Stochastic Calculus of Change-Point Detection with Time-to-Event

J. Sia[1], E. Jonckheere[2,*], L. Shalalfeh[3], and P. Bogdan[4,†]

***Abstract*—Change-Point Detection, in its Cumulative Sum (CUSUM) approach, accumulates real time data from a possibly corrupted process, continuously updates the log-likelihood that the data departs from its normal or expected distribution, and triggers the alarm when the CUSUM reaches a threshold, adjusted so as to compromise between detection delay and false alarm rate. Here, an extra "time-to-event" layer is added that constantly provides the operator or autonomous agent with the expected amount of time it would take for the CUSUM to hit the threshold given the data accumulated up to present. Conceptually, the time-to-event is the dwell time of an Itô process in an interval bounded by the threshold. As main contribution, the Fokker-Planck equation is given a specific approximate solution that is aimed at formulating the time-to-event as the solution to an integrable ordinary differential equation subject to mixed initial-terminal conditions. Moreover, using modern dynamical system theory, we further derive the probability density of the hitting time, in addition to its mere expectation.**

I. Introduction

Consider an i.i.d. sequence $\{X_k\}_{k=1}^N$ with "normal" regime probability density function (PDF) p_0 from $k=1$ up to and including $k=\lambda-1$, and with "abnormal" PDF p_1 as of $k=\lambda$ up to and including N. Change-Point Detection (CPD) endeavors to detect the change-point λ from data $\{X_k\}_{k=1}^n$ with minimum delay $n-\lambda$ subject to an acceptable False Alarm Rate (FAR). There is a vast literature on the subject (see [1] and references cited therein), which can be partitioned into, on the one hand, the Shiryaev (Bayesian) procedure [15] and, on the other hand, Page's minmax CUmulative SUM (CUSUM) procedure [11]. In the Shiryaev procedure, λ is assumed to have an a priori distribution and the goal is to minimize the expected detection delay subject to a false alarm rate. In the CUSUM procedure, λ is deterministic, but unknown, and the goal is to minimize the worst case detection delay subject to an acceptable false alarm rate. Here we follow the CUSUM, consistently with the early work on applications of CPD to anomaly detection approach to security problems [3], [9], [14]. In a companion paper, it is demonstrated that the CUSUM Change-Point Detection applied to Phasor Measurement Units (PMUs) data is able to detect the frequency anomaly, the "event," that ultimately led to the 2012 Indian power grid blackout [17].

1,2,4: Dept. of Elec. Eng., Univ. of Southern California, Los Angeles, CA 90089 {jsia, jonckhee, pbogdan}@usc.edu

3: German Jordanian University, Amman, Jordan, laith.shalalfeh@gju.edu.jo

*: corresp. author, supported by NSF CCF-1423624; †: supported by DARPA

II. From CUSUM Fundamentals to Main Point of the Paper

A. Hypothesis testing

Given a change-point λ, let P_λ denote the probability measure defined as p_0 on $\{X_k\}_{k=1}^{\lambda-1}$ and p_1 on $\{X_k\}_{k=\lambda}^N$. Let $\mathbb{E}_\lambda$ be the corresponding mathematical expectation. Let $\mathbb{E}_{p_{0,1}}$ be the mathematical expectation relative to the probability density $p_{0,1}$ on $\{X_k\}_{k=1}^N$. Note that $\mathbb{E}_\infty = \mathbb{E}_{p_0}$.

The Null Hypothesis H_0 is defined as the absence of changes from λ up to and including n. Quantitatively, it is based on the cumulative sum of the log-likelihood ratios,

$$Z_\lambda^n = \sum_{k=\lambda}^n \log \frac{p_1(x_k)}{p_0(x_k)}. \quad (1)$$

In this security context, the statistic is computed for the worst position of the change-point. For the same robustness reason, it is reset to 0 in case it takes negative values[1]:

$$U^n = \left\{ \max_{0\le\lambda\le n} Z_\lambda^n \right\}_+, \quad (2)$$

where $\{z\}_+ = \max\{0,z\}$. The density of the U^n statistic is given by the solution to the Kolmogorov-Fokker-Planck equation. Rejection of H_0 is not based on a p-value argument, but rather on a False Alarm Rate (FAR) specification. The alarm is triggered, that is, H_0 is rejected, when $U^n \ge h$, where the threshold h achieves a compromise between detection delay $n-\lambda$ and False Alarm Rate.

B. Recursive formulation of statistic

Putting U^n in recursive form:

$$\begin{aligned} U^{n+1} &= \left\{ \max_{0\le\lambda\le n+1} \left(Z_\lambda^n + \log \frac{p_1(x_{n+1})}{p_0(x_{n+1})} \right) \right\}_+ \\ &= \left\{ \max \left\{ U^n + \log \frac{p_1(x_{n+1})}{p_0(x_{n+1})}, \log \frac{p_1(x_{n+1})}{p_0(x_{n+1})} \right\} \right\}_+ . \end{aligned}$$

The first argument of the $\max\{\cdot,\cdot\}$ in the second inequality is the case where the $\max_\lambda$ is reached for $0\le\lambda\le n$ while the second argument is the case where the $\max_\lambda$ is reached for $n+1$. Since U^n is forced to be nonnegative, the first term

[1] Here, the exposition is simplified relative to the traditional one to avoid the formal argument justifying the reset of the recursive algorithm (3) to 0 in case U^{n+1} takes negative values [1, §2.2]. Another reason for this is to prepare the ground for simultaneous detection and identification of p_1.

in the argument of $\max\{\cdot,\cdot\}$ is greater than or equal to the second. Hence, the recursion:

$$U^{n+1} = \max\left\{0, U^n + \log\frac{p_1(X_{n+1})}{p_0(X_{n+1})}\right\}. \tag{3}$$

