Journal of Mathematical Psychology 114 (2023) 102756

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Diffusion models with time-dependent parameters: An analysis of computational effort and accuracy of different numerical methods

Thomas Richter [a,*], Rolf Ulrich [b], Markus Janczyk [c]

[a] Otto-von-Guericke Universität Magdeburg, Magdeburg, Germany
[b] Eberhard Karls Universität Tübingen, Tübingen, Germany
[c] University of Bremen, Bremen, Germany

ARTICLE INFO

Article history:
Received 25 August 2022
Received in revised form 21 February 2023
Accepted 26 February 2023
Available online 27 March 2023

Dataset link: https://doi.org/10.5281/zenodo.6970739

Keywords:
Drift-diffusion model
Cognitive modeling
Numerical approximation
Parameter estimation

ABSTRACT

Drift-diffusion models have become valuable tools in many fields of contemporary psychology and the neurosciences. The present study compares and analyzes different methods (i.e., stochastic differential equation, integral method, Kolmogorov equations, and matrix method) to derive the first-passage time distribution predicted by these models. First, these methods are compared in their accuracy and efficiency. In particular, we address non-standard problems, for example, models with time-dependent drift rates or time-dependent thresholds. Second, a mathematical analysis and a classification of these methods is provided. Finally, we discuss their strengths and caveats.

1. Introduction

Mathematical models describing the time course of decision-making have become an increasingly valuable tool in psychology and neuroscience. One of the most successful variants of such models is the drift-diffusion model (DDM; Ratcliff, 1978), a class of mathematical models for binary decisions (for reviews, see Forstmann, Ratcliff, & Wagenmakers, 2016; Ratcliff, Smith, Brown, & McKoon, 2016; Schwarz, 2022). This section begins with a description of the standard DDM and then introduces two cases where the parameters of the DDM are time-dependent, that is, they vary with the progression of time in an experimental trial. Based on this, we then lay out the purpose of the present paper in more detail and sketch the outline of its remainder.

1.1. The standard DDM

The general idea behind the DDM is that the cognitive system accumulates noisy evidence for one or the other of two response options over time (Fig. 1) . If there are two response options, a selection occurs once the evidence exceeds one of the respective thresholds that are denoted here as b and $-b$ for response options 1 and 2, respectively. Evidence accumulation starts at zero without a bias toward one response option, and is driven toward one threshold with a constant increase per time step, referred to as the drift rate μ. The accumulation process is noisy, and usually modeled as a Wiener process. Because of the noise, the entire diffusion process $\{X(t), t > 0\}$ exceeds the correct threshold at a random point in time, and can also result in an erroneous decision when exceeding the incorrect threshold. While these parameters cover the decision part of the DDM, perceptual and motor processes are captured by the residual (or non-decision) time, which is added to the decision time. The full model comprises several additional parameters. In particular, the starting point, the residual time, and the drift rate can vary from trial to trial, and the starting point can be biased toward one or the other threshold (for overviews of the full model, see Ratcliff, 1978; Voss, Nagler, & Lerche, 2013; Voss, Voss, & Lerche, 2015). The thresholds $(b, -b)$ and the drift rate μ are assumed to be constant within a trial (although, as mentioned, they can vary across trials). In other words, both parameters are *time-independent* (or *stationary*), because their value does not depend on the progression of time within a single trial.

1.2. Time-dependent parameters

While this stationary assumption simplifies mathematical tractability and is sufficient to model many psychological phenomena, variants of the DDM with time-dependency of these parameters are motivated by additional psychological phenomena. For example, Heath (1992) considered that the drift rate may

* Corresponding author.
E-mail addresses: thomas.richter@ovgu.de (T. Richter), rolf.ulrich@uni-tuebingen.de (R. Ulrich), janczyk@uni-bremen.de (M. Janczyk).

https://doi.org/10.1016/j.jmp.2023.102756
0022-2496/

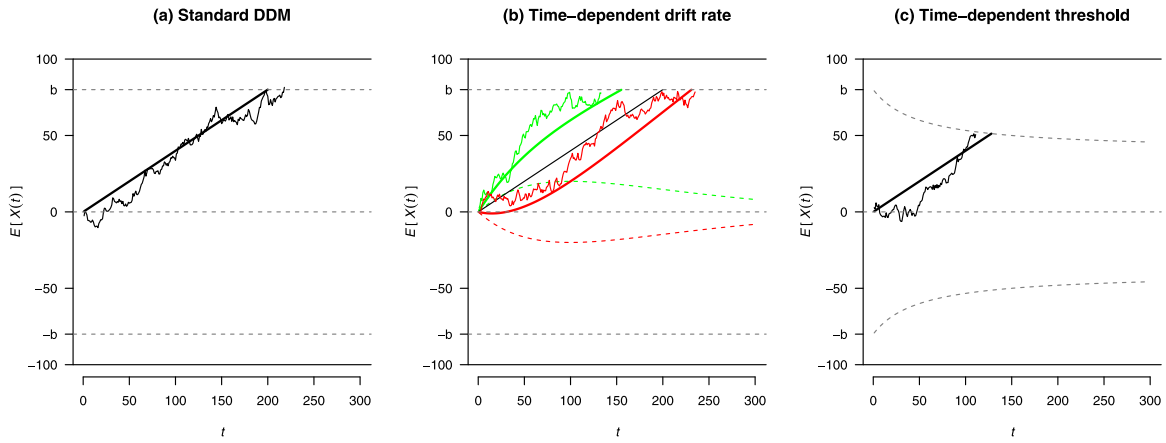


Fig. 1. (a) Illustration of the standard drift-diffusion model (DDM): The straight black line is the expected value $E[X(t)]$ based on the drift rate alone, the jagged black line is an example of a corresponding diffusion process with added Brownian motion. (b) Illustration of time-dependent drift rates according to the Diffusion Model for Conflict (DMC) tasks (Ulrich, Schröter, Leuthold, & Birngruber, 2015): The straight black line is the expected value based on the controlled drift rate alone, the dashed green and red lines are the expected values of the automatic activation modeled by a Gamma function for congruent and incongruent trials, respectively. The solid green and red lines are the expected values based on the superimposition of controlled and automatic processes, and the jagged green and red lines are examples of the corresponding diffusion processes with added Brownian motion. (c) Illustration of time-dependent thresholds: The straight black line is the expected value based on the drift rate alone, and the jagged black line is an example of the corresponding diffusion process with added Brownian motion. Thresholds are modeled with a hyperbolic ratio function.

diminish with increasing processing time. Moreover, Heath also showed how a time-dependent drift rate could arise within the cascade model by McClelland (1979). In the following, however, we will describe more recent modeling work that employed time-dependent changes of the drift rate and time-dependent thresholds.

Time-dependent drift rate. One example for time-dependency of the drift rate $\mu(t)$ is the Diffusion Model for Conflict (DMC) tasks (Ulrich et al., 2015). This model was particularly developed to account for positive and negative delta functions in typical conflict tasks. Consider, for example, an Eriksen flanker task (Eriksen & Eriksen, 1974). In this task, a centrally presented imperative stimulus requires a left or right manual key press, while it is surrounded by flankers that either demand the same response (in congruent trials) or the other response (in incongruent trials). Response times (RTs) are shorter and less error-prone in congruent than in incongruent trials, what is typically referred to as the congruency effect. This congruency effect usually becomes larger with longer RTs, that is, a positive delta function is observed (see de Jong, Liang, & Lauber, 1994; Pratte, Rouder, Morey, & Feng, 2010; Schwarz & Miller, 2012).

The Simon task (Simon, 1969) is another conflict task where participants are to respond to a stimulus (e.g., the identity of a letter or a color), which is presented either in a left or right location. Here, participants respond faster when stimulus location and (correct) response location match (in congruent trials) than when they mismatch (in incongruent trials). However, in this case, the congruency effect becomes smaller with longer RTs, that is, a negative delta function is observed (e.g., Pratte et al., 2010).

Motivated by dual-process accounts for congruency effects, the DMC explains different delta functions by assuming that two diffusion processes are superimposed: (1) a linear drift function representing controlled response selection and (2) a pulse function toward the same response in congruent trials and to the alternative response in incongruent trials (see Fig. 1b for an illustration). The latter process represents the automatic activation induced by the task-irrelevant stimulus features in a conflict task (e.g., the flankers or the location of the stimuli).

The superimposed process tends to hit the upper threshold earlier when the automatic activation turns toward the positive direction (in congruent trials) compared with the case of an activation into the negative direction (in incongruent trials). This reflects the congruency effect (see Fig. 1b). In technical terms, the superimposition results in a time-dependent drift rate, because the drift rate changes with increasing time during a trial. Importantly, different delta functions can also result: The superimposition of the two diffusion processes results in a negative delta function, when the automatic activation reaches its maximum rather early, but it results in a positive delta function, when the maximum is reached later.

Another example of time-dependent drift rates is the Shrinking Spotlight Model (White, Ratcliff, & Starns, 2011) that addresses conflict processing in the flanker task. This model also accounts for the observation that responses with short RTs are especially error-prone in incongruent trials. Such a result suggests that the flankers exhibit their influence especially during the early phases of processing. The model accounts for this by assuming that attention is distributed broadly at the onset of a trial and shrinks over time, thereby decreasing the influence of the flankers the more time has progressed. For a congruent trial, the resulting drift rate is time-independent. In contrast, for incongruent trials, the resulting drift rate is time-dependent and increases with time, as more and more information is processed and the influence of the flankers becomes smaller.

Time-dependent thresholds. The second parameter that may change with time is the threshold b (Fig. 1c). Arguably, in most of the applications of the DDM, thresholds are assumed to be constant within a trial. However, models in which thresholds vary with time—most often they are assumed to collapse with increasing time—have been suggested to account for some phenomena (e.g., Churchland, Kiani, & Shadlen, 2008; Ditterich, 2006a; Evans, Hawkins, & Brown, 2020; Hanks, Mazurek, Kiani, Hopp, & Shadlen, 2011). Collapsing thresholds were even suggested to account for slow errors, thereby rendering the drift rate between-trial variability as unnecessary (e.g., Ditterich, 2006b; Palmer, Huk, & Shadlen, 2005).

Further, a recent study suggested that different experimental methods to induce speed-accuracy tradeoffs affect different parameters of a DDM. More precisely, Katsimpokis, Hawkins, and van Maanen (2020) suggested that instructions affect the initial separation of thresholds, while response deadlines affect the rate of their collapse over time. A very similar idea, though for an

accumulator model, was suggested to account for results obtained with free-choice tasks. These are tasks where the stimulus does not request one particular response, but where the actor should choose among a set of response options (e.g., Berlyne, 1957; Janczyk, Naefgen, & Kunde, 2020). Based on results from priming experiments, Mattler and Palmer (2012) suggested that the decision criterion relaxes in free-choice trials, an assumption similar to a collapsing threshold in a DDM.

It is, however, not clear though whether time-dependent thresholds are psychologically plausible. Moreover, at least three studies (Evans et al., 2020; Hawkins, Forstmann, Wagenmakers, Ratcliff, & Brown, 2015; Voskuilen, Ratcliff, & Smith, 2016) concluded that collapsing thresholds add little if any improvement to model fit when compared to a DDM with time-independent thresholds. For example, Voskuilen et al. (2016) used a hyperbolic ratio function to model a decrease of the thresholds over time (see Fig. 1c).[1] The best estimates for its parameters resulted in a shape resembling a time-independent threshold (see their Fig. 5).

Yet, there may be tasks and situations where collapsing thresholds are more likely to improve model fit. One such situation are difficult tasks with a very small drift rate, thus resulting in long RTs. While the DDM has mainly been applied to rather simple tasks with short RTs, it appears also viable for longer RTs (Lerche & Voss, 2019). In this case, the process will eventually hit a threshold, but this may take much too long with a time-independent threshold (e.g., Voss, Lerche, Mertens, & Voss, 2019). As an additional example, expanded judgment or deferred decision-making tasks are cases where decisions are made based on increasingly smaller amounts of evidence as time progresses (e.g., Busemeyer & Rapoport, 1988), and data from highly practiced participants may better be accounted for with collapsing thresholds as well (Hawkins et al., 2015).

In summary, it is controversial whether or not time-dependent thresholds are valuable to the DDM. Although this controversy is not yet settled, the present article nonetheless includes time-dependent thresholds into its analyses.

1.3. Purpose and outline of the present article

The present article reports a comparison of different methods to solving the underlying mathematics of diffusion models, particularly with time-dependent parameters. More precisely, we investigate, on the one hand, the direct simulation of the model using the stochastic Euler method. On the other hand, we analyze the integral equation method as well as different methods based on a reformulation of the model into partial differential equations (PDEs), such as the matrix method of Diederich and Busemeyer (2003). We focus on the computational efficiency of the different methods; thus we present a theoretical analysis of how much more effort is required for a desired error reduction and investigate the computational times required for the different implementations. In this course, we highlight the critical differences between these methods, their potential advantages and shortcomings, and present solutions to some known issues with particular methods.

Shinn, Lam, and Murray (2020) describe and compare different approximation methods for diffusion models, and the authors also introduce a Python software package that allows for highly efficient computations in a flexible environment. They further discuss the versatility of the approaches, for example, when applied to time-dependent parameters. While their paper focuses on the practical implementation of the methods, the emphasis of our work is to present the theoretical foundations and to explore the significance of the mathematical background for the application.

The remainder of this article is structured in the following way. Section 2 formalizes the DDM and introduces stochastic simulations, random walks, and—with a particular focus—the *Kolmogorov Forward Equation (KFE) and the Kolmogorov Backward Equation (KBE)* as methods to implement the diffusion model. Section 3 introduces several methods of discretizations of the aforementioned methods with emphasis on the quality of the approximations and on the computational cost of these numerical approaches. We also introduce a new approach which is based on the discretization of the KFE and a realization that is able to handle time-dependent drift rates and thresholds without a drop in efficiency (see in particular Section 3.3 and Appendix A.1). The practical dependence of computational costs (or efficiency) on the desired level of error reduction is investigated in Section 4. In Section 5, we fit our approach to empirical data gathered in the context of DMC, that is, a model with time-dependent drift rates (Ulrich et al., 2015). Section 6 summarizes the conclusion of the present work, including advice for researchers using DDMs with time-dependent parameters. The source codes to reproduce all computations are available as a Zenodo repository (Richter, Ulrich, & Janczyk, 2023).

Remark 1 (*Terminology: Model, Method, and Discretization*)**.** In the following, it is necessary to distinguish between model, method, and discretization. A model describes a cognitive process and is formulated as a stochastic differential equation (i.e., Equation (1)). From this model, we deduce the first-passage time distribution, and to do so, different methods are available, that is, stochastic simulations, the KFE (6), the KBE (7), the integral equation (3), and random walks (see Section 2.4). All four methods can be used to compute the predicted probability density function (PDF) of the first-passage times, and if they were solved exactly, they would yield the identical result.

This, however, is not possible in the general case as it would require infinite computational resources. Hence, for each method, different discretizations are possible. For example, the trapezoidal rule (20) is used in approximating the integral equation and finite difference discretizations are applied to the KBE and KFE. Depending on the method and on the specific situation (e.g., time-dependency of parameters), one approach will be superior to the other. A discussion of the subtle differences is the content of the following sections.

1.4. Transparency and openness

This article will describe the mathematical tools to model and discretize diffusion problems. Various numerical approximation techniques will be detailed. The methods are implemented in our own software in C++ and Python. The scripts are deposited for reproducibility of all results (Richter et al., 2023). This archive also contains the data analysis techniques as well as the scripts for generating the figures of this manuscript. This research does not involve human (or animal) experimentation.

2. Reaction time modeling

2.1. The diffusion model

A fundamental assumption of the DDM is that the decision process evolves continuously in time. Given a starting value X_0 at time $t = 0$ in between the lower and upper threshold $X_0 \in$

[1] Hawkins et al. (2015) used the Weibull distribution function as a flexible function allowing to approximate many candidates of how thresholds collapse exactly, for example, whether the large part of the collapse occurs early or rather late during a trial.

$[-b(0), b(0)]$, the decision variable $X(t)$ evolves by the deterministic drift rate $\mu(t)$ and a random influence modeled as Brownian motion by the stochastic differential equation (SDE)

$$dX(t) = \mu(t) + \sigma(t)dB(t), \quad X(0) = X_0 \tag{1}$$

where $\mu(t)$ is the drift rate, $B(t)$ refers to standard Brownian motion, and $\sigma(t)$ is the standard deviation of $X(t)$ (also called "diffusion constant"). If $X(t)$ at time $t > 0$ hits the upper threshold, that is, $X(t) = b(t)$, before hitting the lower threshold, the decision is positive (if $\mu > 0$). However, if it first hits the lower threshold, that is, $X(t) = -b(t)$, the decision is negative. All parameters, drift rate $\mu(t)$, diffusion constant $\sigma(t)$, and thresholds $b(t)$ can be time-dependent, although most of the parameters are usually taken as time-independent.

The stochastic Euler (SE) method to simulate (1) is based on dividing the temporal interval of interest $[0, T]$ into small time steps of size Δt and advancing the initial state $X_0 = X(0)$ via

$$X_{n+1} = X_n + \Delta t \cdot \mu(t_n) + \sigma(t_n) \cdot Z(t_n) \cdot \sqrt{\Delta t}, \quad t_{n+1} = t_n + \Delta t \tag{2}$$

where $Z(t_n)$ follows a Gaussian distribution with zero mean and unit variance. Given an infinite number of such random processes $N_{tr} \to \infty$ simulated with an infinitesimally small time step $\Delta t \to 0$, the positive and negative first passage times define PDFs. As computer simulations will always have to limit to a finite number of trials $N_{tr} \in \mathbb{N}$ and discretize (1) with time step $\Delta t > 0$, the derived PDF will, of course, be an error-prone approximation.

