k} \left[ f((j + 1)k)u((j + 1)k, nh) - f((j - 1)k)u((j - 1)k, nh) \right]. \tag{73}
\]

To solve for \( V^{n+1} \) given \( V^n \), we rewrite (70) as

\[V^{n+1} = A^{-1}BV^n + A^{-1}F^n. \tag{74}\]

We compute \( u(jk, T) \) for \( -M \leq j \leq M \) and denote the resulting vector by \( U^N \). Let \( p_{FF}(x,T) \) denote the vector whose \( j \)-th component is \( p_{FF}(x_j, T) \), the approximation of \( p(x_j, T) \) obtained by solving
the most computationally expensive part of DTQ is the assembly of discretized on the same spatial grid as where we denote convergence theory. For fixed \( n = \text{constant} \), the Fokker-Planck equation numerically. With these definitions, our algorithm for computing \( p_{FP} \) is easily stated: we start with \( \mathbf{V}^0 = \mathbf{0} \), iterate (74) \( N \) times to compute \( \mathbf{V}^N \), and then compute

\[
p_{FP}(\mathbf{x}, T) = \mathbf{U}^N + \mathbf{V}^N.
\]

Note that in our Fokker-Planck solver, the matrices \( A \) and \( B \) defined by (71) and (72) are implemented as sparse tridiagonal matrices. When we use (74) to solve for \( \mathbf{V}^{n+1} \), we use sparse numerical linear algebra to compute both \( A^{-1}B \) and \( A^{-1}F^n \). In particular, \( A^{-1}B \) is precomputed before we loop from \( n = 0 \) to \( n = N - 1 \).

We are now in a position to compare the DTQ and Fokker-Planck methods. For this comparison, we exclusively use the drift and diffusion functions from Example (64a). As described above, among the examples in (64), Example (64a) is the only one that satisfies all of the hypotheses of our DTQ convergence theory.

As mentioned in Section 6, when we implement DTQ in practice, we start with (45) — with \( x \) discretized on the same spatial grid as \( y \), i.e.,

\[
\hat{p}(x_i, t_{n+1}) = k \sum_{j=-M}^{M} G(x_i, y_j) \hat{p}(y_j, t_n)
\]

(75)

For fixed \( n \), as \( j \) varies from \(-M\) to \( M \), the elements \( \hat{p}(y_j, t_n) \) form a \((2M + 1)\)-dimensional vector that we denote \( \mathbf{p}^n \). With this notation, (75) can be written

\[
\mathbf{p}^{n+1} = \mathcal{A} \mathbf{p}^n,
\]

where \( \mathcal{A} \) is the \((2M + 1) \times (2M + 1)\) matrix whose \((i, j)\)-th element is \( kG(x_i, y_j) \). In our experience, the most computationally expensive part of DTQ is the assembly of \( \mathcal{A} \). For the tests presented in this subsection, we have implemented three different methods to compute \( \mathcal{A} \):

1. **DTQ-Naïve.** Here we assemble \( \mathcal{A} \) using dense matrix methods in R. The main advantage of this approach is ease of implementation; the code to compute \( \mathcal{A} \) is only 4 lines long. Incidentally, the convergence tests in the first part of this section use DTQ-Naïve for \( h \geq 0.01 \).
2. **DTQ-CPP.** Implicitly, DTQ-Naïve forces R to loop over the entries of $\mathcal{A}$ serially. In DTQ-CPP, we use Rcpp together with OpenMP directives in C++ to compute and fill in the entries of $\mathcal{A}$ in parallel. In practice, we run this code on a machine with 12 cores, setting the number of OpenMP threads to 12.

