J. Convex Anal. Vol. 19, No. 3, 2012, (609-630)
Research Report No. 2, 2010, *Probab. Statist. Group Manchester* (22 pp)

A Duality Principle for the Legendre Transform

Goran Peskir

We present a duality principle for the Legendre transform that yields the shortest path between the graphs of functions and embodies the underlying Nash equilibrium. A useful feature of the algorithm for the shortest path obtained in this way is that its implementation has a local character in the sense that it is applicable at any point in the domain with no reference to calculations made earlier or elsewhere. The derived results are applied to optimal stopping games of Brownian motion and diffusion processes where the duality principle corresponds to the semiharmonic characterisation of the value function.

1. Introduction

The purpose of the present paper is to formulate and explain a *duality principle* for the Legendre transform that yields the *shortest path* between the graphs of functions and embodies the underlying *Nash equilibrium*. We also explain a canonical role of the von Neumann minimax theorem in this context as well as draw some loose parallels with Fenchel's duality theorem. Unlike the latter theorem, however, the duality principle described below applies to graphs of a very general nature and requires no assumption of convexity or concavity. Another interesting feature of the algorithm for the shortest path obtained in this way is that its implementation has a *local character* in the sense that it is applicable at any point in the domain with no reference to calculations made earlier or elsewhere. In essence this is a consequence of the fact revealed by the duality principle that finding the shortest path between the graphs of functions is *equivalent* to establishing a Nash equilibrium.

The motivation for the developments indicated above comes from optimal stopping problems but could be equivalently restated in the language of free boundary problems. A fundamental result in the optimal stopping theory for a strong Markov process X states that the value function $\hat{V}$ of the optimal stopping problem

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

is the *smallest superharmonic* function that lies above the gain function G, and likewise the

Mathematics Subject Classification 2010. Primary 49J35, 51M25, 60G40. Secondary 53C22, 60J65, 91A15.

Key words and phrases: Legendre transform, von Neumann's minimax theorem, Fenchel's duality theorem, shortest path between graphs/obstacles, geodesic, optimal stopping problem/game, free boundary problem, Brownian motion, diffusion/Markov process, superharmonic/subharmonic/semiharmonic characterisation of the value function, Stackelberg/Nash equilibrium, primal and dual problems of optimal stopping.

value function $\check{V}$ of the optimal stopping problem

$$\check{V}(x) = \inf_{\sigma} \mathsf{E}_x H(X_\sigma) \tag{1.2}$$

is the *largest subharmonic* function that lies below the loss function H . This result dates back to Dynkin [3] and was derived in parallel to the general supermartingale or submartingale characterisation due to Snell [22] (for more details see e.g. [16]). The characterisation leads to the familiar picture where $\hat{V}$ is identified with a rope put above the obstacle G having both ends pulled to the ground (see [15, Figure 1]), and likewise $\check{V}$ is identified with a rope put below the obstacle H having both ends pulled to the sky. Both pictures refer to the case when X is a standard Brownian motion (absorbed at the end points of the interval).

A well-known minimax version of (1.1) and (1.2) is obtained by considering the optimal stopping game where the sup-player chooses a stopping time τ to maximise, and the inf-player chooses a stopping time σ to minimise, the expected payoff

$$\mathsf{M}_x(\tau,\sigma) = \mathsf{E}_x\big[G(X_\tau)\,I(\tau\!<\!\sigma) + H(X_\sigma)\,I(\sigma\!<\!\tau) + K(X_\tau)\,I(\tau\!=\!\sigma)\big] \tag{1.3}$$

where $G \le K \le H$. Defining the upper value and the lower value of the game by

$$V^+(x) = \inf_{\sigma}\sup_{\tau} \mathsf{M}_x(\tau,\sigma) \quad \& \quad V_+(x) = \sup_{\tau}\inf_{\sigma} \mathsf{M}_x(\tau,\sigma) \tag{1.4}$$

one distinguishes (i) *Stackelberg equilibrium*, meaning that $V^+(x) = V_+(x)$ for all x , so that

$$V := V^+ = V_+ \tag{1.5}$$

unambiguously defines the value of the game, and (ii) *Nash equilibrium*, meaning that there exist stopping times τ_* and σ_* such that $\mathsf{M}_x(\tau,\sigma_*) \le \mathsf{M}_x(\tau_*,\sigma_*) \le \mathsf{M}_x(\tau_*,\sigma)$ for all τ and σ and all x (in other words (τ_*,σ_*) is a saddle point). It is easily seen that the Nash equilibrium implies the Stackelberg equilibrium with $V(x) = \mathsf{M}_x(\tau_*,\sigma_*)$ for all x . A variant of the optimal stopping game above was first studied by Dynkin [5] using martingale methods similar to those of Snell [22] (for more details see [7] and the references therein).

If we formally set $H \equiv +\infty$ in (1.3) then the optimal stopping game (1.4) reduces to the optimal stopping problem (1.1) and hence the value function $V = \hat{V}$ admits the superharmonic characterisation. Likewise, if we formally set $G \equiv -\infty$ in (1.3) then the optimal stopping game (1.4) reduces to the optimal stopping problem (1.2) and hence the value function $V = \check{V}$ admits the subharmonic characterisation. The question of the *semiharmonic characterisation* in the general case (when G and H are finite valued) was recently considered in [15]. It was shown there that letting $\hat{V}$ denote the smallest superharmonic function lying between G and H , and letting $\check{V}$ denote the largest subharmonic function lying between G and H , we have $\hat{V} = \check{V}$ *if and only if* the Nash equilibrium holds (see Section 2 for fuller details). This equivalence indicates that finding the value function V is the same as 'pulling a rope' between 'two obstacles' (see [15, Figure 2]) which in turn is equivalent to establishing a Nash equilibrium (a formal proof of these claims will be given below).

The main objective of the present paper is to connect the semiharmonic characterisation of the value function with the *Legendre transform* in the variational sense of Mandelbrojt and

Fenchel (see Section 3 for definitions and further details). Letting F^* denote the *concave conjugate* of F it is well known that the *concave biconjugate*

$$F^{**}(p) = \inf_x \sup_y \big[x(p-y)+F(y)\big] \tag{1.6}$$

defines the *smallest concave* function above F. Likewise, letting F_* denote the *convex conjugate* of F it is well known that the *convex biconjugate*

$$F_{**}(p) = \sup_x \inf_y \big[x(p-y)+F(y)\big] \tag{1.7}$$

defines the *largest convex* function below F. Returning to the optimal stopping problems (1.1) and (1.2) this means that

$$\hat{V} = G^{**} \quad \& \quad \check{V} = H_{**} \tag{1.8}$$

when X is a standard Brownian motion (absorbed at the end points of the interval). The central question to be examined in this paper is whether/how the biconjugate representations (1.8) extend to the setting of the optimal stopping game (1.4) (where obtaining equality between $\hat{V}$ and $\check{V}$ is equivalent to establishing a Nash equilibrium).

