Research Report No. 1, 2022, *Probab. Statist. Group Manchester* (20 pp)

Weak Solutions in the Sense of Schwartz to Dynkin's Characteristic Operator Equation

Goran Peskir

We show that if the transition operator $\mathbb{P}_t$ of a continuous time-homogeneous (strong) Markov process X in $I\!R^n$ maps infinitely differentiable functions with compact support to twice continuously differentiable functions for $t>0$, then the action of the infinitesimal operator $\mathbb{A}$ of X on its domain coincides with the action of the differential operator $\mathbb{D}$ of X understood in the sense of Schwartz distributions. Applying this fact to the process X stopped at the first exit time from a given open subset C of $I\!R^n$, we derive that the solution to Dynkin's characteristic operator equation of X arising from the Dirichlet problem on C can be viewed as a weak solution to the differential operator equation of X in the sense of Schwartz distributions. A useful consequence of this identification is that the solution is infinitely differentiable on C whenever the drift and diffusion coefficients are infinitely differentiable and the differential operator $\mathbb{D}$ of X is hypoelliptic (e.g. when $\mathbb{D}$ satisfies the Hörmander condition). In particular, this is satisfied when X is a degenerate diffusion process in $I\!R^n$ (in the sense that $\mathbb{D}$ is a degenerate parabolic operator) so that the analytic existence results for parabolic PDEs are generally not available. The arguments and results extend to cover more general boundary value problems of this kind including the initial value problems as well.

1. Introduction

The purpose of the paper is to answer the question when solutions to Dynkin's characteristic operator equation associated with a continuous time-homogeneous (strong) Markov process X in $I\!R^n$ can be viewed as weak solutions to the differential operator equation of X understood in the sense of Schwartz distributions. The value of this identification is twofold. Firstly, by constructing the Markov process itself and taking the expected value, we are also establishing the existence of a weak solution. This provides a probabilistic method for constructing (weak) solutions when the analytic existence results of PDE theory may not be available (e.g. when X is a degenerate diffusion process in dimension two or higher). Secondly, having a weak solution provides an opportunity for its upgrade to a strong (classic/smooth) solution. This can be done for instance when the differential operator of X is hypoelliptic (e.g. when it satisfies the Hörmander condition). Such an upgrade would not be possible however if the existence of a weak solution could not be established in the first place (see e.g. [7]).

Mathematics Subject Classification 2020. Primary 60J60, 60J65, 60H20. Secondary 35H10, 35K20, 35K65.

Key words and phrases: Markov process, diffusion process, transition operator, semigroup, infinitesimal operator, Dynkin's characteristic operator, second order differential operator, degenerate, parabolic, elliptic, boundary/initial value problem, weak solutions, Schwartz distribution, adjoint operator, weak (star) topology.

More specifically, we consider the function

$$V(x) = \mathsf{E}_x G(X_{\tau_D}) \tag{1.1}$$

where $X = (X_t)_{t \geq 0}$ is a continuous time-homogeneous (strong) Markov process in $I\!R^n$ starting at x in C under P_x, the set $C \subseteq I\!R^n$ is open, the set D equals $I\!R^n \backslash C$, the stopping time $\tau_D = \inf\{\, t \geq 0 \mid X_t \in D \,\}$ is the first entry time of X into D, and G is a real-valued continuous/measurable function defined on the boundary ∂C between the sets C and D. The strong Markov property of X then implies that

$$\mathbb{L}V = 0 \quad \text{in } C \tag{1.2}$$

where $\mathbb{L}$ is Dynkin's characteristic operator of X (cf. [3, Chapter V]). Note that (1.2) holds in great generality when V from (1.1) is only known to be measurable (and X is not necessarily strong Feller or even Feller). We assume that the action of $\mathbb{L}$ is explicitly known on sufficiently regular functions in the sense that

$$\mathbb{L} = \mathbb{D} \quad \text{on } C_c^\infty \tag{1.3}$$

where $\mathbb{D}$ is a second-order differential operator given by

$$\mathbb{D} = \sum_{i=1}^n \mu_i \partial_i + \frac{1}{2} \sum_{i,j=1}^n (\sigma\sigma^t)_{ij} \partial_{ij}^2 \tag{1.4}$$

for some $\mu = (\mu_i)_{1 \leq i \leq n} : I\!R^n \to I\!R^n$ and $\sigma = (\sigma_{ij})_{1 \leq i,j \leq n} : I\!R^n \to I\!R_+^{n \times n}$ belonging to C^2. The question then arises whether (1.2) can be viewed as

$$\mathbb{D}V \overset{w}{=} 0 \quad \text{in } C \tag{1.5}$$

where the derivatives are understood in the sense of Schwartz distributions (cf. [12]).

We tackle this question in two steps. Firstly, we focus on the infinitesimal operator $\mathbb{A}$ of X and show that if the transition operator $\mathbb{P}_t$ of X maps C_c^∞ into C^2 for $t > 0$, then the action of $\mathbb{A}$ on its domain $D(\mathbb{A})$ coincides with the action of $\mathbb{D}$ understood in the sense of Schwartz distributions. This fact is of independent interest within the semigroup theory of Markov processes. Secondly, we return to the original question by focussing on the process X stopped at τ_D. This regains the most important (localising) feature of $\mathbb{L}$ which ensures that V belongs to its domain $D(\mathbb{L})$ while V fails to belong to $D(\mathbb{A})$ generally. Stopping X at τ_D reduces the state space of the stopped process to the closure of C so that the semigroup arguments from the first step remain applicable. Exploiting the existence of a compact support we then show that the sufficient condition from the first step on the transition operator of the stopped process is satisfied and this establishes that (1.2) yields (1.5) as claimed.

The latter result can then be extended from (1.1) to more general functions

$$V(x) = \mathsf{E}_x \Big(e^{-\int_0^{\tau_D} \lambda(X_t)\, dt} G(X_{\tau_D}) + \int_0^{\tau_D} e^{-\int_0^t \lambda(X_s)\, ds} H(X_t)\, dt \Big) \tag{1.6}$$

where $\lambda : I\!R^n \to [0, \infty)$ is measurable, $G : \partial C \to I\!R$ is continuous/measurable, and $H : I\!R^n \to I\!R$ is continuous, using established arguments of this kind (cf. [8, Section 7]). By the strong Markov property of X we know that the extension of (1.2) then reads as follows

$$\mathbb{L}V - \lambda V = -H \quad \text{in } C. \tag{1.7}$$

Extending (1.5) we obtain

$$\mathbb{D}V - \lambda V \overset{w}{=} -H \quad \text{in } C \tag{1.8}$$

where the derivatives are understood in the sense of Schwartz distributions.

Similarly, in addition to more general boundary value problems (1.6) these extensions also include the initial value problems

$$V(t,x) = \mathsf{E}_x\Big(e^{-\int_0^t \lambda(X_s)\,ds} G(X_t) + \int_0^t e^{-\int_0^s \lambda(X_r)\,dr} H(X_s)\,ds\Big) \tag{1.9}$$

for $(t,x) \in [0,\infty) \times I\!\!R^n$ where $\lambda : I\!\!R^n \to [0,\infty)$ is measurable, $G : \partial C \to I\!\!R$ is continuous/measurable, and $H : I\!\!R^n \to I\!\!R$ is continuous. By the Markov property of X we know that the analogue of (1.7) then reads as follows

$$V_t = \mathbb{L}V - \lambda V + H \quad \text{on } (0,\infty) \times I\!\!R^n\,. \tag{1.10}$$

Extending (1.8) we obtain

$$V_t \overset{w}{=} \mathbb{D}V - \lambda V + H \quad \text{on } (0,\infty) \times I\!\!R^n \tag{1.11}$$

where the derivatives are understood in the sense of Schwartz distributions.

This establishes the existence of weak solutions to equations (1.8) and (1.11) in the sense of Schwartz distributions using purely probabilistic methods. Moreover, if the differential operator in either (1.8) or (1.11) is hypoelliptic (e.g. when the Hörmander condition is verifiable) and μ & σ belong to C^∞, then the weak solutions to either (1.8) or (1.11) are infinitely differentiable respectively. In particular, this is satisfied when X is a degenerate diffusion process in $I\!\!R^n$ (in the sense that $\mathbb{D}$ is a degenerate parabolic operator) so that the analytic existence results for parabolic PDEs are generally not available.

2. Setting

In this section we describe the setting and introduce the notation. For fuller details on the setting and notation we refer to [3, Chapters 1-3] and [4, Chapters II-IV].

1. We consider a continuous time-homogeneous (strong) Markov process $X = (X_t)_{t \ge 0}$ with values in a closed subset E of $I\!\!R^n$ with a non-empty interior E^o. We assume that $X_0 = x$ under the probability measure P_x defined on a measurable space $(\Omega, \mathcal{F})$ for $x \in E$. By $\mathsf{P}(t;x,A) := \mathsf{P}_x(X_t \in A)$ we denote the transition function of X for $t \ge 0$, $x \in E$ and $A \in \mathcal{B}(E)$ where $\mathcal{B}(E)$ denotes the Borel σ-algebra on E. We let $B = B(E)$ denote the Banach space of real-valued bounded measurable functions on E equipped with the supremum norm defined by $\|f\| = \sup\{\,|f(x)| : x \in E\,\}$ for $f \in B$. The (*forward*) *transition operator* $\mathbb{P}_{\rm t} : B \to B$ of X is defined by

$$\mathbb{P}_{\rm t} f(x) = \mathsf{E}_x f(X_t) = \int_E f(y)\, \mathsf{P}(t;x,dy) \tag{2.1}$$

for $t \ge 0$, $f \in B$ and $x \in E$. The Markov property of X implies that $(\mathbb{P}_{\rm t})_{t \ge 0}$ is a (contraction) semigroup of linear operators on B (cf. [3, p. 22]). We let B_0 denote the family of $f \in B$ such that $\lim_{t\to 0} \mathbb{P}_t f = f$ in the supremum norm. It is easily seen that B_0 is a closed (Banach) subspace of B and $\mathbb{P}_t B_0 \subseteq B_0$ for all $t \geq 0$.