3. **DTQ-Sparse.** Here we take advantage of the structure of $\mathcal{A}$. Specifically, we have

\[
\mathcal{A}_{ij} = kG(x_i, y_j) = \frac{k}{\sqrt{2\pi g^2(y_j)h}} \exp\left(-\frac{(x_i - y_j - f(y_j)h)^2}{2g^2(y_j)h}\right).
\]

Let us set $i = j + i'$. Then we have

\[
\mathcal{A}_{j+i', j} = \frac{k}{\sqrt{2\pi g^2(y_j)h}} \exp\left(-\frac{(i'k - f(y_j)h)^2}{2g^2(y_j)h}\right). \tag{77}
\]

We think of $i'$ as indexing the sub-/super-diagonals of $\mathcal{A}$. For each fixed $i' = 0, 1, 2, \ldots$ we evaluate (77) over all $j$ to obtain the $i'$-th subdiagonal of $\mathcal{A}$. For $h$ small, as $i'$ increases, we observe that the entire subdiagonal decays rapidly. In our implementation, we compute subdiagonals until the 1-norm of the subdiagonal drops below $2.2 \times 10^{-16}$ (machine precision in R) multiplied by the 1-norm of the main $i' = 0$ diagonal of $\mathcal{A}$. We then compute the same number of superdiagonals as subdiagonals. The final $\mathcal{A}$ matrix is assembled as a sparse matrix using the CRAN Matrix package (Bates and Maechler 2016).

Given the tridiagonal structure of both $A$ and $B$ in the Fokker-Planck method, we do not believe any reasonable modern implementation would use dense matrices. Similarly, while DTQ-Naïve requires minimal programming effort, a reasonable implementation would look much more like DTQ-CPP or DTQ-Sparse. None of the DTQ methods require more programming effort to implement than the Fokker-Planck method.

**Results for $O(h^{3/4})$ Domain Scaling.** For each $h$ in (61) that satisfies $h \geq 0.01$, we use all three DTQ methods and the Fokker-Planck method to generate numerical approximations of the density function at the final time $T = 1$. For our first set of comparisons, parameters such as $k$ and $M$ are set via (62). In particular, the computational domain is $[-y_M, y_M]$ where $y_M = Mk \propto h^{-3/4}$. We compute the $L^1$ errors between each numerical solution and the exact solution $p(x, T)$. We also record the wall clock time (in seconds) required to compute the solution using each method. Each measurement is repeated 100 times; we report average results.

In the left panel of Figure 2, we have plotted (on log-scaled axes) wall clock time as a function of $L^1$ error for each of the four methods. We see that if one can tolerate a relatively large $L^1$ error, then the fastest method is the DTQ-Naïve method (green); for $L^1$ errors less than 0.03, the fastest method is DTQ-Sparse (purple). The Fokker-Planck method is often the slowest of the four methods. For an error of 0.003, DTQ-Sparse is approximately 100 times faster than the Fokker-Planck method.

**Results for $O(\log h^{-1})$ Domain Scaling.** For our second set of comparisons, we have changed the way that $y_M$ (effectively, the size of the computational domain) scales with $h$. We retain $k = h^{3/4}$ but now set $y_M = (2 + 3/4)(-\log h) \propto (-\log h)$ in accordance with (54). The spatial grid, for all four
Euler-Maruyama method. We restrict \( \Theta \) of the likelihood surfaces. In the right panel of Figure 1 we have plotted (on log-scaled axes) wall clock time as a function of \( L^1 \) error for each of the four methods. Once again, we find that DTQ-Naïve and DTQ-Sparse are the fastest for, respectively, large and small error values. For an error of 0.003, DTQ-Sparse is approximately \( 10^{3/4} \approx 5.62 \) times faster than the Fokker-Planck method.

### 7.3 Likelihood Surface Test

The ability to compute densities accurately should enable accurate computation of likelihoods, an important ingredient for statistical estimation and inference. Let us explore this point through two examples.

First let us modify (1) by making explicit the dependence of the drift \( f \) and diffusion \( g \) on a parameter vector \( \Theta \):

\[
dX_t = f(X_t; \Theta)dt + g(X_t; \Theta)dW_t,
\]

Suppose that we have observed (78) at discrete points in time \( t = \{ j(\Delta t) \}_{j=0}^J \) and recorded the time series \( s = \{ s_0, s_1, \ldots, s_J \} \). To use the data \( s \) and the model (78) to estimate \( \Theta \), we write down the likelihood function \( p(s|\Theta) \). As the solution \( X_t \) of (78) satisfies the Markov property, we have

\[
p(s|\Theta) = p(s_0|\Theta) \prod_{j=1}^J p(s_j|s_{j-1}, \Theta).
\]

We assume that \( X_0 \) is independent of \( \Theta \) and thus that \( p(s_0|\Theta) \) does not influence the \( \Theta \)-dependence of the likelihood function; henceforth we ignore this initial term.