In the following sections, different methods for approximating the PDFs are presented and compared. The focus is on efficient numerical realization, that is, on the question of gaining the best possible accuracy with the least possible computational effort.

2.2. Reformulation as an integral equation

The integral method for computing the first-passage time distribution of Wiener processes in the context of psychological research was first used by Heath (1992). Since then, this method has been successfully applied in numerous studies on RT modeling (Evans et al., 2020; Jones & Dzhafarov, 2014; Smith, 1995, 2023; Smith & Lilburn, 2020; Smith & Ratcliff, 2009, 2021; Smith, Ratcliff, & Sewell, 2014; Voskuilen et al., 2016). A detailed introduction to this method for cognitive modelers is provided by Smith (2000). In more detail, the diffusion model (2) can be reformulated as an Volterra integral equation of first kind (Durbin, 1971). However, this integral equation is not amenable to a simple numerical treatment because singularities complicate the approximation. Buonocore, Nobile, and Ricciardi (1987) reformulated the equation to a Volterra integral equation of second kind to remove the singularities. In a second work, the two-thresholds case has been considered (Buonocore, Giorno, Nobile, & Ricciardi, 1990). For a survey and an overview on the derivations and reformulations into the Volterra equation of second kind, we refer to Smith (2000). Here, by $P_{min}(t, x)$ and $P_{max}(t, x)$ we denote the probability densities at time t for reaching the lower and upper thresholds, when starting with state $x \in [-b, b]$ at initial time 0. The coupled set of integral equations is given by

$$\begin{aligned} P_{min}(t, x) &= -2\Psi[-b(t), t|x, 0] \\ &\quad + 2\int_0^t P_{min}(s, x)\Psi[-b(t), t| - b(s), s] \\ &\quad + P_{max}(s, x)\Psi[-b(t), t|b(s), s]\,\mathrm{d}s \\ P_{max}(t, x) &= -2\Psi[b(t), t|x, 0] \\ &\quad + 2\int_0^t P_{min}(s, x)\Psi[b(t), t| - b(s), s] \\ &\quad + P_{max}(s, x)\Psi[b(t), t|b(s), s]\,\mathrm{d}s \end{aligned} \tag{3}$$

with the Kernel function

$$\Psi(b(t), t|y, s) = \frac{f[b(t), t|y, s]}{2}\left(b'(t) - \mu(t) - \frac{b(t) - y - \int_s^t \mu(r)\,\mathrm{d}r}{t - s}\right), \tag{4}$$

and the transition density $f[x, t|y, s]$ for a process starting in y at time s reaching x at time t, while neglecting all thresholds:

$$f[x, t|y, s] = \frac{1}{\sqrt{2\pi\sigma^2(t - s)}}\exp\left(-\frac{(x - y - \int_s^t \mu(r)\,\mathrm{d}r)^2}{2\sigma^2(t - s)}\right). \tag{5}$$

The reformulation as an integral equation brings along the major advantage of the possibility for a direct discretization without the need for a large number of repeated samples. We give details in Section 3.2 and refer to Hackbusch (1995, Chapter 2) for an overview on the theory and numerical treatment of such integral equations.

2.3. Partial differential equation models

By applying Itô's Lemma (see Øksendal, 2000, Section 4) to the SDE (1), one derives the *Kolmogorov Equations*.[2] These are two partial differential equation (PDEs), namely the *Kolmogorov Forward Equation (KFE)*

$$\partial_t p(x, t) + \mu\partial_x p(x, t) - \frac{\sigma^2}{2}\partial_{xx} p(x, t) = 0 \tag{6}$$

and the *Kolmogorov Backward Equation (KBE)*

$$-\partial_t q(x, t) - \mu\partial_x q(x, t) - \frac{\sigma^2}{2}\partial_{xx} q(x, t) = 0. \tag{7}$$

The connotation of the solutions $p(x, t)$ and $q(x, t)$ to these two PDEs is the following. In case of the KFE, let us assume that $p^0(x)$ is the PDF of the states at initial time $t = 0$. Then, $p(x, t)$ denotes the probability of the process residing in the state x at time t. In contrast, for the KBE, the solution $q(x, t)$ is the probability that a state x at final time t will, at a future time $T > t$, be within the target set denoted by q^T. The names *forward* and *backward* describe the transportation direction of information: The KFE starts at time $t = 0$, where the initial probability $p(x, 0) = p^0(x)$ is prescribed and it evolves forward in time, while the KBE starts at time $t = T$ with terminal values $q(x, T) = q^T(x)$ and evolves backward in time.

Both approaches find their match in numerical realizations of diffusion modeling. The KBE is the starting point of fast-dm (Voss & Voss, 2007), while the random walk approach (Diederich & Busemeyer, 2003; Diederich & Oswald, 2016) is a numerical approximation of the KFE that is also used in Kiani and Shadlen (2009, Supplementary Materials). Further, Shinn et al. (2020) recently presented a computational framework also based on a direct discretization of the KFE.

In principle, the KFE and the KBE approach are interchangeable and can be used equally to calculate PDFs. They differ, however, in details of the application. The KBE is highly efficient if all parameters are time-independent, whereas the KFE approach allows more flexibility. In the following, we briefly discuss the use of the two PDE methods for deriving the PDF of first-passage times.

[2] These are, in statistical mechanics, called the *Fokker–Planck equations*, and they are also closely related to the *Black–Scholes* equations used in computational finance. They are the prototypical *diffusion-transport* problems in plain mathematical language.

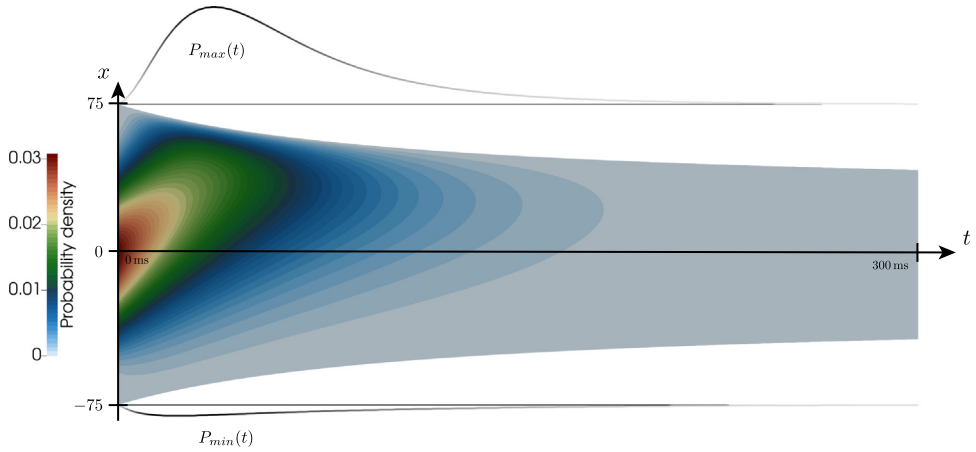


Fig. 2. Visualization of the KFE solution for the case of a time-dependent, collapsing threshold $b(t)$: The solution $p(x,t)$ indicates the probability that, at time t, the state resides in x. Here, one sees the drift towards the upper threshold. The probabilities for reaching upper and lower thresholds, $P_{max}(t)$ and $P_{min}(t)$, are the fluxes of the solution $p(x,t)$ at the thresholds. At time $t=0$, $p(x,0)$ is the probability density of the initial condition. For times $t>0$ the volume is less than one as the thresholds are absorbing (see Section 2.3.1).

2.3.1. Kolmogorov forward equation

Simple RT models describe the following process: At the initial time $t=0$ we start at $X(0)=x_0$, where x_0 might be in the neutral position $x_0=0$ or could be sampled from any probability distribution. We then model the probability of reaching either the lower or upper threshold $-b(t)$ or $b(t)$ at time $t>0$. In the KFE model, this corresponds to the initial value $p^0(\cdot):=\phi_0(\cdot)$, where ϕ_0 is a probability distribution.[3] If $T>0$ is the final time of observation, the domain of interest, where the PDE is solved, is given by

$$\Omega=\{(x,t)\in\mathbb{R}\times\mathbb{R}\mid 0<t<T,\ -b(t)<x<b(t)\}. \tag{8}$$

The thresholds $-b(t)$ and $b(t)$ are absorbing: Once we reach a limit, the process is terminated (i.e. absorbed). Thus, we are interested in the first time reaching the threshold (i.e., the first passage time). Hence, we complete the KFE by homogeneous Dirichlet data, that is, by setting the probabilities to zero on the thresholds $p(-b(t),t)=0$ and $p(b(t),t)=0$. Fig. 2 shows the solution of the KFE method for an example with a time-independent drift rate μ and time-dependent, collapsing thresholds $b(t)$. The PDE is solved in the complete domain Ω and in each point $(x,t)\in\Omega$, the solution $p(x,t)$, specified by its color, gives the probability that state x is reached at time t.

The initial probability $p^0(x)=\phi_0(x)$ is a probability density such that

$$\int_{-b(0)}^{b(0)} p^0(x)\,\mathrm{d}x=1.$$

Assuming for now that $b(t)=b$ is time-independent (the general case is presented in Appendix A.1), the integrated probability evolves by

$$\frac{d}{dt}\int_{-b}^{b} p\,\mathrm{d}x=\int_{-b}^{b}\partial_t p\,\mathrm{d}x=\int_{-b}^{b}\frac{\sigma^2}{2}\partial_{xx}p-\mu\,\partial_x p\,\mathrm{d}x, \tag{9}$$

which is seen by replacing $\partial_t p=\frac{\sigma^2}{2}\partial_{xx}p-\mu\partial_x p$, that is, by using the KFE (6). We integrate by parts[4]

$$\frac{d}{dt}\int_{-b}^{b} p\,\mathrm{d}x=\frac{\sigma^2}{2}\partial_x p(b,t)-\frac{\sigma^2}{2}\partial_x p(-b,t)\underbrace{-\mu p(b,t)+\mu p(-b,t)}_{=0},$$

and observe that the last two terms vanish as $p(b,t)=0$ and $p(-b,t)=0$ on the threshold. This shows that the probability distribution $p(x,t)$ does not always sum up to one but changes in time as

$$\frac{d}{dt}\int_{-b}^{b} p\,\mathrm{d}x=\frac{\sigma^2}{2}\partial_x p(b,t)-\frac{\sigma^2}{2}\partial_x p(-b,t)<0,$$

such that the integrated probability decreases with the states that cross the thresholds. This is illustrated in Fig. 2, where for larger times $t\to\infty$, the integrated probability $\int_{-b}^{b}p(x,t)\,\mathrm{d}x$ will turn to zero. The probability fluxes at the thresholds $-b$ and b are exactly the probabilities of reaching the lower and upper threshold, which we denote by $P_{min}(t)$ and $P_{max}(t)$. They accumulate in time by

$$P_{min}(t)=-\int_0^t\frac{\sigma^2}{2}\partial_x p(-b,t)\,\mathrm{d}t,\quad P_{max}(t)=\int_0^t\frac{\sigma^2}{2}\partial_x p(b,t)\,\mathrm{d}t \tag{10}$$

and these integrals are approximated with the trapezoidal rule. The first-passage time is then obtained by discretizing the KFE (6) in the restricted domain (8) and by integrating the differential Eqs. (10) at the thresholds. This is visualized in Fig. 2, where the fluxes are summed up and give the probability distributions $P_{max}(t)$ and $P_{min}(t)$ of reaching upper and lower thresholds at time t.

2.3.2. The Kolmogorov backward equation and its efficient realization in fast-dm

The KBE is the basis of fast-dm (Voss & Voss, 2007). To simulate the probability of reaching the upper threshold at time T, we define the target set (which will serve as terminal values to the KBE) as

$$q^T(x)=1 \text{ for } x=b \text{ and } q^T(x)=0 \text{ for } -b\le x<b. \tag{11}$$

[3] ϕ_0 could for instance be the Dirac distribution or the symmetric Beta(α,α)-distribution. The *Dirac delta function* $\phi_a(x)$ has the property of concentrating an integral on the point $a\in\mathbb{R}$, that is, $\int_{-\infty}^{\infty}f(x)\phi_a(x)\,\mathrm{d}x=f(a)$. It, therefore, models that we always start in the position $x=a$. Due to its very low regularity (e.g., it is not continuous), it cannot be considered a classical function, but must be regarded as a limit $\phi_\epsilon\to\phi_0$.

[4] Thus, using the simple rule $\int_a^b u'v\,\mathrm{d}x=-\int_a^b uv'\,\mathrm{d}x+u(b)v(b)-u(a)v(a)$ and choosing the constant function $v(x)\equiv 1$ with $v'(t)\equiv 0$ such that the first part vanishes.

Then, on the domain (8), but with time-independent thresholds for the beginning, we solve the KBE (7) backward in time, that is, from $t = T$ back to $t = 0$. The solution $q(x, t)$ denotes the probability that state x at time t will reach the upper threshold at time T (or earlier). For instance, $q(0, 0)$ denotes the probability that the neutral state at time $t = 0$ will reach the threshold latest at time T. The boundary and final-time conditions (11) introduce a discontinuity at $x = b$ for $t = T$ that will lead to reduced convergence of numerical approximation schemes. To deal with this, Boehm, Cox, Gantner, and Stevenson (2021) use splitting the KBE into a time-dependent one that allows for an efficient and accurate solution by a series expansion, and a remaining homogeneous KBE that has a smooth solution and allows for an accurate approximation with finite difference methods.

One benefit of the KBE approach is that no initial probability distribution must be chosen beforehand. Rather, the probability that any initial state $X_0 \in [-b, b]$ will reach the upper threshold at time T is given by $P_{max}(X_0) = q(X_0, 0)$. If the probabilities for reaching the lower threshold at time T are to be estimated, the KBE must be solved again, with the final state $q^T(x) = 1$ for $x = -b$ and $q^T(0) = 0$ for $-b < x \leq b$. Still, with only two KBE discretizations the probabilities for all initial distributions are obtained.

Yet another, and even more striking, problem is that the repeated approximation of the KBE is necessary to obtain the complete distribution $P_{max}(X_0, t)$ of reaching a threshold at a different time t. More precisely, we must solve the KBE backward in time from each initial value $q^t(\cdot)$ for $t \in [0, T]$. Then, by $P_{max}(X_0, t) = q^t(X_0, 0)$, the probability is given that state X_0 reaches the upper threshold at time t.

The success of fast-dm is based on a reformulation that is possible for time-independent drift rates and thresholds. In this case, the KBE is transformed into an equation that runs forward in time. By replacing $t \mapsto s - t$ we define $p^s(x, t)$ as

$$p^s(x, t) := q^s(x, s - t) \quad \Leftrightarrow \quad q^s(x, t) = p^s(x, s - t).$$

Thus, we replace $q^s(x, t)$ by $p^s(x, s - t)$ in the KBE (7) and get[5]

$$\partial_t p^s(x, s - t) - \mu(t)\partial_x p^s(x, s - t) - \frac{\sigma(t)^2}{2}\partial_{xx} p^s(x, s - t) = 0.$$

Finally, we relabel the variable and introduce $r := s - t$

$$\partial_r p^s(x, r) - \mu(s - r)\partial_x p^s(x, r) - \frac{\sigma(s - r)^2}{2}\partial_{xx} p^s(x, r) = 0. \tag{12}$$

If t runs backward in time from s to zero, the new variable $r = s - t$ will run forward from zero to s. This forward running KBE (12) is still equivalent to the original one (7).

The problem of this formulation is that it still depends on the final time $s > 0$ which has to be chosen a priori since it is required to evaluate the drift rate $\mu(s - r)$, the diffusion constant $\sigma(s - r)$, and also the thresholds $b(s - r)$ and $-b(s - r)$. If we, however, assume that these values do not depend on time, they also do not depend on s and the forward running KBE simplifies to

$$\partial_r p^s(x, r) - \mu\partial_x p^s(x, r) - \frac{\sigma^2}{2}\partial_{xx} p^s(x, r) = 0. \tag{13}$$

Since the final time s does not enter the equation, it is sufficient to solve (13) once from $r = 0$ to $r = T$: The probability of reaching the upper threshold when starting at X_0 is given by $P_{max}(X_0, t) = p(X_0, t)$. This is the approach that explains the efficiency of fast-dm.

However, if any parameter is time-dependent, this simplification is not possible and the KBE (7) must be approximated multiple times for deriving the PDF. While not the most beneficial approach, efficient repeated approximations of this equation are still possible, and we refer to Appendix C.2 for a discussion. There, we also discuss the proper approximation of rough initial and threshold values that was raised by Boehm et al. (2021, Section 2).

2.4. Random walks as an explicit discretization of the Kolmogorov forward equation

Diederich and Busemeyer (2003) and Diederich and Oswald (2016) presented a simple and highly efficient approach based on the forward transportation of probabilities by matrix vector multiplications. We only give a very brief introduction here for the purpose of classification and refer to the literature mentioned above for details.