2. We let $M = M(E)$ denote the Banach space of real-valued (finite) measures on $\mathcal{B}(E)$ equipped with the total variation norm defined by $\|\mu\| = \mu^+(E) + \mu^-(E)$ where μ^+ and μ^- are the positive and negative parts of $\mu \in M$ respectively. Letting M^+ denote the family of all non-negative (finite) measures on $\mathcal{B}(E)$ we have $\mu^+ \in M^+$ and $\mu^- \in M^+$ with $\mu = \mu^+ - \mu^-$ for $\mu \in M$. The (*backward*) *transition operator* $\mathbb{Q}_t : M \to M$ of X is defined by

$$\mathbb{Q}_t \mu(A) = \int_E \mathsf{P}_x(X_t \in A)\, \mu(dx) = \int_E \mathsf{P}(t; x, A)\, \mu(dx) \tag{2.2}$$

for $t \geq 0$, $\mu \in M$ and $A \in \mathcal{B}(E)$. The Markov property of X implies that $(\mathbb{Q}_t)_{t\geq 0}$ is a (contraction) semigroup of linear operators on M (cf. [3, p. 49]). We let M_0 denote the family of $\mu \in M$ such that $\lim_{t\to 0} \mathbb{Q}_t \mu = \mu$ in the total variation norm. It is easily seen that M_0 is a closed (Banach) subspace of M and $\mathbb{Q}_t M_0 \subseteq M_0$ for all $t \geq 0$.

3. A natural pairing between B and M is obtained by the scalar product

$$\langle f, \mu \rangle = \int_E f\, d\mu \tag{2.3}$$

for $f \in B$ and $\mu \in M$. Then the semigroups $(\mathbb{P}_t)_{t\geq 0}$ and $(\mathbb{Q}_t)_{t\geq 0}$ become *adjoint* to each other in the sense that that the following identity holds

$$\langle \mathbb{P}_t f, \mu \rangle = \langle f, \mathbb{Q}_t \mu \rangle \tag{2.4}$$

for all $f \in B$ and all $\mu \in M$ with $t \geq 0$ (cf. [9]). Note that (2.4) more explicitly reads

$$\int_E \Big(\int_E f(y)\, \mathsf{P}(t; x, dy) \Big) \mu(dx) = \int_E f(y) \Big(\int_E \mathsf{P}(t; x, dy)\, \mu(dx) \Big) \tag{2.5}$$

for $f \in B$ and $\mu \in M$ from where its validity is evident for $t \geq 0$.

4. The (*forward*) *infinitesimal operator* $\mathbb{A}$ of X is defined by

$$\mathbb{A} f = \lim_{t\to 0} \frac{\mathbb{P}_t f - f}{t} \tag{2.6}$$

for f belonging to the domain $D(\mathbb{A})$ of $\mathbb{A}$ for which the limit in (2.6) exists with respect to the supremum norm. It is well known that $D(\mathbb{A})$ is a dense subspace of B_0 with respect to the supremum norm. The (*backward*) *infinitesimal operator* $\mathbb{B}$ of X is defined by

$$\mathbb{B} \mu = \lim_{t\to 0} \frac{\mathbb{Q}_t \mu - \mu}{t} \tag{2.7}$$

for μ belonging to the domain $D(\mathbb{B})$ of $\mathbb{B}$ for which the limit in (2.7) exists with respect to the total variation norm. It is well known that $D(\mathbb{B})$ is a dense subspace of M_0 with respect to the total variation norm. In view of (2.4) we see that the infinitesimal operators $\mathbb{A}$ and $\mathbb{B}$ are *adjoint* to each other in the sense that the following identity holds

$$\langle \mathbb{A} f, \mu \rangle = \langle f, \mathbb{B} \mu \rangle \tag{2.8}$$

for all $f \in D(\mathbb{A})$ and all $\mu \in D(\mathbb{B})$.

5. It turns out that the convergence with respect to the total variation norm in (2.7) is too strong to be of wider practical use. For this reason we may require that the limits in (2.6) and (2.7) exist in a less demanding sense of the *weak-star* topology (cf. [4, Chapter V, Section 3]). For (2.6) we say that $f_n \in B$ converges *weakly* to $f \in B$ if and only if $\langle f_n, \mu \rangle \to \langle f, \mu \rangle$ for every $\mu \in M$ as $n \to \infty$. For (2.7) we say that $\mu_n \in M$ converges *weakly* to $\mu \in M$ if and only if $\langle f, \mu_n \rangle \to \langle f, \mu \rangle$ for every $f \in B$ as $n \to \infty$. Denoting the actions of the weak limits in (2.6) and (2.7) by $\tilde{\mathbb{A}}$ and $\tilde{\mathbb{B}}$ respectively, and renaming the corresponding B_0 and M_0 to $\tilde{B}_0$ and $\tilde{M}_0$ respectively, we have $D(\mathbb{A}) \subseteq D(\tilde{\mathbb{A}}) \subseteq B_0 \subset \tilde{B}_0 \subseteq B$ and $D(\mathbb{B}) \subseteq D(\tilde{\mathbb{B}}) \subseteq M_0 \subset \tilde{M}_0 \subseteq M$ with $\mathbb{P}_t \tilde{B}_0 \subseteq \tilde{B}_0$ and $\mathbb{Q}_t \tilde{M}_0 \subseteq \tilde{M}_0$ for $t \ge 0$, where $\tilde{B}_0$ is a closed (Banach) subspace of B with respect to the supremum norm, and $\tilde{M}_0$ is a closed (Banach) subspace of M with respect to the total variation norm. Moreover, the weak closures of the sets $D(\mathbb{A})$, $D(\tilde{\mathbb{A}})$, B_0 and $\tilde{B}_0$ coincide in B, and the weak closures of the sets $D(\mathbb{B})$, $D(\tilde{\mathbb{B}})$, M_0 and $\tilde{M}_0$ coincide in M. For proofs of all these claims we refer to [3, p. 40]. Although we will not use them directly in what follows, they do play an important role for a wider understanding of the arguments and results to be presented below.

6. For any $\lambda > 0$ given and fixed, the linear operator $\lambda I - \mathbb{A}$ maps $D(\mathbb{A})$ onto B_0 in a one-to-one way and its inverse $\mathbb{R}_\lambda := (\lambda I - \mathbb{A})^{-1}$ called the *resolvent* of $\mathbb{A}$ is given by $\mathbb{R}_\lambda g = \int_0^\infty e^{-\lambda t} \mathbb{P}_t g \, dt$ for $g \in B_0$. The linear operator $\mathbb{R}_\lambda$ is bounded (i.e. continuous) with its operator norm $\|\mathbb{R}_\lambda\| \le 1/\lambda$ for $\lambda > 0$. The linear operator $\mathbb{A}$ is unbounded (i.e. discontinuous) but *closed* in the sense that if $f_n \in D(\mathbb{A})$ and $f_n \to f$ with $\mathbb{A} f_n \to g$ both in the supremum norm as $n \to \infty$ then $f \in D(\mathbb{A})$ and $\mathbb{A} f = g$. Moreover, if $f \in D(\mathbb{A})$ then $t \mapsto \mathbb{P}_t f$ is differentiable on $[0, \infty)$ with respect to the supremum norm and we have

$$\partial_t \mathbb{P}_t f = \mathbb{A} \mathbb{P}_t f = \mathbb{P}_t \mathbb{A} f \tag{2.9}$$

$$\mathbb{P}_t f - f = \int_0^t \mathbb{A} \mathbb{P}_s f \, ds = \int_0^t \mathbb{P}_s \mathbb{A} f \, ds \tag{2.10}$$

for $t \ge 0$. For proofs of all these claims we refer to [3, pp 23-25]. Note that all these claims (i) hold for $\mathbb{B}$ in place of $\mathbb{A}$ and (ii) extend to $\tilde{\mathbb{A}}$ and $\tilde{\mathbb{B}}$ as well (see [3, p. 40]).