C. Detection delay and False Alarm Rate

The decision that there has been a change is taken at the first time τ that the CUSUM statistic U^n crosses the threshold h:

$$\tau(h) = \min\{n : U^n \geq h\}. \tag{4}$$

With this threshold, assuming that the algorithm is restarted as $U = 0$ at a false alarm and the distribution remains p_0, one would expect the next false alarm at time $T_{\text{FA}} = \mathbb{E}_{p_0}(\tau(h))$, so that the False Alarm Rate, confronted with its upper admissible bound, is

$$\text{FAR} = \frac{1}{\mathbb{E}_{p_0}(\tau(h))} \leq \overline{\text{FAR}}. \tag{5}$$

As proved in the main body of the paper, as a corollary of Eq. (17), $\tau(h)$ is monotone decreasing with h. Therefore, the threshold is selected as $\bar{h} = \min\{h : \text{Eq. (5) holds}\}$. With this threshold, the Average Detection Delay is

$$\text{ADD}_\lambda = \mathbb{E}_\lambda(\tau(\bar{h}) - \lambda : \tau(\bar{h}) \geq \lambda).$$

Theorem 1. *This CPD minimizes* ADD_λ *subject to* $\text{FAR} \leq \overline{\text{FAR}}$.

D. Main point of the paper: time-to-event

The present paper specifically explores the conceptual root of the CUSUM algorithm viewed as an Itô process $U^{t\geq 0}$ running in a domain $\mathcal{D}$ with boundary partitioned into a reflecting barrier and an absorbing barrier. While such a problem is usually treated in the context of a 2-dimensional domain $\mathcal{D}$ bounded by a Jordan curve [7], [10], here, temporarily disregarding a technicality, the domain is taken as $D = (0, h)$. To conform with the CUSUM algorithm, the initial condition is taken as $U^0 = 0$, and a classical problem is the estimation of the dwell time of U^t in $\mathcal{D}$, from which the $\text{FAR}(h)$ relationship (5) is derived.

The novelty here is to provide the (possibly autonomous) system operator observing the CUSUM, currently in the state $U^t = x < h$, with an unbiased estimate of the time-to-go $T(x)$ before the threshold is reached, precisely, $U^{t+T(x)} = h$. Viewing the latter as an "event" makes the problem of a *time-to-event* class [2], [5], [6].

The solution $T(x)$ proceeds via the transition probability $p(U^0, U^t, t)$, solution to the Kolmogorov forward equation subject to Dirichlet-Neumann boundary conditions. The theoretical novelty here is the *explicit* solution $T(x)$ when $x \neq 0$, while usually only the case $x = 0$ is considered [1] using a different procedure that sidesteps the Kolmogorov equation. Moreover, $T(x)$ is obtained from a method that differs from the method of images [7] as the latter fails under nonvanishing drift in the Itô process.

In the present 1-dimensional discrete-time case, the reflection barrier resets the statistic U^{n+1} to 0 when $U^n + \log(p_1(x_{n+1})/p_0(x_{n+1})) < 0$, while the absorption barrier stops the algorithm when $U^n + \log(p_1(x_{n+1})/p_0(x_{n+1})) > h$. While fundamentally the problem at hand is a discrete-time one, it is for the sake of addressing some technicalities in the proper conceptual context that we provide a 1-dimensional Itô model along with the Kolmogorov equation subject to mixed Dirichlet-Neumann conditions. Given the Itô solution, we then proceed to discretize it using the Euler-Maruyama and the Milstein methods. This procedure appears considerably more straightforward than the direct solution to the discrete time problem.

III. Itô stochastic calculus

The idea behind the stochastic calculus approach is to model U^n as the discretization of an Itô diffusion process over a domain $D \subset \mathbb{R}$:

$$dU^t = \boldsymbol{b}(U^t)dt + \boldsymbol{\sigma}(U^t)dB_t, \quad U^0 = x \in D, \tag{6}$$

where B_t is a Brownian motion normalized as $\mathbb{E}(dB_t)^2 = dt$, $\boldsymbol{b}(U) < 0$ is a drift term, $\boldsymbol{\sigma}(U)$ is the intensity of the innovation process $\boldsymbol{\sigma}(U^t)dB_t$, and x is a generalization of the initial condition $x = U^0 = 0$. Such generalization allows the anticipation of the threshold crossing time when the CUSUM process is in the state $U^t = x$. The time-to-escape is given as the solution to an Ordinary Differential Equation (ODE) in the initial condition x with mixed initial/terminal conditions.

A. Justification of Itô model

Such Itô model can be justified from the CUSUM algorithm (3) which, in case $p_{0,1}$ are Gauss densities with means $\mu_{0,1}$ and the common variance σ^2, takes the form

$$U^{n+1} = \left\{U^n - \frac{(\mu_0 - \mu_1)^2}{2\sigma^2} + \frac{\mu_1 - \mu_0}{\sigma}\left(\frac{X_{n+1} - \mu_0}{\sigma}\right)\right\}_+. \tag{7}$$

Consistently with the Itô integral [10], the process (6) away from the reflecting and absorbing barriers is approximated as

$$U^{n+1} = U^n + \boldsymbol{b}(U^n)\underbrace{(n+1) - n}_{\Delta t} + \boldsymbol{\sigma}(U^n)\underbrace{(B_{n+1} - B_n)}_{\Delta B_n}, \tag{8}$$

where n is in Δt units. To identify it with (7), we define

$$\boldsymbol{b}\Delta t = -\frac{(\mu_0 - \mu_1)^2}{2\sigma^2}, \tag{9}$$

$$\boldsymbol{\sigma} = \frac{\mu_1 - \mu_0}{\sigma}, \quad B_{n+1} = \sum_{k=0}^{n+1}\left(\frac{X_k - \mu_0}{\sigma}\right). \tag{10}$$

At the limit $\Delta t \downarrow 0$, we recover the Itô process as

$$dU^t = U^{n+1} - U^n, \quad dB_t = \left(\frac{X_{n+1} - \mu_0}{\sigma}\right)\sqrt{dt}.$$

B. Technicalities

In this simple setting, existence, uniqueness, and t-continuity of the solution of the Itô Stochastic Differential

Equation (SDE) are guaranteed since $\boldsymbol{b}(U)$ and $\boldsymbol{\sigma}(U)$ being independent of U obviously satisfy the sector Lipschitz conditions [10, Th. 5.5],

$$|\boldsymbol{b}(U)|+|\boldsymbol{\sigma}(U)| \leq c(1+U), \quad (11)$$
$$|\boldsymbol{b}(U)-\boldsymbol{b}(V)|+|\boldsymbol{\sigma}(U)-\boldsymbol{\sigma}(V)| \leq d|U-V|, \quad (12)$$

resp., for constants $c, d > 0$.