To answer this question we first show in Section 3 that the Legendre transform admits a dual (geometric/analytic) interpretation for assigning its value at a point. We then show in Section 4 that this interpretation extends to a pair of functions via the *duality* relation

$$\inf_x \sup_{y\in\mathcal{A}_p^H(x)} \big[x(p-y)+G(y)\big] = \sup_x \inf_{y\in\mathcal{A}_G^p(x)} \big[x(p-y)+H(y)\big] \tag{1.9}$$

where $\mathcal{A}_p^H(x)$ and $\mathcal{A}_G^p(x)$ are admissible sets and the joint value equals $V(p)$. We finally show in Section 5 that this value represents the shortest path (geodesic) between the graphs of G and H. The duality relation itself shows that finding the shortest path between the graphs of functions is equivalent to establishing a Nash equilibrium.

The dual (geometric/analytic) interpretation of the Legendre transform reveals that *the superharmonic and subharmonic characterisations* of the value functions $\hat{V}$ and $\check{V}$ represent *dual problems* to the primal problems (1.1) and (1.2) respectively (where *sup/inf* over all stopping times in the primal problem becomes *inf/sup* over all superharmonic/subharmonic functions above/below G/H in the dual problem). The duality relation (1.9) establishes the same fact for the *semiharmonic characterisation* of the value function in the case of the optimal stopping game (1.4). In this case, due to a full symmetry, the primal and dual problems *merge* to form the duality relation itself. These conclusions rest upon the fact that the well-known duality relationship between points and lines in analysis extends to a duality relationship between stopping times and value functions in probability.

The results above extend from Brownain motion to more general diffusion processes using known properties of the fundamental solutions (eigenvalues) to the killed generator equation. This leads to a complete description of geodesics between the graphs of functions associated with such processes. More general Markov processes (such as Lévy processes for example) require separate studies that connect their geodesics to straight lines (i.e. convexity and/or concavity) and these will be undertaken elsewhere. Likewise, for the simplicity of the exposition we present the main results in one dimension only (using the interval $[0,1]$ as a canonical state space). The scope of the extension to higher dimensions will be briefly indicated through the exposition of the general Markovian results in Section 2.

2. Semiharmonic characterisation

In this section we present basic definitions and results on the semiharmonic characterisation of the value function (1.5) that will be used in the proof below.

1. In the setting of the optimal stopping game (1.3)+(1.4) we consider a strong Markov process $X = (X_t)_{t \ge 0}$ defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, \mathsf{P}_x)$ and taking values in a measurable space $(E, \mathcal{B})$, where E is a locally compact Hausdorff space with a countable base, and $\mathcal{B}$ is the Borel σ-algebra on E . It is assumed that the process X starts at x under P_x for $x \in E$ and that the sample paths of X are *right-continuous.* Recall also that X is said to be *left-continuous over stopping times* (quasi-left-continuous) if $X_{\tau_n} \to X_\tau$ P_x-a.s. whenever τ_n and τ are stopping times such that $\tau_n \uparrow \tau$ as $n \to \infty$. It is also assumed that the filtration $(\mathcal{F}_t)_{t\ge 0}$ is right-continuous and that $\mathcal{F}_0$ contains all P_x-null sets from $\mathcal{F}^X_\infty = \sigma(X_t : t \ge 0)$. The main example we have in mind is when $\mathcal{F}_t = \sigma(\mathcal{F}^X_t \cup \mathcal{N})$ where $\mathcal{F}^X_t = \sigma(X_s : 0 \le s \le t)$ and $\mathcal{N} = \{ A \subseteq \Omega : \exists\, B \in \mathcal{F}^X_\infty\, ,\, A \subseteq B\, ,\, \mathsf{P}_x(B) = 0 \}$ for $t \ge 0$ with $\mathcal{F} = \mathcal{F}_\infty$. In addition, it is assumed that the mapping $x \mapsto \mathsf{P}_x(F)$ is (universally) measurable for each $F \in \mathcal{F}$. Finally, without loss of generality we assume that Ω equals the canonical space $E^{[0,\infty)}$ with $X_t(\omega) = \omega(t)$ for $\omega \in \Omega$ and $t \ge 0$ (for further details of these hypotheses see [15, Section 2]).

2. Recall that a measurable function $F : E \to I\!R$ is *finely continuous* (i.e. continuous in the fine topology) if and only if $\lim_{t \downarrow 0} F(X_t) = F(x)$ P_x-a.s. for every $x \in E$. This property is further equivalent to the fact that the sample path $t \mapsto F(X_t(\omega))$ is right-continuous on $I\!R_+$ for every $\omega \in \Omega \setminus N$ where $\mathsf{P}_x(N) = 0$ for all $x \in E$. The functions $G, H, K : E \to I\!R$ satisfying $G \le K \le H$ in (1.3) are assumed to be finely continuous and uniformly integrable in the sense that $\mathsf{E}_x \sup_{t \ge 0} |F(X_t)| < \infty$ where F stands for either G or H . It is also assumed that $\lim_{t \uparrow T} G(X_t) = \lim_{t \uparrow T} H(X_t)$ P_x-a.s. where the horizon T (the upper bound for τ and σ in (1.4) above) may be either finite or infinite. Under these hypotheses it was shown in [7] that if X is *right-continuous* then the *Stackelberg equilibrium* holds, and if X is *right-continuous* and *left-continuous over stopping times* then the *Nash equilibrium* holds. These general results are further refined as follows.

3. Let $F : E \to I\!R$ be a measurable function, let $C \subseteq E$ be a measurable set, and set $D = E \setminus C$. Let $\tau_D = \inf \{ t \ge 0 : X_t \in D \}$ be the first entry time of X into D . The function F is said to be *superharmonic in* C if $\mathsf{E}_x F(X_{\rho \wedge \tau_D}) \le F(x)$ for every stopping time ρ and all $x \in E$. The function F is said to be *subharmonic in* C if $\mathsf{E}_x F(X_{\rho \wedge \tau_D}) \ge F(x)$ for every stopping time ρ and all $x \in E$. The function F is said to be *harmonic in* C if $\mathsf{E}_x F(X_{\rho \wedge \tau_D}) = F(x)$ for every stopping time ρ and all $x \in E$. It is easily verified using the strong Markov property of $(X_{t \wedge \tau_D})_{t \ge 0}$ and the optional sampling theorem that F is *superharmonic/subharmonic/harmonic* in C if and only if $(F(X_{t \wedge \tau_D}))_{t \ge 0}$ is a right-continuous *supermartingale/submartingale/martingale* under P_x whenever F is finely continuous and satisfies $\mathsf{E}_x \sup_{t \ge 0} |F(X_{t \wedge \tau_D})| < \infty$ for $x \in E$.

