# 1 Introduction

We consider two-dimensional correlated Brownian motion with absorbing boundaries:

$$X(t) = x_0 + \mu_x t + \sigma_x W_x(t) \qquad a_x \le X(t) \le b_x \tag{1}$$
$$Y(t) = y_0 + \mu_y t + \sigma_y W_y(t) \qquad a_y \le Y(t) \le b_y \tag{2}$$

where $W_i$ are standard Brownian motions with $\mathrm{Cov}(W_1(t), W_2(t)) = \rho t$. We find the below probability for $(X(t), Y(t))$ and the random variables $M_X(t) = \max_{0\le s\le t} X(t)$, $m_X(t) = \min_{0\le s\le t} X(t)$, $M_Y(t) = \max_{0\le s\le t} Y(t)$, $m_Y(t) = \min_{0\le s\le t} Y(t)$

$$\Pr\left( \begin{array}{ll} X(t) \le x_t, & Y(t) \le y_t, \\ M_X(t) \le b_x, & M_Y(t) \le b_x, \\ m_X(t) \ge a_x, & m_Y(t) \ge a_y, \end{array} \middle| X(0) = x_0, Y(0) = y_0, \theta \right) \tag{3}$$

with $\theta := (\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$. This probability can be expressed as the integral

$$\int_{a_x}^{x_t} \int_{a_y}^{y_t} q(x,y,t)\,dx\,dy,$$

where $q(x,y,t)$ is the solution to the Fokker-Planck equation [Oksendal, 2013]:

$$\frac{\partial}{\partial t'} q(x,y,t') = -\mu_x \frac{\partial}{\partial x} q(x,y,t') - \mu_y \frac{\partial}{\partial y} q(x,y,t') + \frac{1}{2}\sigma_x^2 \frac{\partial^2}{\partial x^2} q(x,y,t') + \rho\sigma_x\sigma_y \frac{\partial^2}{\partial x \partial y} q(x,y,t') + \frac{1}{2}\sigma_y^2 \frac{\partial^2}{\partial y^2} q(x,y,t'), \tag{4}$$

$$q(a_x,y,t') = q(b_x,y,t') = q(x,a_y,t') = q(x,b_y,t') = 0, \qquad 0 < t' \le t. \tag{5}$$

Differentiating $q(x,y,t)$ with respect to the boundaries produces the transition density of a particle beginning and ending at the points $(X(0), Y(0))$ and $(X(t), Y(t))$, respectively, while attaining the minima $a_x/a_y$ and maxima $b_x/b_y$ in each coordinate direction. This transition density is equivalent to the joint density for $(X(t), Y(t), m_X(t), M_X(t), m_Y(t), M_Y(t)) | X(0) = x_0, Y(0) = y_0$, which we denote as $f(x,y,a_x,b_x,a_y,b_y)$:

$$\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} q(x,y,t) = f(x,y,a_x,b_x,a_y,b_y). \tag{6}$$

Probability functions similar to (3) with less than all four boundaries have been used in computing first passage times [Kou et al., 2016, Sacerdote et al., 2016], with application to structural models in credit risk and default correlations [Haworth et al., 2008, Ching et al., 2014]. He et al. [1998] use variants of (6) with respect to some of the boundaries to price financial derivative instruments whose payoff depends on **some** of the observed maxima/minima. Horst et al. [2012] use a full likelihood-based (Bayesian) approach to estimate volatility in *univariate* financial timeseries where open, closing, highest, and lowest prices are included. Their work fits into a body of literature and collection of techniques by practitioners where the observed range of prices is used to make similar estimates. In this paper we will provide the efficient numerical method necessary for carrying out inferential procedures with correlated financial timeseries.

Closed-form solutions to (4) - (5) are available for some parameter regimes. When $\rho = 0$, the transition density of the process is the solution to a well-understood Sturm-Liouville problem where the eigenfunctions of the differential operator are sine functions. When $a_x = -\infty$ and $b_x = \infty$, the method of images can be used to enforce the remaining boundaries. For either $a_x, a_y = -\infty$ or $b_x, b_y = \infty$, eigenfunction of the Fokker-Plank equation can be found in radial coordinates. Both of these techniques are used and detailed by He et al. [1998]. However, to the best of our knowledge, there is no closed-form solution to the general problem in (4) - (5). This also limits the available ways to compute (6), with the most straightforward approach being finite difference with respect to the boundary conditions. This, however, requires one to solve Fokker-Planck equation (4) - (5) accurately for at least 16 slightly different sets of boundaries and combine

the results with numerical differentiation to evaluate the density function for a just single observation (see Section 2.1), motivating the need for an efficient numerical method to solve (4) - (5).

For non-zero correlation, it is still possible to approach the general problem by proposing a biorthogonal expansion in time and space (Risken [1989], sections 6.2), where the eigenfunctions for the differential operator are approximated by the eigenvectors of a linear system obtained using a truncated expansion based on a set of separable basis function, each of which is a product of two sine functions (one in each dimension) satisfying the boundary conditions. However, a drawback of this out-of-the-box solution is that the system matrix for the corresponding eigenvalue problem is dense. Additionally, for nonzero $\rho$, it is necessary to have a large number of basis elements for an accurate approximation. The denseness and size of the resultant system matrices makes the expansion a slow, if not unfeasible, solution. An alternative is to use a finite difference scheme to directly solve the evolution problem after suitable transformations. However, both of these methods need a high degree of numerical resolution to produce practically useful approximations of the transition density. We conjecture that for the trigonometric series expansion, inefficiencies arise from using a separable representation for a differential operator that is intrinsically correlated in the two dimensions; for the finite difference approach, the problem is twofold as the method introduces numerical diffusion in the initial condition and it also uses a separable approximation for the differential operator.

In this paper, we propose a robust and efficient solution to the problems (4) - (5) and (6). The solution is obtained by combining a small-time analytic solution with a Galerkin discretization based on basis functions that are correlated. Our method directly takes into account the correlation parameter present in the differential operator in order to efficiently represent the analytic small-time solution and propagate it forward in time. We apply our computational method to estimate equation parameters with a maximum likelihood approach in settings where the model assumptions of constant $(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$ and Brownian motion driving stochastic evolution are appropriate. Section 2 outlines some methods we considered for this problem. Section 3 describes the Galerkin method we implemented. Section [] includes our numerical experiments.

## 2 Approximate Numerical Solutions

Before considering solutions to the full Fokker-Planck equation (4) - (5), we simplify the PDE by proposing a scaling transformation and an exponential decomposition of the solution, so that we can construct

$$p(x,y,t) = \exp(\alpha x + \beta y + \gamma t) q\left(x(b_x - a_x) + a_x,\ y(b_y - a_y) + a_y,\ t\right), \tag{7}$$

where

$$\begin{aligned}
\alpha &= -\frac{\mu_x}{\sigma_x^2} - \frac{\rho}{\sigma_x \sigma_y (1-\rho^2)} \left( -\frac{\mu_y}{\sigma_y^2} + \frac{\mu_x \rho}{\sigma_x \sigma_y} \right), \\
\beta &= \left( -\frac{\mu_y}{\sigma_y^2} + \frac{\mu_x \rho}{\sigma_x \sigma_y} \right), \\
\gamma &= \frac{1}{2} \left( \frac{\sigma_x}{(b_x - a_x)} \right)^2 \alpha^2 + \frac{1}{2} \left( \frac{\sigma_y}{(b_y - a_y)} \right)^2 \beta^2 + \alpha\beta.
\end{aligned}$$

This new formula satisfies the *normalized* diffusion equation:

$$\frac{\partial}{\partial t} p(x,y,t) = \mathcal{L} p(x,y,t), \quad (x,y) \in (0,1) \times (0,1) := \Omega \tag{8}$$

subject to the constraints

$$p(x,y,t) = 0, \qquad \text{for } (x,y) \in \{x|x=0\} \cup \{y|y=0\},$$
$$p(x,y,0) = \delta\left(x - \frac{x_0 - a_x}{b_x - a_x}\right)\delta\left(y - \frac{y_0 - a_y}{b_y - a_y}\right),$$

where the differential operator $\mathcal{L}$ takes the form

$$\mathcal{L} = \frac{1}{2}\tau_x^2\frac{\partial^2}{\partial x^2} + \rho\tau_x\tau_y\frac{\partial^2}{\partial x \partial y} + \frac{1}{2}\tau_y^2\frac{\partial^2}{\partial y^2},$$

with $\tau_x = \frac{\sigma_x}{(b_x - a_x)}$, $\tau_y = \frac{\sigma_y}{(b_y - a_y)}$. Note here that under this transformation $\rho$ remains the same as in the original problem. Without loss of generality, we will concentrate on the solution of the normalized problem (8) in terms of the diffusion parameters $(\tau_x, \tau_y, \rho)$.