In the specific escape problem of Change-Point Detection, the absorbing barrier is at $h > 0$ and the reflection barrier is at 0, from which it follows that the domain could be chosen as $D = (0, h)$. But choosing the domain this way would depart from the general SDE escape formulation where the initial condition should be in the (open) domain, while here $U^0 = 0$ is on the boundary of D. To circumvent this difficulty, we temporarily enlarge the domain to $D = (-\epsilon, h)$, where the reflecting barrier is at $-\epsilon < 0$, and then show continuity of the escape time as $\epsilon \downarrow 0$.

C. Kolmogorov forward equation and outline of approach

The starting point of it all is the transition probability $p(x, y, t)$ from the initial condition distribution $p(U^0) = \delta_x(y)$ to the CUSUM distribution at some later time $p(y = U^t)$. As is well known, p is solution to the Kolmogorov forward equation (KFE) or Fokker-Planck equation [7, §X.5],[10, §8, p. 153]:

$$\begin{aligned}
\left(\frac{1}{2}\boldsymbol{\sigma}^2\frac{\partial^2}{\partial y^2}-\boldsymbol{\beta}\frac{\partial}{\partial y}\right)p(x,y,t) &= \frac{\partial p}{\partial t},\\
p(x,h,t) &= 0, \quad \text{(Dirichlet)},\\
\left.\frac{\partial p(x,y,t)}{\partial y}\right|_{y=-\epsilon} &= 0, \quad \text{(Neumann)},\\
p(x,y,0) &= \delta(y-x),
\end{aligned}$$

where $\delta(y-x)$ is a unit mass defined on y concentrated at x.

From this transition, we define in Eq. 14 the probability $\pi(x, t)$ that the process has *not yet* reached h by time t starting at x. From there, the probability that the CUSUM process crosses the threshold between times t and $t + dt$ is shown to be $\pi(x, t) - \pi(x, t + dt) = -d\pi$. From there on, in theory, $T(x) = \mathbb{E}_{-d\pi}(t)$, as shown by Eq. 15, but this requires explicit solution to the KFE. To characterized $T(x)$ as the solution to an ODE, rather than a PDE, we introduce the Green function $G(x, y)$ of the KFE, defined by Eq. 13, and relate it to $T(x)$ via Eq. 16, but such argument holds under the condition that the Green function depends more on the difference than the sum of arguments. The whole Appendix A is dedicated to justifying this approximation by constructing an explicit solution to the KFE. Simulation studies on CPD of frequency anomalies in the power grid have shown that the proposed approximation is perfectly acceptable [17]. Appendix B proposes an alternative method based on the method of images, but shown not to be easily applicable in this context.

IV. TIME TO EVENT

The following is the major result of the paper:

Theorem 2. *Consider the process* (6) *running over the domain* $(-\epsilon, h)$*, with the reflecting barrier at* $-\epsilon$ *and the absorbing barrier at* $h > 0$*. Assume both* $\boldsymbol{\sigma}$ *and* $\boldsymbol{b}$ *are independent of* U^t *with drift term* $\boldsymbol{b}(U^t) = \boldsymbol{\beta} < 0$*. Let* $U^0 = x \in (-\epsilon, h)$*. Define* $\tau_{-\epsilon,x}(h)$ *to be the first barrier crossing time, that is,*

$$\tau_{-\epsilon,x}(h) = \min\{\tau > 0 : U^\tau = h;\ U^t \geq -\epsilon, \forall t \leq \tau\}.$$

Then, defining $T(x) := \mathbb{E}(\tau_{-\epsilon,x}(h))$*,* $T(x)$ *is given by the following differential equation subject to mixed Dirichlet-Neumann boundary conditions [13, §1.2], [18, §8.2]:*

$$\begin{aligned}
\left(\frac{1}{2}\boldsymbol{\sigma}^2\frac{\partial^2}{\partial x^2}+\boldsymbol{\beta}\frac{\partial}{\partial x}\right)T(x) &= -1,\\
T(h) &= 0, \quad \text{(Dirichlet)},\\
\left.\frac{\partial T(x)}{\partial x}\right|_{x=-\epsilon} &= 0, \quad \text{(Neumann)}.
\end{aligned}$$

Proof. Let $p(x, y, t)$ be the transition probability from the initial condition $p(U^0) = \delta(U^0 - x)$ to $p(y = U^t)$. As is well known, p is solution to the Kolmogorov forward equation (KFE) [7, §X.5],[10, §8, p. 153]:

$$\begin{aligned}
\left(\frac{1}{2}\boldsymbol{\sigma}^2\frac{\partial^2}{\partial y^2}-\boldsymbol{\beta}\frac{\partial}{\partial y}\right)p(x,y,t) &= \frac{\partial p}{\partial t},\\
p(x,h,t) &= 0, \quad \text{(Dirichlet)},\\
\left.\frac{\partial p(x,y,t)}{\partial y}\right|_{y=-\epsilon} &= 0, \quad \text{(Neumann)},\\
p(x,y,0) &= \delta(y-x),
\end{aligned}$$

where $\delta(y-x)$ is a unit mass defined on y concentrated at x.