4. To state the main result we need let us consider the following two families of functions:

$$\texttt{Sup}[G,H) = \big\{ \, F : E \to [G,H] \text{ is finely continuous and superharmonic in } \{F \!<\! H\} \, \big\} \tag{2.1}$$

$$\texttt{Sub}(G,H] = \big\{ \, F : E \to [G,H] \text{ is finely continuous and subharmonic in } \{F \!>\! G\} \, \big\} \tag{2.2}$$

and let us define the following two functions:

$$\hat{V} = \inf_{F \in \mathtt{Sup}[G,H)} F \qquad \& \qquad \check{V} = \sup_{F \in \mathtt{Sub}(G,H]} F \, . \tag{2.3}$$

Note that $\hat{V}$ represents the smallest superharmonic function lying between G and H, and $\check{V}$ represents the largest subharmonic function lying between G and H. It follows from the results in [15] that if X is *right-continuous* and *left-continuous over stopping times* then

$$V = \hat{V} = \check{V} \, . \tag{2.4}$$

In fact, when X is right-continuous (and not necessarily left-continuous over stopping times), it was shown in [15] that $\hat{V} = \check{V}$ if and only if the Nash equilibrium holds. In this case, however, the families of functions (2.1) and (2.2) also need to meet the requirement that each F from (2.1) is superharmonic in $\{V < H\}$, and each F from (2.2) is subharmonic in $\{V > G\}$, where V is defined by (1.5) above. When X is left-continuous over stopping times (additionally to being right-continuous) the latter requirement is no longer needed. Indeed, this follows from the fact derived in the proof in [15, Theorem 1] that any F from (2.1) or (2.2) satisfies $F \geq V$ or $F \leq V$ respectively (while V belongs to both families). Moreover, setting $D_1 = \{V = G\}$ and $D_2 = \{V = H\}$, letting $\tau_{D_1} = \inf\{\, t \geq 0 : X_t \in D_1 \,\}$ denote the first entry time of X into D_1, and letting $\sigma_{D_2} = \inf\{\, t \geq 0 : X_t \in D_2 \,\}$ denote the first entry time of X into D_2, we then have (see [15, Theorem 1]): (i) The value function V belongs to $\mathtt{Sup}[G,H) \cap \mathtt{Sub}(G,H]$; (ii) The first entry times τ_{D_1} and σ_{D_2} are Nash optimal in the sense that $\mathsf{M}_x(\tau, \sigma_{D_2}) \leq \mathsf{M}_x(\tau_{D_1}, \sigma_{D_2}) \leq \mathsf{M}_x(\tau_{D_1}, \sigma)$ for all stopping times τ and σ and all $x \in E$; (iii) If τ_* and σ_* are Nash optimal stopping times, then $\tau_{D_1} \leq \tau_*$ P_x-a.s. and $\sigma_{D_2} \leq \sigma_*$ P_x-a.s. for all $x \in E$; (iv) The value function V is subharmonic in $C_1 = \{V > G\}$, i.e. the stopped process $(V(X_{t \wedge \tau_{D_1}}))_{t \geq 0}$ is a right-continuous submartingale; (v) The value function V is superharmonic in $C_2 = \{V < H\}$, i.e. the stopped process $(V(X_{t \wedge \sigma_{D_2}}))_{t \geq 0}$ is a right-continuous supermartingale; and (vi) The value function V is harmonic in $C_1 \cap C_2$, i.e. the stopped process $(V(X_{t \wedge \tau_{D_1} \wedge \sigma_{D_2}}))_{t \geq 0}$ is a right-continuous martingale.

5. In order to connect these results to the Legendre transform we first consider the case when either $H \equiv +\infty$ or $G \equiv -\infty$ in (1.3). This formally corresponds to the optimal stopping problems (1.1) and (1.2) where the semiharmonic characterisation reduces to the superharmonic and subharmonic characterisation of the value function respectively. We will see in the next section that this formalism is helpful since it leads to a dual (geometric/analytic) interpretation of the Legendre transform which is instrumental in the formulation of the duality principle to be explained below.

3. Legendre transform

1. The Legendre transform was named after Adrien-Marie Legendre (1752-1833). It represents an application of the duality relation between points on the graph of a function and its tangent/supporting lines specified by their slopes and intercept values. In its classical form the Legendre transform is defined for differentiable (convex/concave) functions F by

$$\mathfrak{L}[F](p) = p x_p - F(x_p) \tag{3.1}$$

where x_p is determined by solving

$$F'(x_p) = p \ . \tag{3.2}$$

Its best known application (in classical mechanics) states that the Hamiltonian (1833) is a Legendre transform of the Lagrangian (1788). While in classical/modern physics (Hamilton's principle) one is seeking a stationary value of the action (the time integral of the Lagrangian) in optimal (stochastic) control one is looking for its minimum or maximum. The former leads to the Euler-Lagrange equations (1740s) and Hamilton's equations (1830s) while the latter leads to the Pontryagin maximum principle (1950s). Their connections are obtained by combining the ideas of Lagrange multipliers with the functional/variational form of the Legendre transform. These form necessary (and sufficient) conditions for the stationarity/optimality. Sufficient (and necessary) conditions are obtained by introducing the value function (of the initial point) which leads to the Hamilton-Jacobi-Bellman equations (1840-1950s). The value function also appears in problems of optimal stopping and this leads to the Wald-Bellman equations (1940s).

2. In parallel to these global developments Friedrichs [10] introduces the idea of *duality* in 1929. In its original form this amounts to associating with the primal problem (P) $\sup_x F(x)$ its dual problem (D) $\inf_y G(y)$ via a judicious choice of the function L such that $F(x) = \inf_y L(x,y)$ and $G(y) = \sup_x L(x,y)$. The equivalence of the problems (P) and (D) is then analogous to the statement of a minimax theorem. Combining the ideas of Lagrange multipliers with the functional/variational form of the Legendre transform this leads to the development of duality methods in optimal control (see [18]) and optimal stochastic control (see [1]) that continues to date. The wide scope of these methods requires that the classic definition of Legendre transform be extended from differentiable (convex/concave) functions to more general ones. Mandelbrojt [12] and Fenchel [8] postulate such variational extensions of (3.1)+(3.2) that remain involutive in the class of convex/concave functions. These extended Legendre transforms are referred to as (convex/concave) conjugate functions of the original function (often they are also referred to as the Legendre-Fenchel transforms). The conjugate functions play a central role in the duality methods referred to above.

3. The purpose of the present section is threefold. Firstly, we explain a canonical role of the von Neumann minimax theorem in the proof of the fact that the (extended) Legendre transform is involutive at each convex/concave function. (The original derivations of this fact given by Mandelbrojt and Fenchel are different.) Secondly, we connect this fact to optimal stopping problems (1.1) and (1.2) by establishing the biconjugate representation for the value function (1.8) when X is a standard Brownian motion (absorbed at the end points of the interval) as well as extending the same representation to more general diffusion processes using known properties of the fundamental solutions (eigenvalues) to the killed generator equation. Thirdly, motivated by the question whether/how these representations extend to the setting of the optimal stopping game (1.4) we show that the (extended) Legendre transform admits a dual (geometric/analytic) interpretation for assigning its value at a point. This will enable us to formulate a duality principle for the (extended) Legendre transform in the next section and answer the question stated above.