To understand the rest of the right-hand side of (79), let us explain how to compute \( p(s_j|s_{j-1}, \Theta) \). With the initial condition \( X_0 = s_{j-1} \), we use the given \( \Theta \) and the model (78) to compute the density of \( X_{\Delta t} \). Evaluating this density at the data point \( s_j \), we would obtain \( p(s_j|s_{j-1}, \Theta) \).

When we cannot compute the required density using an exact solution, we must turn to approximate methods. Suppose we use DTQ. Fix the number \( N \) of internal DTQ steps to take from time \( j(\Delta t) \) to \( (j+1)(\Delta t) \). We then set \( h = (\Delta t)/N \) and choose the grid spacing \( k \) and domain truncation \( M \) in accordance with our convergence results. We initialize by using (8) with \( C = s_{j-1} \); evaluating this on the spatial grid \( \{mk\}_{m=-M}^M \) yields a vector \( \mathbf{p}^1 \). We form the propagator matrix—denoted by \( A \) above—using the parameter vector \( \Theta \) and the functional forms of \( f \) and \( g \). We then apply (76) from \( n = 1 \) to \( n = N - 2 \), producing \( \mathbf{p}^{N-1} \). We do not go all the way to \( n = N - 1 \) because we are not guaranteed that the next data point \( s_j \) belongs to our spatial grid. To avoid interpolation, we directly evaluate the density at \( s_j \); we use (75) with \( x_i \) replaced by the value \( s_j \) and \( n \) set to \( N - 2 \). In this way, we obtain the DTQ approximation of \( p(s_j|s_{j-1}, \Theta) \).

Of course, one is free to choose other methods to approximate \( p(s_j|s_{j-1}, \Theta) \). Here we compare likelihoods computed by DTQ against those computed using the closed-form Hermite approximation method—for more details on how this method works, consult Ait-Sahalia (2002). For each comparison, we first generate simulated data \((t, s)\) by stepping the SDE forward in time using the Euler-Maruyama method. We restrict \( \Theta \) to be two-dimensional, enabling us to compare contour plots of the likelihood surfaces.
**Ornstein-Uhlenbeck.** Consider the SDE

\[ dX_t = \Theta_1(\Theta_2 - X_t) \, dt + \Theta_3 dW_t. \]  

(80)

We fix \( \Theta = (1, 4, 0.25) \) and \( X_0 = 0 \). We apply the Euler-Maruyama method to step forward in time until \( T = 20 \). While time stepping, we use an Euler-Maruyama step size of \( 10^{-3} \), but only save the solution at multiples of \( \Delta t = 0.1 \). Hence we obtain a time series \((t, s)\) of length \( J + 1 = 201 \).

Because the exact transition density of (80) can be computed by hand, we can compute the exact likelihood. We compute the likelihood using the closed-form Hermite function method with parameters \( J = 4 \) and \( K = 3 \). We also apply DTQ to compute the likelihood with, in turn, \( N = 10 \) (\( h = 0.01 \)) and \( N = 20 \) (\( h = 0.005 \)). In both cases, we use the DTQ parameters \( k = h^{0.75} \) and \( M = [\pi k^{-1.5}] \).

For each method, we compute the log likelihood \( \log p(s|\Theta) \) on a \( 50 \times 50 \) equispaced grid of points in the rectangle \([1, 7] \times [0.125, 0.5] \subset \mathbb{R}^2 \). In Figure 3 we show for each method a contour plot of the resulting log likelihood. Successive contours differ by 100 units of log likelihood. For each plot, we have labeled the grid point that maximizes the log likelihood.