## 2.1 Density Calculation

Recalling problem (6), to compute the density we are interested in, we must take derivatives of the solution $p(x,y,t)$ with respect to the boundary parameters $(a_x, b_x, a_y, b_y)$. Because the approximate solutions we will consider below are functions of the parameters without explicit analytic form, $\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} p(x,y,t)$ must be computed numerically using a finite difference approximation. We must therefore find the solution for each of the sixteen perturbed problems

$$p(x,y,t|a_x \pm \varepsilon, b_x \pm \varepsilon, a_y \pm \varepsilon, b_y \pm \varepsilon).$$

The derivative of $p$ with respect to the 4 boundaries is approximated as

$$\begin{aligned}&\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} p(x,y,t) \\ &\quad\approx \frac{\sum_{k_1,k_2,k_3,k_4=\pm 1} c_{\{k_1,k_2,k_3,k_4\}}\, p(x,y,t|\, a_x + k_1\varepsilon, b_x + k_2\varepsilon, a_y + k_3\varepsilon, b_y + k_4\varepsilon)}{(2\varepsilon)^4}\end{aligned}$$

The finite difference approximation, however, introduces fundamental limitations for the type of approximation that can be used and the degree of accuracy admissible for this problem.

Consider an approximate solution $p^{(k)}(x,y,t|\,b)$ where we have included parameter $b$ explicitly as a simplified notation for $[a_x, b_x]$ and $[a_y, b_y]$ and we assume that the approximation approaches the true solution as $k \to \infty$. In general, the truncation error can be represented as

$$p^{(k)}(x,y,t|b) - p(x,y,t|b) = \left(\frac{1}{k}\right)^{\alpha} \mathrm{F}_{reg}(b) + \left(\frac{1}{k}\right)^{\beta} \mathrm{F}_{irreg}(b) + \varepsilon_{mach} F_{round}(b), \tag{9}$$

for some $\alpha, \beta > 0$; $(1/k)^{\alpha}\mathrm{F}_{reg}(b)$ is the regular part of the truncation error, $(1/k)^{\beta}\mathrm{F}_{irreg}(b)$ is the irregular part of the truncation error, and $\varepsilon_{mach}\mathrm{F}_{round}(b)$ is the effect of round-off errors with $\varepsilon_{mach} \sim 10^{-16}$ denoting the machine epsilon for IEEE double precision system.

Note that when expressed using the chain rule, both $\frac{\partial}{\partial a_x}$ and $\frac{\partial}{\partial b_x}$ contain $\frac{\partial}{\partial b}$. As a result, $\frac{\partial^2}{\partial a_x \partial b_x}$ leads to $\frac{\partial^2}{\partial b^2}$. Although in the discussion below, for simplicity, we only illustrate the numerical differentiation on the first derivative, keep in mind that it is the second derivative that is more relevant in the calculation of $\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} p(x,y,t)$.

The coefficient, $\mathrm{F}_{reg}(b)$, of the regular part of truncation error is a smooth function of $b$ with derivative $= O(1)$. The coefficient, $\mathrm{F}_{round}(b)$, in the effect of round-off errors, behaves virtually like a random variable, discontinuous in $b$. For the irregular part of truncation error, the coefficient $\mathrm{F}_{irreg}(b)$ can be thought of in general as continuous in $b$ but not smooth in $b$ where the derivative has discontinuities. Linear interpolation, for example, introduces this kind of irregular truncation error. Based on the expression we wrote out above for the finite difference solution, applying the numerical differentiation with step $\varepsilon$ yields:

$$\begin{aligned}
&\frac{p^{(k)}(x,y,t|\,b+\varepsilon)-p^{(k)}(x,y,t|\,b-\varepsilon)}{2\varepsilon} \\
&\quad=\frac{\partial}{\partial b}p(x,y,t|\,b)+O(\varepsilon^2)+\left(\frac{1}{k}\right)^{\alpha}\frac{\mathrm{F}_{reg}(b+\varepsilon)-\mathrm{F}_{reg}(b-\varepsilon)}{2\varepsilon} \\
&\quad+\left(\frac{1}{k}\right)^{\beta}\frac{\mathrm{F}_{irreg}(b+\varepsilon)-\mathrm{F}_{irreg}(b-\varepsilon)}{2\varepsilon}+\varepsilon_{mach}\frac{\mathrm{F}_{round}(b+\varepsilon)-\mathrm{F}_{round}(b-\varepsilon)}{2\varepsilon}
\end{aligned}$$

In the equation above, as the step in the numerical differentiation is refined, the first line of the RHS is well behaved, converging to the true value $\frac{\partial}{\partial b}p(x,y,t|\,b)$ as $\varepsilon \to 0$. The second line of RHS, however, is problematic. As $\varepsilon \to 0$, the contribution from round-off error blows up to infinity

$$\varepsilon_{mach}\frac{\mathrm{F}_{round}(b+\varepsilon)-\mathrm{F}_{round}(b-\varepsilon)}{2\varepsilon}=O\left(\frac{\varepsilon_{mach}}{\varepsilon}\right)\longrightarrow\infty \quad \text{as } \varepsilon\to 0$$

For a fourth-order derivative, the error contribution of finite machine precision becomes $O(\varepsilon_{mach}/\varepsilon^4)$ and a step size as big as $\varepsilon \sim 10^{-4}$ produces errors $O(1)$. This is catastrophic in instances where values of the density are smaller than 1.

The contribution from the irregular part of truncation error can also be problematic if the finite difference approximation of the derivative of $F_{irreg}(b)$ is $O(k^{\beta}/\varepsilon)$; we will show that this is the case for the finite difference method. As a consequence of both machine precision and possible irregular truncation errors, we observe that there is a fundamental restriction of the size of finite difference steps $\varepsilon$ we can use in the numerical methods below.

## 2.2 Eigenfunction Expansion

Following Section 6.2 of Risken [1989], we may use the biorthogonal decomposition of the solution as a sum of eigenfunctions and time-dependent coefficients determined by eigenvalues:

$$p(x,y,t)=\sum_{\nu}h_{\nu}\phi_{\nu}(x,y)e^{-\lambda_{\nu}t}, \tag{10}$$

where $h_{\nu}$ is the coefficient of $\phi_{\nu}(x,y)$ is the eigenfunction expansion of $p(x,y,0)$. Because the differential operator $\mathcal{L}$ is self-adjoint, the family of eigenfunctions is complete in the Hilbert space $L_2(\Omega)$. Moreover, the eigenvalues are bounded below by a positive constant $c$, so that the solution behaves as expected (see section 6.3 of Risken [1989]).

Since we require $\phi_{\nu}(x,y)$ to be zero on the boundaries, we approximate the eigenfunction using a finite set of orthogonal basis functions satisfying boundary conditions, i.e., a finite sequence of sines,

$$\phi_{\nu}(x,y)=\sum_{l=1}^{L}\sum_{m=1}^{M}c_{l,m,\nu}\sin(2\pi l x)\sin(2\pi m y):=\Psi(x,y)^{T}c_{\nu},$$

where we have truncated the infinite series for some suitably large $L$ and $M$ and defined

$$\begin{aligned}
\psi_{l,m}(x,y)&=\sin(2\pi l x)\sin(2\pi m y),\\
\Psi(x,y)&=(\psi_{1,1}(x,y),\ldots,\psi_{L,M}(x,y))^{T},\\
c_{\nu}&=(c_{1,1,\nu},\ldots,c_{L,M,\nu})^{T}.
\end{aligned}$$

The biorthogonal representation (10) leads to the eigenvalue problem

$$\mathcal{L}\phi_\nu = -\lambda_\nu \phi_\nu, \tag{11}$$

where $\mathcal{L}$ is the differential operator in the normalized Fokker-Planck equation. Applying $L$ to the basis function expansion of $\phi_v$ and again approximating the result using the finite set of basis functions yields

$$\mathcal{L}\phi_\nu = \mathcal{L}(\Psi(x,y)^T c_\nu) = \mathcal{L}(\Psi(x,y)^T)c_\nu = (A\Psi(x,y))^T c_\nu,$$

where $A$ is a matrix analytic in $(\tau_x, \tau_y, \rho)$. In the last part of the equation above, we have truncated the infinite sine series expansion of $L\psi_{l,m}(x,y)$. In the case where $\rho = 0$, $A$ is diagonal because $\{\psi_{l,m}(x,y)\}$ are the eigenfunctions to $\mathcal{L}$. When $\rho \neq 0$, $A$ is no longer diagonal and is in fact dense. Substituting the linear representation of $\mathcal{L}\phi_\nu$ into the eigenvalue problem (11), we arrive at the matrix eigenvalue problem

$$\Psi(x,y)^T A^T c_\nu = -\lambda_\nu \Psi(x,y)^T c_\nu \quad \Leftrightarrow \quad A^T c_\nu = -\lambda_\nu c_\nu$$

whose solution gives the family of orthonormal eigenfunctions. As mentioned already, the efficiency of this approach is dependent on the cost of solving the eigenvalue problem $A^T c_\nu = -\lambda_\nu c_\nu$. With all of the eigenpairs $(c_\nu, \lambda_\nu)$, the approximate solution is then

$$p(x,y,t) \approx \Psi(x,y)^T \sum_\nu h_\nu c_\nu e^{-\lambda_\nu t}.$$

In terms of density calculations for $\frac{\partial^4 p(x,y,t)}{\partial a_x \partial b_x, \partial a_y \partial b_y}$, the eigenfunction expansion does not suffer from having irregular truncation errors, because $A$ is analytic in the diffusion parameters and the eigenpairs $(\lambda_\nu, c_\nu)$ inherit this property (see Theorems 7 and 8 of Chapter 9 in Lax [2007]). However, because a good working approximation of $p(x,y,t)$ requires many terms in the expansion, and because each resultant system matrix is dense, repeated computation of the eigenproblem for the density calculation is unfeasible.