Given time-independent thresholds $-b < x < b$, the state space is discretized into M equidistant steps of size Δx, namely $x_0 < x_1 < \cdots < x_M$. Then, the vector $P_n = (P_{n,1}, \ldots, P_{n,M}) \in \mathbb{R}^M$ indicates the probability distribution at time t_n. More precisely, the entry $P_{n,m}$ indicates the probability that at time t_n the state is within the interval $[x_{m-1}, x_m]$. In each time step $t_n \mapsto t_{n+1}$, these probabilities are transported $P_n \mapsto P_{n+1}$ by fluxes between adjacent states. This results in a very efficient algorithm, each step being equivalent to performing one matrix vector multiplication with a tridiagonal matrix.[6] The boundary fluxes at x_0 and x_M correspond to the probability of reaching upper and lower thresholds.

Denoting the transition probabilities by $a_{n,-}$ and $a_{n,+}$ (see Diederich & Oswald, 2016), the standard DDM is given by the choices

$$a_{n,-} = \frac{1 - \frac{\mu\Delta x}{\sigma^2}}{2\zeta^2}, \quad a_{n,+} = \frac{1 + \frac{\mu\Delta x}{\sigma^2}}{2\zeta^2}, \quad a_{n,0} = 1 - \frac{1}{2\zeta^2}, \tag{14}$$

with a parameter $\zeta > 0$ that controls the stability of the method. In each step $n \mapsto n + 1$ we compute

$$P_{n+1,m} = a_{n,-} \cdot P_{n,m-1} + a_{n,0} \cdot P_{n,m} + a_{n,+} \cdot P_{n,m+1}, \quad m = 1, \ldots, M. \tag{15}$$

This can be realized efficiently by one matrix–vector product with a tridiagonal matrix. Introducing $p_{n,m} := P_{n,m}/\Delta x$, the specific choices of $a_{n,*}$ from (14) allow to write (15) as the equivalent difference equation

$$\frac{p_{n+1,j} - p_{n,j}}{\frac{\Delta x^2}{\sigma^2\zeta^2}} = \frac{\sigma^2}{2} \cdot \frac{p_{n,j+1} - 2p_{n,i} + p_{n,j-1}}{\Delta x^2} - \mu \cdot \frac{p_{n,j+1} - p_{n,j-1}}{2\Delta x}. \tag{16}$$

Finally, introducing a time step Δt as

$$\Delta t = \frac{\Delta x^2}{\sigma^2\zeta^2} \tag{17}$$

reveals the equivalence of this method with the *forward Euler method in time – central difference in space* discretization of the KFE (6), as has already been pointed out in Appendix A.4 of Diederich and Busemeyer (2003). The condition $\zeta \geq 1$ (compare Diederich & Oswald, 2016, and also (17)), corresponds to the classical parabolic stability condition for explicit time stepping schemes $\Delta t \leq \Delta x^2/\sigma^2$ (Fort & Fankel, 1953, and see also Appendix B).

[5] With the chain rule it holds $-\partial_t q^s(x, t) = \partial_t p^s(x, s - t)$, that is, the sign changes.

[6] A matrix $A \in \mathbb{R}^{n\times n}$ is called tridiagonal, if each matrix row $i = 1, \ldots, n$ has at most three values that differ from zero: the diagonal entry A_{ii} and the values left and right of the diagonal $A_{i,i-1}$ and $A_{i,i+1}$, respectively. Linear systems with tridiagonal matrices can be solved efficiently with an effort that is comparable to the multiplication with the matrix itself.

2.5. Discussion on methods for deriving the first-passage time PDF

Section 2 began with introducing the SDE of the diffusion model and its implementation in simulations with the SE method. We then introduced the integral equation method (see also below) and turned more thoroughly to two PDE solutions, that is, the KFE and the KBE. While, in principle exchangeable, the KBE solution is highly efficient with time-independent parameters (as done in fast-dm; Voss & Voss, 2007), whereas the KFE allows for more flexibility (see also Shinn et al., 2020). An explicit discretization of the KFE is also achieved with random walks, as used in the approach by Diederich and Busemeyer (2003) and Diederich and Oswald (2016).

To derive the integral equation method, the SDE (1) is first transformed into a deterministic differential equation, or here into an integral equation, with the help of the Itô calculus. This will allow a direct discretization, that is, without repeated sampling. In contrast to the PDE approaches, the location variable is decoupled in the integral equation. The probabilities of the first passage times are described independently for each initial state $x \in [-b, b]$ at time $t = 0$.

We will now continue by presenting numerical realizations of these methods with a particular emphasis on the effort required to achieve a desired error reduction.

3. Numerical discretizations

The just introduced methods and their numerical realization (which we use synonymously with discretizations) must be considered as two separate matters (see also Remark 1). However, it is often the case that a discrete method is used. This is the case with random walks (Diederich & Busemeyer, 2003) and in general this approach, using discrete arguments, can be called the origin of numerical mathematics (Wanner, 2010).

In the following paragraphs, we will introduce and relate different discretization approaches for the above introduced methods. While there is not much freedom in terms of discretization for the SDE (see Section 3.1), the PDE methods described in Section 3.3 allow for different approaches. Some of them will prove to be equivalent, others have clear advantages or disadvantages.

To describe the differences between the methods quantitatively, we will consider two aspects. First, we address their convergence, that is, the reduction of the error depending on discretization parameters such as the time step size. Second, we will investigate the efficiency, that is, the relation between the achieved error and the required computational effort. In more detail, instead of comparing specific numbers, we will evaluate the methods in terms of their order of convergence. For example, linear convergence means that halving the step size Δt also halves the error, which is denoted as $\mathcal{O}(|\Delta t|)$. Quadratic convergence means that halving the step size Δt leads to a quartering of the error, which is denoted as $\mathcal{O}(|\Delta t|^2)$. More details, for example, which exact quantity is used to measure errors, are given in the following sections and in particular in Remark 2.

3.1. Discretization of the stochastic differential equation by the stochastic Euler method

In general, all discretization will start with a subdivision of the time interval $[0, T]$ into discrete steps of size Δt

$$0 = t_0 < t_1 < \cdots < t_N = T, \quad \Delta t = t_n - t_{n-1}. \tag{18}$$

Given subdivision (18), we aim to approximate the states $X(t)$ in the discrete points only and define

$$X_n := X(t_n), \quad n = 0, 1, \ldots, N.$$

Based on the discrete points in time, the SDE (1) is approximated by the SE method, which is also called the Euler–Maruyama method,

$$X_n = X_{n-1} + \Delta t\mu(t_{n-1}) + \sigma(t_{n-1})Z(t_{n-1})\sqrt{\Delta t}. \tag{19}$$

In a deterministic setting in which forward Euler approximation of the first derivative $X'(t_n)$ is of first order giving linear convergence $\mathcal{O}(\Delta t)$, the resulting SE method is only of order $\mathcal{O}(\Delta t^{\frac{1}{2}})$ due to the discretization of the diffusion term such that, to reduce the error by the factor two, even four times more time steps are required. In addition, by the *law of large numbers*, many samples of (2) are required to obtain accurate probabilities. A detailed analysis of the error is given in Section 3.5. An extension to higher order methods that have a better scaling in terms of the time step size Δt is in general not possible (see Øksendal, 2000, for more details).

3.2. Discretization of the integral equation

The coupled set of integral Eqs. (3) describing the probabilities $P_{min}(t, x)$ and $P_{max}(t, x)$ can be discretized by approximating the integrals with the trapezoidal rule. Hence, as in the case of the SDE (1) we introduce time steps $0 = t_0 < t_1 < \cdots < t_N = T$ (compare (18)). We approximate $P_{min,n}(x) := P_{min}(t_n, x)$ and $P_{max,n} := P_{max}(t_n, x)$. The integrals are approximated with the trapezoidal rule

$$\begin{aligned} P_{min,n}(x) &= -2\Psi^{-}_{n,0} + 2\Delta t \sum_{k=1}^{n-1}\left(P_{min,k}(x)\Psi^{--}_{n,k} + P_{max,k}(x)\Psi^{-+}_{n,k}\right) \\ P_{max,n}(x) &= -2\Psi^{+,0}_{n,0} + 2\Delta t \sum_{k=1}^{n-1}\left(P_{min,k}(x)\Psi^{+-}_{n,k} + P_{max,k}(x)\Psi^{++}_{n,k}\right) \end{aligned} \tag{20}$$

with vectors $\Psi^{-}, \Psi^{+} \in \mathbb{R}^{N+1}$

$$\Psi^{-}_n := \Psi(-b(t_n), t_n|x, 0), \quad \Psi^{+}_n := \Psi(b(t_n), t_n|x, 0), \quad n = 0, \ldots, N \tag{21}$$

and matrices $\Psi^{--}, \Psi^{-+}, \Psi^{+-}, \Psi^{++} \in \mathbb{R}^{(N+1)\times(N+1)}$

$$\begin{aligned} \Psi^{--}_{n,k} &= \Psi(-b(t_n), t_n| - b(t_k), t_k), & \Psi^{-+}_{n,k} &= \Psi(-b(t_n), t_n|b(t_k), t_k), \\ \Psi^{+-}_{n,k} &= \Psi(b(t_n), t_n| - b(t_k), t_k), & \Psi^{++}_{n,k} &= \Psi(b(t_n), t_n|b(t_k), t_k), \\ & n, k = 0, \ldots, N \end{aligned} \tag{22}$$

The probabilities obtained by (20) depend on the initial state $x \in [-b(0), b(0)]$. For a problem with a fixed starting point x, one approximation is sufficient. If the initial value, however, comes from a distribution, we also have to discretize the state space

$$-b = x_0 < x_1 < \cdots < x_M = b, \quad \Delta x = x_{m+1} - x_m. \tag{23}$$

This means that the approximated integral equation (20) must be evaluated $M + 1$ times and $P_{min,n,m} := P_{min}(t_n, x_m)$ and $P_{max,n,m} := P_{max}(t_n, x_m)$ are matrices of size $(M+1)\times(N+1)$. Given that $B_m := B(x_m)$ is the initial distribution, first passage time densities are obtained with help of trapezoidal rule integration in x-direction:

$$P_{min,n} = \Delta x \sum_{m=0}^{M} P_{min,n,m}B_m, \quad P_{max,n} = \Delta x \sum_{m=0}^{M} P_{max,n,m}B_m. \tag{24}$$

In contrast to the PDE methods, where spatial and temporal discretization interact, the integral approach is fully decoupled in space. This means that P_{min,n,m_1} and P_{min,n,m_2} can be computed independently from each other, which allows for efficient parallel implementations.

3.3. Discretizing partial differential equations

KFE (6) and KBE (7) are PDEs and the attribute *partial* refers to the fact that derivatives do not only appear in direction of time t, but also in direction of state x. Therefore, the PDE models further require a discretization of the state space $[-b, b]$. For simplicity, we here discuss time-independent thresholds only and refer to Appendices A.1 and A.2 for further details. Hence, let $-b = x_0 < x_1 < \cdots < x_M = b, \quad \Delta x = x_{m+1} - x_m$ be the splitting of the state space into M uniform steps similar to (23). The set of discrete space–time points (t_n, x_m) is called a mesh and the probabilities are approximated in these mesh points

$$p_{n,m} \approx p(t_n, x_m), \quad n = 0, \ldots, N, \ m = 0, \ldots, M.$$

Briefly stated, discretizations of PDEs require the approximation of time derivatives (t) and of space derivatives (x). Many combinations of such methods are possible and we show the most common ones for the KFE. The most simple *forward Euler – central difference approximation* reads

$$\frac{p_{n+1,m} - p_{n,m}}{\Delta t} + \mu \cdot \frac{p_{n,m+1} - p_{n,m-1}}{2\Delta x} - \frac{\sigma^2}{2} \cdot \frac{p_{n,m+1} - 2p_{n,m} + p_{n,m-1}}{\Delta x^2} = 0. \tag{25}$$

This is the discretization that is equivalent to the random walk approach. We call it *explicit*, as the new time step t_{n+1} only appears once and the new probability $p_{n+1,m}$ could be isolated on the left side of the equation. Explicit methods are very efficient as one step $n \mapsto n+1$ can be realized as one single matrix–vector product. The drawback of the forward Euler method in time is twofold: First, it is not very accurate, as it only converges linearly such that the approximation is limited to

$$p_{n,m} = p(t_n, x_m) + \mathcal{O}(\Delta t) + \mathcal{O}(\Delta x^2). \tag{26}$$

Second, the method lacks stability. Results are only useful if the parabolic stability condition holds for the time step (compare also the discussion at the end of Section 2.4 and in Appendix B)

$$\Delta t \leq \frac{\Delta x^2}{\sigma^2}.$$

Such time step conditions can be avoided by considering *implicit* discretizations. This means that all terms in (25) are evaluated in the unknown new time step t_{n+1} instead of the old time step t_n. The *backward Euler – central difference approximation* of the KFE then is

$$\frac{p_{n+1,m} - p_{n,m}}{\Delta t} + \mu \cdot \frac{p_{n+1,m+1} - p_{n+1,m-1}}{2\Delta x} - \frac{\sigma^2}{2} \cdot \frac{p_{n+1,m+1} - 2p_{n+1,m} + p_{n+1,m-1}}{\Delta x^2} = 0. \tag{27}$$

The direct consequence of an implicit discretization is that each time step now requires the solution of a linear system of equations since all $M+1$ states $p_{n+1,0}, p_{n+1,1}, \ldots, p_{n+1,M}$ must be determined at once. However, for the very simple models considered here, these systems of equations can be solved very efficiently as they are tridiagonal (which means that $p_{n+1,m}$ only depends on the direct neighbors $p_{n+1,m-1}$ and $p_{n+1,m+1}$, see also footnote 2.4) and the additional effort over an explicit discretization is negligible (see also the implementation provided at Richter et al. (2023)). While also being of low order with an error that satisfies (26), no time-step restriction is required. Because of this good stability, Shinn et al. (2020) suggest the backward Euler discretization for the case of time-dependent thresholds.

The classical second-order discretization is the *Crank–Nicolson scheme* that takes the states neither solely in the old time step t_n nor in the new one t_{n+1}, but in the midpoint

$$\frac{p_{n+1,m} - p_{n,m}}{\Delta t} + \mu \cdot \frac{p_{n+1,m+1} - p_{n+1,m-1}}{4\Delta x} - \frac{\sigma^2}{2} \cdot \frac{p_{n,m+1} - 2p_{n,m} + p_{n,m-1}}{2\Delta x^2} = - \mu \cdot \frac{p_{n,m+1} - p_{n,m-1}}{4\Delta x} + \frac{\sigma^2}{2} \cdot \frac{p_{n,m+1} - 2p_{n,m} + p_{n,m-1}}{2\Delta x^2}. \tag{28}$$

This method is implicit and stable such that no time-step restriction is required. Finally, it is second order accurate in space and time

$$p_{n,m} = p(t_n, x_m) + \mathcal{O}(\Delta t^2) + \mathcal{O}(\Delta x^2). \tag{29}$$

Because of these favorable properties, it is our method of choice and it is also implemented in fast-dm (Voss & Voss, 2007) and in the method provided by Shinn et al. (2020). In principle, discretizations of any order in space and time are available.

Fig. 3 visualizes how information is transported from one time-step to the next. To compute $p_{n+1,m}$ with the forward Euler method (left in the figure), only the known information of $p_{n,m}, p_{n,m-1}, p_{n,m+1}$ is required (and this is why we call the method explicit). To compute $p_{n+1,m}$ with the implicit backward Euler (middle in the figure), we need $p_{n,m}$, but also $p_{n+1,m+1}$ and $p_{n+1,m-1}$ that are not known from the last time step, but that have to be computed alongside with $p_{n+1,m}$. As $p_{n+1,m+1}$ will, in turn, depend on $p_{n+1,m+2}$ the complete column $p_{n+1,m}$ for $m = 0, \ldots, M$ must be computed at the same time in a coupled system of equations. The Crank–Nicolson scheme (right in the figure) can be considered as a combination of forward and backward Euler.

A thorough analysis reveals that the Crank–Nicolson method may not require a general time step condition, but still suffers from numerical stability problems since it possesses no sufficient *smoothing property* (Luskin, Rannacher, & Wendland, 1982). Theoretically, the Crank–Nicolson scheme is stable for all time steps $\Delta t > 0$. This, however, is only true if all calculations are done without any error. Yet, a computer simulation will always result in small rounding errors and these can be amplified with the Crank–Nicolson method. It is hence advisable to use a slightly shifted stabilized version of the Crank–Nicolson scheme as discussed in Appendix B.

3.4. Discussion on the methods and their discretizations

Fig. 4 shows the relations between the different methods and their numerical approximations. Note that the four different methods shown in the inner part of the sketch are mathematically equivalent when it comes to the description of the probabilities of reaching the upper or lower threshold. They are, however, the starting points for different discretization schemes, and not all of them are suitable for all possible extensions. Direct discretization of the SDE with the SE method is surely the most flexible approach; it is, however, not competitive in terms of efficiency and accuracy. This will be highlighted in the following section. The discretization of the integral equation by the trapezoidal rule gives highly accurate approximations and has been applied in very general settings and time-dependent configurations. The two formulations based on PDEs both allow for very efficient higher-order discretizations. The advantage of the KBE, which is exploited in fast-dm (Voss & Voss, 2007), is that one solution gives probabilities for all possible initial distributions such that usual parameter fitting tasks will come with a reduced number of free parameters. An application of the KBE to problems with time-dependent drift rates or thresholds will, however, call for multiple solutions and will substantially decrease the efficiency.