4. Let $F : D(F) \to I\!R$ be a measurable function whose domain $D(F)$ is a subset of $I\!R$. To simplify the exposition assume that $D(F)$ equals $[0,1]$ and that F is continuous (and

thus bounded). The *concave conjugate* of F is defined by

$$F^*(p) = \inf_{x \in D(F)} \big[px - F(x)\big] \tag{3.3}$$

for $p \in I\!R$ (see Figure 1). The *convex conjugate* of F is defined by

$$F_*(p) = \sup_{x \in D(F)} \big[px - F(x)\big] \tag{3.4}$$

for $p \in I\!R$. The *concave biconjugate* of F is defined by

$$F^{**}(p) = \inf_{x \in D(F^*)} \big[px - F^*(x)\big] = \inf_{x} \sup_{y \in D(F)} \big[x(p-y) + F(y)\big] \tag{3.5}$$

for $p \in I\!R$ (see Figure 2). The *convex biconjugate* of F is defined by

$$F_{**}(p) = \sup_{x \in D(F^*)} \big[px - F_*(x)\big] = \sup_{x} \inf_{y \in D(F)} \big[x(p-y) + F(y)\big] \tag{3.6}$$

for $p \in I\!R$. Basic properties of the conjugate functions may be summarised as follows:

$$F^* \;\&\; F^{**} \text{ are concave and } F_* \;\&\; F_{**} \text{ are convex} \tag{3.7}$$

$$F_{**}(p) \le F(p) \le F^{**}(p) \text{ for all } p \in D(F) \tag{3.8}$$

$$F \le G \;\Rightarrow\; F^{**} \le G^{**} \text{ and } F_{**} \le G_{**} \tag{3.9}$$

$$F \text{ concave} \;\Rightarrow\; F^{**} = F \tag{3.10}$$

$$F \text{ convex} \;\Rightarrow\; F_{**} = F \tag{3.11}$$

where G is any other function of the same kind as F. While the properties (3.7)-(3.9) are evident from definitions, the involutive properties (3.10) and (3.11) form a key duality relation established by Mandlebrojt [12] and Fenchel [8].

5. To present another proof of (3.10) and (3.11) recall that the *von Neumann minimax theorem* [13] states: If $K \subseteq I\!R^n$ and $L \subseteq I\!R^m$ are compact and convex sets, and a continuous function $f : K \times L \to I\!R$ satisfies (i) $x \mapsto f(x,y)$ is concave on K for every fixed $y \in L$ and (ii) $y \mapsto f(x,y)$ is convex on L for every fixed $x \in K$, then there exists a *saddle point* $(x_*, y_*) \in K \times L$ for f in the sense that $f(x, y_*) \le f(x_*, y_*) \le f(x_*, y)$ for all $(x,y) \in K \times L$. From this it follows in particular that $\sup_{x \in K} \inf_{y \in L} f(x,y) = \inf_{y \in L} \sup_{x \in K} f(x,y) = f(x_*, y_*)$ (i.e. the sup and inf commute).

Replacing F by $-F$ in (3.5) it is easily seen that (3.10) reduces to (3.11). To derive (3.11) we may note that the following inequality is always satisfied

$$F_{**}(p) = \sup_{x} \inf_{y \in D(F)} \big[x(p-y) + F(y)\big] \le \inf_{y \in D(F)} \sup_{x} \big[x(p-y) + F(y)\big] = F(p) \tag{3.12}$$

for $p \in D(F)$ where the infimum in the final equality is attained at $y = p$ since otherwise the supremum over all x would be $+\infty$. The implication (3.11) therefore reduces to showing that the inequality in (3.12) is an equality (i.e. the sup and inf commute). Setting $f(x,y) = x(p-y) + F(y)$ we see that all hypotheses of the von Neumann minimax theorem are satisfied but one ($I\!R$ is not compact). Replacing the supremum over all x by the supremum over all $x \in [-n,n]$ and applying the von Neumann minimax theorem in this setting we find

$$
\begin{aligned}
F_{**}(p) &= \lim_{n\to\infty} \sup_{x\in[-n,n]} \inf_{y\in D(F)} \big[x(p-y)+F(y) \big] \\
&= \lim_{n\to\infty} \inf_{y\in D(F)} \sup_{x\in[-n,n]} \big[x(p-y)+F(y) \big] \\
&= \lim_{n\to\infty} \inf_{y\in D(F)} \big[n|p-y|+F(y) \big] \\
&= \lim_{n\to\infty} \big[n|p-y_n|+F(y_n) \big] = F(p)
\end{aligned} \tag{3.13}
$$

where the final equality follows from the fact that the (approximate) minima points y_n must converge to p since otherwise the 'penalisation' term $n|p-y_n|$ would explode as $n \to \infty$. Note also that $n|p-y_n|$ cannot converge to a strictly positive number since then (3.13) would violate the inequality in (3.12). This completes the proof of (3.11).

6. The train of thought just exposed can also be applied in more general settings where the concave/convex conjugates make sense. We refer to [21] and the references therein for further extensions of the von Neumann minimax theorem that may be useful in this context. Omitting further details in this direction we now turn to the following well-known corollary which establishes a remarkable link between the Legendre transform and optimal stopping in Theorem 3.1 below. Assuming that $F : D(F) \to I\!R$ is measurable (and bounded) where $D(F)$ is a (compact and convex) subset of $I\!R$ we have:

(3.14) $\quad F^{**}$ is the *smallest concave* function that lies above F;

(3.15) $\quad F_{**}$ is the *largest convex* function that lies below F.

Indeed, if G is a concave function such that $G \geq F$ on $D(F)$, then by (3.9) and (3.10) we have $G^{**} \geq F^{**}$ and $G^{**} = G$, so that $G \geq F^{**}$ on $D(F)$. The claim (3.14) then follows by (3.7) and (3.8). The claim (3.15) can be derived analogously.

Theorem 3.1. *Consider the optimal stopping problems* (1.1) *and* (1.2) *where* X *is a standard Brownian motion in* $[0,1]$ *absorbed at either* 0 *or* 1, *the functions* $G : [0,1] \to I\!R$ *and* $H : [0,1] \to I\!R$ *are measurable (and bounded), and the supremum and infimum are taken over all stopping times* τ *of* X. *Then*

$$\hat{V} = G^{**} \quad \& \quad \check{V} = H_{**} \tag{3.16}$$

i.e. the value function can be identified as the concave/convex biconjugate of the gain/loss function. More explicitly, this reads