## 2.3 Finite Difference

A finite difference method which approximates the spatial derivatives in problem (8) requires the solution a system of differential equations

$$\dot{c}(t) = Bc(t) \Rightarrow \qquad\qquad c(t) = \exp(Bt)\, c(0), \tag{12}$$

which reduces to the eigenvalue decomposition of a matrix $B$. Here, $c(t)$ is a vector consisting of values of the solution in (8) on a set of grid points over $\Omega$ at time $t$, and the product $Bc(t)$ approximates $\mathcal{L}p(x,y,t)$. The system matrix $B$ is dependent on the discretization scheme used to approximate $\mathcal{L}$. Using a central-in-space scheme over a **regular** grid on $\Omega$ with $\Delta x = \Delta y = h$,

$$B = \frac{1}{2}\tau_x^2 \frac{1}{h^2} B_{x,x} + \rho\tau_x\tau_y \frac{1}{4h^2} B_{x,y} + \frac{1}{2}\tau_y^2 \frac{1}{h^2} B_{y,y},$$

where each of the matrices $B_{x,x}, B_{x,y}, B_{y,y}$ approximate $\frac{\partial^2}{\partial x^2}, \frac{\partial^2}{\partial x \partial y}, \frac{\partial^2}{\partial y^2}$, respectively. It should be noted here that a regular grid approach with a constant $h$ independent of parameters is appealing, because it allows us to construct once and store the matrices $B_{x,x}, B_{x,y}, B_{y,y}$, which saves valuable computational resources if we are to solve the finite difference eigenproblem (12) repeatedly for different parameter values $(\tau_x, \tau_y, \rho)$.

Unlike the system matrix for the trigonometric expansion, the system matrix $B$ here is sparse: each row of $(B_{x,x}, B_{x,y}, B_{y,y})$ is composed of all zeros except for three or four entries. This structure does not change as $h \to 0$. The eigenvalue problem is therefore much cheaper to solve.

However, the fundamental limitation of using a finite difference method is that it introduces irregular truncation error, because of the linear interpolation necessary for function arguments not on grid points.

Referring to the notation in (9), and setting $k := 1/h$, the interpolation error introduced by a finite difference scheme at position $b$ for the boundary parameters using step $h$ has the general form of

$$\text{Interpolation error } = O(h^2)(1-\text{rem}(b/h,1))\text{rem}(b/h,1)$$

The coefficient part $\text{F}_{irreg}(b) = (1-\text{rem}(b/h,1))\text{rem}(b/h,1)$ is continuous in $b$ but not differentiable. Its first derivative has the behavior of

$$\frac{\partial}{\partial b}\text{F}_{irreg}(b) = \frac{1}{h}\left(1-2\text{rem}(b/h,1)\right).$$

The contribution from the irregular part of truncation error is

$$h^2\frac{\text{F}_{irreg}(b+\varepsilon)-\text{F}_{irreg}(b-\varepsilon)}{2\varepsilon} = O\left(\frac{h^2}{\max(\varepsilon,h)}\right)$$

In the second order numerical differentiation, however, the contribution from the irregular part of truncation error behaves like

$$h^2\frac{\text{F}_{irreg}(b+\varepsilon)-2\text{F}_{irreg}(b)+\text{F}_{irreg}(b-\varepsilon)}{\varepsilon^2} = O\left(\frac{h^2}{\max(\varepsilon^2,h^2)}\right)$$

The interplay between $h$ and $\varepsilon$ limits the size $\varepsilon$ we can use to perform density calculations for a fixed step size $h$.

# 3 A Novel Semidiscrete Galerkin Method

We propose a semidiscrete Galerkin (continuous in time, discrete in space) solution to the general diffusion problem (8). The method relies on an analytic approximation for the solution $p(x,y,t)$ for small time $t$ and a basic convergence estimate from approximation theory for semidiscrete Galerkin-type solutions to parabolic problems.

Our numerical method produces a functional representation for the approximate solution which 1) imposes a computational burden comparable to or better than that of the finite difference method and 2) is infinitely differentiable with respect to the boundary parameters, allowing us to perform the crucial numerical differentiation with respect to these parameters.

As described in Section 2.2, the solution to the model problem has the eigenfunction expansion

$$p(x,y,t) = \sum_{\nu=0}^{\infty} h_\nu \phi_\nu(x,y)e^{-\lambda_\nu t},$$

where $h_\nu$ is the coefficient of $\phi_\nu(x,y)$ in the eigenfunction expansion of $p(x,y,0)$. Proceeding with the standandard Galerkin approach, we propose a solution $p^{(k)}(x,y,t)$ of similar form

$$p^{(k)}(x,y,t) = \sum_{i=0}^{k} c_i(t)\psi_i(x,y),$$

where the basis functions $\psi_i(x,y)$ satisfy the boundary conditions. We also require that all first- and second-order derivatives of $\psi_i(x,y)$ are in $L_2(\Omega)$, i.e. $\psi_k(x,y) \in W_2^2(\Omega)$ where $\Omega = (0,1)\times(0,1)$ is the domain of the normalized problem. This will allow us to take derivatives of the approximate solution with respect to the boundaries for the problem. Finally, we designate the set of basis functions as

$$S_k := \{\psi_1(x,y),\ldots,\psi_k(x,y)\}$$

Since $p^{(k)}$ is an approximation to the solution $p$, it does not follow the differential equation exactly nor can it represent the initial condition fully. We capture this by defining residuals

$$\frac{\partial}{\partial t}p^{(k)}(x,y,t) - \mathcal{L}p^{(k)}(x,y,t) := R_e(k),$$
$$p(x,y,0) - p^{(k)}(x,y,0) := R_0(k).$$

There are various conditions that could be imposed on the residual functions (see Section 2.10.3 of Norrie and De Vries [1973] for a summary). The *orthogonality* condition coincides with the Galerkin procedure:

$$\int_\Omega R_e(k)\psi_i(x,y)dxdy = 0, \qquad \int_\Omega R_0(k)\psi_i(x,y)dxdy = 0, \qquad i = 0,\dots k, \tag{13}$$

which is equivalent to the weak formulation of the heat problem

$$\left\langle \partial_t p^{(k)}(x,y,t), \psi_i \right\rangle = \left\langle \mathcal{L}p^{(k)}(x,y,t), \psi_i \right\rangle,$$
$$\left\langle p^{(k)}(x,y,0), \psi_i \right\rangle = \left\langle p(x,y,0), \psi_i \right\rangle,$$

where $<\cdot,\cdot>$ is the usual inner product in $L_2(\Omega)$. The orthogonality conditions (13) lead to the system of equations

$$M\dot{\mathbf{c}}(t) = S\mathbf{c}(t), \tag{14}$$
$$M\mathbf{c}(0) = \mathbf{p}(0), \tag{15}$$

where $M$ is the mass matrix, $S$ is the stiffness matrix, and $\mathbf{p}(0)$ is the vector projection of $p(x,y,0)$ onto the span of $S_k$. Matrix $M$, matrix $S$, and vector $\mathbf{p}(\mathbf{0})$ are expressed in elements as:

$$[M]_{ij} = \int_\Omega \psi_i\psi_j dxdy,$$
$$[S]_{ij} = -\frac{1}{2}\tau_x^2\int_\Omega \left(\frac{\partial}{\partial x}\psi_i(x,y)\right)\left(\frac{\partial}{\partial x}\psi_j(x,y)\right)dxdy - \rho\tau_x\tau_y\int_\Omega \left(\frac{\partial}{\partial x}\psi_i(x,y)\right)\left(\frac{\partial}{\partial y}\psi_j(x,y)\right)dxdy$$
$$-\frac{1}{2}\tau_y^2\int_\Omega \left(\frac{\partial}{\partial y}\psi_i(x,y)\right)\left(\frac{\partial}{\partial y}\psi_j(x,y)\right)dxdy$$
$$[\mathbf{p}(0)]_i = \int_\Omega p(x,y,0)\psi_i(x,y)dxdy.$$

The entries of $S_{ij}$ are computed with integration by parts to enforce the boundary conditions. The semidiscrete Galerkin approximation becomes

$$p^{(k)}(x,y,t) = \psi(x,y)^T \exp\left(M^{-1}St\right)\mathbf{c}(0),$$

with $\psi(x,y) = (\psi_0(x,y),\dots,\psi_k(x,y))^T$. Since both $M$ and $S$ are symmetric and infinitely differentiable with respect $(\tau_x,\tau_y,\rho)$ (each entry in the matrices has this property, which can be seen in the defining expressions above), the system matrix $M^{-1}S$ is symmetric, and it is infinitely differentiable with respect to the parameters $(\tau_x,\tau_y,\rho)$ as well. As a consequence, the matrix exponential

$$\exp\left(M^{-1}St\right) = \sum_{n=0}^{\infty}\frac{1}{n!}\left(M^{-1}S\right)^n t^n$$

is also infinitely differentiable with respect to the diffusion parameters. Hence, we can safely apply $\frac{\partial}{\partial a_x \partial b_x \partial a_y \partial b_y}$ to the Gelerkin approximation $p^{(k)}(x,y,t)$ as long as the derivatives of the vectors $\psi(x,y)^T$ and $\mathbf{c}(0)$ exist. This is ensured as long as each $\psi_i(x,y)$ is differentiable with respect to the normalized problem parameters.