7. Recall that $C_c^\infty = C_c^\infty(E)$ denotes the space of all infinitely differentiable functions from E into $I\!R$ having compact supports contained in E^o. Recall also that $C^2 = C^2(E)$ denotes the space of all twice continuously differentiable functions from E into $I\!R$. Recall finally that $L^1_{\text{loc}} = L^1_{\text{loc}}(E)$ denotes the space of all locally integrable functions from E into $I\!R$ with respect to Lebesque measure λ. The distributional action (in the sense of Schwartz) of F from L^1_{loc} on φ from C_c^∞ is defined by

$$\langle F, \varphi \rangle = \int_E F \varphi \, d\lambda \tag{2.11}$$

making F a continuous linear functional on C_c^∞ with respect to the supremum norm. For $\alpha = (\alpha_1, \ldots, \alpha_n) \in I\!N_0^n$ we set $\partial^\alpha = \partial^{|\alpha|}/(\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n})$ where $|\alpha| := \sum_{i=1}^n \alpha_i$. The distributional partial derivative $\partial^\alpha F$ of F from L^1_{loc} is defined through its action on φ from C_c^∞ by

$$\langle \partial^\alpha F, \varphi \rangle = (-1)^{|\alpha|} \langle F, \partial^\alpha \varphi \rangle \tag{2.12}$$

making $\partial^\alpha F$ a continuous linear functional on C_c^∞ with respect to the supremum norm. Note that (2.12) follows using integration by parts when F is sufficiently smooth so that $\partial^\alpha F$ exists in the sense of classic partial derivatives. For fuller details see [12] and the references therein.

3. Distributional action of the infinitesimal operator

In this section we state and prove the main result of the paper (Theorem 1). This result will be further refined and extended in the subsequent two sections.

1. We assume throughout that the setting and notation of Section 2 remain in place. Thus we consider a continuous time-homogeneous (strong) Markov process $X = (X_t)_{t \geq 0}$ with values in a closed subset E of $I\!R^n$ with a non-empty interior E^o and having the infinitesimal operator $\mathbb{A}$ defined on $D(\mathbb{A})$ by (2.6) above. Recall that f belonging to $D(\mathbb{A})$ is generally known to be measurable only (as X may not be strong Feller or even Feller). We assume that the action of $\mathbb{A}$ is explicitly known on sufficiently regular functions in the sense that

$$\mathbb{A} = \mathbb{D} \quad \text{on } C_c^\infty \tag{3.1}$$

where $\mathbb{D}$ is a second-order differential operator given by

$$\mathbb{D} = \sum_{i=1}^{n} \mu_i \partial_i + \frac{1}{2} \sum_{i,j=1}^{n} (\sigma\sigma^t)_{ij} \partial_{ij}^2 \tag{3.2}$$

for some $\mu = (\mu_i)_{1 \leq i \leq n} : E \rightarrow I\!R^n$ and $\sigma = (\sigma_{ij})_{1 \leq i,j \leq n} : E \rightarrow I\!R_+^{n \times n}$ belonging to C^2. Note that σ_{ij} can be zero for some and many of i and j although not for all of them (to exclude trivialities). It follows from (2.12) that the distributional value $\mathbb{D}F$ of F from L^1_{loc} is defined through its action on φ from C_c^∞ by

$$\langle \mathbb{D}F, \varphi \rangle = \langle F, \mathbb{D}^* \varphi \rangle \tag{3.3}$$

where $\mathbb{D}^*$ is the *adjoint* of $\mathbb{D}$ defined by

$$\mathbb{D}^* \varphi = -\sum_{i=1}^{n} \partial_i (\mu_i \varphi) + \frac{1}{2} \sum_{i,j=1}^{n} \partial_{ij}^2 \big((\sigma\sigma^t)_{ij} \varphi \big) \tag{3.4}$$

for $\varphi \in C^2$. When the equality (3.3) holds for all $\varphi \in C_c^\infty$ we will say that the action of $\mathbb{D}$ on F from L^1_{loc} is understood in the sense of Schwartz distributions. The main result of the paper can now be stated as follows (see also Remark 2 below).

Theorem 1. *In the setting of Section 2 recalled above, suppose that*

$$\mathbb{P}_t C_c^\infty \subseteq C^2 \tag{3.5}$$

for every $t > 0$. Then the action of $\mathbb{A}$ on its domain $D(\mathbb{A})$ coincides with the action of $\mathbb{D}$ understood in the sense of Schwartz distributions, i.e. we have

$$\langle \mathbb{A} f, \varphi \rangle = \langle f, \mathbb{D}^* \varphi \rangle \tag{3.6}$$

for all $f \in D(\mathbb{A})$ and all $\varphi \in C_c^\infty$.

Proof. Let $\varphi \in C_c^\infty$ be given and fixed. Set $d\mu = \varphi \, d\lambda$ where λ is Lebesgue measure on $I\!R^n$ and note that μ belongs to M. The proof will be carried out in 5 steps as follows. Note that the steps 2-4 are mainly included for fuller understanding and could be omitted.

1. By (2.4) and (2.6) we have

$$\langle \mathbb{A} f, \mu \rangle = \lim_{t \to 0} \Big\langle \frac{\mathbb{P}_t f - f}{t}, \mu \Big\rangle = \lim_{t \to 0} \Big\langle f, \frac{\mathbb{Q}_t \mu - \mu}{t} \Big\rangle = \lim_{t \to 0} \langle f, \nu_t \rangle = \lim_{t \to 0} \nu_t(f) =: \nu(f) \tag{3.7}$$

for all $f \in D(\mathbb{A})$ where we set $\nu_t := (\mathbb{Q}_t \mu - \mu)/t$ for $t > 0$. Note that ν_t belongs to M and $f \mapsto \nu_t(f) := \langle f, \nu_t \rangle = \int_E f \, d\nu_t$ defines a continuous linear functional on B_0 for every $t > 0$.

2. We claim that under (3.5) we have

$$\sup_{t>0} \|\nu_t\| \le \int_E |\varphi^*| \, d\lambda < \infty \tag{3.8}$$

where we set $\varphi^* = \mathbb{D}^* \varphi$ upon recaling (3.4) above. For this, fix $t > 0$ and recall that

$$\|\nu_t\| = \sup \Big\{ \int_E f \, d\nu_t \mid f \text{ is measurable and } -1 \le f \le 1 \Big\} . \tag{3.9}$$

Setting $\bar{\nu}_t := \nu_t^+ + \nu_t^-$ and recalling that C_c^∞ is dense in $L^1_{\text{loc}}(\bar{\nu}_t)$ (see [11, p. 69]) we find that (3.9) can be simplified as follows

$$\|\nu_t\| = \sup \Big\{ \int_E f \, d\nu_t \mid f \in C_c^\infty \text{ and } -1 \le f \le 1 \Big\} . \tag{3.10}$$

Using (2.4) and (2.10) with $f \in C_c^\infty$ satisfying $-1 \le f \le 1$ given and fixed, we find that

$$\begin{aligned} \int_E f \, d\nu_t &= \langle f, \nu_t \rangle = \Big\langle f, \frac{\mathbb{Q}_t \mu - \mu}{t} \Big\rangle = \Big\langle \frac{\mathbb{P}_t f - f}{t}, \mu \Big\rangle = \Big\langle \frac{1}{t} \int_0^t \mathbb{A} \mathbb{P}_s f \, ds, \mu \Big\rangle \\ &= \frac{1}{t} \int_0^t \big\langle \mathbb{A} \mathbb{P}_s f, \mu \big\rangle \, ds = \frac{1}{t} \int_0^t \big\langle \mathbb{D} \mathbb{P}_s f, \varphi \big\rangle \, ds = \frac{1}{t} \int_0^t \big\langle \mathbb{P}_s f, \mathbb{D}^* \varphi \big\rangle \, ds \\ &= \frac{1}{t} \int_0^t \int_E (\mathbb{P}_s f) \, \varphi^* \, d\lambda \, ds \le \frac{1}{t} \int_0^t \int_E |\mathbb{P}_s f| \, |\varphi^*| \, d\lambda \, ds \le \|f\| \int_E |\varphi^*| \, d\lambda \\ &\le \int_E |\varphi^*| \, d\lambda < \infty \end{aligned} \tag{3.11}$$

where in the sixth and seventh equality we use that $\mathbb{P}_s f \in C^2$ for all $s \in [0, t]$ by (3.5) above combined with (3.1) and integration by parts respectively. Combining (3.10) and (3.11) we see that (3.8) holds as claimed.

3. Recalling that $D(\mathbb{A})$ is dense in B_0 (with respect to the supremum norm) we see from (3.7) and (3.8) above that the *principle of uniform boundedness* (see [4, Theorem 6, p. 60]) is applicable to $(\nu_t)_{t \ge 0}$ so that

$$\lim_{t \to 0} \nu_t(f) =: \nu(f) \tag{3.12}$$

exists for every $f \in B_0$ and defines a continuous linear functional ν on B_0 satisfying

$$\|\nu\| \le \sup_{t>0} \|\nu_t\| < \infty . \tag{3.13}$$

In the standard notation this means that $\nu \in B_0^*$ and by the Hahn-Banach theorem (see [4, pp 62-63]) there exists $\hat{\nu} \in B^*$ with $\|\hat{\nu}\| = \|\nu\|$ such that $\hat{\nu} = \nu$ on B_0. Recalling that B^* is isometrically isomorphic to the space ba consisting of all finitely additive real-valued (finite) measures on $I\!R^n$ equipped with the total variation norm (see [4, pp 258-259]), we can conclude that $\hat{\nu}(f) = \int_E f\, d\hat{\nu}$ for $f \in B$ with some $\hat{\nu} \in ba$. In particular this shows that $\nu(f) = \int_E f\, d\hat{\nu}$ for $f \in B_0$ where $\hat{\nu} \in ba$. While certainly revealing this line of argument seems to be inconclusive however due to not knowing a priory whether the finitely additive measure $\hat{\nu}$ is countably additive. Motivated by this question we will now take a different tack and examine the limiting functional ν more directly by focusing on the left-hand side of its definition in (3.7) above.