What we take away from Figure 3 is that both the closed-form and DTQ method are in close agreement regarding the overall shape of the likelihood surface. Moreover, the maximum likelihood estimates that we would obtain from either method are both close to one another and close to the ground truth value of \((4, 0.25)\). Note that the exact transition density of (80) is Gaussian, and hence ought to be well-represented using either Hermite function expansion or using a DTQ method that features a Gaussian kernel.

**Bistable SDE.** Let us now turn to the nonlinear SDE with drift and diffusion functions

\[ f(x, \Theta) = \Theta_1 x (\Theta_2 - x^2), \]  

(81a)

\[ g(x, \Theta) = \frac{1}{4} \sqrt{1 + 9x^2}. \]  

(81b)

Assume that \( \Theta_1, \Theta_2 > 0 \). The corresponding deterministic ODE \( dX_t = f(X_t, \Theta)dt \) has an unstable fixed point at 0 and stable fixed points at \( \pm \sqrt{\Theta_2} \). The stochastic model, with both \( f \) and \( g \) present, combines this bistability with state-dependent diffusion. Because of the bistability, an initial condition of \( X_0 = 0 \) leads to a density that eventually becomes bimodal.

To generate data from this model, we fix the parameters \((\Theta_1, \Theta_2) = (1, 8)\) and \( X_0 = 0 \). We apply precisely the same procedure as in the Ornstein-Uhlenbeck case above, but this time we generate two trajectories, one that moves from \( X_0 \) toward \( +\sqrt{\Theta_2} \), and another that moves from \( X_0 \) toward \( -\sqrt{\Theta_2} \). We treat both trajectories as independent. Together, these trajectories form our data \((t, s)\) for this test.

For the bistable model, we do not have an exact transition density. Hence we compare the likelihood surface computed using DTQ against that computed using the closed-form Hermite function method. For DTQ, we use the parameter \( N = 10 \) (\( h = 0.01 \)) with \( k = h^{0.75} \) but \( M = [0.25 \pi k^{-1.5}] \). For the closed-form method, we use \((J, K) \in \{(4, 3), (8, 4), (24, 4)\}\). These correspond to successively higher-order series approximations.

For each method, we compute the log likelihood \( \log p(s|\Theta) \) on a \( 50 \times 50 \) equispaced grid of points in the rectangle \([-1, 4] \times [5, 12.5] \subset \mathbb{R}^2 \). In Figure 4 we show for each method a contour plot of the resulting log likelihood. Successive contours differ by 100 units of log likelihood, except for the plot in the lower-right, for which successive contours differ by 1000 units.

Examining Figure 4, we see that the likelihood surfaces produced by the Hermite function method are unreasonable. They all reach their maxima along the boundaries of the plot—note that this situation does not improve if we recompute and replot on a larger domain. As we increase the order of
approximation, we do see the emergence of a mode that eventually reaches \((\Theta_1, \Theta_2) \approx (1, 11)\) when \(J = 24\). We speculate that with sufficiently many terms, this mode would move closer to the ground truth. However, we were not successful in computing likelihood surfaces with this method for large values of \(J \geq 32\) in less than 24 hours of wall clock time.

The DTQ method’s likelihood surface, in contrast, reaches a maximum at \((1.14, 8.1)\), not far from ground truth. This surface takes less than one minute to generate on a modern desktop computer. Importantly for the purposes of maximum likelihood estimation (MLE), we see that the likelihood surface appears to be convex in a neighborhood of the maximum. For this example, repeated numerical tests using quasi-Newton optimizers initialized at random points in the domain show rapid convergence to the MLE.

Our hypothesis is that the closed-form method will encounter difficulty approximating multimodal densities, especially if we restrict attention to Hermite approximations with a small number of terms. While we expect this can be remedied with a priori knowledge of the shape of the transition density, if no such knowledge is available, we believe the DTQ method is competitive. DTQ makes no prior assumptions on the shape of the density.