### 3.1 Small-time Solution

Before solving the Galerkin equations (14) - (15), we develop a small-time analytic solution to the problem. The small-time solution is derived by considering the fundamental solution $G(x,y,t)$ for the unbounded problem in (8), which is the bivariate Gaussian density with mean and covariance determined by the initial condition and the diffusion parameters [Stakgold and Holst, 2011]. We can then find a small enough $t_\varepsilon$ such that $G\left(x,y,t_\varepsilon \left| \frac{x_0-a_x}{b_x-a_x}, \frac{y_0-a_y}{b_y-a_y} \right.\right)$ is numerically zero on the boundaries of $\Omega$. We select the largest largest among all time instances satisfying this condition. This is done in the following way.

1. Scale and rotate the coordinate axes by

$$\begin{pmatrix} \xi \\ \eta \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{1}{\tau_x\sqrt{1-\rho}} & -\frac{1}{\tau_y\sqrt{1-\rho}} \\ \frac{1}{\tau_x\sqrt{1+\rho}} & \frac{1}{\tau_y\sqrt{1+\rho}} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

   so that the problem obeys the standard diffusion equation

$$\frac{\partial}{\partial t}p(\xi,\eta,t) = \frac{1}{2}\frac{\partial^2}{\partial^2\xi}p(\xi,\eta,t) + \frac{1}{2}\frac{\partial^2}{\partial^2\eta}p(\xi,\eta,t)$$

   on a computational domain now transformed to a parallelogram $\tilde{\Omega}$ (see left panel of Figure (1)). The transformed initial condition will be denoted as $(\xi_0,\eta_0)$.

2. The fundamental solution $G(\xi,\eta,t|\xi_0,\eta_0)$ in this coordinate frame which does not take into account boundaries follows the bivariate Gaussian probability density function

$$G(\xi,\eta,t|\xi_0,\eta_0) = \frac{1}{2\pi t}\exp\left(-\frac{(\xi-\xi_0)^2+(\eta-\eta_0)^2}{2t}\right).$$

   We define the **distance** between $G(\xi,\eta,t|\xi_0,\eta_0)$ and any of the linear boundaries of $\tilde{\Omega}$ as the shortest distance between $(\xi_0,\eta_0)$ to each of the boundary segments (blue line segments in the left panel of Figure (1)). There are four such distances $d_1,d_2,d_3,d_4$; and assume that they are listed in increasing magnitude. Note that the perpendicular intersections may not fall within the line segment when the 4 line segments form a very skewed diamond. However, the shortest distance to a line segment will be well-behaved since we only need to enforce the boundary condition over a single segment.

   Setting $t_\varepsilon = d_2/8$ ensures that the fundamental solution $G(\xi,\eta,t_\varepsilon|\xi_0,\eta_0)$ is *at most* $\approx 10^{-15}$ on the second-farthest boundary, as well as the other two farthest boundaries. In this way, $G(\xi,\eta,t_\varepsilon|xi_0,\eta_0)$ satisfies the boundary condition on the three farthest boundaries numerically.

3. Reflect the point $(\xi_0,\eta_0) \to (\xi_0',\eta_0')$ about the closest boundary. The image function $G(\xi,\eta,t_\varepsilon|\xi_0',\eta_0')$ satisfies the diffusion equation and is equal to $G(\xi,\eta,t_\varepsilon|\xi_0,\eta_0)$ on the closest boundary. Further, $G(\xi,\eta,t_\varepsilon|\xi_0',\eta_0')$ takes on values less than $10^{-15}$ on all other boundaries, because it is outside of $\bar{\Omega}$. For this same reason, the system of images satisfies the initial condition for the problem.

   Considering the difference of the two images, the small-time solution

$$p(\xi,\eta,t_\varepsilon) := G(\xi,\eta,t_\varepsilon|\xi_0,\eta_0) - G(\xi,\eta,t_\varepsilon|\xi_0',\eta_0')$$

   satisfies all of the boundaries numerically and also satisfies the governing diffusion equation. Performing a change of variables produces the small-time solution $p(x,y,t_\varepsilon)$, as shown in the right panel of Figure (1).

Using Theorem 5.E of Zeidler [1995], we can solve for $p(x,y,t)$ by considering the smooth $p(x,y,t_\varepsilon)$ as an initial condition and evolving it forward in time by $t-t_\varepsilon$. This replaces initial condition orthogonality

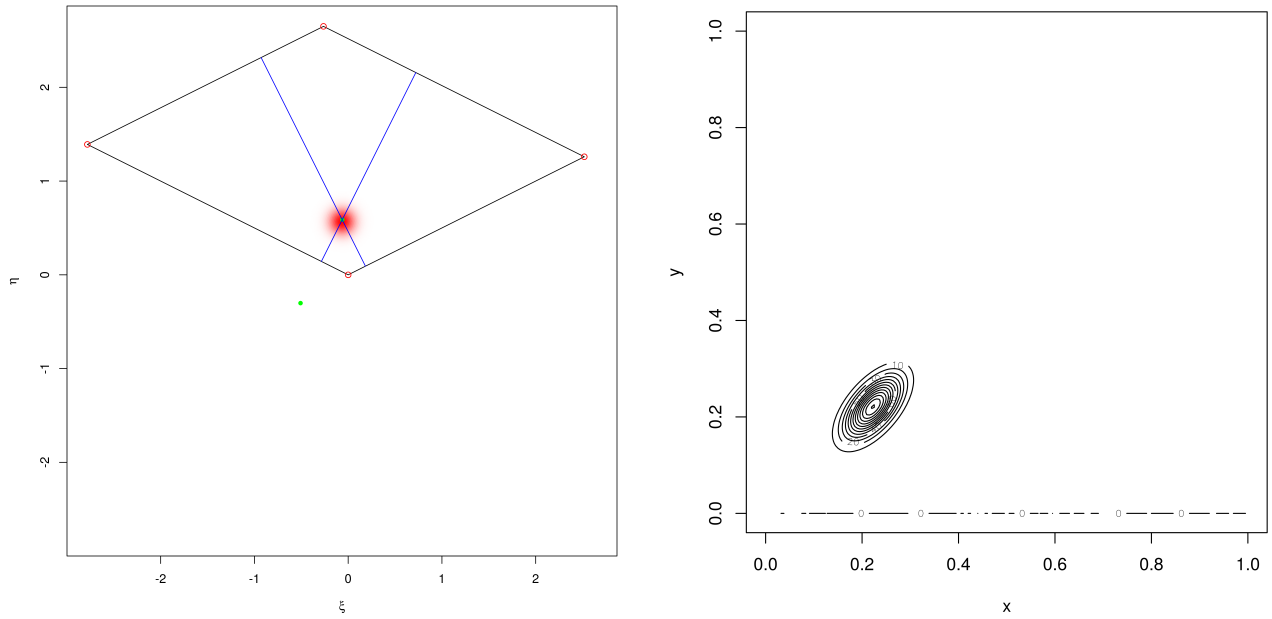


Figure 1: An example of the small-time solution $p(\xi, \eta, t_\varepsilon)$ on the transformed domain $\tilde{\Omega}$ with $\tau_x = \tau_y = 1$ and $\rho = 0.6$. Right: The shaded red region is a heatmap of the small-time solution in the transformed coordinate frame, while the blue line segments represent the distance between the boundaries and the initial condition coordinate. The green point outside of the computational domain is the center of the reflected image $(\xi_0', \eta_0')$ about the closest boundary. Left: The small-time solution transformed back to the original coordinate system. Here, the contours denote the level-sets for the function. They very closely approximate the level sets for the fundamental solution of the unbounded problem (8)