$$\hat{V}(p) = \inf_{x} \sup_{y\in[0,1]} \big[x(p-y)+G(y) \big] \tag{3.17}$$

$$\check{V}(p) = \sup_{x} \inf_{y\in[0,1]} \big[x(p-y)+H(y) \big] \tag{3.18}$$

for any $p \in [0,1]$ *given and fixed.*

Proof. It is well known and easily verified (using Jensen's inequality and the optional sampling theorem) that superharmonic/subharmonic functions of X coincide with concave/convex functions (recall that a measurable function $F : [0,1] \rightarrow I\!R$ is superharmonic/subharmonic if $\mathsf{E}_x F(X_\tau)$ is smaller/larger than $F(x)$ for all stopping times τ of X and all $x \in [0,1]$). It is also well known that the value function $\hat{V}$ is concave and the value function $\check{V}$ is convex (see e.g. (2.5) in [14] for a standard argument dating back to [6, p. 115]). Since each superharmonic function above G remains above $\hat{V}$ as well, and each subharmonic function below H remains below $\check{V}$ as well, we see by (3.14) and (3.15) that (3.16) holds as claimed. From (3.14) and (3.15) we also see that (3.16)-(3.18) embody the classic superharmonic/subharmonic characterisation of the value function (see Chapter 1 in [16] and the references therein). An early proof of the latter fact in the case of standard Brownian motion is given in [6, pp. 112-126]. One may note that the 'non-negativity' of the concave majorant is not needed in this proof and the statement of this fact (see Figure 28 on p. 115 in [6] and the claim following it) unless both $G(0) \geq 0$ and $G(1) \geq 0$. The extra requirement appears to be rooted in the implication (stated on p. 100 in [6]) that if $G \leq 0$ then it is never optimal to stop (and thus $\hat{V} \equiv 0$). A possible way of interpreting the latter conclusion is to assume that 0 and 1 are killing boundary points (not belonging to the state space) so that G is set to be zero at 0 and 1 by the usual (cemetery) convention. In this case, however, it is clear that $\hat{V}$ cannot be seen as the shortest path from $G(0)$ to $G(1)$ lying above G unless both $G(0) = 0$ and $G(1) = 0$ (assuming that G is continuous). $\square$

The biconjugate representations (3.17) and (3.18) extend from Brownian motion to more general diffusion processes using known properties of the fundamental solutions (eigenvalues) to the killed generator equation. Focusing only on the case when the boundaries are absorbing and leaving other cases to similar arguments this can be done as follows.

7. Let $X = (X_t)_{t \geq 0}$ be a regular diffusion process in $[0,1]$ absorbed at either 0 or 1 , and let $\lambda \geq 0$ be given and fixed. Consider the optimal stopping problems

$$\hat{V}(x) = \sup_\tau \mathsf{E}_x e^{-\lambda\tau} G(X_\tau) \quad \& \quad \check{V}(x) = \inf_\sigma \mathsf{E}_x e^{-\lambda\sigma} H(X_\sigma) \tag{3.19}$$

for $x \in [0,1]$, where $G : [0,1] \rightarrow I\!R$ and $H : [0,1] \rightarrow I\!R$ are measurable (and bounded) functions, and the supremum and infimum are taken over all stopping times τ of X . Let $I\!L_X$ be the infinitesimal generator of X , and let φ and ψ be continuous solutions to

$$I\!L_X F = \lambda F \tag{3.20}$$

on $[0,1]$ such that φ is increasing with $\varphi(0) > 0$ and ψ is decreasing with $\psi(1) > 0$. It is well known that such solutions exist (possibly in a generalised sense) and that they are unique up to a multiplicative constant. Recall also that under regularity conditions we have

$$I\!L_X F(x) = \mu(x) F'(x) + D(x) F''(x) \tag{3.21}$$

for $x \in (0,1)$ where $\mu \in I\!R$ is the drift and $D > 0$ is the diffusion coefficient of X (see e.g. [2, Chapter 2] and [11, Section 4.6]). Note that when $\lambda = 0$ we can take $\varphi = S$ and $\psi \equiv 1$ where S is the scale function of X .

Theorem 3.2. *Consider the optimal stopping problems* (3.19)*, and let* φ *and* ψ *be the solutions to* (3.20) *defined above. Then*

$$\hat{V}(p) = \inf_x \sup_{y\in[0,1]} \Big[x\big[\tfrac{\varphi}{\psi}(p) - \tfrac{\varphi}{\psi}(y)\big] + \tfrac{G}{\psi}(y)\Big]\psi(p) \tag{3.22}$$

$$= \inf_x \sup_{y\in[0,1]} \Big[x\big[\tfrac{\psi}{\varphi}(p) - \tfrac{\psi}{\varphi}(y)\big] + \tfrac{G}{\varphi}(y)\Big]\varphi(p)$$

$$\check{V}(p) = \sup_x \inf_{y\in[0,1]} \Big[x\big[\tfrac{\varphi}{\psi}(p) - \tfrac{\varphi}{\psi}(y)\big] + \tfrac{H}{\psi}(y)\Big]\psi(p) \tag{3.23}$$

$$= \sup_x \inf_{y\in[0,1]} \Big[x\big[\tfrac{\psi}{\varphi}(p) - \tfrac{\psi}{\varphi}(y)\big] + \tfrac{H}{\varphi}(y)\Big]\varphi(p)$$

for any $p \in [0,1]$ *given and fixed.*

Proof. It is well known (see [4, Theorem 16.4]) that λ-superharmonic/subharmonic functions F of X can be characterised by the condition that F/ψ is (φ/ψ)-concave/convex or equivalently that F/φ is $(-\psi/\varphi)$-concave/convex (recall that a measurable function $F : [0,1] \to I\!R$ is λ-superharmonic/subharmonic if $\mathsf{E}_x\, e^{-\lambda\tau} F(X_\tau)$ is smaller/larger than $F(x)$ for all stopping times τ of X and all $x \in [0,1]$). While the necessity of the latter condition is easily verified by taking τ in the preceding definition to be the first exit time of X from a bounded interval, the sufficiency can be verified by a direct argument as follows. By Jensen's inequality and the optional sampling theorem we have

$$\begin{aligned} \mathsf{E}_x\, e^{-\lambda\tau} F(X_\tau) &= \mathsf{E}_x\, e^{-\lambda\tau} \psi(X_\tau)\, (F/\psi)\circ(X_\tau) \\ &= \psi(x)\, \tilde{\mathsf{E}}_x\, (F/\psi)\circ(\varphi/\psi)^{-1}\circ(\varphi/\psi)(X_\tau) \\ &\le \psi(x)\, (F/\psi)\circ(\varphi/\psi)^{-1}\big(\tilde{\mathsf{E}}_x\, (\varphi/\psi)(X_\tau)\big) \\ &= \psi(x)\, (F/\psi)\circ(\varphi/\psi)^{-1}\big((1/\psi(x)\, \mathsf{E}_x\, e^{-\lambda\tau}\varphi(X_\tau)\big) \\ &= \psi(x)\, (F/\psi)\circ(\varphi/\psi)^{-1}\big((\varphi/\psi)(x)\big) = F(x) \end{aligned} \tag{3.24}$$