4. Applying (3.7) to $f \in C_c^\infty$ from $D(\mathbb{A})$ and making use of (3.1) we find that

$$\nu(f) = \lim_{t \to 0} \nu_t(f) = \langle \mathbb{A}f, \mu \rangle = \langle \mathbb{D}f, \mu \rangle = \langle f, \mathbb{D}^*\varphi \rangle = \int_E f \varphi^*\, d\lambda \tag{3.14}$$

where in the penultimate equality we use integration by parts. It follows from (3.14) that

$$|\nu(f)| \le \|f\| \int_E |\varphi^*|\, d\lambda \tag{3.15}$$

for all $f \in C_c^\infty$. Recalling that C_c^∞ is dense in C_c (with respect to the supremum norm) and ν is continuous on $D(\mathbb{A})$ we find by (3.14) and (3.15) that (3.14) extends as follows

$$\nu(f) = \int_E f \varphi^*\, d\lambda \tag{3.16}$$

for all $f \in D(\mathbb{A}) \cap C$. Combining (3.7) and (3.16) we see that (3.6) holds for all $f \in D(\mathbb{A}) \cap C$.

5. In the final step we focus on f belonging to $D(\mathbb{A}) \setminus C$ when the arguments from the previous step are not applicable. Note moreover that the general arguments developed in this step are applicable in the previous step as well. For this, fix $t > 0$ and recall that $\nu_t \in M$ so that $\nu_t = \nu_t^+ - \nu_t^-$ with $\nu_t^\pm \in M^+$ and $\rho_t := (1/t) \int_0^t \mathbb{Q}_s \mu^* ds \in M$ where $d\mu^* = \mathbb{D}^*\varphi\, d\lambda$ so that $\rho_t = \rho_t^+ - \rho_t^-$ with $\rho_t^\pm \in M^+$. Then $\kappa_t := \nu_t^+ + \nu_t^- + \rho_t^+ + \rho_t^- \in M^+$. Using that C_c^∞ is dense in $L^1(\kappa_t)$ we know that for every $f \in B_0$ there exists $f_\varepsilon \in C_c^\infty$ such that $f_\varepsilon \to f$ in $L^1(\kappa_t)$ and therefore in $L^1(\sigma_t)$ as well when $\varepsilon \downarrow 0$ for any $\sigma_t \in \{\nu_t^+, \nu_t^-, \rho_t^+, \rho_t^-\}$. It follows therefore using the same arguments as above that

$$\begin{aligned} \nu_t(f) = \langle f, \nu_t \rangle &= \int_E f\, d\nu_t = \int_E f\, d\nu_t^+ - \int_E f\, d\nu_t^- = \int_E f_\varepsilon\, d\nu_t^+ - \int_E f_\varepsilon\, d\nu_t^- + d(\varepsilon) \\ &= \int_E f_\varepsilon\, d\nu_t + d(\varepsilon) = \langle f_\varepsilon, \nu_t \rangle + d(\varepsilon) = \Big\langle f_\varepsilon, \frac{\mathbb{Q}_t \mu - \mu}{t} \Big\rangle + d(\varepsilon) \\ &= \Big\langle \frac{\mathbb{P}_t f_\varepsilon - f_\varepsilon}{t}, \mu \Big\rangle + d(\varepsilon) = \Big\langle \frac{1}{t} \int_0^t \mathbb{A}\mathbb{P}_s f_\varepsilon\, ds, \mu \Big\rangle + d(\varepsilon) \\ &= \frac{1}{t} \int_0^t \big\langle \mathbb{D}\mathbb{P}_s f_\varepsilon, \varphi \big\rangle\, ds + d(\varepsilon) = \frac{1}{t} \int_0^t \big\langle \mathbb{P}_s f_\varepsilon, \mathbb{D}^*\varphi \big\rangle\, ds + d(\varepsilon) \\ &= \frac{1}{t} \int_0^t \big\langle f_\varepsilon, \mathbb{Q}_s \mu^* \big\rangle\, ds + d(\varepsilon) = \Big\langle f_\varepsilon, \frac{1}{t} \int_0^t \mathbb{Q}_s \mu^*\, ds \Big\rangle + d(\varepsilon) \end{aligned} \tag{3.17}$$

$$= \langle f_\varepsilon, \rho_t \rangle + d(\varepsilon) = \int_E f_\varepsilon \, d\rho_t + d(\varepsilon)$$
$$= \int_E f_\varepsilon \, d\rho_t^+ - \int_E f_\varepsilon \, d\rho_t^- + d(\varepsilon) \longrightarrow \int_E f \, d\rho_t^+ - \int_E f \, d\rho_t^-$$
$$= \int_E f \, d\rho_t = \Big\langle f, \frac{1}{t}\int_0^t \mathbb{Q}_s \mu^* \, ds \Big\rangle = \Big\langle \frac{1}{t}\int_0^t \mathbb{P}_s f \, ds, \mu^* \Big\rangle$$

for $f \in B_0$ as $\varepsilon \downarrow 0$. Letting $t \downarrow 0$ and using that $\mathbb{P}_t f \to f$ (with respect to the supremum norm) for $f \in B_0$ we see from (3.17) that

$$\nu(f) = \lim_{t \to 0} \nu_t(f) = \lim_{t \to 0} \Big\langle \frac{1}{t}\int_0^t \mathbb{P}_s f \, ds, \mu^* \Big\rangle = \langle f, \mu^* \rangle = \langle f, \mathbb{D}^* \varphi \rangle \tag{3.18}$$

for all $f \in B_0$. Combining (3.7) and (3.18) we see that (3.6) holds for all $f \in D(\mathbb{A})$ as claimed and this completes the proof. $\square$

Remark 2. On more careful inspection of the proof above we see that the sufficient condition (3.5) in Theorem 1 can be made more precise as follows

$$\mathbb{P}_t f \text{ restricted to } E^o \text{ is } C^2 \text{ for every } f \text{ in } C_c^\infty \tag{3.19}$$

whenever $t > 0$ is given and fixed. In other words, in (3.5) there is no need to establish that $x \mapsto \mathbb{P}_t f(x)$ is C^2 at the boundary points of E when $f \in C_c^\infty$ is given and fixed. This is sometimes handy as we shall see in the statement and proof of Lemma 4 below.

Example 3. Suppose that $X = (X_t)_{t \ge 0}$ is a unique weak solution to the SDE system

$$dX_t = \mu(X_t)\, dt + \sigma(X_t)\, dB_t \quad (X_0 = x) \tag{3.20}$$

where X_t belongs to $I\!R^n$, $\mu = (\mu_i)_{1 \le i \le n}$ belongs to $I\!R^n$, $\sigma = (\sigma_{ij})_{1 \le i,j \le n}$ belongs to $I\!R_+^{n \times n}$, and B_t is a standard Brownian motion in $I\!R^n$ for $t \ge 0$ with $n \ge 1$ given and fixed. Then X is a (strong) Markov process (cf. [1, Theorem 5.1, p. 14]) having the infinitesimal operator $\mathbb{A}$ satisfying (3.1) with $\mathbb{D}$ given in (3.2) above (as is readily verified by Itô's formula for instance). Moreover, if μ and σ are C^2 with Lipschitz μ'' and σ'', then it is well known (cf. [1, Section 10, p. 28] and/or [10, Theorem 40, p. 310]) that the solution X to (3.20) can be constructed such that the flow $x \mapsto X_t^x$ is C^2 on $I\!R^n$ for $t > 0$ given and fixed. It follows therefore that

$$x \mapsto \mathbb{P}_t f(x) = \mathsf{E}_x f(X_t) = \mathsf{E} f(X_t^x) \text{ is } C^2 \text{ on } I\!R^n \tag{3.21}$$

whenever $f \in C_c^\infty$. This shows that the sufficient condition (3.5) in Theorem 1 is satisfied in this case. Consequently, by the result of Theorem 1 we can conclude that the action of $\mathbb{A}$ on its domain $D(\mathbb{A})$ coincides with the action of $\mathbb{D}$ understood in the sense of Schwartz distributions (i.e. (3.6) holds). Setting

$$V(t,x) = \mathsf{E}_x f(X_t) \tag{3.22}$$

for $t \ge 0$ and $x \in E$ we see by combining the previous conclusion with (2.9) above that

$$V_t \overset{w}{=} \mathbb{D}V \text{ on } (0,\infty) \times I\!R^n \tag{3.23}$$

with $V(0,x) = f(x)$ for $x \in E$ whenever $f \in D(\mathbb{A})$. Note that V_t in (3.23) exists in the strong (classic) sense. We will extend (3.23) beyond f in $D(\mathbb{A})$ in Section 5 below.

4. Extension to boundary value problems

In this section we show how the result of Theorem 1 can be applied to boundary value problems as discussed in Section 1 above.

1. Consider the function

$$V(x) = \mathsf{E}_x G(X_{\tau_D}) \tag{4.1}$$

where $X = (X_t)_{t \ge 0}$ is a continuous time-homogeneous (strong) Markov process in $I\!R^n$ starting at x in C under P_x, the set $C \subseteq I\!R^n$ is open, the set D equals $I\!R^n \backslash C$, the stopping time $\tau_D = \inf\{\, t \ge 0 \mid X_t \in D \,\}$ is the first entry time of X into D, and G is a real-valued continuous/measurable function defined on the boundary ∂C between the sets C and D. We will assume moreover that X is a unique weak solution to the SDE system (3.20) above where μ and σ are C^2 with Lipschitz μ'' and σ''. By the arguments recalled in Example 3 above we then know that the flow

$$x \mapsto X_t^x \;\text{ is }\; C^2 \;\text{ on }\; \bar{C} \tag{4.2}$$

for every $t > 0$ given and fixed.