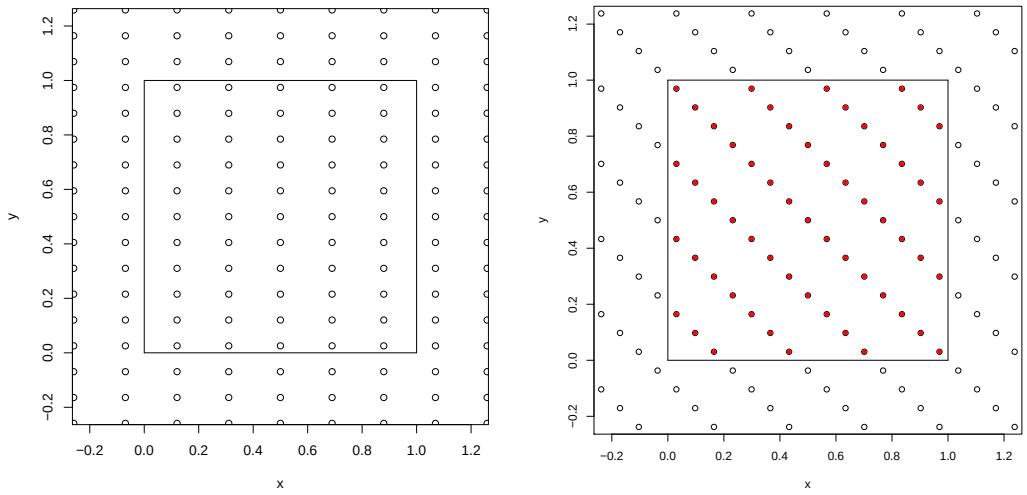


Figure 2: A sample grid design for $l = 1$, $\sigma = 0.3$ and $\rho = 0.6$. The left panel corresponds to the initial grid $\{(x'_j, y'_j)\}_{j=0}^{k'}$ over $\Omega$ (solid black square). The right panel depicts the rotated initial grid. The set of final node points $\{(x_i, y_i)\}_{i=0}^{k}$ is contained within $\Omega$ and is denoted by the red solid points.

condition in (15) with

$$M\mathbf{c}(t_\varepsilon) = \mathbf{p}(t_\varepsilon), \tag{16}$$
$$[\mathbf{p}(t_\varepsilon)]_i = \int_\Omega p(x,y,t_\varepsilon)\psi_i(x,y)dxdy.$$

Intuitively, the bigger $t_\varepsilon$, the smaller both $R_e(k)$ and $R_0(k)$ will be, and the better our method will do.

## 3.2 Orthonormal Basis Family

We motivate the construction of the orthonormal basis functions by once again considering the fundamental solution for the unbounded problem (8). In the absence of boundaries, (8) is solved by the function

$$G(x,y,t|x'_0,y'_0) = \frac{1}{2\pi\, t\, \tau_x\tau_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2t(1-\rho^2)}\left(\frac{(x-x'_0)^2}{\tau_x^2} - 2\rho\frac{(x-x'_0)(y-y'_0)}{\tau_x\tau_y} + \frac{(y-y'_0)^2}{\tau_y^2}\right)\right\},$$

with $x'_0 = (x_0-a_x)/(b_x-a_x)$, $\quad y'_0 = (y_0-a_y)/(b_y-a_y)$. We choose the family of basis functions $S_k = \{\psi_i(x,y),\ 0 \le i \le k\}$

$$\psi_i(x,y) = \frac{1}{2\pi\sigma^2\sqrt{1-\tilde\rho^2}} \exp\left\{-\frac{1}{2(1-\tilde\rho^2)\sigma^2}\left((x-x_i)^2 - 2\tilde\rho(x-x_i)(y-y_i) + (y-y_i)^2\right)\right\} x(1-x)\, y(1-y) \tag{17}$$

for some parameters $(\tilde\rho, \sigma)$ and a collection of nodes $\{(x_i, y_i)\}_{i=0}^{k}$ which form a grid over $\Omega$.

This grid is determined by the choice of kernel parameters $(\tilde\rho, \sigma)$ with a scaling parameter $l$, and it is defined in the following way. For $(\tilde\rho, \sigma, l)$, overlay a grid $\{(x'_j, y'_j)\}_{j=0}^{k'}$ such that $(x'_0, y'_0) = (0.5, 0.5)$ and the subsequent grid points are $l\sigma(1+\tilde\rho)$ units apart in the $x-$direction and $l\sigma(1-\tilde\rho)$ units apart in the $y-$direction over $[1/2 - 1/\sqrt{2}, 1/2 + 1/\sqrt{2}] \times [1/2 - 1/\sqrt{2}, 1/2 + 1/\sqrt{2}]$ (see left panel of Figure (2)). Next, apply the clockwise, $\pi/4$ rotation centered on $(1/2, 1/2)$ to each node

$$\begin{pmatrix} x_j \\ y_j \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} x'_j - 1/2 \\ y'_j - 1/2 \end{pmatrix} + \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}.$$

The grid is comprised of all nodes within $\Omega$: $\{(x_i,y_i)\}_{i=0}^{k} = \{(x_j,y_j)|(x_j,y_j) \in \Omega, j = 0,\ldots,k'\}$ (see right panel of Figure (2)). It should be noted that the level sets of the heat kernel for the basis functions are ellipses with major and minor axes aligned with the node points. Further, there is more resolution (ie. the node layout is denser) along the direction corresponding to the smaller standard deviation of the basis heat kernel in the principal coordinate frame. Finally, for a given $l$, nodes in either principal direction are separated by $l$ standard deviations of the basis heat kernel. This layout naturally takes into account some degree of correlation $\tilde{\rho}$ so that the small-time solution, having level sets similar to those of the kernels, can be better resolved. Essentially, the collection $\{\psi_i(x,y|\tilde{\rho},\sigma)\}_{i=0}^{k}$ is composed of fundamental solutions to a heat diffusion problem tuned by $\sigma$ and $\tilde{\rho}$, tampered such that their support is on $\Omega$, zero on the boundaries, still smooth, and inheriting the correlation structure of the fundamental solution to the problem. More basis elements are added by decreasing $\sigma$ and altering $l$, while the degree of correlation is controlled by $\tilde{\rho}$. For the purposes of this paper, we found that keeping fixed $l = 1$ with a moderate $\tilde{\rho}$ ($|\tilde{\rho}| \sim 0.6$) yields reasonable results. It is important, however, that $\text{sign}(\tilde{\rho}) = \text{sign}(\rho)$ so that the problematic, narrow component of the small-time solution (which can be seen in the minor axis of the contours of the small-time solution in the right panel of Figure (1)) can be resolved.

At this point, having explicitly defined $\psi_i(x,y|\sigma,\tilde{\rho})$, it should be noted that these basis functions are constant with respect to the normalized problem parameters $(\tau_x,\tau_y,\rho)$. Thus, the derivative of Galerkin approximation $p^{(k)}$ with respect to the boundaries takes on the form

$$\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} p^{(k)}(x,y,t) = \psi(x,y)^T \frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} \left\{\exp\left(M^{-1}St\right)\mathbf{c}(0)\right\},$$

which is well-defined given the discussion above.

### 3.3 Error Bound

A bound on the closeness of the approximate solution $p^{(k)}(x,y,t)$ to the strong solution $p(x,y,t)$ is developed in Bramble et al. [1977]. Their result shows that the Galerkin approximation we use converges to the strong solution in $L_2(\Omega)$, and it motivates the thrust of our numerical solution. First, we define the *error* term

$$e^{(k)}(t) = p(x,y,t) - p^{(k)}(x,y,t),$$

as well as the norm

$$\|w\|_2 = \sum_{j=0}^{\infty} \lambda_j^2 \left\langle w, \phi_j \right\rangle^2$$

for the eigenpairs $(\lambda_j,\phi_j)$ of the operator $\mathcal{L}$. As referred to in Bramble et al. [1977], functions $w \in L_2(\Omega)$ with $\|w\|_2 < \infty$ are also in $W_2^2(\Omega)$. Finally, if we have the condition (corresponding to equation 2.1 in Bramble et al. [1977])

$$\|p(x,y,t_\varepsilon) - p^{(k)}(x,y,t_\varepsilon)\|_{L_2(\Omega)} \leq Ch(k)^2 \|p(x,y,t_\varepsilon)\|_2, \tag{18}$$

where $h(\cdot)$ is a decreasing function of $k > 0$, Theorem 2.1 in Bramble et al. [1977] applies and we have the error estimate

$$\|e^{(k)}(t)\|_{L_2(\Omega)} \leq Ch(k)^2 \|p(x,y,t_\varepsilon)\|_2. \tag{19}$$

Here, the constant $C$ and the function $h(k)$ are the same as in (18). The implication is that if the basis functions in $S_k$ represent the small-time solution $p(x,y,t_\varepsilon)$ with no error, the Galerkin solution forward in time is also without error.