Next, we show that $p(x, y, t)$ decays exponentially fast as $t \to \infty$. Consider an elementary solution of the form $p(x, y, t) = f(y)g(t)$. Substitute $f(y)g(t)$ for $p(x, y, t)$ in the KFE equation and we get

$$\begin{aligned}
\frac{1}{2}\boldsymbol{\sigma}^2 f''(y)-\boldsymbol{\beta}f'(y) &= \lambda f(y),\\
g'(t) &= \lambda g(t),
\end{aligned}$$

where λ is a constant, an eigenvalue of the Kolmogorov Forward partial differential operator. Multiply both sides of the equation for f by $f(y)dy$, integrate by parts, and use the boundary conditions to get

$$-\frac{1}{2}\boldsymbol{\sigma}^2\|f'\|_D^2+\frac{\boldsymbol{\beta}}{2}f(0)^2=\lambda\|f\|_D^2,$$

where $\|\cdot\|_D$ denotes the L^2-norm over D. It follows that $\lambda < 0$. The equation for g hence yields the result of asymptotic decay.

Using the (nonuniform!) convergence of p to 0 as $t \to \infty$, let

$$G(x,y)=-\int_0^\infty p(x,y,t)\,dt.$$

By integrating the Kolmogorov forward equation, we get

$$\left(\frac{1}{2}\boldsymbol{\sigma}^2\frac{\partial^2}{\partial y^2}-\boldsymbol{\beta}\frac{\partial}{\partial y}\right)G(x,y)=\delta(x-y),$$

subject to the boundary conditions,

$$G(x,h)=0, \quad \frac{\partial}{\partial y}G(x,y)\Big|_{y=-\epsilon}=0.$$

It is shown in Appendices A-B and A-C that $G(x,y)$ depending on the difference of arguments, $G(x,y)=g(|x-y|)$, is a reasonable approximation. Adopting this approximation, $G(x,y)$ becomes a formal Green function as per [13, Def. 5.59]:

$$\left(\frac{1}{2}\boldsymbol{\sigma}^2\frac{\partial^2}{\partial x^2}+\boldsymbol{\beta}\frac{\partial}{\partial x}\right)G(x,y)=\delta(x-y). \tag{13}$$

Following [7, Sec. X.5, p. 341], the probability that the motion has not reached the threshold h as of epoch t is

$$\pi(x,t)=\int_{-\epsilon}^{h}p(x,y,t)\,dy. \tag{14}$$

It then follows that, starting at x, the probability density of the threshold crossing time $p_T(x,t)$ is given by $p_T(x,t)dt=\pi(x,t)-\pi(x,t+dt)$. Hence the expected lifetime of the motion before it reaches h is

$$\begin{aligned} T(x) &:= \mathbb{E}(\tau_{-\epsilon,x}(h)) \\ &= \int_0^\infty t(\pi(x,t)-\pi(x,t+dt)) \quad (15)\\ &= -\int_{-\epsilon}^{h}G(x,y)dy. \quad (16)\end{aligned}$$

Finally, integrating both sides of Eq. (13) from $y=-\epsilon$ to h and using the preceding yields the result. □

Corollary 1. *The solution to the differential equation for $T(x)$ is given by*

$$T(x)=-\frac{2x+e^{-\frac{2\boldsymbol{\beta}x}{\boldsymbol{\sigma}^2}}\boldsymbol{\sigma}^2c_1-2\boldsymbol{\beta}c_2}{2\boldsymbol{\beta}}, \tag{17}$$

where the integration constants c_1, c_2 are evaluated from the boundary conditions:

$$c_1=\frac{1}{\boldsymbol{\beta}}e^{-\frac{2\boldsymbol{\beta}\epsilon}{\boldsymbol{\sigma}^2}}, \quad c_2=\frac{2h\boldsymbol{\beta}+e^{-\frac{2\boldsymbol{\beta}(h+\epsilon)}{\boldsymbol{\sigma}^2}}\boldsymbol{\sigma}^2}{2\boldsymbol{\beta}^2}.$$

Proof. This explicit form of the solution $T(x)$ was computed symbolically by Mathematica. □

The following corollary should be obvious.

Corollary 2. $\mathbb{E}(\tau_{-\epsilon,x}(h))$ *is continuous as* $\epsilon\downarrow 0$, $\forall x\geq 0$.

Finally, setting $x=0$ and $\epsilon=0$ yields

$$T(0)=\frac{h}{2\boldsymbol{\beta}}\left(2+\frac{\boldsymbol{\sigma}^2}{\boldsymbol{\beta}h}\left(e^{-\frac{2\boldsymbol{\beta}h}{\boldsymbol{\sigma}^2}}-1\right)\right). \tag{18}$$

Remark 1. *If we denote by $\mathcal{L}(y)$ the partial differential operator of the left-hand side of the Kolmogorov* forward *equation, it is observed that the partial differential operator for $T(x)$ is the formal adjoint $\mathcal{L}^*(x)$ of $\mathcal{L}(y)$; in other words, it is relevant to the Kolmogorov* backward *equation. This is not surprising. $p(x,y,t)$ as the transition from a present Dirac delta to the future distribution is a matter of the* forward *equation. But $T(x)$ anticipates a future event, the crossing of the threshold, and endeavors to determine what warning in the past was appropriate. Hence it is a matter of the* backward *equation.*

Remark 2. *The proof of Th. 2 is a generalization of [4, Th. 1] to the case where the drift $\boldsymbol{\beta}$ is nonvanishing. Moreover, here, the proof is based on the* explicit *construction of the transition probability $p(x,y,t)$, with the $\lim_{t\to\infty}p(x,y,t)$ convergence different from the one of [4].*

Remark 3. *Eq.* (18) *is available in [1, Eq. 3.1.105], but proved via moment generating methods. However, here, Eq.* (18) *is derived from the general Eq.* (17)*. The latter is of interest for change-point detection as $T(U^t)$ tells the operator in how much time the alarm is expected to ring.*

V. First hitting time distribution

In dynamical system theory, the first hitting time refers to the first time a dynamical process hits the boundary of some subset of the sample space starting from a precise initial condition x in such sample space.