3.5. Comparing effort and accuracy of the models

Next, we compare the different discretization schemes shown in the outer layer of Fig. 4 in terms of their computational complexity. In other words, and more precisely, we ask: *What is*

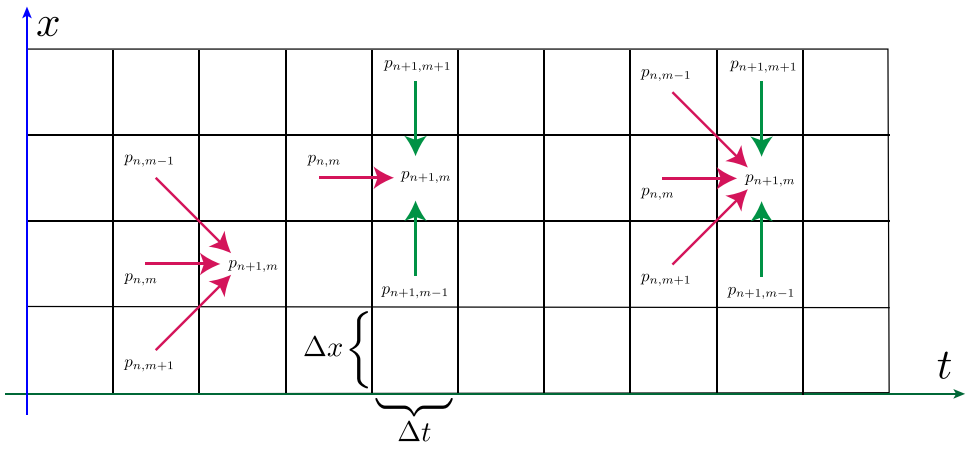


Fig. 3. Mesh of the space–time domain Ω and visualization of the flow of information in different discretizations of PDEs. Left: explicit forward Euler, Middle: implicit backward Euler, Right: implicit Crank–Nicolson. Explicit couplings are shown in red, implicit connections in green. If a scheme is implicit, all states $p_{n+1,m}$ within one column (i.e., for $m = 0, \ldots, M$) must be solved at once.

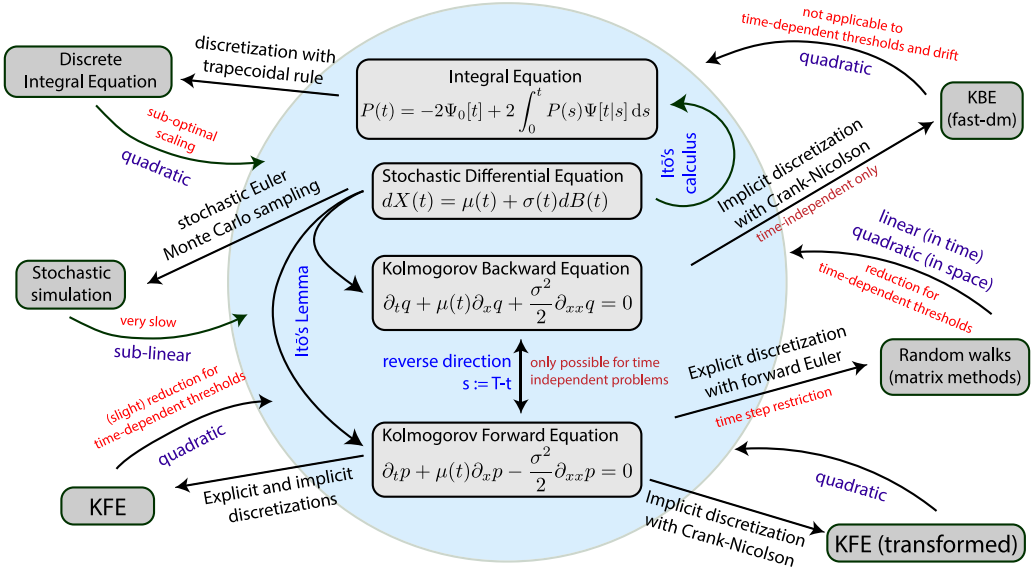


Fig. 4. Different methods (center) and their numerical approximations. By random processes we denote the direct discretization of the SDE, by KFE the approach presented in Shinn et al. (2020), by random walks the approach by Diederich and Busemeyer (2003), and by fast-dm, the framework of Voss and Voss (2007). By KFE (transformed) we denote the modification of the KFE solution described in the present work. The integral equation method is the one introduced in Buonocore et al. (1990). Blue labels indicate the technique that is used to show the relation between different models, black arrows refer to the discretization technique. In violet, we indicate the accuracy of the discretization and finally, in red, we highlight possible limitations.

the effort, measured in elementary operations, to determine the probability density with an error that does not exceed a bound $\epsilon > 0$? When we talk about errors, we always refer to the relative maximum error of the cumulative distribution function (CDF), see Remark 2 for details.

Instead of giving exact computing times, which would depend on the implementation and the hardware, we analyze the complexity with respect to the error ϵ. To be precise, the question is: *How much does the required effort grow if we aim at reducing the error ϵ by a factor of 10?* The results of the following analyses are summarized in Fig. 5.

We begin with the case of time-independent parameters. The SDE (1) is converging slowly for decreasing step sizes $\Delta t \to 0$ and increasing trial counts $N_{tr} \to \infty$. The error behaves like (Kloeden & Platen, 1999, Theorem 10.2.2)

$$E_{SDE}(\Delta t, N_{tr}) = \mathcal{O}\left(\Delta t^{\frac{1}{2}} + N_{tr}^{-\frac{1}{2}}\right). \tag{30}$$

The optimal balance between time step and number of trials is to choose $\Delta t = \epsilon^{\frac{1}{2}}$ and $N_{tr} = \epsilon^{-\frac{1}{2}}$. The effort of each trial is bound by $T/\Delta t$ such that for a total of N_{tr} trials we expect the complexity

$$C_{SDE}(\epsilon) = \mathcal{O}\left(T \cdot \epsilon^{-4}\right). \tag{31}$$

Reducing the desired tolerance ϵ by a factor of 10 thus calls for $10^4 = 10\,000$ times higher effort.

The integral equation method described in Sections 2.2 and 3.2 is based on the trapezoidal rule. It converges quadratically in space and time (Hackbusch, 1995, Theorem 2.2.12)

$$E_{IE}(\Delta t, \Delta x) = \mathcal{O}(\Delta t^2 + \Delta x^2). \tag{32}$$

Discretization of the integral equation calls for a nested time-loop such that the complexity is quadratic in time $N = (1/\Delta t)^2$ and linear in space $M = 1/\Delta x$. For the balanced discretization $\Delta x = \Delta t = \epsilon^{\frac{1}{2}}$ the complexity for reaching the accuracy $\epsilon > 0$ is given by

$$C_{IE}(\epsilon) = \mathcal{O}(T \cdot (b_{max} - b_{min}) \cdot \epsilon^{-\frac{3}{2}}). \tag{33}$$

To increase the accuracy by a factor of 10, the effort will increase by a factor of $10^{\frac{3}{2}} \approx 32$.

The integral equation method benefits from configurations with concentrated initial densities given by a Dirac distribution. In this case, no space discretization is required and the effort decreases to $(1/\Delta t)^2$. Hereby the complexity is reduced to

$$C_{IE}(\epsilon) = \mathcal{O}(T \cdot \epsilon^{-1})$$

which proves to be optimal. Decreasing the error by the factor 10 calls for 10 times higher effort.

The random walk approach can be interpreted as a *forward Euler – central difference approximation* of the KFE, and it converges to the PDE solution like

$$E_{RW}(\Delta t, \Delta x) = \mathcal{O}\big(\Delta t + \Delta x^2\big). \tag{34}$$

A simple proof in the time-independent case is provided by Grossmann, Roos, and Stynes (2007, Theorem 2.74). The stability restriction $\Delta t = \mathcal{O}(\Delta x^2)$ is optimal in terms of reaching the desired tolerance $\epsilon > 0$ with $\Delta t = \epsilon$ and $\Delta x = \epsilon^{\frac{1}{2}}$. The effort in each time step consists of 3 multiplications for transporting the probabilities and $N = T/\Delta t$ steps are required. This gives the overall complexity in terms of $\epsilon > 0$

$$C_{RW}(\epsilon) = \mathcal{O}\big(T \cdot (b_{max} - b_{min}) \cdot \epsilon^{-\frac{3}{2}}\big). \tag{35}$$

To increase the accuracy by a factor of 10, the computational effort increases by a factor of about 32.

In fast-dm, the trapezoidal rule, which is of second order, is used for time discretization combined with central second-order finite differences in space converging like

$$E_{fast-dm}(\Delta t, \Delta x) = \mathcal{O}\big(\Delta t^2 + \Delta x^2\big), \tag{36}$$

see Grossmann et al. (2007, Theorem 2.64) for a simplified setting and Rannacher (1984, Theorem 2) for a detailed discussion including the proper handling of the KBE's threshold and initial conditions. The error analysis given here also applies to the KFE. The trapezoidal rule is unconditionally stable and allows to pick the optimally balanced step sizes $\Delta x = \Delta t = \epsilon^{\frac{1}{2}}$. In each step, a tridiagonal system of equations must be solved, which is possible in six operations per spatial unknown (the values $p_{n+1,0}, \ldots, p_{n+1,M}$), twice as expensive as the explicit random walk process. As the trapezoidal rule allows for larger time steps Δt, the complexity scales like

$$C_{fast-dm}(\epsilon) = \mathcal{O}\big(T \cdot (b_{max} - b_{min}) \cdot \epsilon^{-1}\big). \tag{37}$$

Increasing the accuracy by a factor of 10 requires 10 times more effort. The KFE approach of Shinn et al. (2020) has the same complexity as fast-dm, which has the additional advantage over the KFE that it allows for choosing the initial distribution a posteriori.

The above conclusions no longer hold for time-dependent drift rate or thresholds. Nothing changes if the SDE is used to simulate discrete trials. Also the integral equation method can directly be applied to the general case (Smith, 2000) without a reduction of accuracy or an increased effort. Time-dependent thresholds are also covered in the random walk approach of Diederich and Oswald (2016). These authors observed a drop of the spatial convergence rate (34) to first order, such that the error only behaves like $\mathcal{O}(\Delta x + \Delta t)$. This collapse of accuracy comes from the brick-like discretization of the thresholds. Second-order convergence could be restored by resorting to extrapolation techniques (see also Grossmann et al., 2007, Section 2.4.4). The stability requirement $\Delta t = \mathcal{O}(\Delta x^2)$ leads to the sub-optimal scaling between desired tolerance and effort (the superscript *"td"* means time-dependent)

$$E^{td}_{RW}(\epsilon) = \mathcal{O}\big(T \cdot (b_{max} - b_{min}) \cdot \epsilon^{-3}\big).$$

Decreasing ϵ by a factor of 10 requires 1 000 times the effort.

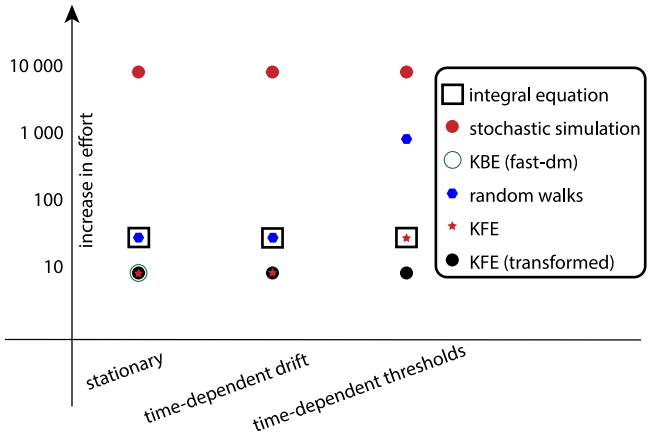


Fig. 5. Computational complexity of the approaches to compute the first-passage time PDFs in different settings. We indicate the increase in computational effort that comes with decreasing the desired accuracy by a factor of 10. By KFE (transformed) we refer to an alternative KFE approach using the modifications shown in Appendix A.1 for the time-dependent case.

While time-dependent thresholds and drift rates (or even the diffusion constant) will not worsen the convergence of the KBE methods used in fast-dm (36), they will hinder the efficient evaluation based on reverting the time. It will not be possible to solve the backward problem in one sweep. Instead, multiple solutions to each discrete point in time would be required, that is, from $1 \cdot \Delta t$ to 0, from $2 \cdot \Delta t$ to 0, etc. A total of $N = T/\Delta t$ solutions is required and the complexity becomes

$$C^{td}_{KBE} = \mathcal{O}\big(T \cdot (b_{max} - b_{min}) \cdot \epsilon^{-\frac{3}{2}}\big). \tag{38}$$

In Shinn et al. (2020), time-dependent thresholds combined with the second order Crank–Nicolson discretization of the KFE seems to cause an oscillatory behavior such that the first order Euler scheme (Grossmann et al., 2007, Theorem 2.74) must be used as fall-back option. The resulting overall complexity then reads

$$C^{td}_{KFE} = \mathcal{O}(T \cdot (b_{max} - b_{min}) \cdot \epsilon^{-\frac{3}{2}})$$

and increasing the accuracy by a factor of 10 calls for 32 times the effort.

In Appendix A.1, we demonstrate how the discretization of the KFE can be efficiently extended to all cases (including time-dependent drift rates and thresholds) without any degeneration of the complexity. This improvement might be purely academic as all approaches based on the two PDEs are highly efficient and the differences are not relevant in most practical applications.

4. A numerical study: Accuracy and efficiency

In this section, we demonstrate the numerical approximation properties of the KFE, random walks, the integral equation method, and the stochastic Euler approach in different cases and show that the computational effort is always linear with respect to the desired accuracy $\epsilon > 0$. This makes this approach robust in general settings.

We consider four different configurations. In the first case, drift rate μ, diffusion constant σ, and threshold b are all considered time-independent; in the second case, drift rate and threshold are time-dependent. The third test case is very similar to the second one, however, we consider the special case of time-dependent drift rates with large derivatives, that is, $|\mu'(t)| \gg 0$. Further, at $t = 0$ the initial probability is given as Beta-distribution over $[b_{min}(0), b_{max}(0)]$ with shape parameters $(5, 5)$.

In all cases, we consider the time limit $T = 1000$ and do not take decision times $t > T$ into account. Finally, in a fourth case, we take a Dirac distribution as initial value.

Case I For this case, all parameters are assumed as time-independent with the following values:

$$\mu = 0.5, \quad b_{min} = -75, \quad b_{max} = 75, \quad \sigma = 4$$

Case II For this case both drift rate and the threshold are time-dependent at the same time. The time-dependent drift rate $\mu(t)$ was modeled as in DMC (see Ulrich et al., 2015) and is the sum of the (time-independent) drift rate of the controlled process, μ_c, and the (time-dependent) drift rate of the automatic process, μ_a, thus:

$$\mu(t) = \mu_c + \mu_a(t)$$

More precisely, the (expected) time-course of the automatic process is modeled as a (rescaled) Gamma function, that is,

$$E[X_a(t)] = A \cdot \mathrm{e}^{-\frac{t}{\tau}} \left(\frac{t \cdot \mathrm{e}}{(a-1)\tau} \right)^{a-1}$$

with A being positive and negative in congruent and incongruent trials, respectively. The first derivative is then the drift rate $\mu_a(t)$:

$$\mu_a(t) = \frac{dE[X_a(t)]}{dt} = A \cdot \mathrm{e}^{-\frac{t}{\tau}} \left(\frac{t \cdot \mathrm{e}}{(a-1)\tau} \right)^{a-1} \cdot \left(\frac{a-1}{t} - \frac{1}{\tau} \right) \tag{39}$$

With $a = 2$ the drift rate then becomes

$$\mu(t) = \mu_c + A \cdot \mathrm{e}^{-\frac{t}{\tau}} \left(\frac{t \cdot \mathrm{e}}{\tau} \right) \cdot \left(\frac{1}{t} - \frac{1}{\tau} \right) \quad \text{with}$$
$$\mu_c = 0.5\,, A = 20,\ \tau = 150,\ \sigma = 4.$$

The time-dependent thresholds were modeled with a hyperbolic ratio function (see Voskuilen et al., 2016), that is, as:

$$b_{max}(t) = b_0 \left(1 - \kappa \cdot \frac{t}{t + t_{0.5}} \right) \quad \text{with}$$
$$b_0 = 75,\ \kappa = 0.6,\ t_{0.5} = 150\,\text{ms} \tag{40}$$

and $b_{min}(t) = -b_{max}(t)$.

Case III This test case is identical to **Case II** apart from the parameter τ. Instead of the fixed value $\tau = 150$, here we will consider different values in the interval $[5, 150]$. It turns out that especially small values of τ make the approximation more difficult. We will give an explanation for this behavior later.

Case IV The fourth case is again similar to **Case II**, but the initial value is fixed to $X_0 = 0$, whereas **Cases I**–**III** picked the initial from the smooth Beta-distribution. This irregular case is important in application (Kiani & Shadlen, 2009) and we also include this test case as the highly non-regular Dirac at time $t = 0$ is known to be troublesome to approximate (see also Boehm et al., 2021).

To compare the different approaches, we compute the PDF $p(t)$ of reaching the upper threshold at time $t > 0$. The "true" solution $p(t)$ is obtained by solving the KFE on a very fine discretization. For the different approaches, we compare the relative maximum error, as is detailed in Remark 2.

Remark 2 (*Realization*)**.** The KFE, the random walk approach, and also the integral equation method are implemented in Python and no parallelization is used. Coding is based on NumPy (Harris et al., 2020) and SciPy (Virtanen et al., 2020) such that all inner loops are performed efficiently in the corresponding C-backend (see also Appendix C.1). All these methods are run on a Macbook Pro with M1 Max CPU.