2. Recall that *Dynkin's characteristic operator* $\mathbb{L}$ of X is defined by

$$\mathbb{L} f(x) = \lim_{U \downarrow x} \frac{\mathsf{E}_x f(X_{\sigma_U}) - f(x)}{\mathsf{E}_x \sigma_U} \tag{4.3}$$

for f belonging to the domain $D(\mathbb{L})$ of $\mathbb{L}$ where the limit is taken over a given family of open sets $U \subset C$ shrinking down to $x \in C$ and $\sigma_U = \inf\{\, t \ge 0 \mid X_t \notin U \,\}$ is the first exit time of X from U. The strong Markov property of X applied at σ_U implies that

$$\begin{aligned} \mathsf{E}_x V(X_{\sigma_U}) &= \mathsf{E}_x \mathsf{E}_{X_{\sigma_U}} G(X_{\tau_D}) = \mathsf{E}_x \mathsf{E}_x \big(G(X_{\tau_D}) \circ \theta_{\sigma_U} \,|\, \mathcal{F}^X_{\sigma_U} \big) \\ &= \mathsf{E}_x G(X_{\sigma_U + \tau_D \circ \theta_{\sigma_U}}) = \mathsf{E}_x G(X_{\tau_D}) = V(x) \end{aligned} \tag{4.4}$$

for every open set $U \subseteq C$ containing any given and fixed $x \in C$. Using (4.4) in (4.3) we see that V belongs to $D(\mathbb{L})$ and we have

$$\mathbb{L} V = 0 \;\text{ in }\; C\,. \tag{4.5}$$

Note however that V does not belong to the domain $D(\mathbb{A})$ of the infinitesimal operator $\mathbb{A}$ of X so that the result of Theorem 1 above is not directly applicable to V. A key advantage of (4.3) in comparison with (2.6) is that X_{σ_U} appearing in (4.3) belongs to $\bar{C}$ while X_t appearing in (2.6) through (2.1) may also be outside $\bar{C}$ so that the strong Markov property of X at t can no longer be used to derive that $\mathbb{P}_t f = f$ for all $t > 0$ close to zero (which would imply that the limit in (2.6) exists and equals zero as in (4.5) above).

3. Motivated by the simple observations stated above, let a compact set $K \subseteq C$ be given and fixed, set $L = I\!R^n \backslash K$, and consider the stopped process $X^{\tau_L} = (X_{t \wedge \tau_L})_{t \ge 0}$ where $\tau_L = \inf\{\, t \ge 0 \,|\, X_t \in L \,\}$ is the first entry time of X into L under P_x for $x \in K$. Then X^{τ_L}

is a continuous time-homogeneous (strong) Markov process in $E := K$ starting at x in K under P_x . By $\mathbb{P}_t^{\tau_L}$ we denote the (forward) transition operator of X^{τ_L} defined by

$$\mathbb{P}_t^{\tau_L} f(x) = \mathsf{E}_x f(X_{t\wedge\tau_L}) \tag{4.6}$$

for $t \geq 0$, $f \in B$ and $x \in K$ (recall (2.1) above). By $\mathbb{A}^{\tau_L}$ we denote the (forward) infinitesimal operator of X^{τ_L} defined by

$$\mathbb{A}^{\tau_L} f = \lim_{t\to 0} \frac{\mathbb{P}_t^{\tau_L} f - f}{t} \tag{4.7}$$

for f belonging to $D(\mathbb{A}^{\tau_L})$ (recall (2.6) above). The strong Markov property of X applied at $t \wedge \tau_L$ implies that

$$\begin{aligned} \mathbb{P}_t^{\tau_L} V(x) &= \mathsf{E}_x V(X_{t\wedge\tau_L}) = \mathsf{E}_x \, \mathsf{E}_{t\wedge\tau_L} G(X_{\tau_D}) = \mathsf{E}_x \mathsf{E}_x \big(G(X_{\tau_D}) \circ \theta_{t\wedge\tau_L} \,|\, \mathcal{F}^X_{t\wedge\tau_L} \big) \\ &= \mathsf{E}_x G(X_{t\wedge\tau_L + \tau_D\circ\theta_{t\wedge\tau_L}}) = \mathsf{E}_x G(X_{\tau_D}) = V(x) \end{aligned} \tag{4.8}$$

for all $x \in K$ and $t > 0$. Using (4.8) in (4.7) we see that V belongs to $D(\mathbb{A}^{\tau_L})$ and we have

$$\mathbb{A}^{\tau_L} V = 0 \quad \text{in } K \, . \tag{4.9}$$

Note that (4.9) can be seen as a localised version of (4.5) expressed in terms of an infinitesimal operator rather than the characteristic operator. Moreover, it is readily verified (using Itô's formula for instance) that

$$\mathbb{A}^{\tau_L} = \mathbb{D} \quad \text{on } C_c^\infty \tag{4.10}$$

where $\mathbb{D}$ is the second-order differential operator given by (3.2) above. The question whether (4.5) can be viewed as

$$\mathbb{D} V \overset{w}{=} 0 \quad \text{in } C \tag{4.11}$$

where the derivatives are understood in the sense of Schwartz distributions has therefore been reduced to the question whether (4.9) can be viewed in this way on any compact set $K \subseteq C$ given and fixed. This reduction makes the result of Theorem 1 above applicable if we can verify that the sufficient condition (3.5) is satisfied. Recalling from Remark 2 above that (3.5) reads more precisely as (3.19) we now show that this is the case.

Lemma 4. *In the setting of Section 2 recalled above, we have*

$$\mathbb{P}_t^{\tau_L} C_c^\infty \subseteq C^2 \tag{4.12}$$

for every $t > 0$ *.*

Proof. Let $\varphi \in C_c^\infty$ be given and fixed. Without loss of generality we can assume that $\varphi = 0$ on $K \backslash K_\varepsilon$ where $K_\varepsilon = \{ x \in K \,|\, d(x, \partial K) \geq \varepsilon \}$ for some $\varepsilon > 0$ given and fixed where d denotes the Euclidean distance on $I\!R^n$. Recalling (3.19) let $x \in K^o$ be given and fixed. Then for any $t > 0$ given and fixed we have

$$\mathbb{P}_t^{\tau_L} \varphi(x) = \mathsf{E}_x \varphi(X_{t\wedge\tau_L}) = \mathsf{E}_x \varphi(X_t) I(t < \tau_L) + \mathsf{E}_x \varphi(X_{\tau_L}) I(t \geq \tau_L) = \mathsf{E} \varphi(X_t^x) I(t < \tau_L^x) \tag{4.13}$$

where we use that $X_{\tau_L} \in \partial K$ so that $\varphi(X_{\tau_L}) = 0$. Reducing dimension n to 1 for simplicity of the notation in what follows, we see from (4.13) that for $h > 0$ sufficiently small so that $x+h \in K^o$ we have

$$\mathbb{P}_t^{\tau_L}\varphi(x+h) - \mathbb{P}_t^{\tau_L}\varphi(x) = \mathsf{E}\big[\varphi(X_t^{x+h})I(t<\tau_L^{x+h})-\varphi(X_t^x)I(t<\tau_L^x)\big]\,. \tag{4.14}$$

Note that the integrand on the right-hand side of (4.14) can be written as follows

$$\begin{aligned}\varphi(X_t^{x+h})I(t<\tau_L^{x+h})-\varphi(X_t^x)I(t<\tau_L^x) = \;&\varphi(X_t^{x+h})I(t<\tau_L^{x+h})I(\tau_L^x<t)\\ &+\varphi(X_t^{x+h})I(t<\tau_L^{x+h})I(\tau_L^x=t)\\ &+\varphi(X_t^{x+h})I(t<\tau_L^{x+h})I(\tau_L^x>t)\\ &-\varphi(X_t^x)I(t<\tau_L^x)\,.\end{aligned} \tag{4.15}$$