Note that the drift function \(f\) in (81a) does not satisfy the hypotheses of our convergence theory. However, this is not essential; we can multiply the drift by \(\exp(-x^2/Z)\) for sufficiently large \(Z\), changing practically nothing about the numerical values of \(f\) actually encountered in this test while reshaping \(f\) to be admissible. As for \(g\) in (81b), we have discussed above the fact that the convergence theory still holds—see (65).

8 Conclusion and Future Directions

In this paper, we have established fundamental properties of the DTQ method, including theoretical and empirical convergence results. The present research motivates four main questions that we seek to answer in future work. Before getting into these questions, let us make three concluding remarks regarding our results.

First, we have not yet mentioned that DTQ features two properties that are not always easy to establish for numerical methods for the Fokker-Planck equation (2): (i) DTQ automatically preserves the nonnegativity of the computed density \(\hat{p}\), and (ii) the DTQ density \(\hat{p}\) has a normalization constant that can be estimated for finite \(h, k > 0\). In practice, we find that \(\hat{p}\) is very close to being correctly normalized.

Second, \(p(x, T)\) and \(\hat{p}(x, t_N)\) correspond to, respectively, the random variables \(X_T\) and \(x_N\). Convergence in \(L^1\) of \(\hat{p}\) to \(p\) is equivalent to convergence in total variation of \(x_N\) to \(X_T\). Note that

\[
\int_{x \in \mathbb{R}} \hat{p}(x, t_{n+1}) \, dx = k \sum_{j=-\infty}^{\infty} \hat{p}(y_j, t_n) = kK_n, \tag{82}
\]

implying that \(\hat{q}(x, t_{n+1}) = \hat{p}(x, t_{n+1})/(kK_n)\) is the density of a continuous random variable \(y_n\). An easy consequence of our results is that \(\hat{q}\) converges to \(\hat{p}\) in \(L^1\), implying convergence of \(y_N\) to \(x_N\) in total variation.

Third, if we trace back the crux of our convergence proof, a key step is estimating the \(L^1\) error of \(\tau\) starting from the trapezoidal rule error estimate (41). To do this, it was essential that we have an estimate of \(N\) that is an \(L^1\) function of \(x\). It was to obtain such an estimate that we put our efforts into Lemma[5]. We have tried to replicate this analysis using more conventional error estimates for the
trapezoidal rule—estimates that require less regularity of the integrand than we have assumed. Thus far, these other attempts have failed because they do not yield an upper bound on $\tau$ that is itself an $L^1$ function of $x$. The approach in the present work is the only one that we have gotten to work.

As mentioned above, there are four main questions that the present work motivates. These questions are the subject of future work:

1. When we derived the DTQ method, we used three approximations: (i) an Euler-Maruyama approximation of the original SDE, (ii) a trapezoidal quadrature rule, and (iii) a finite dimensionalization of $\tilde{p}$ that consists of sampling the function on a truncated grid. The first question to ask is: what happens to the DTQ method if we improve upon these initial approximations?

Regarding (ii), we can say that we have written a test code in which we use Gauss-Hermite quadrature instead of the trapezoidal rule. This does not yield better convergence. Given the exponential convergence of $\hat{p}$ to $\tilde{p}$ established here, this should not be a surprise.

Regarding (iii), rather than sampling the function $\tilde{p}(x, t_n)$ on a discrete grid, we could have instead chosen to represent $\tilde{p}(x, t_n)$ as a linear combination of functions—for instance, a linear combination of Gaussian densities, where each density is centered at a grid point $x_j$. In a collocation scheme, we would then insert these approximations of $\tilde{p}$ into (6) and enforce equality at a finite number of points. We have tried this as well in a test code. While such a scheme does not yield better numerical behavior, it may be easier to analyze.

If we had to choose one approximation (among (i), (ii), or (iii)) to target, we would choose (i). Suppose we replace the Euler-Maruyama method with a higher-order method. The higher-order method then induces a new conditional density function $\tilde{p}_{n+1|n}$ that replaces the Gaussian kernel $G$. Using this new $\tilde{p}_{n+1|n}$ in place of $G$, the evolution equation (45) for $\hat{p}$ remains the same. Preliminary results with the weak trapezoidal method (Anderson and Mattingly 2011) indicate that, in this way, we can obtain a version of the DTQ method that features $O(h^2)$ convergence of $\hat{p}$ to $p$. Note that if we instead retain approximation (i) and replace (ii) and/or (iii), we will be stuck with the $O(h)$ convergence rate of $\hat{p}$ to $p$, thereby blocking improvements to the overall convergence rate of $\hat{p}$ to $p$.