where we use that $\mathsf{E}_x\, e^{-\lambda\tau}\psi(X_\tau) = \psi(x)$ and $\mathsf{E}_x\, e^{-\lambda\tau}\varphi(X_\tau) = \varphi(x)$ [since $(e^{-\lambda t}\psi(X_t))_{t\ge0}$ and $(e^{-\lambda t}\varphi(X_t))_{t\ge0}$ are (bounded) martingales] and $\tilde{\mathsf{E}}_x$ denotes the expectation under the probability measure defined by $\tilde{\mathsf{P}}_x(A) = (1/\psi(x))\, \mathsf{E}_x\, 1_A\, e^{-\lambda\tau}\psi(X_\tau)$ for A belonging to the σ-algebra where P_x is defined. This verifies the sufficiency in the case of λ-superharmonic functions, and in the case of λ-subharmonic functions the inequality only needs to be reversed. Moreover, it is also well known that the re-scaled value function $\hat{V}/\psi$ is (φ/ψ)-concave and the re-scaled value function $\check{V}/\psi$ is (φ/ψ)-convex (see e.g. (2.7) in [20] for a standard argument dating back to [6, p. 115]). Since each λ-superharmonic function above G remains above $\hat{V}$ as well, and each λ-subharmonic function below H remains below H as well, we see that (3.17) is applicable to $(\hat{V}/\psi)\circ(\varphi/\psi)^{-1}$ and $(G/\psi)\circ(\varphi/\psi)^{-1}$ in place of $\hat{V}$ and G respectively, and (3.18) is applicable to $(\check{V}/\psi)\circ(\varphi/\psi)^{-1}$ and $(H/\psi)\circ(\varphi/\psi)^{-1}$ in place of $\check{V}$ and H respectively. It can then be verified using direct calculations that this yields the representations (3.22) and (3.23). From these implications we also see that (3.22) and (3.23) embody the classic superharmonic/subharmonic characterisations of the value functions (see Chapter 1 in [16] and the references therein). $\square$

Remark 3.3. If the functions G and H in Theorems 3.1 and 3.2 are continuous, then the first entry times of the process X into the closed sets $\{\hat{V}=G\}$ and $\{\check{V}=H\}$ are optimal (i.e. the supremum and infimum are attained at these stopping times). This can be derived using standard optimal stopping techniques (see e.g. [16, Corollary 2.9]). The main focus of Theorems 3.1 and 3.2 rests on establishing the variational (deterministic) representations for $\hat{V}$ and $\check{V}$ bearing in mind that this also yields the optimal stopping times.

Remark 3.4. We assumed in Theorems 3.1 and 3.2 that the state space of the process X equals $[0,1]$ for simplicity and the results of these theorems extend to more general state spaces (bounded or unbounded) using similar arguments. The same remark applies to the boundary behaviour of the process X at the 'end' of the state space. It should be noted, however, that not every boundary behaviour leads immediately to the same conclusions. For example, if the boundary point 0 is a point of normal/instantaneous reflection for the process X, then the value function $\check{V}$ is no longer the smallest concave function above G. In this case, however, one can extend the (old) state space $[0,1]$ to a (new) state space $[-1,1]$ by symmetry and apply the results of Theorem 3.1 to the (new) evenly extended G and X. The restriction of the resulting (new) value function to $[0,1]$ is then the (old) value function in the initial problem. Similarly, if the state space of X equals $I\!R$, then quite often $e^{-\lambda t}G(X_t) \to 0$ as $t \to \infty$ so that the boundary behaviour at the 'end' of time is reminiscent of the absorbtion at 0 or 1 and the same conclusions as in Theorems 3.1 and 3.2 can be drawn (given that other technical/boundedness conditions are satisfied). This can also be done in the absence of such limiting conditions if a fuller attention is given to the technical/boundedness conditions themselves. As this programme appears to be clear and no crucial insight is to be gained from the increased generality itself we shall omit further details in this direction.

We now turn to the question whether/how the biconjugate representations (3.17)+(3.18) and (3.22)+(3.23) extend to the setting of the optimal stopping game (1.4). A closer analysis of this question has revealed that the Legendre transform admits a dual (geometric/analytic) interpretation for assigning its value at a point that will now be described.

8. *Dual interpretation.* Consider the concave conjugate function F^* defined in (3.3) above (where $D(F)$ equals $[0,1]$ for simplicity). Let $p \in I\!R$ be given and fixed. To find the value $F^*(p)$ we may proceed in two equivalent (dual) ways as follows. Firstly, note that p represents the slope of the straight line $x \mapsto px$ (passing through the origin) and that its value remains constant throughout. To find the infimum over all $x \in [0,1]$ in (3.3) we may thus take the vertical line passing through 0 as the 'reader' of the intercepts c produced by $x \mapsto px+c$ when c runs over $I\!R$ (see Figure 1). Then note that there are those lines $x \mapsto px+c$ which after starting at 0 (the reader) meet the graph of F at some point in $[0,1]$. Let us denote the set of all c satisfying this property by A_1. Next note that there also are those lines $x \mapsto px+c$ which after starting at 0 (the reader) do not meet the graph of F at any point in $[0,1]$. Let us denote the set of all c satisfying this property by A_2. The fact is that $\sup A_1$ equals $\inf A_2$ and their joint value coincides with $-F^*(p)$. This is a (mutually) dual way of looking at the Legendre transform referred to above. Both claims appear to be evident. Indeed, to see that $\sup A_1 = -F^*(p)$ one may observe that to each $c \in A_1$ there corresponds $x_c \in [0,1]$ at which $x \mapsto px+c$ meets $x \mapsto F(x)$ for the first time on $[0,1]$ (when x runs from 0

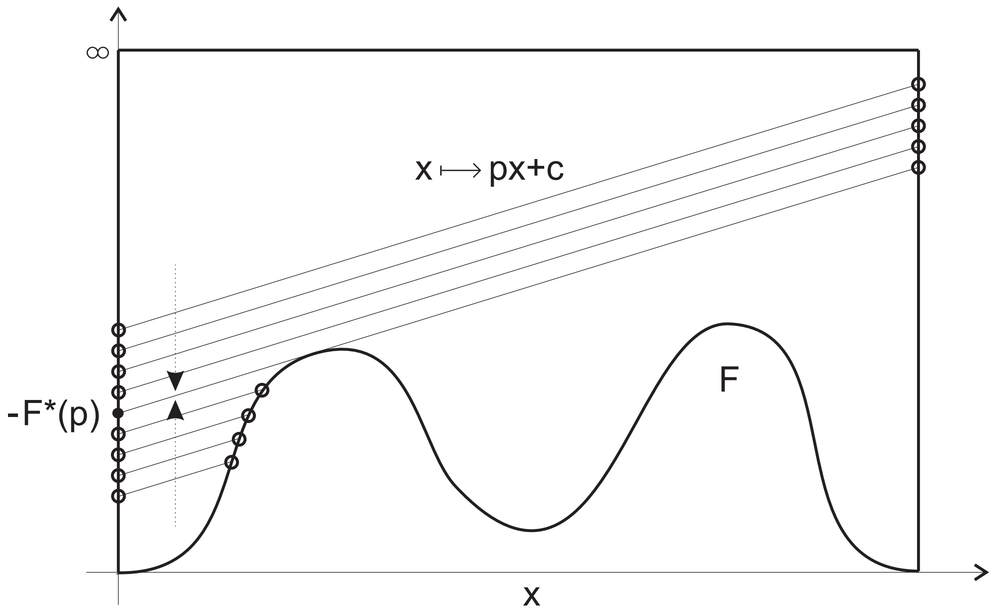

Figure 1. A dual (geometric/analytic) interpretation of the concave conjugate function $F^*(p) = \inf_x[px - F(x)]$. The analogous interpretation holds for the convex conjugate function $F_*(p) = \sup_x[px - F(x)]$.