We now claim that the first two terms on the right-hand side of (4.15) are equal to zero for all sufficiently small $h > 0$. For this, recall the known fact (cf. [10, Theorem 37, p. 301]) that the flow $x \mapsto X^x$ can be chosen to be continuous in the topology of uniform convergence on compacts in the sense that

$$\sup_{0\le s\le t}\big|X_s^{x_n}-X_s^x\big| \rightarrow 0 \tag{4.16}$$

whenever $x_n \rightarrow x$ in K as $n \rightarrow \infty$. Combining (4.16) with the fact that $X_{\tau_L}^x$ belongs to the set of points at ∂K that are regular for L (cf. [2, Theorem 11.4, p. 62]) we see that $I(t < \tau_L^{x+h})I(\tau_L^x < t)I(X_t^x \in K^\varepsilon) = 0$ for all sufficiently small $h > 0$. Moreover, using the fact that $\varphi = 0$ on $K\setminus K_\varepsilon$ combined with that fact that $x \mapsto X_t^x$ is continuous on K , we also see that $\varphi(X_t^{x+h})I(t < \tau_L^{x+h})I(\tau_L^x < t)I(X_t^x \notin K^\varepsilon) = 0$ for all sufficiently small $h > 0$. This shows that the first term on the right-hand side of (4.16) equals zero for all sufficiently small $h > 0$. Similarly, using the fact that $\varphi = 0$ on $K\setminus K_\varepsilon$ combined with the fact that $X_t^x \in \partial K$ when $\tau_L^x = t$ and the fact that $x \mapsto X_t^x$ is continuous on K , we see that $\varphi(X_t^{x+h})I(t<\tau_L^{x+h})I(\tau_L^x=t) = 0$ for all sufficiently small $h > 0$. This shows that the second term on the right-hand side of (4.16) equals zero for all sufficiently small $h > 0$. Finally, using the fact that $\varphi = 0$ on $K\setminus K_\varepsilon$ combined with the fact that $x \mapsto X_t^x$ is continuous on K , we see that $\varphi(X_t^{x+h})I(t<\tau_L^{x+h})I(\tau_L^x>t)I(X_t^x \notin K_\varepsilon) = 0$ for all sufficiently small $h > 0$. On the other hand, if $\tau_L^x > t$ with $X_t^x \in K_\varepsilon$ then $d((X_s)_{0\le s\le t}, \partial K) \ge \delta > 0$ for some δ small enough because the points at ∂K can be assumed to be regular for L (by the choice of K in the first place). Using (4.16) this shows that if $\tau_L^x > t$ with $X_t^x \in K_\varepsilon$ then $\tau_L^{x+h} > t$ so that $\varphi(X_t^{x+h})I(t<\tau_L^{x+h})I(\tau_L^x>t)I(X_t^x\in K_\varepsilon) = \varphi(X_t^{x+h})I(\tau_L^x>t)I(X_t^x\in K_\varepsilon) = \varphi(X_t^{x+h})I(\tau_L^x>t)$ for all sufficiently small $h > 0$. Inserting these conclusions back into (4.15) we get

$$\varphi(X_t^{x+h})I(t<\tau_L^{x+h})-\varphi(X_t^x)I(t<\tau_L^x) = \big(\varphi(X_t^{x+h})-\varphi(X_t^x)\big)I(t<\tau_L^x) \tag{4.17}$$

for all sufficiently small $h > 0$. This shows that

$$\begin{aligned}\lim_{h\rightarrow 0}\frac{\varphi(X_t^{x+h})I(t<\tau_L^{x+h})-\varphi(X_t^x)I(t<\tau_L^x)}{h} &= \lim_{h\rightarrow 0}\frac{\varphi(X_t^{x+h})-\varphi(X_t^x)}{h}I(t<\tau_L^x)\\ &= \varphi'(X_t^x)\,\partial_x X_t^x\,I(t<\tau_L^x)\end{aligned} \tag{4.18}$$

with $\partial_x X_0^x = 1$. Using this fact in (4.14) we find by the mean value theorem and the dominated convergence theorem that $\partial_x \mathbb{P}_t^{\tau_L} \varphi(x)$ exists and is given by

$$(4.19) \qquad \partial_x \mathbb{P}_t^{\tau_L} \varphi(x) = \lim_{h \to 0} \frac{\mathbb{P}_t^{\tau_L} \varphi(x+h) - \mathbb{P}_t^{\tau_L} \varphi(x)}{h} = \mathsf{E}\big[\varphi'(X_t^x)\, \partial_x X_t^x\, I(t<\tau_L^x)\big]$$
$$= \mathsf{E}_x\big[\varphi'(X_t)\, \partial_x X_t\, I(t<\tau_L)\big] = \mathsf{E}_x\big[\varphi'(X_{t\wedge\tau_L})\, \partial_x X_{t\wedge\tau_L}\big]$$

for all $x \in K^0$. In particular, this shows that $x \mapsto \mathbb{P}_t^{\tau_L} \varphi(x)$ is C^1 on K^o for every $\varphi \in C_c^\infty$. Using the same arguments we similarly find that $\partial_x^2 \mathbb{P}_t^{\tau_L} \varphi(x)$ exists and is given by

$$(4.20) \qquad \partial_x^2 \mathbb{P}_t^{\tau_L} \varphi(x) = \lim_{h \to 0} \frac{\partial_x \mathbb{P}_t^{\tau_L} \varphi(x+h) - \partial_x \mathbb{P}_t^{\tau_L} \varphi(x)}{h}$$
$$= \mathsf{E}\big[\big(\varphi''(X_t^x)(\partial_x X_t^x)^2 + \varphi'(X_t^x)\partial_x^2 X_t^x\big)\, I(t<\tau_L^x)\big]$$
$$= \mathsf{E}_x\big[\big(\varphi''(X_t)(\partial_x X_t)^2 + \varphi'(X_t)\partial_x^2 X_t\big)\, I(t<\tau_L)\big]$$
$$= \mathsf{E}_x\big[\varphi''(X_{t\wedge\tau_L})(\partial_x X_{t\wedge\tau_L})^2 + \varphi'(X_{t\wedge\tau_L})\partial_x^2 X_{t\wedge\tau_L}\big]$$

for all $x \in K^0$. In particular, this shows that $x \mapsto \mathbb{P}_t^{\tau_L} \varphi(x)$ is C^2 on K^o for every $\varphi \in C_c^\infty$ as claimed and the proof is complete. □

We can now answer the question addressed prior to Lemma 4 above.

Corollary 5. *In the setting of Section 2 recalled above, consider the function* V *defined in* (4.1) *above. We then have*

$$(4.21) \qquad \mathbb{D}V \overset{w}{=} 0 \quad in\ C$$

where $\mathbb{D}$ *is the second-order differential operator given by* (3.2) *above and the derivatives in* (4.21) *are understood in the sense of Schwartz distributions.*

Proof. For any $\varphi \in C_c^\infty$ given and fixed we know that the support of φ is contained in K for some compact set K . Setting $L = I\!R^n \setminus K$ and considering the stopped process X^{τ_L} as in (4.6) above, we know by (4.12) in Lemma 4 that the sufficient condition (3.5) in Theorem 1 is satisfied with $\mathbb{P}_t^{\tau_L}$ in place of $\mathbb{P}_t$ for $t > 0$. By the result of Theorem 1 we therefore know that the equality (3.6) holds with $\mathbb{A}^{\tau_L}$ in place of $\mathbb{A}$. Recalling (4.9) this in turn means that (4.21) is satisfied as claimed and the proof is complete. □

4. The conclusion (4.21) of Corollary 5 extends from (4.1) to more general functions

$$(4.22) \qquad V(x) = \mathsf{E}_x\bigg(e^{-\int_0^{\tau_D} \lambda(X_t)\, dt} G(X_{\tau_D}) + \int_0^{\tau_D} e^{-\int_0^t \lambda(X_s)\, ds} H(X_t)\, dt\bigg)$$

for $x \in C$ where the setting is the same as following (4.1) above and $\lambda : I\!R^n \to [0,\infty)$ is measurable, $G : \partial C \to I\!R$ is continuous/measurable, and $H : \bar{C} \to I\!R$ is continuous. This can be done by introducing the killed process $\hat{X} = (\hat{X}_t)_{t \ge 0}$ defined by

$$(4.23) \qquad \hat{X}_t = X_t \ \text{ if } \ t < \zeta$$

$$= \Delta \quad \text{if } \; t \geq \zeta$$

for $t \geq 0$ where $\zeta := \inf \{ \, t \geq 0 \, | \, A_t \geq e \, \}$ for $A_t := \int_0^t \lambda(X_s) \, ds$ and $e \sim \mathrm{Exp}(1)$ realised independently from X (i.e. the driving Brownian motion B in (3.20) above) while Δ is a cemetery point (coffin state) added externally to E and all functions from the extended state space $E \cup \{\Delta\}$ to $\mathbb{R}$ take value 0 on Δ by definition. The killed process $\hat{X}$ is (strong) Markov and the characteristic operator $\hat{\mathbb{L}}$ of $\hat{X}$ is given by

$$\hat{\mathbb{L}} = \mathbb{L} - \lambda I \tag{4.24}$$

where $\mathbb{L}$ is the characteristic operator of X and I is the identity operator. From (4.23) we see that (4.22) can be rewritten as follows

$$V(x) = \mathsf{E}_x \Big(G(\hat{X}_{\tau_D}) + \int_0^{\tau_D} H(\hat{X}_t) \, dt \Big) =: V^G(x) + V^H(x) \tag{4.25}$$

where $V^G(x) = \mathsf{E}_x G(\hat{X}_{\tau_D})$ and $V^H(x) = \mathsf{E}_x \int_0^{\tau_D} H(\hat{X}_t) \, dt$ for $x \in C$. Using the strong Markov property of $\hat{X}$ as in (4.4) we find that V^G belongs to $D(\hat{\mathbb{L}})$ and by (4.5) we have

$$\hat{\mathbb{L}} V^G = \mathbb{L} V^G - \lambda V^G = 0 \;\text{ in }\; C\,. \tag{4.26}$$