We can ensure condition (18) is met if $S_k$ is complete in $L_2(\Omega)$ as $k$ grows. The other two conditions necessary for the error bound to apply are demonstrated by Bramble et al. [1977] for the Galerkin method. Equation (19) can be summarized in a simple way: the *error* of the method is controlled by how much

variation the small-time solution has; the *rate* of decrease of the error is controlled by how well the span of $S_k$ represents the small-time solution compared to the variation of the initial condition as $k$ increases. In the context of (19), our method, with its small-time analytic solution and choice of basis functions, is specifically tailored to minimize the error between the strong solution and its Galerkin approximation under $L_2(\Omega)$.

# 4 Estimation

Consider the problem of estimating the parameters $(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$ from an i.i.d. set of samples $(Z_1, \ldots, Z_n)$ from random variable $Z(t)$ at a given value of $t$ in the process in (1)-(2). Specifically, $Z(t)$ is formed as

$$Z(t) = (X(t), Y(t), m_x(t), M_x(t), m_y(t), M_y(t))$$

where

$$m_x(t) = \min_{0 \le t' \le t} X(t), \;\; M_x(t) = \max_{0 \le t' \le t} X(t), \;\; m_y(t) = \min_{0 \le t' \le t} Y(t), \;\; M_y(t) = \max_{0 \le t' \le t} Y(t).$$

We say that $Z_i$ is sampled from the distribution corresponding to the probability density function (6)

$$Z_i \sim F(\theta),$$

where the cumulative distribution function $F$ has the usual interpretation

$$F(z = (x, y, a_x, b_x, a_y, b_y)|\theta) = \Pr\left(X(t) \le x, Y(t) \le y, m_x \le a_x, M_x \le b_x, m_y \le a_y, M_y \le b_y\right).$$

This estimation problem is of particular importance in quantitative finance where the model equations (1) - (2) (with various bells and whistles attached) are widely used. However, to the best of our knowledge, all current *likelihood* methods in the literature either ignore the observed maximum/minimum information or use only some of it. Likelihood-free approaches, like that of Rogers and Satchell [1991], on the other hand suffer from not being able to be easily integrated into inferential frameworks that require explicit estimates of probability.

Since we do not have a closed-form solution for the likelihood, we will use an iterative derivative-free maximization algorithm (the Nelder-Mead method; see Lagarias et al. [1998] for review and convergence properties) which requires repeated evauluation of the likelihood. For moderate to large samples sizes this is feasible, because our numerical method is specifically designed for computational efficiency for repeatedly evaluating the density function (6). The maximum likelihood estimator (MLE) for the true parameters based on $n$ samples, which we call $\hat{\theta}_n := (\hat{\mu}_x, \hat{\mu}_y, \hat{\sigma}_x, \hat{\sigma}_y, \hat{\rho})$, is especially useful in practical settings when it exhibits *consistency*, *i.e.*: the MLE gets closer to the true parameter vector $\theta$ as more data is collected and included in the likelihood (assuming the model and its parameters remain constant during the data collection). More precisely, the estimator $\hat{\theta}_n$ is consistent if it converges in probability to the true parameter:

$$\Pr(|\hat{\theta}_n - \theta|) \to 0 \qquad \text{as} \qquad n \to \infty.$$

## 4.1 Consistency

In this section we prove that the MLE based on the Galerkin approximation $p^{(k)}(x, y, t)$ to the governing Fokker-Planck equation (4) is consistent. To do so, we first show that the distribution on $Z$ based on the approximate $p^{(k)}(x, y, t)$ converges to the true distribution $F(\cdot|\theta)$. Define the probability densities

$$f(z) = f(x, y, a_x, b_x, a_y, b_y) := \Pr\left(X(t) \in dx, Y(t) \in dy, m_x \in da_x, M_x \in db_x, m_y \in da_y, M_y \in db_y \,|\, \theta\right) \tag{20}$$

$$f^{(k)}(z) = f^{(k)}(x, y, a_x, b_x, a_y, b_y) := \frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} q^{(k)}(x, y, t | a_x, b_x, a_y, b_y). \tag{21}$$

The function $f(z)$ is the probability density function corresponding to $q(x,y,t)$, the strong solution to the governing Fokker-Planck equation; $f^{(k)}(z)$ corresponds to the approximate solution $q^{(k)}$ of the unnormalized equation obtained from the Galerkin approximation $p^{(k)}$ of the normalized equation. Before proceeding, we should note that both $f(z)$ and $f^{(k)}(z)$ exist. So see the case for $f(z)$, consider the relation

$$
\begin{aligned}
&f(x,y,a_x,b_x,a_y,b_y) = \\
&\Pr\left(X(t)\in dx, Y(t)\in dy, m_x\in da_x, M_x\in db_x, m_y\in a_y, M_y\in db_y \,|\theta\right) = \\
&\mathbb{P}_W\left(\underbrace{\left\{\omega\in W \,\middle|\, X_\omega(t)=x, Y_\omega(t)=y, \inf_{t'\in[0,1]} X_\omega(t')=a_x, \sup_{t'\in[0,1]} X_\omega(t')=b_x, \inf_{t'\in[0,1]} Y_\omega(t')=a_y, \sup_{t'\in[0,1]} Y_\omega(t')=b_y\right\}}_{A(x,y,a_x,b_x,a_y,b_y)=A(z)}\right)
\end{aligned}
$$

where $\mathbb{P}_W$ is the Wiener measure on the sample space $W$ of all realizations (paths) $(X_\omega(t), Y_\omega(t))$ from the stochastic process (1) - (2) defined in the usal way using Kolmogorov's extension of measure over cylinder sets on $t\to\mathbb{R}^2$ (see Freidlin [1985], Section 1.2). Sets of the form $A(x,y,a_x,b_x,a_y,b_y)$ can be defined as a countable intersection/union of cyliner sets on $t\to\mathbb{R}^2$, hence they are measurable under $\mathbb{P}_W$; therefore $f(z)$ exists and is bounded above by 1. Moreover, since $q(x,y,t|a_x,b_x,a_y,b_y)$ represents the integral of $f(z)$ with respect to the boundary variables, we have

$$f(x,y,a_x,b_x,a_y,b_y) = \frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} q(x,y,t|a_x,b_x,a_y,b_y).$$

The existence of $f^{(k)}(z)$ as defined in (21) is guaranteed by Theorem 4.1 in Singler [2008]. The result states that weak solutions to parabolic problems are differentiable with respect to parameters as long as the weak (Galerkin-form) operator $\left\langle \mathcal{L}\psi_i(x,y), \psi_j(x,y)\right\rangle$ is differentiable with respect to the parameters. The condition certainly holds for the normalized problem (8) with our choice of basis functions $S_k$, as they are infinitely differentiable with repsect to $x,y$, the boundaries, and the diffusion parameters.

At this point we will assume that for a sufficiently large $k$, $f^{(k)}(z)$ is positivie for all $z\in Z$ and is integrable over $Z$. As such, we may regard it as a proper probability density function with the cumulative probability density and probability measure over $Z$ being

$$F^{(k)}(z|\theta) = \int_{-\infty}^{a_x}\int_{-\infty}^{a_y}\int_{-\infty}^{b_x}\int_{-\infty}^{b_y}\int_{-\infty}^{x}\int_{-\infty}^{y} \frac{\partial^4}{\partial a'_x \partial b'_x \partial a'_y \partial b'_y} q^{(k)}(x',y',t|a'_x,b'_x,a'_y,b'_y)\; dx'\,dy'\,da'_x\,db'_x\,da'_y\,db'_y, \tag{22}$$

$$\Pr_k(A) := \int_A f^{(k)}(z)\,dz, \quad \text{for any measurable } A\subset Z. \tag{23}$$

We will prove that for every $z\in Z$,

$$\lim_{k\to\infty} F^{(k)}(z|\theta) = F(z|\theta).$$

First, we prove the Lemma

**Lemma 1.** *For $z=(x,y,a_x,b_x,a_y,b_y)$,*

$$\lim_{k\to\infty}\int_{a_x}^{b_x}\int_{a_y}^{b_y} f^{(k)}(x,y,a_x,b_x,a_y,b_y)\,dx\,dy = \int_{a_x}^{b_x}\int_{a_y}^{b_y} f(x,y,a_x,b_x,a_y,b_y)\,dx\,dy.$$

*Proof.* Define the sets of form, for $z=(x,y,a_x,b_x,a_y,b_y)$,

$$B(a_x,b_x,a_y,b_y) = \left\{z'\in Z \,\middle|\, z'\in[a_x,b_x]\times[a_y,b_y]\times[a_x,\infty)\times(-\infty,b_x]\times[a_y,\infty)\times(-\infty,b_y]\right\}.$$

Elements within $B(a_x,b_x,a_y,b_y)$ are equivalent to sample paths that stay within the region $[a_x,b_x]\times[a_y,b_y]$:

$$B(a_x,b_x,a_y,b_y) = \{\omega\in W | X_\omega(t)\in[a_x,b_x], Y_\omega(t)\in[a_y,b_y], m_x\in[a_x,b_x), M_x\in(a_x,b_x]\}$$