All SDE-based computations were coded in C++ and run on an AMD EPYC 7662 CPU and using 64 parallel cores. In the figures, we show the average error of 64 repetitions with the indicated number of $N_{tr} = \mathcal{O}(1/\Delta t)$ trials each.

To compare the accuracy of the different methods, the relative maximum-norm error of the CDFs for reaching the two thresholds is computed. This is the error norm used in all figures.

$$\text{Error} = \frac{\max_{n=0,\dots,N} |p_{CDF}(t_n) - p_{CDF}^{\text{exact}}(t_n)|}{\max_{n=0,\dots,N} |p_{CDF}^{\text{exact}}(t_n)|}. \tag{41}$$

The "exact" CDFs $p_{CDF}^{\text{exact}}(t_n)$ are approximated using the KFE approach on meshes that are refined both in space Δx and time Δt. The code to reproduce all examples is published on Zenodo (Richter et al., 2023).

Remark 3 (*Computing CDFs from PDFs*)**.** It appears trivial to transform a PDF to a CDF. Usually, it suffices to call the *cumsum* command of Matlab, Python, or similar packages and to multiply with the step size Δt. This, however, corresponds to the box rule, also called rectangle formula for numerical quadrature (Quarteroni, Sacco, & Saleri, 2007, Section 9.2.1) which is of first order only. Where second-order discretizations are used to discretize the KFE (as we do in the present approach) or the KBE (as done in fast-dm), it is also advisable to use higher-order quadrature (like the trapezoidal rule) when transforming the PDF into a CDF. Otherwise, the CDF accuracy would be reduced to first order only. Given that PDF(i) for $i = 0, 1, \dots, n$ is the probability in the time steps t_i, the trapezoidal rule for computing the CDF is:

$$\text{CDF}(n) = \frac{\Delta t}{2}\,\text{PDF}(0) + \sum_{i=1}^{n-1} \Delta t \cdot \text{PDF}(i) + \frac{\Delta t}{2}\,\text{PDF}(n).$$

In the following, we present the results of the four test cases described above to compare the performance of the different methods.

Case I. Fig. 6 visualizes the resulting error in the CDF versus the time step size Δt (upper row) and the computational efficiency, which is the computational effort (measured in seconds), as a function of the error (lower row). For each of the four methods, we choose the spatial discretization parameter respectively the number of trials such that an optimal balance is given, that is, for the SE method we choose $N_{tr} = 10\,000/\Delta t$, $\Delta x \approx \Delta t$ for the KFE and the integral equation method, and $\Delta x = \sigma/\sqrt{\Delta t}$ for the random walk.

The results show the expected order of convergence (upper row) with respect to the discretization parameter Δt. If we compare the complexity of the approaches, that is, the resulting error plotted over the computational effort, we observe that the SDE scales as expected (compare (31)). However, both the random walk approach and the KFE behave better than expected. The random walk scales linearly in $1/\epsilon$ instead of the expected $\mathcal{O}(\epsilon^{-\frac{3}{2}})$, while the KFE scales nearly like $\mathcal{O}(\epsilon^{-\frac{1}{2}})$ instead of the expected linear behavior $\mathcal{O}(\epsilon^{-1})$. We attribute this slight discrepancy to the very efficient C-backend of the NumPy library and refer to Appendix C.1 for a discussion.[7]

[7] The equations considered here are very small. For instance, $b_{max} = 75$, $b_{min} = -75$ and $\Delta x = 2.5$ means that only 60 probabilities $p_{n,m}$ must be identified in each time step. For such problems, a large fraction of the computational time will be spent on the interface between Python and the C-backend of NumPy, whereas the actual computations are very fast.

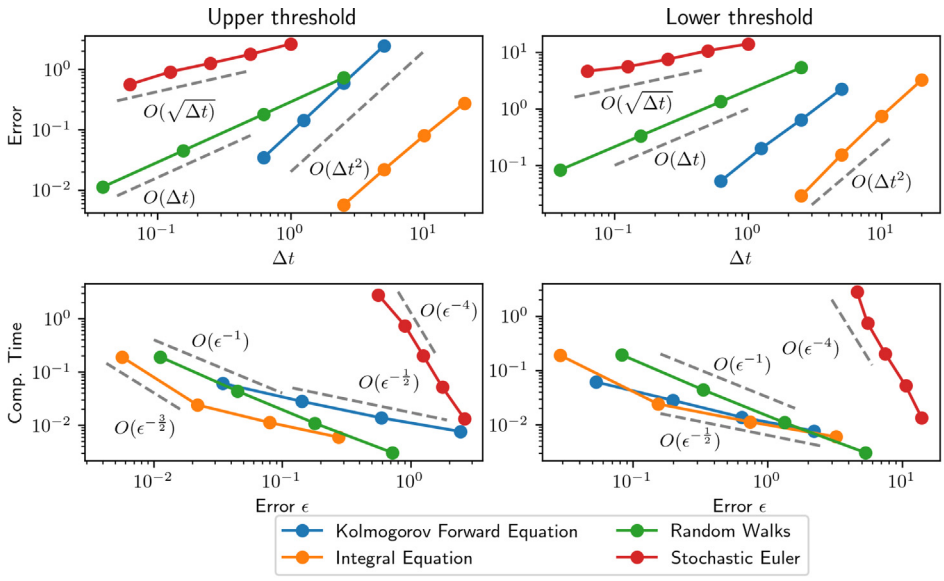


Fig. 6. Visualization of the results for **Case I**: Comparison of the error versus time step Δt (upper row) and the computational time required to reach the error ϵ (lower row). The dashed lines show the complexity.

It is clearly evident that the SE method for approximating the SDE is not competitive, particularly as the C++ implementation runs on a server with 64 parallel threads, whereas the other three approaches are sequential. Even though the integral equation and the KFE approach have a slight advantage over random walks, the differences are not substantial. Both, as well as fast-dm, which would behave here according to the KFE approach, provide an accuracy of 0.1% in less than a second.

Case II. This test with both time-dependent thresholds and drift rate will show that all methods except for the KFE discretization developed here will suffer from a reduced accuracy. We have argued in Section 2.3.1 that applying the KBE to time-dependent problems will call for multiple solutions and an increased effort. Hence, we do not consider the KBE. As noted in Diederich and Oswald (2016, Section 5.3), the random walk approach suffers from stability problems if time-dependent thresholds are considered and shows spikes in the PDFs whenever the threshold $b(t)$ crosses a mesh layer of size Δx. In Appendix A.2, we present our approach to remove these spikes. However, as noted in Diederich and Oswald (2016), the convergence rate drops to $\mathcal{O}(\Delta x)$. Since the parabolic time step restriction $\Delta t = \mathcal{O}(\Delta x^2)$ must still be satisfied, the random walk approach is not a balanced discretization and thus yields the non-optimal efficiency (see also Diederich & Busemeyer, 2003; Diederich & Oswald, 2016).

The KFE also calls for adjustments along the thresholds if these are time-dependent. In Shinn et al. (2020), the authors modified the space discretization in a way that is similar to the approach noted in Diederich and Oswald (2016). They also report an oscillatory behavior that is cured by replacing the second-order Crank–Nicolson method with the first-order backward Euler discretization. The benefit over the random walks is that the implicit backward Euler method is unconditionally stable and no time-step restriction must be satisfied.

In Appendix A.1, we introduce a more subtle modification of the KFE method for time-dependent thresholds that is based on a transformation of the problem onto one that has fixed thresholds. Although being mathematically more challenging, this approach is able to give accurate results without compromising efficiency. This trick is well-established when handling PDEs and the same transformation has been applied to the KBE by Boehm et al. (2021).

Fig. 7 visualizes the resulting accuracy for estimating the PDFs of reaching the upper threshold $b(t)$ and the corresponding computational efficiency. The upper panel shows the approximation properties of the different approaches and the results support the theoretical predictions: While the integral method and the KFE approaches maintain the approximation properties that they show for the case of time-independent problems, the random walk method falls back to linear convergence with respect to Δx which, by the time step restriction $\Delta t \approx \Delta x^2$, relates to $\Delta x^{\frac{1}{2}}$ convergence with respect to time. The lower panel in the figure shows the computational time that is required to reach the accuracy ϵ. A comparison of the results with that presented in the lower row of Fig. 6 shows the advantage of the modified KFE method, which is the only method which gives optimal complexity in case of time-dependent parameters.

So far, we have seen that in case all parameters are time-independent (**Case I**), the SDE approach is much less efficient than the other discretizations based on the integral equation or the PDEs, which are hardly distinguishable. Yet, with time-dependent parameters (**Case II**), the random walk approach becomes noticeably less efficient than the KFE approach. Here, both thresholds and drift rate are time-dependent simultaneously, however, the drop in accuracy of the random walk discretization is only due to the time-dependence of the thresholds. In fact, it has been shown in Diederich and Oswald (2016) and also in Shinn et al. (2020) that time-dependent drift rates alone do not cause any problems.

Case III. Time-dependent drift rates with a non-vanishing derivative at initial time $\mu'(0) \neq 0$ lead to excessive discretization errors. Stated briefly, this effect is not due to a special shape of the drift rate, and not due to non-vanishing derivatives at time $t = 0$, but only due to its maximum derivative in the whole interval. It is well known that the deterministic Euler method applied to the drift problem (no diffusion) $dX(t) = \mu(t)$ will yield an approximation with error behaving like

$$|X_n - X(t_n)| \leq C \cdot |\Delta t| \cdot \max_{t \in [0, 1000]} |\mu'(t)|,$$

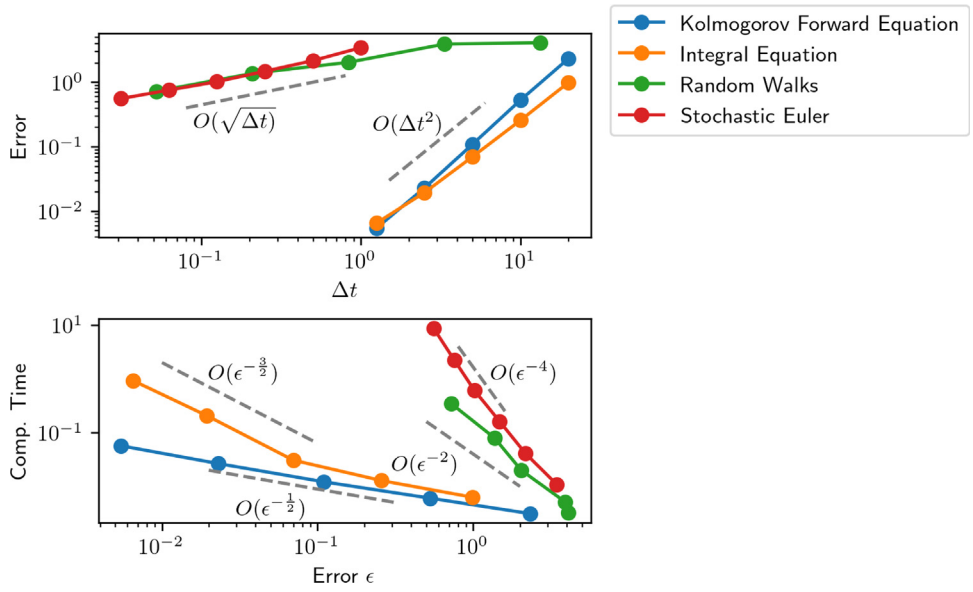


Fig. 7. Visualization of the results for **Case II**: Comparison of the error versus time step Δt (upper panel) and the computational efficiency, that is, required computational time for reaching the error tolerance ϵ (lower panel).

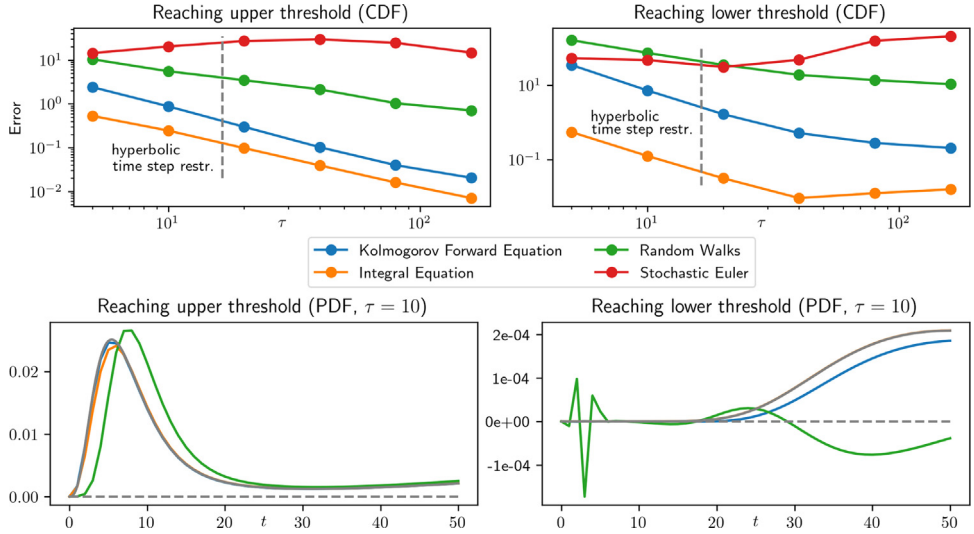


Fig. 8. Visualization of the results of **Case III**: Comparison of the performance of all approaches using time step size $\Delta t = 1$ for decreasing values of τ, which corresponds to an increasing derivative of the drift rate $|\mu'(t)|$. The upper row shows the error in the CDFs for reaching the thresholds as a function of τ. The lower row shows the PDFs for $\tau = 10$. The gray lines are the PDFs as derived from a highly resolved discretization. Note that the explicit random walk violates the hyperbolic time step condition $\Delta t \leq \Delta x/2 \cdot \mu(t)$ (for details, see Appendix B), what results in negative probabilities for the explicit random walk approach. For the lower threshold the results obtained with the integral equation method are indistinguishable from the exact solution. (Note that the SE method is not included in the lower row plots).

that depends on the maximum derivative of the drift rate over the complete interval (see Hairer, Nørsett, & Wanner, 2008, Theorem 3.1). The same error estimate will apply to the random walk approach which is also based on the forward Euler method. The KFE simulations are based on the second-order trapezoidal rule and thus, the error should behave as $|\Delta t|^2 \cdot \max_{t\in[0,1000]} |\mu''(t)|$ such that the second derivative of $\mu(t)$ is involved. The special form of $\mu(t)$ (via (39)) yields the dependency $|\mu'| = \mathcal{O}(\tau^{-1})$ and $|\mu''| = \mathcal{O}(\tau^{-2})$ such that small values of τ lead to large derivatives and larger errors.

Fig. 8 visualizes the error in reaching the thresholds depending on the choice of τ, that is, the parameter that determines the peak time of the Gamma function in DMC (Ulrich et al., 2015). The impact of small values of τ is clearly visible. For random walks, too small values of τ lead to a violation of the hyperbolic time step restriction

$$\Delta t \leq \frac{\Delta x}{2}\mu(t),$$

which results in an unstable discretization that even gives negative probabilities. The 2nd order implicit discretization used for KFE computations has no time-step restriction. However, dominant drift leads to what is called a *singularly disturbed transport diffusion problem* (Johnson, 2009, Chapter 9) and central differences (such as used in fast-dm and our mapped KFE approach) will call for a restriction on the spatial step size

$$\Delta x \leq \frac{\sigma}{2\max|\mu|}.$$

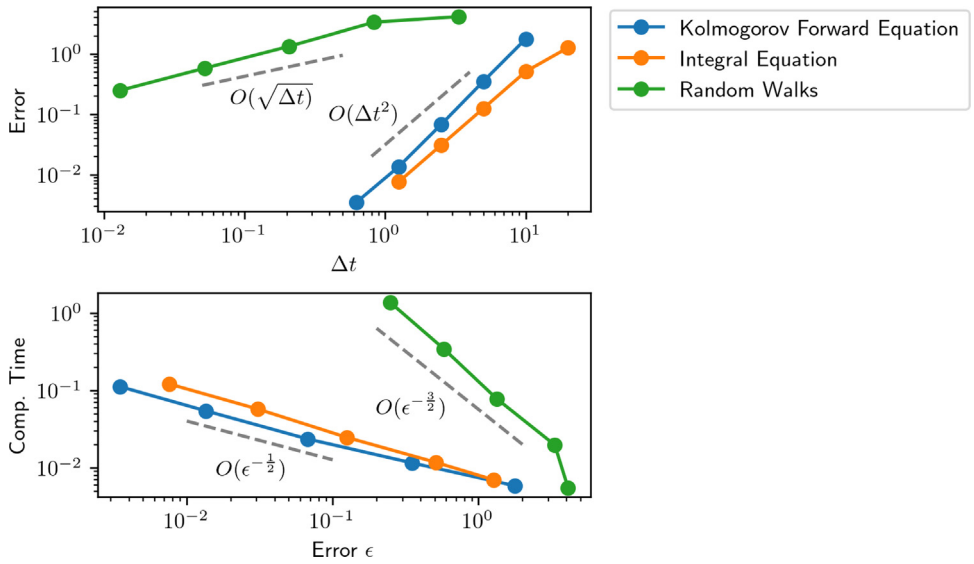


Fig. 9. Visualization of the results for **Case IV**: Comparison of the error versus time step Δt (upper panel) and the computational efficiency, that is, required computational time for reaching the error tolerance ϵ (lower panel).