Similarly, the strong Markov property of X applied at σ_U as in (4.4) above implies that

$$\begin{aligned} \mathsf{E}_x V^H(\hat{X}_{\sigma_U}) &= \mathsf{E}_x \, \mathsf{E}_{\hat{X}_{\sigma_U}} \Big(\int_0^{\tau_D} H(\hat{X}_t) \, dt \Big) = \mathsf{E}_x \, \mathsf{E}_x \Big(\int_0^{\tau_D} H(\hat{X}_t) \, dt \circ \theta_{\sigma_U} \, \big| \, \mathcal{F}^X_{\sigma_U} \Big) \\ &= \mathsf{E}_x \Big(\int_0^{\tau_D \circ \theta_{\sigma_U}} H(\hat{X}_{t+\sigma_U}) \, dt \Big) = \mathsf{E}_x \Big(\int_{\sigma_U}^{\sigma_U + \tau_D \circ \theta_{\sigma_U}} H(\hat{X}_t) \, dt \Big) \\ &= \mathsf{E}_x \Big(\int_0^{\tau_D} H(\hat{X}_t) \, dt - \int_0^{\sigma_U} H(\hat{X}_t) \, dt \Big) = V^H(x) - \mathsf{E}_x \Big(\int_0^{\sigma_U} H(\hat{X}_t) \, dt \Big) \end{aligned} \tag{4.27}$$

for every open set $U \subseteq C$ containing any given and fixed $x \in C$. Using (4.27) in (4.3) we see that V^H belongs to $D(\hat{\mathbb{L}})$ and we have

$$\hat{\mathbb{L}} V^H(x) = \lim_{U \downarrow x} \frac{\mathsf{E}_x V^H(\hat{X}_{\sigma_U}) - V^H(x)}{\mathsf{E}_x \sigma_U} = - \lim_{U \downarrow x} \frac{\mathsf{E}_x \big(\int_0^{\sigma_U} H(\hat{X}_t) \, dt \big)}{\mathsf{E}_x \sigma_U} = -H(x) \tag{4.28}$$

for $x \in C$ where in the last equality we use that H is continuous. This shows that

$$\hat{\mathbb{L}} V^H = \mathbb{L} V^H - \lambda V^H = -H \quad \text{in } \; C\,. \tag{4.29}$$

Combining (4.25) with (4.26) and (4.29) we see that V belongs to $D(\mathbb{L})$ and we have

$$\mathbb{L} V - \lambda V = -H \quad \text{in } \; C\,. \tag{4.30}$$

Note as in (4.5) above however that V does not belong to the domain $D(\mathbb{A})$ of the infinitesimal operator $\mathbb{A}$ of X so that the result of Theorem 1 above is not directly applicable to V.

5. Proceeding in the same way as in (4.6)-(4.9) above with $\hat{X}$ in place of X , we find that V^G belongs to $D(\hat{\mathbb{A}}^{\tau_L})$ and we have

$$\mathbb{A}^{\tau_L} V^G = \mathbb{A}^{\tau_L} V^G - \lambda V^G = 0 \quad \text{in } K \tag{4.31}$$

where $K \subseteq C$ is any compact set and $L = I\!R^n \setminus K$. Similarly, applying the same arguments as in (4.27) and (4.28) above we find that V^H belongs to $D(\hat{\mathbb{A}}^{\tau_L})$ and we have

$$\mathbb{A}^{\tau_L} V^H = \mathbb{A}^{\tau_L} V^H - \lambda V^H = -H \quad \text{in } K\,. \tag{4.32}$$

Combining (4.25) with (4.31) and (4.32) we see that V belongs to $D(\mathbb{A}^{\tau_L})$ and we have

$$\mathbb{A}^{\tau_L} V - \lambda V = -H \quad \text{in } K\,. \tag{4.33}$$

Recalling (4.10) hence we see that the question whether (4.30) can be viewed as

$$\mathbb{D} V - \lambda V \overset{w}{=} -H \quad \text{in } C \tag{4.34}$$

where the derivatives are understood in the sense of Schwartz distributions has therefore been reduced to the question whether (4.33) can be viewed in this way on any compact set $K \subseteq C$ given and fixed. This reduction makes the result of Theorem 1 above applicable if we can verify that the sufficient condition (3.5) is satisfied. The latter can be done by applying the result of Lemma 4 as in the proof of Corollary 5 above.

Corollary 6. *In the setting of Section 2 recalled above, consider the function V defined in* (4.22) *above. We then have*

$$\mathbb{D} V - \lambda V \overset{w}{=} -H \quad \text{in } C \tag{4.35}$$

where $\mathbb{D}$ is the second-order differential operator given by (3.2) *above and the derivatives in* (4.35) *are understood in the sense of Schwartz distributions.*

Proof. For any $\varphi \in C_c^\infty$ given and fixed we know that the support of φ is contained in K for some compact set K . Setting $L = I\!R^n \setminus K$ and considering the stopped process X^{τ_L} as in (4.6) above, we know by (4.12) in Lemma 4 that the sufficient condition (3.5) in Theorem 1 is satisfied with $\mathbb{P}_t^{\tau_L}$ in place of $\mathbb{P}_t$ for $t > 0$. By the result of Theorem 1 we therefore know that the equality (3.6) holds with $\mathbb{A}^{\tau_L}$ in place of $\mathbb{A}$. Recalling (4.33) this in turn means that (4.35) is satisfied as claimed and the proof is complete. □

Remark 7. The fact that V from (4.22) satisfies (4.35) is claimed to be 'obvious' in the proof of Corollary 8.2 in [13]. The background arguments given in the proof of Theorem 8.1 in [13] leading to this conclusion seem to invoke approximations of the degenerate parabolic PDEs by the (non-degenerate) elliptic PDEs upon making use of the previously established bounds in the PDE literature on the derivatives of the solutions that are independent of the ellipticity of the approximating PDEs (cf. Theorem 10.1 in [13]). The present paper partly grew out from our inability to follow the arguments in [13] and desire to produce simpler and/or more canonical (probabilistic) arguments which would also be applicable in greater generality. The derivation of (4.35) outlined in the proof of Corollary 6 above relies upon the general result of Theorem 1 above. This approach is different from the approach undertaken in [13].

6. Having a weak solution V from (4.22) to the equation (4.35) provides an opportunity for its upgrade to a strong (classic/smooth) solution. This can be done when the differential operator $\mathbb{D}$ of X given by (3.2) above is hypoelliptic, e.g. when it satisfies the Hörmander condition (cf. [5]). For this, note that $\mathbb{D}$ can be rewritten as the 'sum of squares' as follows

$$\mathbb{D} = \sum_{i=1}^n \mu_i \partial_i + \frac{1}{2} \sum_{i,j=1}^n (\sigma\sigma^t)_{ij} \partial_{ij}^2 = D_0 + \sum_{i=1}^m D_i^2 \tag{4.36}$$

where D_i is a first-order differential operator given by

$$D_i = \sum_{j=1}^n \beta_{ij} \partial_j \tag{4.37}$$

for $0 \le i \le m$ with the coefficients β_{ij} expressed explicitly as

$$\beta_{0j} = \mu_j - \frac{1}{2} \sum_{k,l=1}^n \sigma_{lk} \partial_l \sigma_{jk} \;\; (1 \le j \le n) \tag{4.38}$$

$$\beta_{ij} = \tfrac{1}{\sqrt{2}} \sigma_{ji} \;\; (1 \le i \le m) \; (1 \le j \le n) \tag{4.39}$$

and $m = n$ (generally m could also differ from n in (4.36) above). Identifying D_i with $(\beta_{i1}, \ldots, \beta_{in})$ we see that each D_i may be viewed as a function from C to $I\!R^n$ defined by $D_i(x) = (\beta_{i1}(x), \ldots, \beta_{in}(x))$ for $x \in C$ and $0 \le i \le n$. The *Lie bracket* of D_i and D_j understood as differential operators is defined by

$$[D_i, D_j] = D_i D_j - D_j D_i \tag{4.40}$$

for $0 \le i, j \le n$. The smallest vector space in $I\!R^n$ that (i) contains all $D_0, D_1, \ldots, D_n$ understood as vectors in $I\!R^n$ and (ii) is closed under the Lie bracket operation (4.40) is referred to as the *Lie algebra* generated by $D_0, D_1, \ldots, D_n$ and is denoted by $Lie(D_0, D_1, \ldots, D_n)$. In other words $Lie(D_0, D_1, \ldots, D_n) = \text{span}\, \{D_i, [D_i, D_j], [[D_i, D_j], D_k], \ldots \,|\, 0 \le i, j, k, \ldots \le n\}$. Note that $Lie(D_0, D_1, \ldots, D_n)$ may be viewed as a function from C into the family of linear subspaces of $I\!R^n$ whose (algebraic) dimensions could also be strictly smaller than n.

Corollary 8. *In the setting of Section 2 recalled above, consider the function V defined in* (4.22) *above. Suppose moreover that $\mu = (\mu_i)_{1 \le i \le n}$ and $\sigma = (\sigma_{ij})_{1 \le i,j \le n}$ belong to C^∞ on C. If the Hörmander condition*

$$\dim Lie(D_0, D_1, \ldots, D_n) = n \tag{4.41}$$

holds on C, then V belongs to C^∞ on C whenever H does so. In this case we have

$$\mathbb{D}V - \lambda V = -H \quad \text{in } C \tag{4.42}$$

where $\mathbb{D}$ is the second-order differential operator given by (3.2) *above.*

Proof. By (4.35) in Corollary 6 we know that V is a weak solution to the equation (4.42) in the sense of Schwartz distributions. By (4.41) and Theorem 1.1 in [5] we know that the differential operator $\mathbb{D}-\lambda I$ is hypoelliptic in C. Combining these two facts we can then conclude that the weak solution V to the equation (4.42) belongs to C^∞ on C whenever H does so. This completes the proof. $\square$

5. Extension to initial value problems

In this section we show how the result of Theorem 1 can be applied to initial value problems as discussed in Section 1 above.