Here we consider the first hitting time $\tau_{-\epsilon,x}(h)$ of the boundary h starting at $x\in[0,h]$. The expected value of this hitting time was already computed as $\mathbb{E}(\tau_{-\epsilon,x}(h))$ in Eq. (15), from which it follows that the density of $\tau_{-\epsilon,x}(h)$ is $\partial\pi/\partial t$. Using (14), this yields

$$p_T(x,t)=\int_{-\epsilon}^{h}\frac{\partial p(x,y,t)}{\partial t}\,dy, \tag{19}$$

where $p(x,y,t)$ is the solution to the KFE subject to mixed Dirichlet-Neumann conditions.

VI. Conclusion

Classical CPD defines the "event" to be the CUSUM crossing the threshold h set up consistently with an admissible false alarm rate. This event could be sign of imminent danger [16]. The CUSUM could be monitored in real time by a system operator (or an autonomous agent), but an observed up-trend does not address the problem of the expected time to the event. This is precisely the problem addressed here, formally the dwell time of an Itô process in a compact set. The difficulty in solving the diffusion equation of an Itô process subject to mixed Dirichlet-Neumann boundary conditions and a Dirac measure initial condition is the nonvanishing drift term β. Here we have proposed an approximate solution to this problem, based on an assumption on the Green function. An example of its relevance in power grid frequency stability is available in a companion paper [17].

Next to the elusiveness of the exact solution to the Kolmogorov equation[2], several other questions questions remain

[2]The consensus of the Mathematica stack exchange is that such equation cannot be solved neither symbolically nor numerically.

to be investigated. First, the direct solution of the discrete-time case (rather than the indirect discretization of the solution to the continuous-time case) remains open but preliminary studies have shown that it relies on compounded convolution of densities making its implementation nontrivial. Next, the issues of correlated observations, and lastly the combined identification of the abnormal density p_1, need to be addressed. Regarding this last challenge, a heuristic approach based on sufficient statistics is proposed in [17].

Appendix A
Explicit solution to the Kolmogorov Forward Equation with drift under Dirichlet-Neumann boundary conditions

A. Boundary conditions

We seek a solution $p(x,y,t)$ to the Kolmogorov Forward Equation as a superposition of elementary solutions of the form $f_k(y)g_k(t)$. Note that the variable x is temporarily discarded in the solution, as we temporarily disregard the initial condition and focus on the PDE and the Dirichlet-Neumann conditions. Injecting such elementary solution in the PDE yields

$$\frac{1}{2}\boldsymbol{\sigma}^2 f_k''(y) - \boldsymbol{\beta} f_k'(y) = \lambda_k f_k(y), \tag{20a}$$

$$g_k'(t) = \lambda_k g_k(t), \tag{20b}$$

where, as demonstrated in the proof of Th. 2, $\lambda_k < 0$ is a real constant. The characteristic polynomial of the first equation is obviously $(1/2)\boldsymbol{\sigma}^2 s^2 - \boldsymbol{\beta} s - \lambda_k = 0$ with characteristic roots

$$s_{k,\pm} = \frac{\boldsymbol{\beta} \pm \sqrt{\boldsymbol{\beta}^2 + 2\lambda_k \boldsymbol{\sigma}^2}}{\boldsymbol{\sigma}^2}.$$

Consider the case where $\beta^2 + 2\lambda_k \boldsymbol{\sigma}^2 < 0$. Set

$$\alpha := \frac{\boldsymbol{\beta}}{\boldsymbol{\sigma}^2} < 0, \quad \omega_k := \frac{\sqrt{-\boldsymbol{\beta}^2 - 2\lambda_k \boldsymbol{\sigma}^2}}{\boldsymbol{\sigma}^2} > 0.$$

The solutions to the first and second differential equations (20) are of the form

$$f_k(y) = e^{\alpha y}(a_k \cos(\omega_k y) + b_k \sin(\omega_k y)), \tag{21a}$$

$$g_k(t) = e^{\lambda_k t}. \tag{21b}$$

The Dirichlet and Neumann boundary conditions read, resp.,

$$f_k(y = h) = 0 \Leftrightarrow a_k \cos(\omega_k h) + b_k \sin(\omega_k h) = 0, \tag{22}$$

$$f_k'(y = 0) = 0 \Leftrightarrow b_k \omega_k + \alpha a_k = 0. \tag{23}$$

The Neumann condition yields

$$\omega_k = -\alpha \left(\frac{a_k}{b_k}\right). \tag{24}$$

This together with the Dirichlet condition indicates that the set $\{\omega_k\}_{k=1}^{\infty}$ of eigenfrequencies is given as solutions to the fixed point problem

$$\tan(\omega_k h) = \frac{\omega_k}{\alpha}. \tag{25}$$

The solution $f_k(y)$ can therefore be rewritten as

$$f_k(y) = e^{\alpha y}(a_k \cos(\omega_k y) + b_k \sin(\omega_k y)),$$

where in addition to the Neumann condition (24) we have the option to normalize this solution. Regarding the function $g_k(t)$, it is easily found that

$$\lambda_k = -\frac{\boldsymbol{\sigma}^2 \omega_k^2}{2}.$$

Hence, at this stage, the solution to the KFE along with the Dirichlet-Neumann boundary conditions is

$$p(\cdot, y, t) = \sum_k e^{\alpha y}(a_k \cos(\omega_k y) + b_k \sin(\omega_k y))e^{\lambda_k t}. \tag{26}$$

B. Initial conditions: The case of vanishing drift

We take a short interlude to examine the case $\alpha = 0$, which will clearly indicate the difficulties the drift term $\boldsymbol{\beta}$ brings about. First observe that in case of vanishing drift, the differential operator $(\boldsymbol{\sigma}^2/2)\partial^2/\partial^2 y$ is self-adjoint. The boundary condition solution (26) reduces to

$$p(\cdot, y, t) = \sum_k a_k \cos(\omega_k y) e^{\lambda_k t}$$

where $\omega_k h = (2k+1)\pi/2$ and a_k is a normalization condition such that $\|a_k \cos(\omega_k y)\|_{L^2[0,h]} = 1$, that is, $a_k = \sqrt{h/2}$. The crucial feature that enormously simplifies the incorporation of a x-dependent term in $p(\cdot, y, t)$ to satisfy the initial condition is that the set $\{a_k \cos(\omega_k y)\}$ is orthonormal—but not complete—in $L^2[0,h]$. The relevance of this observation is as follows:

Lemma 1. *If a set* $\{\phi_k\}_{k=1}^{\infty}$ *is complete and* orthonormal *in* $L^2[0,h]$ *then* $\sum_k \phi_k(y)\phi_k(x) = \delta(x-y)$.