2. Can we patch DTQ to handle diffusion functions $g$ that equal zero at, say, a finite number of discrete points in the computational domain? We believe there should be some way of doing this by subtracting out singularities of $G$ inside the Chapman-Kolmogorov equation (7).

3. Can we derive DTQ-like methods for stochastic differential equations driven by stochastic processes other than the Wiener process? In ongoing work, we are studying how to derive such methods to solve for the density in the case when we replace $dW_t$ by a process whose increments follow a Lévy $\alpha$-stable distribution. For such an SDE, current methods for computing the density involve numerical solution of a fractional Fokker-Planck equation. We expect DTQ-like methods to be highly competitive for such problems.

4. How can we further apply DTQ to problems of statistical inference? In a typical inference problem, we seek to use data to infer parameters in the drift and diffusion functions. In preliminary work (Bhat and Madushani 2016, Bhat et al. 2016), we have shown how DTQ can be used to efficiently compute two quantities that are important for inference: the likelihood function and its gradient with respect to the parameters. Further improvements to and generalizations of DTQ, as described above, will yield improved inference algorithms.
Acknowledgements

H.S.B. acknowledges support from the US National Science Foundation (NSF), both through DMS-1723272 and through computational time on the MERCED cluster (ACI-1429783). H.S.B. and R.W.M.A.M. acknowledge support for this work from UC Merced, through a UC Merced Committee on Research grant, USAP summer fellowships, and Applied Mathematics Graduate Group travel grants. Both authors are grateful for the careful and detailed anonymous review of this paper.

References

Aït-Sahalia Y (2002) Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation. Econometrica 70(1):223–262

Aït-Sahalia Y (2008) Closed-form likelihood expansions for multivariate diffusions. Annals of Statistics 36(2):906–937

Anderson DF, Mattingly JC (2011) A weak trapezoidal method for a class of stochastic differential equations. Communications in the Mathematical Sciences 9(1):301–318

Bally V, Talay D (1996) The law of the Euler scheme for stochastic differential equations. II. Convergence rate of the density. Monte Carlo Methods and Applications 2(2):93–128

Bates D, Maechler M (2016) Matrix: Sparse and Dense Matrix Classes and Methods. URL https://CRAN.R-project.org/package=Matrix, r package version 1.2-4

Bhat HS (2014) Algorithms for linear stochastic delay differential equations. In: Melas VB, Mignani S, Monari P, Salmaso L (eds) Topics in Statistical Simulation, Springer Proceedings in Mathematics & Statistics, vol 114, Springer New York, pp 57–65

Bhat HS, Kumar N (2012) Spectral solution of delayed random walks. Phys Rev E 86:045,701

Bhat HS, Madushani RWMA (2015) Computing the density function for a nonlinear stochastic delay system. IFAC-PapersOnLine 48(12):316–321, 12th IFAC Workshop on Time Delay Systems (TDS 2015), Ann Arbor, Michigan, USA, 28-30 June 2015

Bhat HS, Madushani RWMA (2016) Nonparametric adjoint-based inference for stochastic differential equations. In: Proceedings of the 3rd IEEE International Conference on Data Science and Advanced Analytics. pp 798–807

Bhat HS, Madushani RWMA, Rawat S (2016) Scalable SDE filtering and inference with Apache Spark. Journal of Machine Learning Research: Workshop and Conference Proceedings 53:18–34

Cai Y (2003) Convergence theory of a numerical method for solving the Chapman-Kolmogorov equation. SIAM Journal on Numerical Analysis 40(6):2337–2351