If this condition is violated, also implicit discretizations might produce oscillatory solutions with negative probabilities. By modifying the spatial discretization, for example, using an upwind method instead of central differences (Johnson, 2009, Chapter 9), stability is preserved at the cost of reduced convergence, yielding $\mathcal{O}(|\Delta x|)$ instead of $\mathcal{O}(|\Delta x|^2)$. Fig. 8 shows that the approximation properties of random walks, KFE, and the integral equation method show large differences. Especially the error of the integral method is far below the errors of the other two approaches (and of course far below the error of the stochastic simulation). However, all calculations were performed at very small step size $\Delta t = 1$ ms and the picture changes when simulation times are taken into account. While a simulation with random walks needs about 0.0075 s, it is 0.058 s for the KFE and 0.72 s for the integral method.

Case IV. In this test case, the initial distribution is a Dirac distribution. In terms of the PDE methods, this means that the initial value is a highly irregular function, and the solution will have a singularity at time $t = 0$. As discussed in Boehm et al. (2021), such irregular initial might lead to problems in the discretization and an increased error. For the KBE, Boehm et al. (2021) suggested splitting the PDE into a regular part which can be easily discretized with standard finite difference approaches and into the irregular part which is approximated in a fast converging sequence. Another reason to consider this test case is the construction of the integral equation method, which benefits from concentrated initial values, because here a discretization of the location variable can be omitted.

In Fig. 9 we show the approximation error and the computational efficiency for the integral equation method, random walks, and the KFE. The results seem surprising at first glance, as they do not reflect the assumed difficulties of the KFE method due to the irregular initial distribution. In the following, we will explain this behavior.

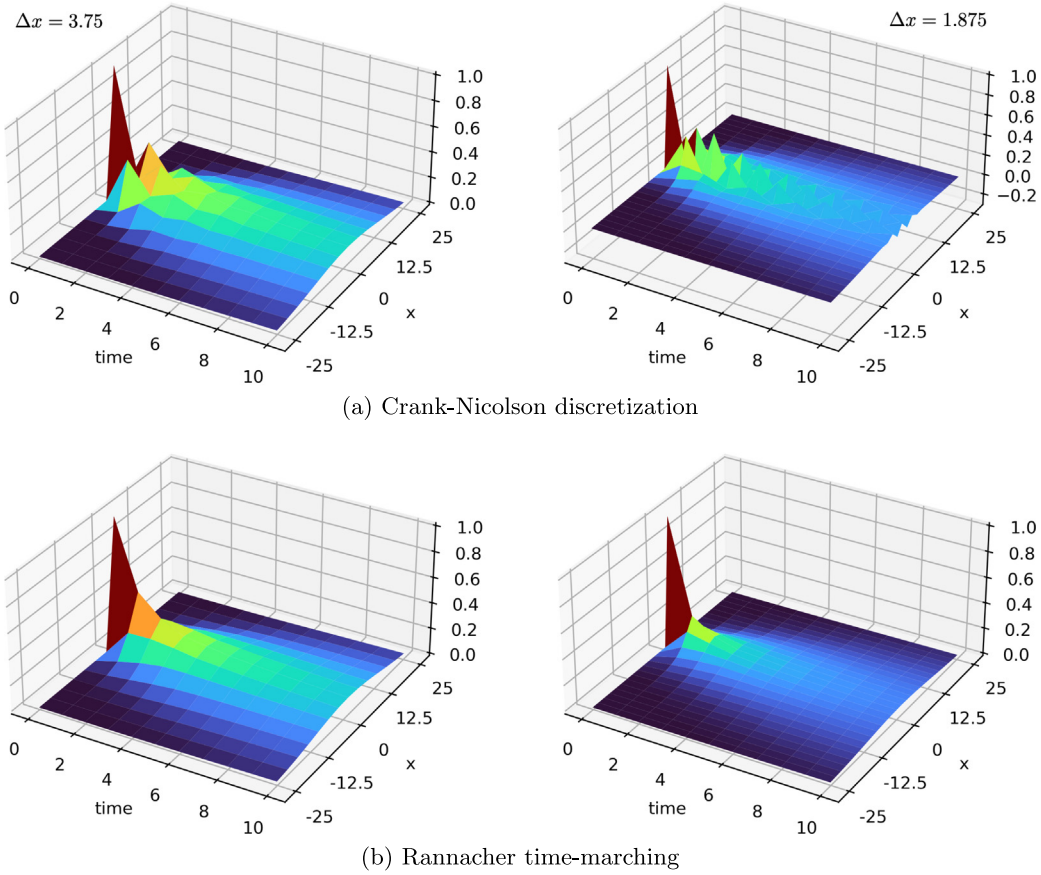
The KFE, as a simple linear elliptic problem, has a smoothing property (Evans, 2010, Section 2.3). Even if the initial value $p_0(x)$ is irregular (e.g., a Dirac as considered here), the solution $p(x, t)$ is smooth for all $t \geq t_0 > 0$. This means that the solution will be differentiable and easy to approximate as long as we "stay away from zero".

A numerical time-stepping scheme must also have this smoothing property. The Crank–Nicolson method does not have this smoothing property and requires additional stabilization. This can be achieved either by the variant discussed in Appendix B or, even better, by a modification called *Rannacher time-marching* (Rannacher, 1984): The first couple of steps (usually two or four) are replaced by approximations with the implicit backward Euler discretization. It can then be shown that this modified Crank–Nicolson approximation applied to rough initial data gives the error estimate

$$\|p(t_n) - p_n\| \leq \frac{1}{t_n^2}\Delta t^2. \tag{42}$$

The error can still "explode" for $t_n \to 0$. The convergence is, however, of second order for all positive times. This approach has been refined by Giles and Carter (2006) and extended to Dirac initials. They suggest replacing the first Crank–Nicolson step by four backward Euler steps with step size $\Delta t/4$. In the context of the drift-diffusion problem, the error singularity for $t_n \to 0$ indicated by (42) plays no role as the solution $p(x, t)$ is only evaluated along the thresholds and high frequent errors are quickly damped before those are reached. The Rannacher time-marching procedure is the standard method in the field of computational finance for approximating the Black–Scholes equation, a variant of the KBE. The analysis for the most simple case that would refer to zero-drift and constant thresholds is given in Rannacher (1982), and the general case that directly applies to the KFE is discussed in Rannacher (1984).

In Fig. 10 we demonstrate that such stabilizing modifications are not even required for the typical application of the KFE. In the lower row, we show the smooth results for Rannacher time-marching. The time step is chosen as $\Delta t = 1$ and the spatial step size as $\Delta x = 3.75$ on the left and $\Delta x = 1.875$ on the right. Although the Crank–Nicolson method (upper row) shows highly oscillating errors, these are quickly damped. This is why also the non-stabilized Crank–Nicolson method provides the full approximation order: The probabilities are only needed at the thresholds $-b(t)$ and $b(t)$ and until the information from the initial value has substantially reached these margins, the standard Crank–Nicolson method has sufficiently smoothed out the oscillations coming from the initial disturbance. The explicit Euler scheme, which is the basis of the random walks approach is only conditionally stable when $\Delta t \leq \Delta x^2/\sigma^2$.



(a) Crank-Nicolson discretization

(b) Rannacher time-marching

Fig. 10. Case IV: Smoothing property for a Dirac initial distribution. We show the solution $p(x, t)$ on a small subset of the domain for time $t \in [0, 10]$ and $x \in [-25, 25]$. The oscillations have no effect along the thresholds as the solution is already smoothed out there.

4.1. Summary and discussion of the numerical study

The results of the numerical study show that, in principle, all methods and discretizations are suitable to efficiently determine the PDF of the first-passage time. As one could expect, the only exception was the SE method, which clearly fell behind the other methods regarding accuracy and efficiency.

Yet, some differences also need to be noted for the other methods and discretizations. In terms of accuracy, the KFE, random walks, and the integral equation method are comparable if the discretization parameters, that is, the time step size Δt and the space step size Δx are chosen optimally. However, this picture changes when using time-dependent thresholds. In this case, only the KFE and the integral equation method maintain accuracy.

A second aspect is the efficiency, that is, the required computing time to achieve the desired accuracy. In case of a smooth initial distribution, only the KFE is optimal and the effort scales linearly with the error. For the other methods, the computational time increases faster when increasing the accuracy requirements. For the integral equation method, however, this reduced efficiency is marginal and only of theoretical importance.

In the next section, we will use the (transformed) KFE method to estimate parameters to empirical data.

5. Data fitting

In this section, we apply the KFE discretization developed in the present article to fit DMC (Ulrich et al., 2015), a diffusion model with a time-dependent drift rate, to empirical data. As the thresholds are time-independent, the KFE method described in Shinn et al. (2020) is also applicable without any restriction.

Strictly speaking, we have so far only considered the decision part of a response, and thus the PDF as derived from the considered methods is that of the decision time. However, RTs properly comprise perceptual and motor response parts as well. These two contributions are typically summarized as the residual (or non-decision) time. DMC assumes the residual time as normally distributed with expected value μ_R and standard deviation σ_R which is added to the decision time to yield RTs. Thus, the PDF of RTs in DMC is the convolution of the PDF of the decision time and a normal distribution.

5.1. Simon and Eriksen task data from Ulrich et al. (2015)

Ulrich et al. (2015) reported an experiment with $n = 16$ participants performing both a Simon task and an Eriksen flanker task. Briefly, the letters H and S served as the imperative stimuli requiring a left or right response. In the Simon task, the letters were presented to the left or right of the fixation cross, and a trial was congruent when the stimulus and the correct response were on the same side. In the flanker task, the imperative stimulus (i.e., the target) appeared centrally, and two letters appeared on its left and right side as flankers. A trial was congruent, if the target and flankers were the same letters.

Overall, a congruency effect was observed for RTs and error rates, and it was roughly of the same size for both tasks. However, the flanker congruency effect became larger with longer RTs (i.e., a positive delta function) and the Simon congruency effect became smaller with longer RTs (i.e., a negative delta function). These results are visualized as filled circles in Fig. 11. Accordingly, both tasks differed in their estimated values of the peak of the Gamma function, that is, the Simon data yielded a smaller value of τ.

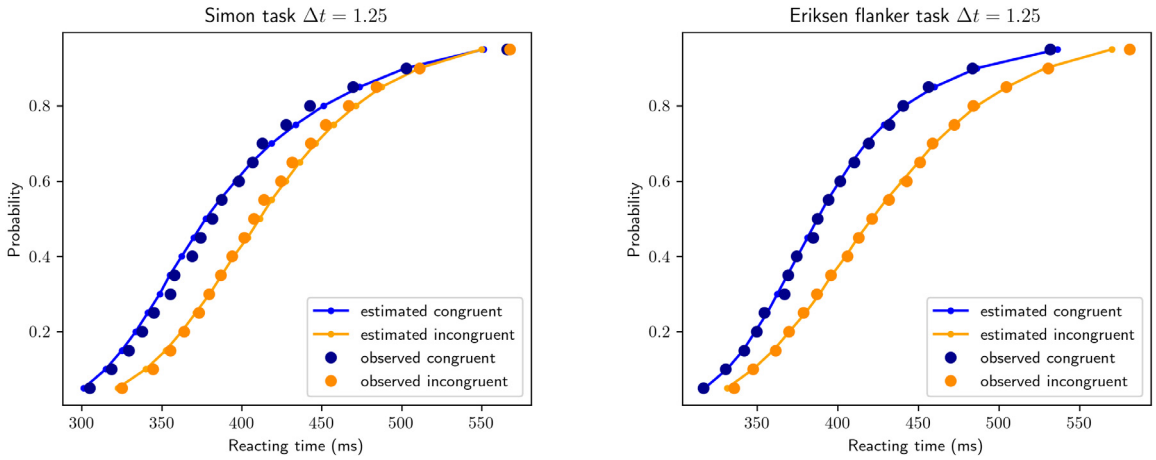


Fig. 11. Visualization of the observed data provided by Ulrich et al. (2015) (filled circles) separately for congruent (blue) and incongruent (orange) trials. The solid lines are the estimated CDFs that result from the fitting procedure described in Section 5.2. The computations are obtained with the KFE using the time step size $\Delta t = 1.25$ ms (see also Tables 1 and 2).

Table 1
Fitting results for the Simon task: The table shows the dependency of the identified parameters on the discretization parameter Δt (with $\Delta x = \Delta t$ and provides the required computational time in the last column. The results are in good agreement with the ones published in Ulrich et al. (2015), here provided in the last row labeled *reference*.

Δt	σ	a	b	μ_R	σ_R	A	μ_c	τ	α	Time
5	4.25	2	52.4	321.0	30.3	21.0	0.63	49.8	1.99	44 s
2.5	4.25	2	54.1	321.1	29.9	19.0	0.62	48.1	2.06	90 s
1.25	4.25	2	58.8	319.2	33.0	14.5	0.69	46.9	1.95	188 s
0.625	4.25	2	58.1	321.2	33.4	14.6	0.69	50.2	1.73	434 s
0.3125	4.25	2	57.2	322.4	35.7	15.0	0.69	45.0	1.97	909 s
reference	4.25	2	54.6	322.8	38.6	16.0	0.69	34.9	2.80	

Table 2
Fitting results for the Eriksen flanker task: The table shows the dependency of the identified parameters on the discretization parameter Δt (with $\Delta x = \Delta t$ and provides the required computational time in the last column. The results are in good agreement with the ones published in Ulrich et al. (2015), here provided in the last row labeled *reference*.

Δt	σ	a	b	μ_R	σ_R	A	μ_c	τ	α	Time
5	3.98	2	49.9	336.6	29.0	20.2	0.56	57.4	2.07	48 s
2.5	3.98	2	52.4	335.2	30.4	18.7	0.57	55.0	2.13	92 s
1.25	3.98	2	54.6	334.6	30.0	19.2	0.59	52.8	1.98	182 s
0.625	3.98	2	45.4	339.6	28.9	17.0	0.44	90.3	1.77	459 s
0.3125	3.98	2	49.1	338.1	27.4	17.8	0.50	96.3	1.27	949 s
reference	3.98	2	51.3	331.8	36.6	19.2	0.69	118.3	2.15	

5.2. Fitting procedure and results

To fit the parameters to the data, we simplified the procedure of Ulrich et al. (2015) and minimized the weighted least squares errors of the CDF measured in 20 equally spaced bins, but did not additionally identify the conditional accuracy function (which can be derived from the full CDF). The problem parameters are similar to those of **Case II** described in Section 4. In particular, the drift rate is parameterized as described in (39). The exponent was fixed as $a = 2$. The thresholds are time-independent. The diffusion constant was fixed to those values provided by Ulrich et al. (2015, Table 2), that is, $\sigma = 4.25$ for the Simon task and $\sigma = 3.98$ for the Eriksen flanker task. The parameter σ cannot be identified along with the threshold b and the drift rate μ as the resulting system would be under-determined, for this see Ulrich, Schröter, Leuthold, and Birngruber (2016) and also Remark 4 in Appendix A.1. The initial distribution is a Beta-distribution with parameter α. α and all further parameters, that is, the threshold b, residual time parameters μ_R and σ_R, as well as μ_c, A and τ, are left free for the optimization procedure, which is a simple Nelder–Mead solver as implemented in Python's SciPy library.

The identified parameters confirm those reported by Ulrich et al. (2015), however, with slight differences in the exact values (particularly regarding α and τ). To understand the possible origin of this discrepancy, we focus on the dependency of the optimization results on the accuracy of the computations. Tables 1 and 2 list the identified parameters for discretizations of increasing fineness. The last row of each table gives the reference parameters as identified by Ulrich et al. (2015). While the identified parameters still change drastically for $\Delta t > 1$, they converge for finer discretizations. The last column of both tables indicates the complete time used for the optimization procedure. In both cases, Simon and Erikson flanker, the solver required about 1 000 evaluations of the goal functional, corresponding to 1 000 approximations of the underlying KFE.

In Fig. 6, we can see that for obtaining the KFE-accuracy at $\Delta t \approx 1$ ms, the random walk approach would require $\Delta t \approx 0.1$ ms and stochastic simulations would even require $\Delta t \ll 0.01$ ms with a corresponding increase in computational time. In particular, using the SE method is not feasible to obtain a sufficiently accurate PDF for accurate parameter identification.

6. Conclusion

The present study's purpose was to compare various approaches to approximate the first-passage time distribution, with

a focus on contemporary diffusion models with time-dependent parameters, usually time-dependent drift rates or time-dependent thresholds. On the one hand, we discussed different methods for obtaining the PDF of the first-passage times of a diffusion process; on the other hand, we discussed the corresponding discretizations of these methods.

The most simple method is the SE, which simulates single trials and whose repeated application yields the PDF. In contrast, the transition with Ito's Lemma leads to the two Kolmogorov equations, two PDEs that can also be used to approximate the PDF. Alternatively, the SDE can be reformulated as an integral equation.

The SE method is the most common and flexible discretization technique for SDEs. Although it is well established and has been used often, the present paper shows that there are always better alternatives that provide a good approximation far more efficiently, at least for those applications discussed here.

The two PDE approaches differ in several respects. First, they differ in their character: In an explicit method, each time step can be written as a matrix–vector multiplication (e.g., random walks), while implicit methods require solving a linear system of equations (e.g., fast-dm or the method we pursued in the present paper). The advantage of implicit methods is that no condition for spatial and temporal step sizes has to be respected; the discretization is always stable. Second, the methods differ in the approximation order. While the Euler methods converge linearly (halving the step size Δt halves the error), the trapezoidal rule is quadratic (halving the step size Δt quarters the error). The explicit Euler method is used in the case of random walks, and the implicit Euler method is the fallback option in Shinn et al. (2020) when applied to time-dependent thresholds. fast-dm, KFE in the case of not time-dependent thresholds, and our KFE approach based on a remapping of the time-dependent domain use the second-order trapezoidal rule. This gives smaller errors for the same number of time steps, resulting in higher computational efficiency.