Then

$$\begin{aligned}\Pr(B(a_x,b_x,a_y,b_y)) &= \int_{B(a_x,b_x,a_y,b_y)} f(z)\,dz \\ &= \int_{a_x}^{\infty}\int_{-\infty}^{b_x}\int_{a_y}^{\infty}\int_{-\infty}^{b_y}\int_{-\infty}^{b_x}\int_{-\infty}^{b_y} f(x',y',a'_x,b'_x,a'_y,b'_y)dx'dy'da'_xdb'_xda'_ydb'_y \\ &= \int_{a_x}^{b_x}\int_{a_y}^{b_y} q(x,y,a_x,b_x,a_y,b_y)\,dxdy,\end{aligned} \tag{24}$$

where the last equality employs the interpretation of $q$ in (3) and we can freely change the order of integration as $f$ is bounded for all $z \in Z$. Similarly,

$$\Pr_k(B(a_x,b_x,a_y,b_y)) = \int_{a_x}^{b_x}\int_{a_y}^{b_y} q^{(k)}(x,y,a_x,b_x,a_y,b_y)\,dxdy.$$

From (24), the partial derivative of the probability $\Pr(B(a_x,b_x,a_y,b_y))$ is well-defined and can be written as

$$\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y}\Pr(B(a_x,b_x,a_y,b_y)) = \int_{a_x}^{b_x}\int_{a_y}^{b_y} f(x,y,a_x,b_x,a_y,b_y)\,dxdy.$$

The second-order **finite difference** approximation the above expression can be expressed as a linear combination of probabilities of perturbed sets $B(\cdot)$:

$$\lim_{\varepsilon\to 0}\quad \frac{1}{\varepsilon^4}\sum_{i=1}^{16} c(i)\Pr(B(a_x+k_1(i)\varepsilon, b_x+k_2(i)\varepsilon, a_y+k_3(i)\varepsilon, b_y+k_4(i)\varepsilon) = \frac{\partial^4}{\partial a_x\partial b_x\partial a_y\partial b_y}\Pr(B(a_x,b_x,a_y,b_y)),$$

where $c$ and $k$ are functions from $i$ to the appropriate coefficients in the second-order finite difference approximation $c(i) \to \{-1,1\}$, $k_j(i) \to \{-1,1\}$. Using Big-O notation, for a sufficiently small $\varepsilon$

$$\sum_{i=1}^{16} c(i)\Pr(B(a_x+k_1(i)\varepsilon, b_x+k_2(i)\varepsilon, a_y+k_3(i)\varepsilon, b_y+k_4(i)\varepsilon) = \varepsilon^4\frac{\partial^4}{\partial a_x\partial b_x\partial a_y\partial b_y}\Pr(B(a_x,b_x,a_y,b_y)) + O(\varepsilon^6; a_x,b_x,a_y,b_y). \tag{25}$$

The convergence result in Section 3.3 implies that

$$\begin{aligned}\Pr_k(B(a_x+k_1(i)\varepsilon, b_x+k_2(i)\varepsilon, a_y+k_3(i)\varepsilon, b_y+k_4(i)\varepsilon) &= \int_{a_x+k_1(i)\varepsilon}^{b_x+k_2(i)\varepsilon}\int_{a_y+k_3(i)\varepsilon}^{b_y+k_4(i)\varepsilon} q^{(k)}(x,y,a_x,b_x,a_y,b_y)\,dxdy \to \\ \int_{a_x+k_1(i)\varepsilon}^{b_x+k_2(i)\varepsilon}\int_{a_y+k_3(i)\varepsilon}^{b_y+k_4(i)\varepsilon} q(x,y,a_x,b_x,a_y,b_y)\,dxdy &= \Pr(B(a_x+k_1(i)\varepsilon, b_x+k_2(i)\varepsilon, a_y+k_3(i)\varepsilon, b_y+k_4(i)\varepsilon) \\ &\qquad \text{as } k\to\infty \text{ in } L_2(\Omega).\end{aligned}$$

Hence, for a sufficiently large $k$ dependent on the supremum over $i$, and given the error estimate in (19), we have the relation

$$\begin{aligned}\Pr_k(B(a_x+k_1(i)\varepsilon, b_x+k_2(i)\varepsilon, a_y+k_3(i)\varepsilon, b_y+k_4(i)\varepsilon) &= \Pr(B(a_x+k_1(i)\varepsilon, b_x+k_2(i)\varepsilon, a_y+k_3(i)\varepsilon, b_y+k_4(i)\varepsilon) \\ &\quad + O(h(k)^2; a_x+k_1(i)\varepsilon, b_x+k_2(i)\varepsilon, a_y+k_3(i)\varepsilon, b_y+k_4(i)\varepsilon)\end{aligned}$$

Note here that the dominating terms $O(h(k)^2;\cdot)$ are differentiable with respect to the boundary parameters

$(a_x, b_x, a_y, b_y)$ since $q$ and $q^{(k)}$ have this property. Therefore, if we replace $\Pr(\cdot)$ in (25) with $\Pr_k(\cdot)$

$$\begin{aligned}\sum_{i=1}^{16} c(i) \Pr_k(B(a_x + k_1(i)\varepsilon, b_x + k_2(i)\varepsilon, a_y + k_3(i)\varepsilon, b_y + k_4(i)\varepsilon) &= \sum_{i=1}^{16} c(i) \Pr(B(a_x + k_1(i)\varepsilon, b_x + k_2(i)\varepsilon, a_y + k_3(i)\varepsilon, b_y + k_4(i)\varepsilon) \\ &+ \sum_{i=1}^{16} c(i) O(h(k)^2; a_x + k_1(i)\varepsilon, b_x + k_2(i)\varepsilon, a_y + k_3(i)\varepsilon, b_y + k_4(i)\varepsilon) \\ &= \varepsilon^4 \frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} \Pr(B(a_x, b_x, a_y, b_y)) + O(\varepsilon^6; a_x, b_x, a_y, b_y) \\ &+ \sum_{i=1}^{16} c(i) O(h(k)^2; a_x + k_1(i)\varepsilon, b_x + k_2(i)\varepsilon, a_y + k_3(i)\varepsilon, b_y + k_4(i)\varepsilon).\end{aligned}$$

Dividing both sides by $\varepsilon^4$ produces

$$\begin{aligned}\frac{1}{\varepsilon^4} \sum_{i=1}^{16} c(i) \Pr_k(B(a_x + k_1(i)\varepsilon, b_x + k_2(i)\varepsilon, a_y + k_3(i)\varepsilon, b_y + k_4(i)\varepsilon) &= \frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} \Pr(B(a_x, b_x, a_y, b_y)) + O(\varepsilon^2; a_x, b_x, a_y, b_y) \\ &+ \frac{1}{\varepsilon^4} \sum_{i=1}^{16} c(i) O(h(k)^2; a_x + k_1(i)\varepsilon, b_x + k_2(i)\varepsilon, a_y + k_3(i)\varepsilon, b_y + k_4(i)\varepsilon)\end{aligned}$$

As mentioned above $O(h(k)^2; \cdot)$ is differentiable with respect to the boundary parameters, so that the rightmost term is still $O(h(k)^2)$ as $\varepsilon \to 0$. Taking the limit in $\varepsilon$, we have

$$\begin{aligned}\int_{a_x}^{b_x} \int_{a_y}^{b_y} f^{(k)}(x, y, a_x, b_x, a_y, b_y)\, dx\, dy &= \frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} \Pr_k(B(a_x, b_x, a_y, b_y)) \\ &= \frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} \Pr(B(a_x, b_x, a_y, b_y)) + O(h(k)^2; a_x, b_x, a_y, b_y) \\ &= \int_{a_x}^{b_x} \int_{a_y}^{b_y} f(x, y, a_x, b_x, a_y, b_y)\, dx\, dy + O(h(k)^2; a_x, b_x, a_y, b_y).\end{aligned}$$

Therefore, we have the desired result:

$$\lim_{k \to \infty} \int_{a_x}^{b_x} \int_{a_y}^{b_y} f^{(k)}(x, y, a_x, b_x, a_y, b_y)\, dx\, dy = \int_{a_x}^{b_x} \int_{a_y}^{b_y} f(x, y, a_x, b_x, a_y, b_y)\, dx\, dy.$$

□

**Lemma 2** (Convergence in distribution)**.** *For any $z \in Z$, $\lim_{k \to \infty} F^{(k)}(z|\theta) = F(z)$.*

*Proof.* Let $I^{(k)}(z) = \int_{a_x}^{x} \int_{a_y}^{y} f^{(k)}(u, v, a_x, b_x, a_y, b_y)\, du\, dv$ and let $I(z) = \int_{a_x}^{x} \int_{a_y}^{y} f(u, v, a_x, b_x, a_y, b_y)\, du\, dv$. It is possible to show that $\lim_{k \to \infty} I^{(k)}(z) = I(z)$ as a consequence of Lemma 1 by considering some $\chi_k(z) \in S_k$ approximating the indicator $1(u \le x, v \le y)$ as $k \to \infty$ and setting up a triangle inequality. However, we will omit this technical detail here.