Canor T, Denoël V (2013) Transient Fokker-Planck-Kolmogorov equation solved with smoothed particle hydrodynamics method. International Journal for Numerical Methods in Engineering 94:535–553
Chen L, Jakobsen ER, Naess A (2017) On numerical density approximations of solutions of sdes with unbounded coefficients. Advances in Computational Mathematics URL https://doi.org/10.1007/s10444-017-9558-4

Eddelbuettel D (2013) Seamless R and C++ Integration with Rcpp. Springer, New York

Eddelbuettel D, François R (2011) Rcpp: Seamless R and C++ integration. Journal of Statistical Software 40(8):1–18

Eddelbuettel D, Sanderson C (2014) RcppArmadillo: Accelerating R with high-performance C++ linear algebra. Computational Statistics and Data Analysis 71:1054–1063

Fuchs C (2013) Inference for Diffusion Processes: With Applications in Life Sciences. Springer, Berlin

Giles MB, Nagapetyan T, Ritter K (2015) Multilevel Monte Carlo approximation of distribution functions and densities. SIAM/ASA J Uncertainty Quantification 3(1):267–295

Gobet E, Labart C (2008) Sharp estimates for the convergence of the density of the Euler scheme in small time. Electronic Communications in Probability 13:352–363

Hu Y, Watanabe S (1996) Donsker delta functions and approximations of heat kernels by the time discretization method. J Math Kyoto Univ 36:494–518

Hurn AS, Jeisman JI, Lindsay KA (2007) Seeing the wood for the trees: A critical evaluation of methods to estimate the parameters of stochastic differential equations. Journal of Financial Econometrics 5(3):390–455

Kohatsu-Higa A (1997) High order Ito-Taylor approximations to heat kernels. J Math Kyoto Univ 37:129–150

Kou SC, Olding BP, Lysy M, Liu JS (2012) A multiresolution method for parameter estimation of diffusion processes. Journal of the American Statistical Association 107(500):1558–1574

Kushner HJ (1974) On the weak convergence of interpolated Markov chains to a diffusion. Ann Probability 2:40–50

Linetsky V (1997) The path integral approach to financial modeling and options pricing. Computational Economics 11(1):129–163

Lund J, Bowers KL (1992) Sinc methods for quadrature and differential equations. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA

Luo X, Yau SST (2013) Hermite spectral method to 1D forward Kolmogorov equation and its application to nonlinear filtering problems. IEEE Transactions on Automatic Control 58(10):2495–2507

Milstein GN, Schoenmakers JGM, Spokoiny V (2004) Transition density estimation for stochastic differential equations via forward-reverse representations. Bernoulli 10(2):281–312

Naess A, Johnsen JM (1993) Response statistics of nonlinear, compliant offshore structures by the path integral solution method. Probabilistic Engineering Mechanics 8(2):91–106
Paola MD, Sofi A (2002) Approximate solution of the Fokker-Planck-Kolmogorov equation. Probabilistic Engineering Mechanics 17:369–384

Pedersen AR (1995) A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. Scandinavian Journal of Statistics 22(1):55–71

Pichler L, Masud A, Bergman LA (2013) Numerical solution of the Fokker-Planck equation by finite differences and finite element methods—a comparative study. In: Computational Methods in Stochastic Dynamics, vol 2, Springer, pp 69–85

Rogers LCG (1985) Smooth transition densities for one-dimensional diffusions. Bull London Math Soc 17(2):157–161

Rosa-Clot M, Taddei S (2002) A path integral approach to derivative security pricing II: Numerical methods. International Journal of Theoretical and Applied Finance 05(02):123–146

Sanderson C, Curtin R (2016) Armadillo: a template-based C++ library for linear algebra. Journal of Open Source Software 1:26

Santa-Clara P (1997) Simulated likelihood estimation of diffusions with an application to the short term interest rate. Tech. Rep. 12-97, Anderson School of Management, UCLA, Los Angeles, California

Skaug C, Naess A (2007) Fast and accurate pricing of discretely monitored barrier options by numerical path integration. Computational Economics 30(2):143–151

Stenger F (2012) Numerical Methods Based on Sinc and Analytic Functions. Springer Series in Computational Mathematics, Springer, New York