Lemma 2. *The set* $\left\{a_k \cos\left(\frac{(2k+1)\pi}{2h}x\right)\right\}_{k=0}^{\infty}$ *is orthonormal, but not complete, in* $L^2[0,h]$.

We use $\sum_k a_k^2 \cos(\omega_k y)\cos(\omega_k x)$ as $\delta(x-y)$ and construct the solution at initial $t = 0$ condition

$$p(x, y, 0) = \sum_k a_k^2 \cos(\omega_k x)\cos(\omega_k y)e^{\lambda_k 0}, \tag{27}$$

subject to some restrictions emanating from the lack of completeness of the orthonormal set. The Dirac δ, as a functional $\overline{\text{span}}_{L^2[0,h]}\{\cos(\omega_k y)\} \to \mathbb{R}$, would be restricted to testing functions in the L^2-closure, $\overline{\text{span}}_{L^2}$, of the orthonormal set. This entails the same restrictions on the Green function,

$$G(x, y) = -\int_0^\infty p(x, y, t)dt, \tag{28}$$

which is used to compute forcing solutions to the partial differential operator $\frac{1}{2}\boldsymbol{\sigma}^2 \frac{\partial^2}{\partial y^2}$ via

$$\frac{1}{2}\boldsymbol{\sigma}^2 \frac{\partial^2}{\partial y^2} G(x, y) = \delta(x - y).$$

More relevant, however, is the need to restrict the closure to the Sobolev space $W^{2,2}[0,h]$ of functions admitting (distributional) derivatives up to and including order two and square-integrable along with their derivatives up to and including order two[3].

In order to extend the initial condition solution (27) to $t > 0$ and to allow the Dirac δ to diffuse, we add phases to the arguments $\omega_k x$ of the cosines:

$$p(x,y,t) = \sum_k a_k^2 \cos(\omega_k x + \varphi_k(t)) \cos(\omega_k y + \varphi(t)) e^{\lambda_k t},$$

with as primary conditions to maintain the initial conditions $\varphi_k(0) = 0$, $\forall k$, and $\varphi_k(t > 0) \neq 0$ for at least one k. To keep the boundary conditions holding approximately, we require $\varphi_k(t) \ll \omega_k y$. But injecting this perturbed solution in the diffusion PDE creates additional $\dot{\varphi}_k$ terms that should ideally vanish. A phase factor satisfying $\varphi_k(0) = 0$, $\varphi_k(t > 0) > 0$, along with $\dot{\varphi_k} = 0$ almost everywhere is typically constructed as the Lebesgue singular distribution. We will keep those singular phase factors as ideal solutions, regardless of how they are implemented.

Coming back to the Green function, mandated by the proof of Th. 2 to depend mostly on the difference of arguments, observe the following:

$$p(x,y,t) = \frac{1}{2} \sum_k a_k^2 (\cos(\omega_k(y-x)) + \cos(\omega_k(y+x) + 2\varphi(t)) e^{\lambda_k t}.$$

The key point is to observe that the time integration (28) damps the second component carrying the $(y+x)$ dependency of the Green function, while the $(y-x)$ term remains unaffected. This gives giving $G(x,y)$ a stronger dependency on $x - y$ than on $y + x$.

C. Initial conditions: the case of nonvanishing drift

We now generalize the preceding subsection to the case of a drift. We proceed from Eq. (26), where now by Eq. (23) the b_k terms are nonvanishing, and we enforce the initial condition that will bring the x-argument in $p(\cdot, y, t)$. Besides b_k non-vanishing, another difficulty is that $\omega_k h$ is no longer an odd multiple of $\pi/2$, but solution to the fixed point problem (25). Nevertheless, from the geometry of the fixed point problem, it is easily seen that $\lim_{k\to\infty} \omega_k = (2k+1)\pi/2$. Another factor contributing to the departure from orthogonality of the set $\{e^{\alpha y}(a_k \cos(\omega_k y) + \boldsymbol{b}_k \sin(\omega_k y))\}_{k=0}^{\infty}$ is the damping term $\exp(\alpha y)$, but this effect is mitigated if $|\alpha| h \ll 1$.

1) Approximate Dirac delta: Instead of enforcing the exact initial condition $\delta(x-y)$, we replace it with

$$\delta(x-y) \approx \sum_{k=0}^{\infty} (a_k \cos(\omega_k x) + \boldsymbol{b}_k \sin(\omega_k x)) \\ e^{\alpha(x+y)} (a_k \cos(\omega_k y) + \boldsymbol{b}_k \sin(\omega_k y)), \quad (29)$$

[3] $W^{\ell,p}(\Omega)$ denotes the space of p-integrable functions over Ω admitting (distributional) derivatives up to and including order ℓ. We use the old fashioned $W^{\ell,2}$ rather than H^2 since the latter modern notation is too easy to be confused with Hardy spaces [13, Sec. 6.4.1].

which, as we show in Subsection A-C2, is a viable approximation under specific conditions. From there on, introduction of Lebesgue-singular phase factors $\{\phi_k\}$ as in the preceding subsection provides an approximate solution of $p(x,y,t)$ to the Kolmogorov equation:

$$p(x,y,t) \approx \\ e^{\alpha(x+y)} \sum_{k=0}^{\infty} (a_k \cos(\omega_k x + \phi_k(t)) + \boldsymbol{b}_k \sin(\omega_k x + \phi_k(t))) \\ (a_k \cos(\omega_k y + \phi_k(t)) + \boldsymbol{b}_k \sin(\omega_k y + \phi_k(t))) \\ e^{\lambda_k t}. \quad (30)$$

This will be taken as final solution.