Next, we know that $I(z)$ is integrable over $(a_x, b_x, a_y, b_y)$ as $\Pr(Z) = \mathbb{P}_W(W) = 1$. The Dominated Convergence Theorem applies, and we therefore have

$$\lim_{k \to \infty} \int_{-\infty}^{a_x} \int_{-\infty}^{b_x} \int_{-\infty}^{a_y} \int_{-\infty}^{b_y} I^{(k)}(z)\, da'_x\, db'_x\, da'_y\, db'_y = \int_{-\infty}^{a_x} \int_{-\infty}^{b_x} \int_{-\infty}^{a_y} \int_{-\infty}^{b_y} I(z)\, da'_x\, db'_x\, da'_y\, db'_y,$$

which implies the result of the Lemma. □

**Lemma 3.** *The maximum likelihood estimator is consistent as $n \to \infty$ and $m \to \infty$:*

$$\hat{\theta}_{n,k} \to \theta$$

.

*Proof.* By Lemma 2

$$Z_k \xrightarrow{d} Z \text{ as } k \to \infty.$$

Next, given Theorem 4.1 in Singler [2008], we know that, for each $k$, $q_k$ is analytic in both the diffusion parameters and boundary parameters. Hence, the probability density function satisfies the criteria A1 - A6 in Casella and Berger [2002] to guarantee that, for data $Z_k \sim F_k(\theta)$,

$$\hat{\theta}_{n,k}(Z_k) \xrightarrow{p} \theta \text{ as } n \to \infty.$$

Now we need to show that the same holds for data sampled from $F$ as $k \to \infty$. To do this, we will use Chebyshev's inequality:

$$\Pr_Z\left(\left|\hat{\theta}_{n,k}(Z) - \theta\right| \geq \varepsilon\right) \leq \frac{\mathrm{E}_Z\left[(\hat{\theta}_{n,k}(Z) - \theta)^2\right]}{\varepsilon^2}.$$

By the Maximum theorem [REFERENCE], $\hat{\theta}_{n,k}(x)$ is a continuous function with respect to $x$, and further because we have bounded $\hat{\theta}$ from below and above,

$$\mathrm{E}_{Z_k}\left[(\hat{\theta}_{n,k}(Z_k) - \theta)^2\right] \to \mathrm{E}_Z\left[(\hat{\theta}_{n,k}(Z) - \theta)^2\right] \text{ as } k \to \infty$$

by the portmanteau lemma. Finally, we can show that

$$\mathrm{E}_{Z_k}\left[(\hat{\theta}_{n,k}(Z_k) - \theta)^2\right] \to 0 \text{ as } n \to \infty, \tag{26}$$

since the expected value of the estimator tends to $\theta$ and its variance goes to 0 when $n \to \infty$. Therefore, given any $\varepsilon > 0$ and $\delta > 0$, we can find a sufficiently large $n$ and $k$ such that

$$\Pr_Z\left(\left|\hat{\theta}_{n,k}(Z) - \theta\right| \geq \varepsilon\right) \leq \frac{\mathrm{E}_Z\left[(\hat{\theta}_{n,k}(Z) - \theta)^2\right]}{\varepsilon^2} < \delta$$

□

## 4.2 Simulation Study

Next we present some numerical simulation results for the Galerkin method. The finite-difference step used to derive the likelihood function is of size $\Delta = 1/16$ with the $O(\Delta^2)$ order accuracy. If the inital/final condition is forced outside of the computation domain, the $O(\Delta)$ finite difference approximation is used instead. The inner products that need to be computed to establish the system of equations for the method are computed using the trapezoidal rule on a regular $400 \times 400$ grid.

Data is generated from the model with zero drift via forward Euler discretization where the obtained discrete-time extrema are recorded and used as the realized extrema of the process. We use 500 data sets, each comprised of 128 simulations generated with the same parameters

$$\mu_x = 0, \quad \mu_y = 0, \quad \sigma_x = 1, \quad \sigma_y = 1, \quad \rho = 0.6.$$

The drift parameters are assumed known, so that the MLE is comprised of the diffusion and correlation parameters: $\hat{\theta} = (\hat{\sigma}_x, \hat{\sigma}_y, \hat{\rho})$. For each dataset, the MLE found is a sample from the repeated-sampling distrubtion for $\hat{\theta}$. Results are compared to the repeated-sampling distribution of the MLE based on the usual bivariate normal likelihood which does not take into account the boundaries. Results are shown in Figure (3), including the root-mean-square-errors of the two estimators.

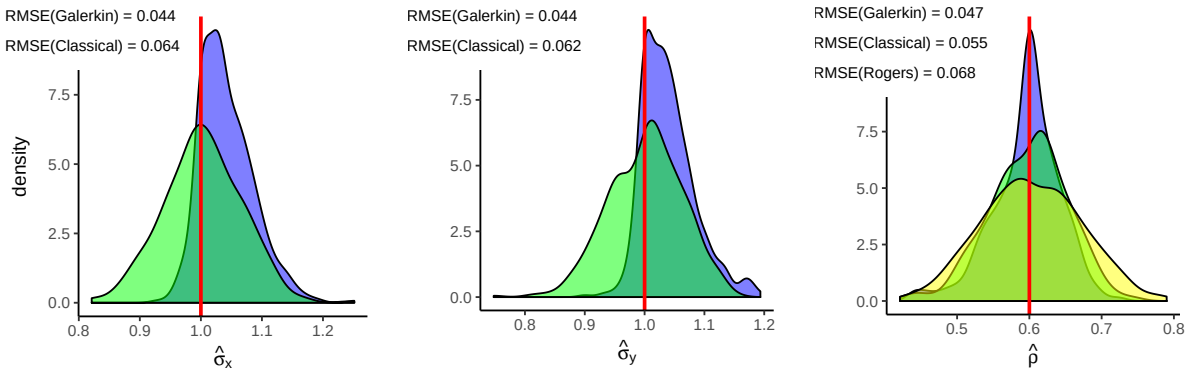


Figure 3: Kernel-density plots of the MLE samples obtained from the Galerkin likelihood (blue) and the classical likelihood (green). The data-generating parameters are denoted with the vertical solid line. The root-mean-square errors for the pairs of estimators are included. As expected, the Galerkin-based likelihood outperforms the Gaussian likelihood for each parameter.

# References


JH Bramble, AH Schatz, V Thomée, and LB Wahlbin. Some convergence estimates for semidiscrete galerkin type approximations for parabolic equations. *SIAM Journal on Numerical Analysis*, 14(2):218–241, 1977.

George Casella and Roger L Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002.

Wai-Ki Ching, Jia-Wen Gu, and Harry Zheng. On correlated defaults and incomplete information. *arXiv preprint arXiv:1409.1393*, 2014.

Mark Iosifovich Freidlin. *Functional Integration and Partial Differential Equations.(AM-109)*, volume 109. Princeton university press, 1985.

Helen Haworth, Christoph Reisinger, and William Shaw. Modelling bonds and credit default swaps using a structural model with contagion. *Quantitative Finance*, 8(7):669–680, 2008.

Hua He, William P Keirstead, and Joachim Rebholz. Double lookbacks. *Mathematical Finance*, 8(3):201–228, 1998.

Enrique Ter Horst, Abel Rodriguez, Henryk Gzyl, and German Molina. Stochastic volatility models including open, close, high and low prices. *Quantitative Finance*, 12(2):199–212, 2012.

Steven Kou, Haowen Zhong, et al. First-passage times of two-dimensional brownian motion. *Advances in Applied Probability*, 48(4):1045–1060, 2016.

Jeffrey C Lagarias, James A Reeds, Margaret H Wright, and Paul E Wright. Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on optimization*, 9(1):112–147, 1998.

Peter Lax. *Linear Algebra and Its Applications*. John Wiley & Sons, 2nd edition, 2007.

Douglas H Norrie and GJ De Vries. *The finite element method: fundamentals and applications*. Academic Press, 1973.

Bernt Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

Hannes Risken. *The Fokker-Planck Equation: Methods of Solution and Applications*. Springer-Verlag, 1989.

L Christopher G Rogers and Stephen E Satchell. Estimating variance from high, low and closing prices. *The Annals of Applied Probability*, pages 504–512, 1991.

Laura Sacerdote, Massimiliano Tamborrino, and Cristina Zucca. First passage times of two-dimensional correlated processes: Analytical results for the wiener process and a numerical method for diffusion processes. *Journal of Computational and Applied Mathematics*, 296:275–292, 2016.

John R Singler. Differentiability with respect to parameters of weak solutions of linear parabolic equations. *Mathematical and Computer Modelling*, 47(3):422–430, 2008.

Ivar Stakgold and Michael J Holst. *Green's functions and boundary value problems*, volume 99. John Wiley & Sons, 2011.

Eberhard Zeidler. Applied functional analysis: main principles and their applications. *Applied mathematical sciences*, 109, 1995.