to 1). Choosing the largest c in A_1 thus corresponds to approaching the infimum over all $x \in [0,1]$ in (3.3) arbitrarily close from above (see Figure 1). On the other hand, to see that $\inf A_2 = -F^*(p)$ one may argue oppositely and note that to each $c \in A_2$ there corresponds no $x_c \in [0,1]$ at which $x \mapsto px+c$ meets $x \mapsto F(x)$ on $[0,1]$. Choosing the smallest c in A_2 thus corresponds to approaching the infimum over all $x \in [0,1]$ in (3.3) arbitrarily close from below (see Figure 1). From these arguments we clearly see that the two values must be equal indeed. It is also clear that each c can be identified with the straight line $x \mapsto px+c$ when the slope p is given and fixed (as well as that these straight lines need to be replaced by hyperplanes in higher dimensions). In this context there is another useful aspect which we wish to highlight now. This is the fact clearly seen from Figure 1 that the two vertical lines at 0 and 1 (containing the holding points with $x \mapsto px+c$ at both ends) taken together with the horizontal line at ∞ can be viewed as the graph of a (multi-valued) function Π . This multi-valued function can in turn be obtained as the limit of (single-valued) functions Λ_n that lie above F on $[0,1]$ and tend to ∞ as $n \to \infty$. The point of this approximation is that if such a function Λ above F is given itself, then choosing the holding points with the straight lines $x \mapsto px+c$ to lie on the graph of Λ instead of the (limiting) vertical lines at 0 and 1 (on the graph of Π), one obtains a definition of the Legendre transform of F in the presence of Λ . Although we will not make use of this definition below we will see that the formal replacement of the imaginary (multi-valued) function Π with a given (single-valued) function Λ plays a helpful role in the formulation and understanding of the duality principle for the double Legendre transform to be presented below. To this end we now turn to describing a dual (geometric/analytic) interpretation of the double Legendre transform itself.

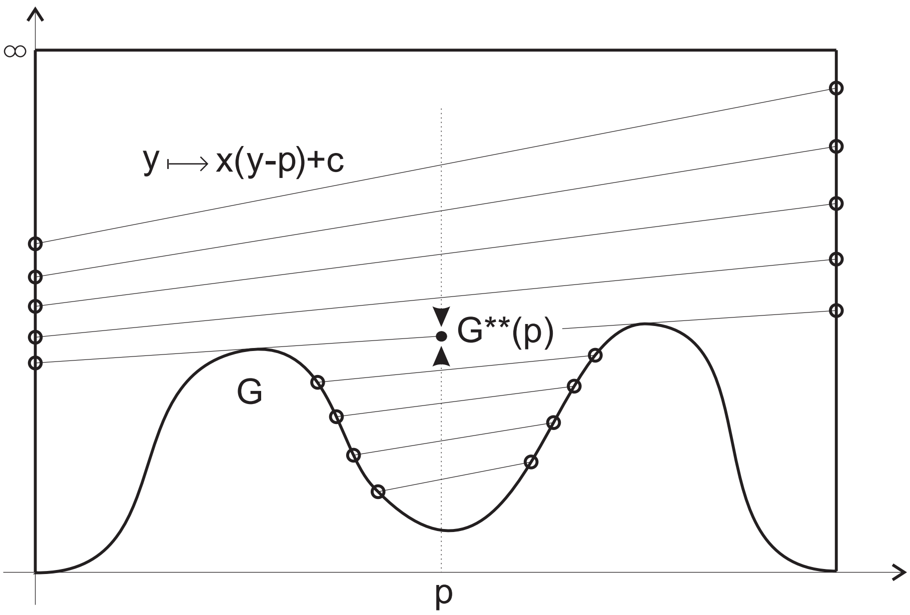

Figure 2. A dual (geometric/analytic) interpretation of the concave biconjugate function $G^{**}(p) = \inf_x \sup_y [x(p{-}y){+}G(y)]$. The analogous interpretation holds for the convex biconjugate function $H_{**}(p) = \sup_x \inf_y [x(p{-}y){+}H(y)]$.

Consider the concave biconjugate function G^{**} defined in (3.5) above (where F is replaced by G for notational convenience and $D(F)$ equals $[0,1]$ for simplicity). Let $p \in I\!R$ be given and fixed. To find the value $G^{**}(p)$ we may proceed in two equivalent (dual) ways as follows (resembling but also differing from the arguments above). Firstly, note that p no longer represents a slope but the position of the 'reader' (i.e. the vertical line passing through p) having the same role as the vertical line passing through 0 above. To find the infimum over all x and the supermum over all $y \in [0,1]$ in (3.5) we may first fix $x \in I\!R$ that now represents the slope of the straight line $y \mapsto x(y{-}p){+}c$ which figures out in the definition (3.5) after replacing the original expression $x(p{-}y){+}F(y)$ with the more intuitive expression $F(y){-}x(y{-}p)$ for $y \in [0,1]$. Now the rationale of the argument is the same as above with one notable exception: The slope x is no longer constant but needs to be chosen so to minimise the maximum of $F(y){-}x(y{-}p)$ over $y \in [0,1]$. Having understood this difference we can then proceed as before and note that there are those lines $y \mapsto x(p{-}y){+}c$ which after starting at p (the reader) meet the graph of G both before 0 and 1 (when y runs from p backwards and forwards). Let us denote the set of all c satisfying this property by A_1 . Next note as before that there also are those lines $y \mapsto x(p{-}y){+}c$ which after starting at p (the reader) do not meet the graph of G before 0 or 1 (in the previous sense). Let us denote the set of all c satisfying this property by A_2 . The fact again is that $\sup A_1$ equals $\inf A_2$ and their joint value coincides with $G^{**}(p)$. This is a (mutually) dual way of looking at the double Legendre transform referred to above. Both claims can be established in a similar way as for $-F^*(p)$ above (see Section 4 below). A crucial difference needs to be remembered, however, and this is that the slope x is no longer constant but needs to be chosen so to minimise the maximum of $F(y){-}x(y{-}p)$ over all $y \in [0,1]$. The result is shown in Figure 2 and the mapping $p \mapsto G^{**}(p)$ represents the smallest concave function that lies above $p \mapsto G(p)$ (recall (3.14) above). The comments on the hyperplanes (in higher dimensions) and the imaginary (multi-valued) function Π carry over to the present case unchanged, and it is especially the latter (through the change of Π to H) that is instrumental in revealing the duality principle to be presented next.