It remains to examine the extent to which the Green function $G(x,y)$ depends mostly on the difference of arguments. First, observe the $\exp(\alpha(x+y))$ contribution to the sum, but under small drift, $|\alpha| 2h \ll 1$, its contribution remains minor. After trigonometric manipulations, the coefficient of the $\exp(\alpha(x + y) \exp(\lambda_k t) a_k^2$ is found to be

$$\frac{1}{2}(\cos(\omega_k(x+y) + 2\phi_k(t)) + cos(\omega_k(x-y))). \quad (31)$$

Likewise, the coefficient of the $\exp(\alpha(x+y) \exp(\lambda_k t) b_k^2$ is found to be

$$\frac{1}{2}(-\cos(\omega_k(x+y) + 2\phi_k(t)) + cos(\omega_k(x-y))). \quad (32)$$

Finally, the coefficient of the $\exp(\alpha(x+y)) \exp(\lambda_k t) a_k b_k$ is found to be

$$(\sin(\omega_k(x+y) + 2\phi_k(t)) + \sin(\omega_k(x-y))). \quad (33)$$

The time-integration of all such terms damps the dependency on the sum $x + y$, while the dependency on the difference $x - y$ remains unaffected, giving the Green function a stronger dependency on the difference of argument, $x - y$. Because of the $\sin(\omega_k(x-y))$ term in (33), $G(x,y)$ doesn't yet solely depend on $|x-y|$. To further show that it depends mostly on the absolute value, $|x-y|$, we need to invoke the argument that under small drift, a_k is dominant relative to b_k and hence (31) is dominant relative to (32) and (33).

2) Validity of the $\delta(x-y)$ approximation: To show that (29) has tractable approximation error, remember that the left-hand side $\delta_x(y)$ is a density on y parameterized by x; moreover, the right-hand side, denote it $\mathrm{RHS}_x(y)$, is also a y-density parameterized by x, where the a_k, b_k's are normalized so that $\int_0^h \mathrm{RHS}_x(y) dy = 1$. (This is possible since the boundary conditions only specify the a_k/b_k ratio in view of (24).) Since the initial distribution $\delta_x(y)$ is "transported" to $\mathrm{RHS}_x(y)$ by the Itô process, it is natural to validate the approximation by the 1-Wasserstein distance $\mathcal{W}_1(\delta_x(y), \mathrm{RHS}_x(y))$, which precisely endeavors to find the minimum Kantorovich cost [19] of transporting $\delta_x(y)$ to $\mathrm{RHS}_x(y)$. Through the Kantorovich-Rubinstein duality, the 1-Wasserstein distance can be computed as follows:

Theorem 3 ([12])**.** *Let Q_{δ_x} and Q_{RHS_x} be the quantiles of*

δ_x *and* RHS_x*, resp. Then*

$$\mathcal{W}_1(\delta_x, \mathrm{RHS}_x) = \int_0^1 |Q_{\delta_x}(y) - Q_{\mathrm{RHS}_x}(y)| dy.$$

Obviously, $Q_{\delta_x}(y) = x$. From there on, the following can be shown:

Corollary 3 ([8])**.**

$$\mathcal{W}_1(\delta_x, \mathrm{RHS}_x) = x - \mathbb{E}_{y \sim \mathrm{RFH}_x(y)} y.$$

Computationally,

$$\mathcal{W}_1(\delta_x, \mathrm{RHS}_x) = x - \sum_k \Gamma_k (a_k \cos(\omega_k x) + \boldsymbol{b}_k \sin(\omega_k), \tag{34a}$$

where

$$\Gamma_k = \int_0^h y e^{\alpha y} (a_k \cos(\omega_k y) + \boldsymbol{b}_k \sin(\omega_k y)) dy. \tag{34b}$$

Using this criterion, it is verified that for the examples of [17], the approximation is valid.

3) Exact Dirac delta: The set

$$\{e^{\alpha y}(a_k \cos(\omega_k y) + \boldsymbol{b}_k \sin(\omega_k y))\}_{k=0}^{\infty}$$

is not orthonormal, but it can be made L^2-orthonormal by the Gram-Schmidt procedure. From there on, the procedure to construct a Green function depending more strongly on the difference rather than the sum of arguments is essentially the same as that of the $\boldsymbol{\beta} = 0$ case.

More specifically, to simplify notation, let

$$\boldsymbol{f}(y, \Phi) = \mathrm{Col}\{e^{\alpha y}(a_k \cos(\omega_k y + \phi_k) + b_k \sin(\omega_k y + \phi_k))\}_{k=0}^{\infty}$$

where, to avoid clutter, we have omitted the dependency of Φ and ϕ_k on t. Because of the drift related α term, the basis $\boldsymbol{f}$ need not be orthonormal, but this can be corrected via a lower triangular operator $A = \{a_{i,j} : 1 \leq j < i \leq \infty\}$ that makes $\mathbf{f} = A\boldsymbol{f}$ orthonormal. As such, $\mathbf{f}(y, \Phi(0))^T \mathbf{f}(x, \Phi(0) = \delta(x - y)$. Lastly, define $\Lambda(t) = \mathrm{diag}\{\exp(\lambda_k t)\}_{k=1}^{\infty}$ and the solution is

$$p(x, y, t) = \boldsymbol{f}^T(y, \Phi(t)) A^T \Lambda(t) A \boldsymbol{f}(x, \Phi(t)), \tag{35}$$

$$= \mathbf{f}^T(y, \Phi(t)) \Lambda(t) \mathbf{f}(x, \Phi(t)). \tag{36}$$

APPENDIX B
SOLUTION OF NO-DRIFT KOLMOGOROV FORWARD EQUATION BY METHOD OF IMAGES

In case of no drift ($\boldsymbol{\beta} = 0$), which from (7) is a reasonable approximation iff $|\mu_0 - \mu_1|/\sigma << 1$, we show that the solution to the Kolmogorov forward equation is given by an infinite linear combination of the fundamental diffusion