In the case of time-independent parameters, the methods based on PDEs (i.e., random walks, fast-dm, KFE, our approach) provide almost the same efficiency. Only the SE method falls far behind and is not competitive. Consider, for instance, Fig. 6. There we show that all PDE methods generate approximations with less than 1% error within 0.01 s computational time, whereas the SE method requires more than 1 s. If we aim at reducing the error to 0.1%, the PDE approaches still manage to produce the result in less than 0.1 s, whereas the SE method was not able to reach such low errors at all. An extrapolation of the results from Fig. 6 suggests that the SE would take more than 1 000 s. Finally, identifying parameters requires many computations, in our case, a good 1 000 runs each. We will now turn to the point of time-dependent parameters.

6.1. The efficiency of different methods to realize DDMs with time-dependent parameters

According to our analyses, especially the discretization of the integral equation and the KFE can be used flexibly for applications with time-dependent parameters while providing very high accuracy simultaneously. This, however, requires a suitable modification of the KFE discretization to cope with time-dependent thresholds. Modifications of the finite difference approximations as suggested in Diederich and Oswald (2016) or Shinn et al. (2020) were, in fact, not able to fully preserve the approximation property and provide a stable result at the same time. Hence, we describe in Appendix A.1 a transformation of the KFE that allows for approximations with optimal accuracy (and efficiency) in that case as well.

In contrast, if only the drift rate is time-dependent, no further modification is necessary. In this case, all PDE approaches (including random walks) can approximate the diffusion problem with the same efficiency and accuracy.

The integral equation gives a robust approximation with high accuracy. Theoretically, its efficiency is non-optimal as its effort scales quadratically with the inverse of the time step size. In usual applications, however, such small time steps have little relevance.

6.2. Advice for practical applications

The analyses in this study provide information about what methods can or should be used for specific problems. The most straightforward point in this regard is that it is always advisable to use an approach based on one of the two PDEs or the integral equation. The SDE, in contrast, should only be the fallback option in cases, where no PDE or integral equation is known. In general, the PDE should be discretized with an implicit time-stepping scheme. The additional effort is negligible, as only linear systems with tridiagonal matrices must be solved. The benefit of implicit methods is the stable discretization independent of time and spatial step sizes and the possibility of smooth approximations for rough initial and boundary conditions.

Some differences apply, however, about whether parameters are time-independent or not. If the parameters are time-independent, the KBE approach, such as implemented in fast-dm, is optimal and has the additional advantage over the KFE methods that the initial distribution can be chosen a posteriori. If, however, this feature of the initial distribution is not required, both the KBE and KFE are equivalent. In general, with time-independent parameters, all methods under consideration in the present study are highly efficient and well-suited, except for the SE methods. Even if random walks are theoretically somewhat less well-performing, this is not very important in practice since the computation times are short anyway. Furthermore, the method impresses with its direct access close to the mathematical model, which allows a very simple intuitive implementation.

The integral method is the one with the smallest error constant and it gives the best results for a certain time step size Δt. This largely compensates for its non-optimal efficiency in applications.

In the case of time-dependent parameters, however, the integral approach and the KFE are more flexible and efficient than the other methods. By adjusting the discretization scheme along the thresholds, the KFE preserves its good accuracy and stability and is the only method with theoretically optimal efficiency in all cases.

Parameter fitting applications should have the accuracy of the underlying discretization procedure in mind. As shown in Tables 1 and 2, the values of the identified parameters strongly depend on the quality of the approximation. As a rule, parameter fitting should always be repeated on three discretizations with increasing fineness. The result should only be accepted if the identified parameters are within acceptable ranges.

6.3. Outlook

In this paper, we have focused on solving a single SDE $X'(t) = f(t) + dW(t)$. However, in some cases, the solution of a system with multiple stochastic differential equations is sought, for example, in the leaky competing accumulator model (Usher & McClelland, 2001). In addition, other application cases do not fall into the framework covered by this article. For instance, Koob, Ulrich, and Janczyk (2023) used the diffusion model in a multitasking context and had a second diffusion process start once the first process exceeded the threshold.

In such cases, there is no immediate transition to a PDE that directly gives the PDF. Instead, one will probably need to continue relying on the SE method to analyze the properties of such a system, which, however, comes with extremely long computation times and reduced accuracy. In this case, substantial acceleration would be possible using modern computer hardware such as accelerator cards (GPU).

A further point for future development concerns the optimization algorithms used. Often, and here as well, simple optimization schemes are used in parameter fitting tasks, such as the Nelder–Mead method (Nelder & Mead, 1965). As gradient-free methods, these methods are straightforward and versatile to use, but they converge slowly and require many repeated calculations of the PDF. Alternatively, the gradient descent method or Newton-type methods can be used, for example, BFGS (Kelley, 1999), which can solve the data fitting problem with substantially fewer iterations. On the other hand, these optimization methods require the derivatives of the PDF concerning the parameters given by further PDEs, so-called adjoint and tangent equations that must be approximated. Such an approach is discussed by Hartmann and Klauer (2021) using the integral approximation of the PDF.

Data availability

Data and software for the reproduction of results is available in a Zenodo repository (Richter et al., 2023) at https://doi.org/10.5281/zenodo.6970739.

Appendix A. Handling time-dependent thresholds

Using time-dependent thresholds $b_{min}(t)$ and $b_{max}(t)$ gives rise to several technical difficulties if any method is used, except for the SE discretization of the SDE (this, however, is always highly inefficient). The problems of the other approaches result from the domain Ω, which is not of tensor-product type, but rather the state-range $b_{min}(t) < x < b_{max}(t)$ changes for every t. Diederich and Oswald (2016) tackled this problem by considering a brick-like discretization of the domain, including uniform state steps of size Δx until the threshold is reached. A technical modification is required if the bricks do not precisely represent the threshold. One such approach is provided in Section 5.3 of Diederich and Oswald (2016), and a brief explanation is provided in Appendix A.2. Although robust results can be achieved, we still observe a drop in accuracy, as discussed in Section 3.5. First, in Appendix A.1, we will show how the KFE can be applied to the case of time-dependent thresholds without any loss in approximation quality.

A.1. Solving the Kolmogorov forward equation with time-dependent thresholds in optimal complexity by a reference map

A uniform finite difference approximation of the KFE or the KBE would naturally be prone to the same reduction of convergence rates. Instead of applying complex techniques of higher order threshold reconstruction (see Grossmann et al., 2007, Section 2.4.4), we suggest an alternative based on reformulating the KFE on a simple and fixed domain. The same procedure has been applied to the KBE by Boehm et al. (2021) to handle collapsing thresholds. For time-dependent settings, the KBE still requires multiple solutions to the backward-in-time problem (as was mentioned in Section 3.5).

For the following, we assume that $-b(t) < x < b(t)$ defines the state space at time t. We introduce the map

$$\hat{T}(\hat{x}, t) = \hat{x} \cdot b(t),$$

going from the simple rectangular domain $(-1, 1) \times (0, T)$ to the real domain (see Fig. 12). Further, we define the probabilities $\hat{p}(\hat{x}, t)$ on the rectangular domain as

$$\hat{p}(\hat{x}, t) = p(\hat{T}(\hat{x}, t), t) = p(\hat{x} \cdot b(t), t) \;\Leftrightarrow\; p(x, t) = \hat{p}(\hat{x}, t) = \hat{p}(x/b(t), t).$$

By the chain rule, we can express the derivatives of p with respect to $\hat{x}$ and formulate the KFE on the rectangle

$$\partial_t \hat{p}(\hat{x}, t) - \Big(\frac{\mu(t)}{b(t)} - \frac{\hat{x} b'(t)}{b(t)}\Big)\partial_{\hat{x}} \hat{p}(\hat{x}, t) - \frac{\sigma(t)^2}{2b^2(t)} \partial_{\hat{x}\hat{x}} \hat{p}(\hat{x}, t) = 0. \quad (43)$$

This transformed equation can be approximated by standard finite differences on a uniform spatial and temporal mesh without loss of accuracy. Such approaches are called *Arbitrary Lagrangian Eulerian* coordinates and are standard in the handling of free-boundary value problems (Richter, 2017).

Remark 4. The mapping of the system to reference coordinates (43) reveals that drift rate $\mu(t)$, thresholds $b(t)$, and diffusion constant $\sigma(t)$ cannot be identified at the same time, as (43) is equivalent to

$$\partial_t \hat{p} - \tilde{\mu}(t) \partial_{\hat{x}} \hat{p} - \frac{\tilde{\sigma}(t)^2}{2} \partial_{\hat{x}\hat{x}} \hat{p} = 0.$$

with

$$\tilde{\mu}(t) := \frac{\mu(t)}{b(t)} - \frac{\hat{x} b'(t)}{b(t)}, \quad \tilde{\sigma}(t) := \frac{\sigma(t)}{b(t)}$$

determined by only two independent quantities. Fixing one of $\mu(t)$, $b(t)$, or $\sigma(t)$ still allows to freely choose both $\tilde{\mu}(t)$ and $\tilde{\sigma}(t)$.

The impossibility to identify threshold, drift rate, and diffusion constant at the same time was also noted by Ulrich et al. (2016).

A.2. Random walks on variable domains

As discussed in Section 5.3 of Diederich and Oswald (2016), the random walk approach must be modified to handle time-dependent thresholds. Without such a modification, the PDFs will show spikes whenever the threshold $b(t)$ crosses the next mesh layer of size Δx (see Fig. 13 for a visualization of this defect).

Similar to the procedure described in Diederich and Oswald (2016), we propose a simple modification that is based on splitting those control volumes of size $\Delta t \times \Delta x$ that are cut by the threshold (see Fig. 14) into a fraction of size $\delta_n \Delta x$ that is within the domain and a fraction $(1 - \delta_n)\Delta x$ outside. The transport of probabilities is then achieved in two steps (see Fig. 14 for a visualization). First, we step forward $n \mapsto n+1$ such as described in Section 2.4 in all elements that are within the domain or that are cut by the threshold $b(t)$, that is,

$$P_{n+1,m} = a_{n,-} \cdot P_{n,m-1} + a_{n,0} \cdot P_{n,m} + a_{n,+} \cdot P_{n,m+1}.$$

Second, we adjust the probabilities in all those elements that are cut, shaded in turquoise in Fig. 14. Here, we attribute the fraction $(1 - \delta_{n+1})$ to the outside and keep only δ_{n+1}, that is,

$$P_{n+1,m+1} = P_{n+1,m+1} + (1 - \delta_{n+1}) \cdot P_{n+1,m}, \quad P_{n+1,m} = \delta_{n+1} \cdot P_{n+1,m}.$$

As becomes clear from Fig. 13 (in particular from its right panel), this procedure cannot avoid the spikes completely, but it gives a smooth solution of acceptable quality. Also, the resulting method is of first order in space $\mathcal{O}(\Delta x)$ only, such that the overall accuracy drops to $\mathcal{O}(\Delta t + \Delta x)$, which, due to the parabolic stability condition $\Delta t = \Delta x^2$, leads to the non-optimal accuracy $\mathcal{O}(\sqrt{\Delta t})$. A higher-order splitting of the control volumes along the threshold might be suited to restore the full order of convergences.

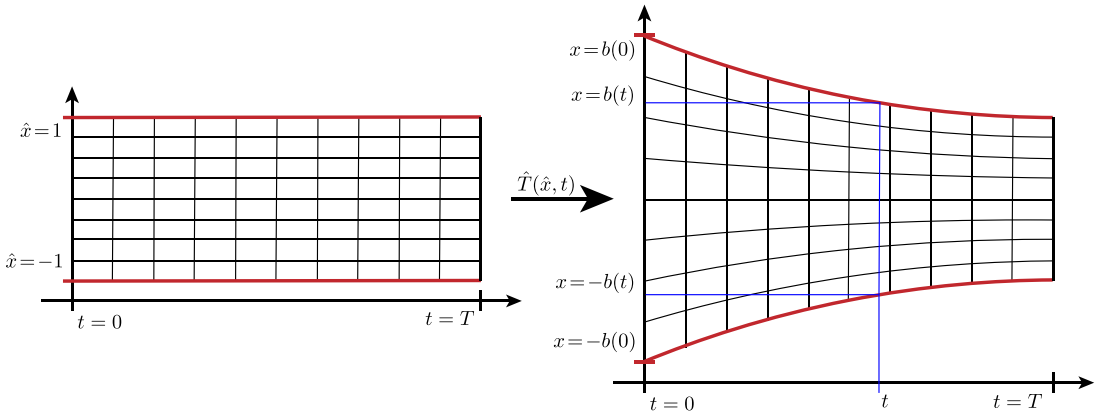


Fig. 12. Mapping from the fixed reference domain (left) with thresholds $(-1, 1)$ to the real domain with variable threshold $(b_{min}(t), b_{max}(t))$. The KFE is discretized in the fixed domain on the left side, but the solution is transformed into the variable domain on the right.

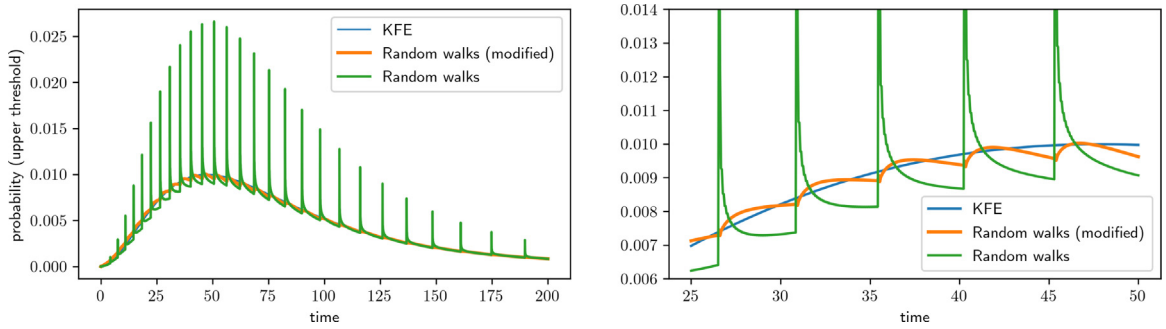


Fig. 13. Known defect of the random walk approach that causes spikes whenever the time-dependent threshold $b(t)$ crosses a discretization line. The left panel visualizes this for the unmodified random walk, the KFE solution (that does not suffer from this defect), and a simple modification presented in the present Appendix A.2, that yields rather smooth results. The right panel provides a magnification of the interval $t \in [25, 50]$.

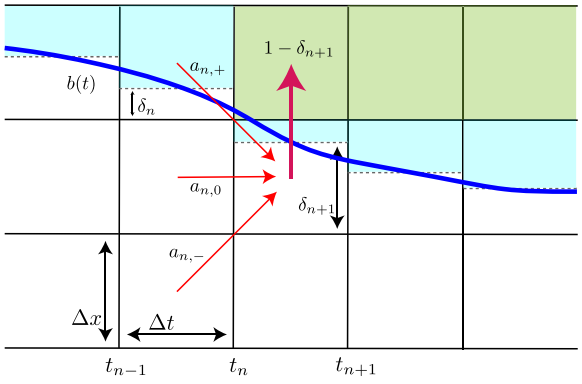


Fig. 14. Modification of the random walk approach to time-dependent thresholds. First, probabilities are transported to the next time step via the usual factors p_n^l, p_n^m, and p_n^u. Then, the cut elements are modified. Only δ_{n+1} of the probability remains and $(1-\delta_{n+1})$ is transported across the threshold.

Appendix B. Stability and efficiency of time-stepping methods

In this appendix, we briefly consider different time discretizations of the KFE with time-independent parameters. By $\mathbf{p}_n = (p_{n,0}, \ldots, p_{n,M}) \in \mathbb{R}^{M+1}$ we denote the probabilities at discrete time t_n. Then, one step of all PDE methods can be uniformly written as

$$\mathbf{p}_n + \theta\Delta t\mathbf{A}\mathbf{p}_n = \mathbf{p}_{n-1} + (\theta-1)\Delta t\mathbf{A}\mathbf{p}_{n-1},$$

with a parameter $\theta \in [0, 1]$ and a matrix $\mathbf{A}$ given as

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ \alpha_1 & \alpha_2 & \alpha_3 & 0 & \cdots & 0 \\ 0 & \alpha_1 & \alpha_2 & \alpha_3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \alpha_1 & \alpha_2 & \alpha_3 \\ 0 & \cdots & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \begin{aligned} \alpha_1 &= -\frac{\sigma^2}{2\Delta x^2} - \frac{\mu}{2\Delta x} \\ \alpha_2 &= \frac{\sigma^2}{\Delta x^2} \\ \alpha_3 &= -\frac{\sigma^2}{2\Delta x^2} + \frac{\mu}{2\Delta x} \end{aligned} \tag{44}$$

The matrix is tridiagonal, meaning that each row has only the values α_2 on the diagonal, α_1 left of it, and α_3 on the right of the diagonal (the first and the last rows differ due to the thresholds). For $\theta = 0$, this is the explicit Euler method that yields the random walk approach, and the method is called explicit as the new probability vector $\mathbf{p}_n$ is given by a simple matrix–vector product (I is the unit matrix in $\mathbb{R}^{M+1}$)

The method is implicit whenever $\theta > 0$, which means that a linear system of equations has to be solved in each step

$$(I + \theta\Delta t\mathbf{A})\mathbf{p}_n = (I + (\theta-1)\Delta t\mathbf{A})\mathbf{p}_{n-1}.$$

For $\theta = 1$, this is the implicit Euler method used by Shinn et al. (2020) (for time-dependent thresholds), and for $\theta = \frac{1}{2}$, this is the trapezoidal rule, sometimes also called Crank–Nicolson method, that is used in the present paper and by fast-dm (for

the KBE). While it is generally more costly to solve an implicit equation, the tridiagonal form of the matrix **A** makes it simple here, and the effort of the implicit methods using the *Thomas algorithm* (Quarteroni et al., 2007, Section 3.7.1) is just twice the effort of the explicit ones.