Trefethen LN, Weideman JAC (2014) The exponentially convergent trapezoidal rule. SIAM Review 56(3):385–458

Wehner MF, Wolfer WG (1983) Numerical evaluation of path-integral solutions to Fokker-Planck equations II. Restricted stochastic processes. Physical Review A 28(5):3003–3011

Yu JS, Cai GQ, Lin YK (1997) A new path integration procedure based on Gauss-Legendre scheme. International Journal of Non-Linear Mechanics 32(4):759–768
Figure 1: For each of the six examples in (64), we test the DTQ method’s convergence. For each example, we plot errors between DTQ and exact solutions on log-scaled axes as a function of $h$, the temporal step size; all other parameters are given by (62). We compute errors in each of the three norms given by (63). The horizontal axes (labels and tick mark locations) are the same for all plots and correspond to the $h$ values in (61). Least-squares fits to the $L^1$ error data are indicated by black lines and corresponding slope values. For all examples, we observe first-order convergence, consistent with our $O(h)$ theoretical result.
Figure 2: For a particular SDE, Example (64a), suppose we are interested in computing the density \( p(x, T) \) at time \( T = 1 \). When we compute this density, we will incur some error, measured here in the \( L^1 \) norm. The plotted results show that for a fixed value of this error, the DTQ methods require less computational time (measured in wall clock seconds) than a method for numerically solving the Fokker-Planck PDE. In all simulations, we use a domain \([-y_M, y_M]\). For the simulations in the left (respectively, right) plot, we have scaled the domain according to \( y_M \propto h^{-3/4} \) (respectively, \( y_M \propto \log h^{-1} \)), where \( h > 0 \) is the time step. In both plots, we see that for smaller values of the error, the fastest method is DTQ-Sparse; for larger values of the error, the fastest method is DTQ-Naïve. In particular, for the smallest error of 0.003, DTQ-Sparse is over \( 10^2 \) (respectively, \( 10^{3/4} \)) times faster than the Fokker-Planck method in the left (respectively, right) plot. Despite the fact that our Fokker-Planck solver uses the same sparse numerical linear algebra as DTQ-Sparse, it is often the slowest of the four methods. For details regarding the three implementations of DTQ (DTQ-Naïve, DTQ-CPP, and DTQ-Sparse) as well as the implementation of our Fokker-Planck solver, please see Section 7.2.
Figure 3: Contour plots of the log likelihood $\log p(s|\Theta)$ computed using four different methods: the exact transition density (upper-left), the closed-form Hermite function method (upper-right), DTQ with $h = 0.01$ (lower-left) and DTQ with $h = 0.005$ (lower-right). All four plots are in qualitative agreement, and the maxima (labeled by black points) are all near each other. The data for this test was simulated using the Ornstein-Uhlenbeck SDE (80) with ground truth parameters $\Theta_2 = 4$, $\Theta_3 = 0.25$, and $\Theta_1 = 1$—this last parameter is kept constant in the results plotted here.
Figure 4: Contour plots of the log likelihood $\log p(s|\Theta)$ computed using four different methods: DTQ with $h = 0.01$ (upper-left) and the closed-form Hermite function method with $(J, K) = (4, 3)$ (upper-right), $(8, 4)$ (lower-left), and $(24, 4)$ (lower-right). The data for this test was simulated using the bistable SDE with drift and diffusion given by (81) and ground truth parameters $\Theta_1 = 1$ and $\Theta_2 = 8$. We find that the DTQ method produces a reasonable likelihood surface with a maximizer close to ground truth. The likelihood surfaces generated by the Hermite function method all reach their maxima on the boundary. One can see, however, that as the order of approximation increases, the Hermite function method’s likelihood surface gives rise to a mode that starts at $(\Theta_1, \Theta_2) \approx (2.5, 10.5)$ (at $J = 8$, lower-left) and moves to $\approx (1, 11)$ (at $J = 24$, lower-right). We speculate that with sufficiently many terms, this mode would move closer to the ground truth. We hypothesize that the Hermite function method has difficulty with this problem because of the bimodal nature of the underlying density.