1. Consider the function

$$V(t,x) = \mathsf{E}_x\Big(e^{-\int_0^t \lambda(X_s)\,ds} G(X_t) + \int_0^t e^{-\int_0^s \lambda(X_r)\,dr} H(X_s)\,ds\Big) \tag{5.1}$$

for $(t,x) \in [0,\infty)\times I\!R^n$ where $X = (X_t)_{t\ge 0}$ is a continuous time-homogeneous (strong) Markov process in $I\!R^n$ starting at x under P_x, the function $\lambda : I\!R^n \rightarrow [0,\infty)$ is measurable, the function $G : \partial C \rightarrow I\!R$ is continuous/measurable, and the function $H : I\!R^n \rightarrow I\!R$ is continuous. We will moreover assume that X is a unique weak solution to the SDE system (3.20) above where μ and σ are C^2 with Lipschitz μ'' and σ''.

2. The conclusion (4.35) of Corollary 6, which in turn relies upon the result of Theorem 1, can be applied to the function V from (5.1). This can be done by introducing a backward time-space process $\bar{X} = (\bar{X}_t)_{t\ge 0}$ defined as a flow by

$$\bar{X}_s^{t,x} = (t-s, X_s^x) \tag{5.2}$$

for $s \ge 0$ where $(t,x) \in [0,\infty)\times I\!R^n$ is given and fixed. Setting $C = (0,\infty)\times I\!R^n$ so that $D = I\!R^{n+1}\backslash C = (-\infty,0]\times I\!R^n$ we see that

$$\tau_D^{t,x} = \inf\{\, s \ge 0 \,|\, \bar{X}_s \in D\,\} \tag{5.3}$$

equals t so that (5.1) can be rewritten as follows

$$V(\bar{x}) = \mathsf{E}_{\bar{x}}\Big(e^{-\int_0^{\tau_D} \lambda(\bar{X}_s)\,ds} G(\bar{X}_{\tau_D}) + \int_0^{\tau_D} e^{-\int_0^s \lambda(\bar{X}_r)\,dr} H(\bar{X}_s)\,ds\Big) \tag{5.4}$$

for $\bar{x} = (t,x) \in C$ where we extend definitions of λ, G and H by setting $\lambda(\bar{x}) := \lambda(x)$, $G(\bar{x}) := G(x)$ and $H(\bar{x}) := H(x)$ for $\bar{x} = (t,x) \in C$. Note that the function V in (5.4) has the same structure as the function V in (4.22) above. From (5.2) it is clear that the differential operator $\bar{D}$ of $\bar{X}$ (in the sense of (3.1) above) is given by

$$\bar{\mathbb{D}} = -\partial_t + \mathbb{D} \tag{5.5}$$

where $\mathbb{D}$ is the second-order differential operator given by (3.2) above. Combining these facts then leads to the following conclusion.

Corollary 9. *In the setting of Section 2 recalled above, consider the function* V *defined in* (5.1) *above. We then have*

$$V_t \overset{w}{=} \mathbb{D}V - \lambda V + H \quad \textit{in } C \tag{5.6}$$

where $\mathbb{D}$ *is the second-order differential operator given by* (3.2) *above and the derivatives in* (5.6) *are understood in the sense of Schwartz distributions.*

Proof. Recall that (5.1) can be rewritten as (5.4) which in turn has the same structure as (4.22) above. It follows therefore from (4.35) in Corollary 6 above upon recalling (5.5) that

$$\bar{\mathbb{D}} - \lambda V = -V_t + \mathbb{D}V - \lambda V \overset{w}{=} -H \quad \text{in } C\,. \tag{5.7}$$

Rearranging terms in (5.7) we obtain (5.6) as claimed and the proof is complete. □

3. To upgrade the weak solution (5.1) to the equation (5.6) using the Hörmander condition as in Corollary 8 above, note that the differential operator $\bar{\mathbb{D}}$ of $\bar{X}$ given in (5.5) above can be rewritten as the 'sum of squares' as follows

$$\bar{\mathbb{D}} = -\partial_0 + \sum_{i=1}^{n} \mu_i \partial_i + \frac{1}{2} \sum_{i,j=1}^{n} (\sigma\sigma^t)_{ij} \partial_{ij}^2 = \bar{D}_0 + \sum_{i=1}^{n} \bar{D}_i^2 \tag{5.8}$$

where $\bar{D}_i$ is a first-order differential operator given by

$$\bar{D}_i = \sum_{j=0}^{n} \beta_{ij} \partial_j \tag{5.9}$$

for $0 \le i \le n$ with the coefficients β_{ij} expressed explicitly as

$$\beta_{00} = -1 \;\; \& \;\; \beta_{i0} = 0 \;\; (1 \le i \le n) \tag{5.10}$$

in addition to (4.38) and (4.39) above (with $m = n$). Viewing $\bar{D}_i$ as functions from C to $I\!R^{n+1}$ this amounts to setting

$$\bar{D}_0 = (-1, \beta_{01}, \ldots \beta_{0n}) \;\; \& \;\; \bar{D}_i = (0, \beta_{i1}, \ldots, \beta_{in}) \;\; (1 \le i \le n)\,. \tag{5.11}$$

Note that $Lie(\bar{D}_0, \bar{D}_1, \ldots, \bar{D}_n)$ can be viewed as a function from C into the family of linear subspaces of $I\!R^{n+1}$ whose (algebraic) dimensions could also be strictly smaller than $n+1$.

Corollary 10. *In the setting of Section 2 recalled above, consider the function* V *defined in* (5.1) *above. Suppose moreover that* $\mu = (\mu_i)_{1 \le i \le n}$ *and* $\sigma = (\sigma_{ij})_{1 \le i,j \le n}$ *belong to* C^∞ *on* C *. If the (parabolic) Hörmander condition*

$$\dim Lie(\bar{D}_0, \bar{D}_1, \ldots, \bar{D}_n) = n+1 \tag{5.12}$$

holds on C *, then* V *belongs to* C^∞ *on* C *whenever* H *does so. In this case we have*

$$V_t = \mathbb{D}V - \lambda V + H \quad \textit{in } C \tag{5.13}$$

where $\mathbb{D}$ *is the second-order differential operator given by* (3.2) *above.*

Proof. By (5.6) in Corollary 9 we know that V is a weak solution to the equation (5.13) in the sense of Schwartz distributions. By (5.12) and Theorem 1.1 in [5] we know that the differential operator $\bar{\mathbb{D}}-\lambda I = -\partial_t+\mathbb{D}-\lambda I$ is hypoelliptic in C. Combining these two facts we can then conclude that the weak solution V to the equation (5.13) belongs to C^∞ on C whenever H does so. This completes the proof. □

Remark 11. Note that the conclusions (5.6) and (5.13) in Corollary 9 and Corollary 10 respectively do not require that the function G belongs to the domain $D(\mathbb{A})$ of the infinitesimal operator $\mathbb{A}$ of X. These conclusions thus improve upon those derived at the end of Example 3 using the general semigroup theory. Note however that V_t in (5.6) is only claimed to exist in the weak sense while V_t in (5.13) exists in the strong (classic) sense as well.

Example 12. A rich family of degenerate two-dimensional diffusion processes arising in quickest detection problems to which the results of the present paper are applicable can be found in [6] and further related papers. These examples served as a main motivation for deriving the results of the present paper.

Acknowledgements. The author gratefully acknowledges support from the United States Army Research Office Grant ARO-YIP-71636-MA.

References

[1] Bass, R. F. (1998). *Diffusions and Elliptic Operators.* Springer.

[2] Blumenthal, R. M. and Getoor, R. K. (1968). *Markov Processes and Potential Theory.* Academic Press.

[3] Dynkin, E. B. (1965). *Markov Processes.* Vols I & II. Springer.

[4] Dunford, N. *and* Schwartz, J. T. (1958). *Linear Operators. Part I: General Theory.* Interscience Publishers.

[5] Hörmander, L. (1967). Hypoelliptic second order differential equations. *Acta Math.* 119 (147–171).

[6] Johnson, P. *and* Peskir, G. (2017). Quickest detection problems for Bessel processes. *Ann. Appl. Probab.* 27 (1003–1056).

[7] Kannai, Y. (1971). An unsolvable hypoelliptic differential operator. *Israel J. Math.* 9 (306–315).

[8] Peskir, G. *and* Shiryaev, A. (2006). *Optimal Stopping and Free-Boundary Problems.* Lectures in Mathematics, ETH Zürich, Birkhäuser.

[9] Phillips, R. S. (1955). The adjoint semi-group. *Pacific J. Math.* 5 (269–283).

[10] Protter, P. E. (2004). *Stochastic Integration and Differential Equations.* Springer.

[11] Rudin, W. (1987). *Real and Complex Analysis.* McGraw-Hill.

[12] Schwartz, L. (1951). *Théory des Distributions.* Hermann.

[13] Stroock, D. *and* Varadhan, S. R. S. (1972). On degenerate elliptic-parabolic operators of second order and their associated diffusions. *Comm. Pure Appl. Math.* 25 (651–713).

Goran Peskir
Department of Mathematics
The University of Manchester
Oxford Road
Manchester M13 9PL
United Kingdom
goran@maths.man.ac.uk