$$f(y) := (1/\sqrt{2\pi t}) \exp(-(y - (x + \boldsymbol{\beta} t))^2 / 2t)$$

of $\delta(y - x)$ shifted infinitely many times to enforce the boundary conditions in a successive approximation scheme [7, Sec. X.5]. Start by observing that

$$F_1(y) := f(y) + f(-y)$$

satisfies the diffusion equation for $\boldsymbol{\beta} = 0$ and the Neumann condition. Observe that it fails to satisfy the diffusion equation whenever $\boldsymbol{\beta} \neq 0$. This is the reason why the method of images [7] is restricted to $\boldsymbol{\beta} = 0$. With this restriction, we proceed by noting that $F_1(y)$ fails to satisfy the Dirichlet condition, but this can be corrected as

$$F_2(y) := (f(y) + f(-y)) - (f(2h - y) + f(-2h + y)),$$

which satisfies the Dirichlet but not the Neumann condition, resulting in a sequential correction taking the form

$$\begin{aligned} F_{2k} &= F_{2k-1} + (-1)^k (f(y - 2kh) + f(-y + 2kh)), \\ F_{2k+1} &= F_{2k} + (-1)^k (f(-y - 2kh) + f(y + 2kh)), \end{aligned} \tag{37}$$

starting with F_1 and F_2 as already explicitly defined. To demonstrate that F_{2k} satisfies the Dirichlet condition, observe that F_2 does and the recursive step consists in observing that $F_{2(k+1)}(y) - F_{2k}(y)$ vanishes for $y = h$. Regarding the Neumann condition, take F_1 to be the initial step and the recursive part of the proof is based on the observation that $F_{2k+1}(y) - F_{2(k-1)}(y)$ is invariant under the substitution $y \mapsto -y$, which implies that $\frac{d}{dy}\left(F_{2k+1}(y) - F_{2(k-1)}(y)\right)\Big|_{y=0}$, that is, the Neumann condition. Then we define the transition density as

$$p(x, y, t) := \lim_{m \to \infty} F_m(y).$$

To prove uniform convergence in the compact set $[-\epsilon, h]$, observe that at every step, the correction term $F_m(y) - F_{m-1}(y)$ decays exponentially as $m \to \infty$ uniformly over the compact set $[0, h]$. Therefore, $\{F_m\}_{m=1}^{\infty}$ is a Cauchy sequence and hence converges, but not uniformly.

Note that $f(y)$ depends only on the difference $x - y$ and that $f(-y)$ depends on $x + y$. Hence, $F_m(y)$ and its limit depend on both $x - y$ and $x + y$.

REFERENCES

[1] Michèle Basseville and Igor V. Nikiforov. *Detection of Abrupt Changes - Theory and Application*. Englewood Cliffs, New Jersey, 2010.

[2] A. Bender, A. Groll, and F. Scheipl. A generalized additive model approach to time-to-event analysis. *Statistical Modelling*, 18(3-4):299–321, 2018.

[3] R. Blazek, H. Kim, B. Tozovskii, and A. Tartakovsky. A novel approach to detection of 'denial-of-service' attacks via adaptive sequential and batch sequential change-point detection. In *Proceedings of the 2001 IEEE Workshop on Information Assurance and Security*, pages 220–226, U.S. Military Academy, West Point, NY, June 5-6 2001.

[4] C. Caginalp and X. Chen. Analytical and nuerical results for an escape problem. *Arch. Rational Mech. Anal.*, 203:329–342, 2012.

[5] M. Catalano, A. Lijoi, and I. Prunster. Approximation of bayesian models for time-to-event data. *Electronic Journal of Statistics*, 14:3366–3395, 2020.

[6] P. Chapfuwa, C. Tao, C. Li, and et al. Adversarial time-to-event modeling. *Proc Mach Learn Res.*, 80:735–744, 2018.

[7] W. Feller. *An introduction to probability theory and its applications—volume 2*. John Wiley & Sons, 1971.

[8] I. Khalid, C. Weidner, E. Jonckheere, S. Shermer, and F. Langbein. Statistically characterising robustness and fidelity of quantum controls. *Physical Review A, submitted*, 2022. Available at arXiv.

[9] H. Kim, B. L. Rozovskii, and A. G. Tartakovsky. A nonparametric multichart CUSUM test for rapid detection of DOS attacks in computer networks. *International Journal of Computing & Information Sciences*, 2(3):149–158, December 2004.

[10] B. Oksendal. *Stochastic Differential Equations*. Springer, Berlin, 1995.

[11] E. Page. Continuous inspection schemes. *Biometrika*, 41:100–115, 1954.

[12] Aaditya Ramdas, Nicolás García Trillos, and Marco Cuturi. On Wasserstein two-sample testing and related families of nonparametric tests. *Entropy*, 19(2), 2017.

[13] M. Renardy and R. C. Rogers. *An introduction to partial differential equations*, volume 13 of *Texts in Applied Mathematics*. Springer, 1993.

[14] Khushboo Shah, Edmond Jonckheere, and Stephan Bohacek. Dynamic modeling of internet traffic for intrusion detection. *EURASIP Journal on Advances in Signal Processing*, 2007, 2007.

[15] A. N. Shiryaev. On optimum methods in quickest detection problems. *Theory Probab. Appli.*, pages 22–46, 1963.

[16] J. Sia, E. Jonckheere, L. Shalalfeh, and P. Bogdan. PMU change point detection of imminent voltage collapse and stealthy attacks. In *IEEE CDC*, pages 6812–6817, Miami Beach, FL, 2018.

[17] J. Sia, E. Jonckheere, L. Shalalfeh, and P. Bogdan. Pmu change-point detection of frequency anomaly avenue to voltage collapse. Available in arXiv???, 2022.

[18] M. E. Taylor. *Partial Differential Equations II—Qualitative Studies of Linear Equations*, volume 116 of *Applied Mathematical Sciences*. Springer, New York, Berlin, 1997.

[19] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.