To obtain stable solutions without any oscillations, the methods have to satisfy certain conditions on their parameters. Most important, the explicit Euler method must fulfill the parabolic time-step requirement, with (44)

$$1 - \Delta t\alpha_2 > 0 \quad \Leftrightarrow \quad \Delta t < \frac{\Delta x^2}{\sigma^2}.$$

Further, all methods must satisfy the sign condition that

$$1 + \theta\Delta t\alpha_2 > 0 \text{ and } \theta\Delta t\alpha_i \leq 0 \text{ for } i = 1, 3.$$

While the first condition is satisfied for all $\theta \geq 0$ and $\Delta t > 0$, the second condition is only satisfied when

$$\mu \leq \frac{\sigma^2}{\Delta x}.$$

Violation of this condition means that the problem is too much drift-dominated which leads to the instabilities described in **Case III** in Section 4.

Finally, implicit methods are stable for all time steps when $\theta \geq \frac{1}{2}$ which includes the implicit Euler method and the trapezoidal rule. However, the limit case $\theta = \frac{1}{2}$ is still not sufficiently stable to reduce oscillations coming from rounding errors, which are unavoidable in computer simulations. At the same time, enhanced stability and second-order convergence are given for the choice $\theta = \frac{1}{2} + \Delta t$. This modified trapezoidal rule is used in our KFE approach (see also Luskin et al., 1982).

Appendix C. Realization in python

C.1. Efficient implementation

In this Appendix, we briefly comment on the better-than-expected performance of several approximation methods used in Section 4. For instance, both random walks and the Crank–Nicolson discretization of the KFE presented in Fig. 6 show a complexity that is by far superior to the postulated $\mathcal{O}(\epsilon^{-\frac{3}{2}})$ in the case of random walks and $\mathcal{O}(\epsilon^{-1})$ for the KFE. The same is true for the integral equation method that only for very small step sizes exhibits the expected $\mathcal{O}(\epsilon^{-\frac{3}{2}})$ complexity. The reason for this lies in using NumPy (Harris et al., 2020) for all numerical operations acting on matrices and vectors. We explain this behavior for the integral equation method. For simplicity, we show the single-threshold case (compare (20)):

$$P_n(x) = -2\Psi_0^n(x) + 2\Delta t \sum_{k=1}^{n-1} P_k(x)\Psi_k^n, \quad n = 1, 2, \ldots, N. \tag{45}$$

Here, $P_n(x)$ is the probability at time t_n for the initial x and

$$\Psi_0^n(x) = \Psi[b(t_n), t_n|x, 0], \quad \Psi_k^n(x) = \Psi[b(t_n), t_n|b(t_k), t_k],$$
$$n = 1, \ldots, N, \ k = 1, \ldots, n-1. \tag{46}$$

The effort lies in the repeated evaluation of the kernel function (4) for each n, k, and x. The sum over k in (45) brings along a quadratic scaling in $\mathcal{O}(1/\Delta t^2)$. However, if we treat $P \in \mathbb{R}^{N+1\times M+1}$ and $\Psi^n \in \mathbb{R}^{M+1\times N+1}$ as matrices, the entries of (46) can be computed by just one single call to NumPy. Then, the sum is given by a matrix–vector product

$$P_n = -2\Psi_0^n + 2\Delta t\Psi^n P\big|_{1,\ldots,k}.$$

The benefit lies in the high-speed processing of the C++-backend. For moderate vector sizes, the actual NumPy computations require less time than the instantiation of the NumPy backend from Python. This is the reason behind the apparently linear (instead of quadratic) scaling of (45) for moderate values of n. While all implementations would benefit from direct coding in C or C++, the advantage is only small compared to using NumPy.

C.2. Implementation of the KBE in the time-dependent setting

In Section 2.3.1 we have detailed on the disadvantage of the KBE for time-dependent problems (see also Boehm et al., 2021, for a discussion). To determine the whole distribution up to the final time $T = t_N$, N problems of type (12) must be approximated up to $t_1, t_2, \ldots, t_N$, respectively. This would result in a quadratic scaling concerning $1/\Delta t$. While this non-optimal scaling cannot be circumvented in the general case, we sketch the idea for an efficient implementation based on the fast NumPy backend as discussed in the previous section. Denoting by p^s for $s \in \{1, \ldots, N\}$ the solution for final time t_s, a Crank–Nicolson discretization will call for the solution of the linear systems backward in time from t_s to t_0

$$s = 1, 2, \ldots, N: \quad A_{s-n}p_n^s = B_{s-n}p_{n-1}^s, \quad n = 1, \ldots, s, \tag{47}$$

where the matrices $A_{s-n}, B_{s-n} \in \mathbb{R}^{M\times M}$ come from the discretization of the state space. Eq. (47) reveals the quadratic (in $N = \mathcal{O}(1/\Delta t)$) effort with ever-changing matrices that need to be reassembled. By introducing the reversed index $r = s - n$ we can rewrite (47) as forward in time problems from t_0 to t_r

$$r = N-1, \ldots, 1, 0: \quad A_r p_n^{r+n} = B_r p_{n-1}^{r+n}, \quad n = 1, \ldots, N-r. \tag{48}$$

Now, the matrices A_r, B_r are reused to solve $N - r$ linear systems each. Eq. (48) can be written as a sequence of the linear matrix equation.

$$r = N-1, \ldots, 1, 0:$$
$$A_r[p_1^{r+1}, p_2^{r+2}, \ldots, p_N^N] = B_r[p_0^{r+1}, p_1^{r+2}, \ldots, p_{N-1}^N].$$

After N steps, the solution matrix $(p_1^1, p_2^2, \ldots, p_N^N)$ contains the solutions to the KBE for all final times $t_1, t_2, \ldots, t_N$. Formally the complexity has not changed, but the impact on a practical implementation is substantial, as NumPy will solve each matrix equation with one call within the C++-backend. The effective runtime will appear linear in $N \cdot M$ as long as M and N are of moderate size, which is the optimal complexity that is also given for the Crank–Nicolson discretization of the KFE.

References

Berlyne, D. (1957). Conflict and choice time. *British Journal of Psychology*, *48*, 106–118.

Boehm, U., Cox, S., Gantner, G., & Stevenson, R. (2021). Fast solutions for the first-passage distribution of diffusion models with space-time-dependent drift functions and time-dependent boundaries. *Journal of Mathematical Psychology*, *105*, Article 102613. http://dx.doi.org/10.1016/j.jmp.2021.102613.

Buonocore, A., Giorno, V., Nobile, A., & Ricciardi, L. (1990). On the two-boundary first-crossing-time problem for diffusion processes. *Journal of Applied Probability*, *27*, 102–114.

Buonocore, A., Nobile, A., & Ricciardi, L. (1987). A new integral equation for the evaluation of first-passage-time probability densities. *Advances in Applied Probability*, *19*, 784–800.

Busemeyer, J., & Rapoport, A. (1988). Psychological models of deferred decision making. *Journal of Mathematical Psychology*, *32*, 91–134.

Churchland, A., Kiani, R., & Shadlen, M. (2008). Decision-making with multiple alternatives. *Nature Neuroscience*, *11*, 693–702.

de Jong, R., Liang, C.-C., & Lauber, E. (1994). Conditional and unconditional automaticity: A dual-process model of effects of spatial stimulus-response correspondence. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 731–750.

Diederich, A., & Busemeyer, J. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time. *Journal of Mathematical Psychology*, *47*, 304–322.
Diederich, A., & Oswald, P. (2016). Multi-stage sequential sampling models with finite or infinite time horizon and variable boundaries. *Journal of Mathematical Psychology*, *74*, 128–145.
Ditterich, J. (2006a). Evidence for time-variant decision making. *European Journal of Neuroscience*, *24*, 3628–3641.
Ditterich, J. (2006b). Stochastic models of decisions about motion direction: Behavior and physiology. *Neural Networks*, *19*, 981–1012.
Durbin, J. (1971). Boundary-crossing probabilities for the Brownian motion and Poisson processes and techniques for computing the power of the Kolmogorov-Smirnov test. *Journal of Applied Probability*, *8*, 431–453. http://dx.doi.org/10.2307/3212169.
Eriksen, B., & Eriksen, C. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *1*, 143–149.
Evans, L. (2010). *Graduate studies in mathematics: vol. 19, Partial differential equations* (2nd ed.). Americal Mathematical Society.
Evans, N. J., Hawkins, G. E., & Brown, S. D. (2020). The role of passing time in decision-making. *Journal of Experimental Psychology: Learning Memory and Cognition*, *46*, 316–326. http://dx.doi.org/10.1037/xlm0000725.
Forstmann, B., Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual Review of Psychology*, *67*, 641–666.
Fort, E., & Fankel, S. (1953). Stability conditions in the numerical treatment of parabolic differential equations. *Mathematical Tables and Other Aids to Computation*, *7*, 135–152.
Giles, M., & Carter, R. (2006). Convergence analysis of Crank–Nicolson and Rannacher time-marching. *Journal of Computational Finance*, *9*, 89–112. http://dx.doi.org/10.21314/JCF.2006.152.
Grossmann, C., Roos, H.-G., & Stynes, M. (2007). *Numerical treatment of partial differential equations*. Springer.
Hackbusch, W. (1995). Integral equations. Theory and numerical treatment. In *International series of numerical mathematics: vol. 120*, Birkhäuser.
Hairer, E., Nørsett, S., & Wanner, G. (2008). *Solving ordinary differential equations I* (3rd ed.). Springer.
Hanks, T., Mazurek, M., Kiani, R., Hopp, E., & Shadlen, M. (2011). Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *The Journal of Neuroscience*, *27*, 6339–6352.
Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., et al. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. http://dx.doi.org/10.1038/s41586-020-2649-2.
Hartmann, R., & Klauer, K. (2021). Partial derivatives for the first-passage time distribution in Wiener diffusion models. *Journal of Mathematical Psychology*, *103*, Article 102550.
Hawkins, G., Forstmann, B., Wagenmakers, E., Ratcliff, R., & Brown, S. (2015). Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *The Journal of Neuroscience*, *35*, 2476–2484.
Heath, R. A. (1992). A general nonstationary diffusion model for two-choice decision-making. *Mathematical Social Sciences*, *23*, 283–309. http://dx.doi.org/10.1016/0165-4896(92)90044-6.
Janczyk, M., Naefgen, C., & Kunde, W. (2020). Are freely chosen actions generated by stimulus codes or effect codes? attention. *Perception, & Psychophysics*, *82*, 3767–3773.
Johnson, C. (2009). *Numerical solution of partial differential equations by the finite element method*. Dover Publications.
Jones, M., & Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review*, *121*, 1–32. http://dx.doi.org/10.1037/a0034190.
Katsimpokis, D., Hawkins, G., & van Maanen, L. (2020). Not all speed-accuracy trade-off manipulations have the same psychological effect. *Computational Brain & Behavior*, *3*, 252–268.
Kelley, C. (1999). *Iterative methods for optimization*. SIAM, http://dx.doi.org/10.1137/1.9781611970920.
Kiani, R., & Shadlen, M. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*, *324*, 759–764. http://dx.doi.org/10.1126/science.1169405.
Kloeden, P., & Platen, E. (1999). *Applications of mathematics: vol. 23, Numerical solution of stochastic differential equations*. Springer, Corrected Third Printing.
Koob, V., Ulrich, R., & Janczyk, M. (2023). Response activation and activation-transmission in response-based backward crosstalk: Analyses and simulations with an extended diffusion model. *Psychological Review*, *130*, 102–136.
Lerche, V., & Voss, A. (2019). Experimental validation of the diffusion model based on a slow response time paradigm. *Psychological Research*, *83*, 1194–1209.
Luskin, M., Rannacher, R., & Wendland, W. (1982). On the smoothing property of the Crank-Nicholson scheme. *Applicable Analysis*, *14*, 117–135.
Mattler, U., & Palmer, S. (2012). Time course of free-choice priming effects explained by a simple accumulator model. *Cognition*, *123*, 347–360.
McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*, 287–330.
Nelder, J., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, *7*, 308–313. http://dx.doi.org/10.1093/comjnl/7.4.308.
Palmer, J., Huk, A., & Shadlen, M. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, *5*, 376–404.
Pratte, M., Rouder, J., Morey, R., & Feng, C. (2010). Exploring the differences in distributional properties between Stroop and Simon effects using delta plots. *Attention, Perception, & Psychophysics*, *72*, 2013–2025.
Quarteroni, A., Sacco, R., & Saleri, F. (2007). *Numerical mathematics* (2nd ed.). Springer.
Rannacher, R. (1982). Discretization of the heat equation with singular initial data. *Zeitschrift für Angewandte Mathematik und Mechanik*, *62*, T346 – T348.
Rannacher, R. (1984). Finite element solution of diffusion problems with irregular data. *Numerische Mathematik*, *43*, 309–327.
Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.
Ratcliff, R., Smith, P., Brown, S., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*, 260–281.
Richter, T. (2017). Fluid-structure interactions. Models, analysis and finite elements. *Lecture notes in computational science and engineering: vol. 118*, Springer.
Richter, T., Ulrich, R., & Janczyk, M. (2023). *Diffusion models with time-dependent parameters: An analysis of computational effort and accuracy of different numerical methods*. Zenodo, http://dx.doi.org/10.5281/zenodo.6970739.
Schwarz, W. (2022). *Random walk and diffusion models: An introduction for life and behavioral scientists*. Springer.
Schwarz, W., & Miller, J. (2012). Response time models of delta plots with negative-going slopes. *Psychonomic Bulletin & Review*, *19*, 555–574.
Shinn, M., Lam, N., & Murray, J. (2020). A flexible framework for simulating and fitting generalized drift-diffusion models. *ELife*, *9*, Article e56938. http://dx.doi.org/10.7554/eLife.56938.
Simon, J. R. (1969). Reactions towards the source of stimulation. *Journal of Experimental Psychology*, *55*, 270–279.
Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review*, *102*, 567–593.
Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology*, *44*, 408–463.
Smith, P. L. (2023). "Reliable organisms from unreliable components" revisited: The linear drift, linear infinitesimal variance model of decision making. *Psychonomic Bulletin & Review*, http://dx.doi.org/10.3758/s13423-022-02237-3.
Smith, P. L., & Lilburn, S. D. (2020). Vision for the blind: Visual psychophysics and blinded inference for decision models. *Psychonomic Bulletin & Review*, *27*, 882–910. http://dx.doi.org/10.3758/s13423-020-01742-7.
Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. *Psychological Review*, *116*, 283–317.
Smith, P. L., & Ratcliff, R. (2021). Modeling evidence accumulation decision processes using integral equations: Urgency-gating and collapsing boundaries. *Psychological Review*, *129*, 235–267. http://dx.doi.org/10.1037/rev0000301.
Smith, P. L., Ratcliff, R., & Sewell, D. K. (2014). Modeling perceptual discrimination in dynamic noise: Time-changed diffusion and release from inhibition. *Journal of Mathematical Psychology*, *59*, 95–113. http://dx.doi.org/10.1016/j.jmp.2013.05.007.
Ulrich, R., Schröter, H., Leuthold, H., & Birngruber, T. (2015). Automatic and controlled stimulus processing in conflict tasks: Superimposed diffusion processes and delta functions. *Cognitive Psychology*, *78*, 148–174.
Ulrich, R., Schröter, H., Leuthold, H., & Birngruber, T. (2016). Corrigendum to Automatic and controlled stimulus processing in conflict tasks: Superimposed diffusion processes and delta functions (Cognitive Psychology, 2015, 78, 148–174). *Cognitive Psychology*, *91*, 150.
Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*, 550–592.
Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*, 261–272. http://dx.doi.org/10.1038/s41592-019-0686-2.
Voskuilen, C., Ratcliff, R., & Smith, P. L. (2016). Comparing fixed and collapsing boundary versions of the diffusion model. *Journal of Mathematical Psychology*, *73*, 59–79. http://dx.doi.org/10.1016/j.jmp.2016.04.008.
Voss, A., Lerche, V., Mertens, U., & Voss, J. (2019). Sequential sampling models with variable boundaries and non-normal noise: A comparison of six models. *Psychonomic Bulletin and Review*, *26*, 813–832.

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology. A practical introduction.. *Experimental Psychology*, *60*, 385–402.
Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods*, *39*, 767–775.
Voss, A., Voss, J., & Lerche, V. (2015). Assessing cognitive processes with diffusion model analyses: A tutorial based on fast-dm-30. *Frontiers in Psychology*, *6*(336).
Wanner, G. (2010). Kepler, Newton and numerical analysis. *Acta Numerica*, 561–598.
White, C., Ratcliff, R., & Starns, J. (2011). Diffusion models of the flanker task: Discrete versus gradual attentional selection. *Cognitive Psychology*, *63*, 201–238.
Øksendal, B. (2000). *Stochastic differential equations: An introduction with applications* (Fifth Edition, Corrected Printing ed.). Springer.