4. Duality principle

1. Let $G : [0,1] \rightarrow I\!R$ and $H : [0,1] \rightarrow I\!R$ be continuous functions satisfying $G \le H$ with $G(0) = H(0)$ and $G(1) = H(1)$, and let $p \in [0,1]$ be given and fixed. For $x \in I\!R$ (slope) and $c \in [G(p), H(p)]$ (height) define

$$\ell_F^p(x,c) = \sup\{\, y \in [0,p] : x(y-p)+c = F(y)\,\} \tag{4.1}$$

$$r_F^p(x,c) = \inf\{\, y \in [p,1] : x(y-p)+c = F(y)\,\} \tag{4.2}$$

where F stands for either G or H (with $\sup \emptyset = 0$ and $\inf \emptyset = 1$). Given $x \in I\!R$ define the admissible sets

$$\begin{aligned} \mathcal{A}_p^H(x) = \bigcup_{c \in [G(p),H(p)]} \Big(&\{\, y \in [0,1] :\ \ell_G^p(x,c) \le \ell_H^p(x,c) \le y \le r_H^p(x,c) \le r_G^p(x,c)\,\} \\ &\cup \{\, y \in [0,1] :\ \ell_G^p(x,c) \le y < \ell_H^p(x,c) \le r_H^p(x,c) \le r_G^p(x,c) \\ &\qquad \text{if } x(y'-p)+c \le H(y') \text{ for all } y' \in [\ell_G^p(x,c), \ell_H^p(x,c)]\,\} \\ &\cup \{\, y \in [0,1] :\ \ell_G^p(x,c) \le \ell_H^p(x,c) \le r_H^p(x,c) < y \le r_G^p(x,c) \\ &\qquad \text{if } x(y'-p)+c \le H(y') \text{ for all } y' \in [r_H^p(x,c), r_G^p(x,c)]\,\}\Big) \end{aligned} \tag{4.3}$$

$$\begin{aligned} \mathcal{A}_G^p(x) = \bigcup_{c \in [G(p),H(p)]} \Big(&\{\, y \in [0,1] :\ \ell_H^p(x,c) \le \ell_G^p(x,c) \le y \le r_G^p(x,c) \le r_H^p(x,c)\,\} \\ &\cup \{\, y \in [0,1] :\ \ell_H^p(x,c) \le y < \ell_G^p(x,c) \le r_G^p(x,c) \le r_H^p(x,c) \\ &\qquad \text{if } x(y'-p)+c \ge G(y') \text{ for all } y' \in [\ell_H^p(x,c), \ell_G^p(x,c)]\,\} \\ &\cup \{\, y \in [0,1] :\ \ell_H^p(x,c) \le \ell_G^p(x,c) \le r_G^p(x,c) < y \le r_H^p(x,c) \\ &\qquad \text{if } x(y'-p)+c \ge G(y') \text{ for all } y' \in [r_G^p(x,c), r_H^p(x,c)]\,\}\Big) \end{aligned} \tag{4.4}$$

as indicated in Figure 3 and Figure 4 respectively. The biconjugate Legendre transform of G in the presence of H is defined by

$$G_H^{**}(p) = \inf_x \sup_{y \in \mathcal{A}_p^H(x)} \big[x(p-y)+G(y)\big] \tag{4.5}$$

and the biconjugate Legendre transform of H in the presence of G is defined by

$$H_{**}^G(p) = \sup_x \inf_{y \in \mathcal{A}_G^p(x)} \big[x(p-y)+H(y)\big] \tag{4.6}$$

for $p \in [0,1]$. The $\inf A_2 / \sup A_1$ *algorithm* presented in the final paragraph of Section 3 above (applied to single-valued functions G and H analogously) provides a close *alternative* way for deriving the values (4.5) and (4.6). This is indicated in Figure 3 and Figure 4 respectively.

To see that the resulting values are the same, consider Figure 3 and note that the straight line $y \mapsto x(y-p)+c$ passing through any given height c (black dot) lying strictly above the resulting value $\inf A_2$ (the lowest black dot) can be *rotated* (clockwise or anticlockwise) until it hits G (at either side of p). The resulting angle of rotation determines the slope x at which the value of the supremum in (4.5) (taken over the resulting interval containing the second/third set in the union (4.3) above) coincides with the given height c (showing that each such height c is attained at some slope x). Taking the infimum over all x in (4.5) corresponds to moving the given height c downwards until it reaches the resulting value $\inf A_2$. Note that it cannot go strictly below $\inf A_2$ since each straight line passing through a given height c yielding a non-empty interval in the union (4.3) for some slope x can always be *translated* downwards (if needed) to create the same effect as the rotating straight line above. This shows that the resulting value $\inf A_2$ (the lowest black dot) coincides with $G_H^{**}(p)$ in (4.5). Similarly, consider Figure 4 and note that the straight line $y \mapsto x(y-p)+c$ passing through any given height c (black dot) lying strictly below the resulting value $\sup A_1$ (the highest black dot) can be *rotated* (clockwise or anticlockwise) until it hits H (at either side of p). The resulting angle of rotation determines the slope x at which the value of the infimum in (4.6) (taken over the resulting interval containing the second/third set in the union (4.4) above) coincides with the given height c (showing that each such height c is attained at some slope x). Taking the supremum over all x in (4.6) corresponds to moving the given height c upwards until it reaches the resulting value $\sup A_1$. Note that it cannot go strictly above $\sup A_1$ since each straight line passing through a given height c yielding a non-empty interval in the union (4.4) for some slope x can always be *translated* upwards (if needed) to create the same effect as the rotating straight line above. This shows that the resulting value $\sup A_1$ (the highest black dot) coincides with $H_{**}^G(p)$ in (4.6). With reference to the optimal stopping game in the proof below we remark that each straight line in the $\inf A_2/\sup A_1$ algorithm represents the value function associated with the first exit time of the process X from the interval. Alternatively these straight lines (geodesics) can also be obtained as solutions to the boundary value problem associated with the infinitesimal generator of the process X on the interval. These interpretations extend to more general diffusion/Markov processes.

Theorem 4.1 (Duality principle). *We have*

$$G_H^{**}(p) = H_{**}^G(p) \tag{4.7}$$

for all $p \in [0,1]$ *(see Figures 3-6).*

Proof. Associate with G and H the optimal stopping game (1.3)+(1.4) where X is a standard Brownian motion in $[0,1]$ absorbed at either 0 or 1. Since X is continuous we know by the results in Section 2 that the Stackelberg and Nash equilibria are satisfied in this setting. In particular, the value of the game is unambiguously defined by (1.5) and this value satisfies the identities (2.4). Recalling that finely continuous functions for X coincide with continuous functions (in the Euclidean topology), and that superharmonic/subharmonic functions for X coincide with concave/convex functions, we will now show that

$$G_H^{**} = \hat{V} \quad \& \quad H_{**}^G = \check{V} \tag{4.8}$$

on $[0,1]$. Note that after this is done the duality relation (4.7) will follow by combining the identities (4.8) with the identities (2.4) above.

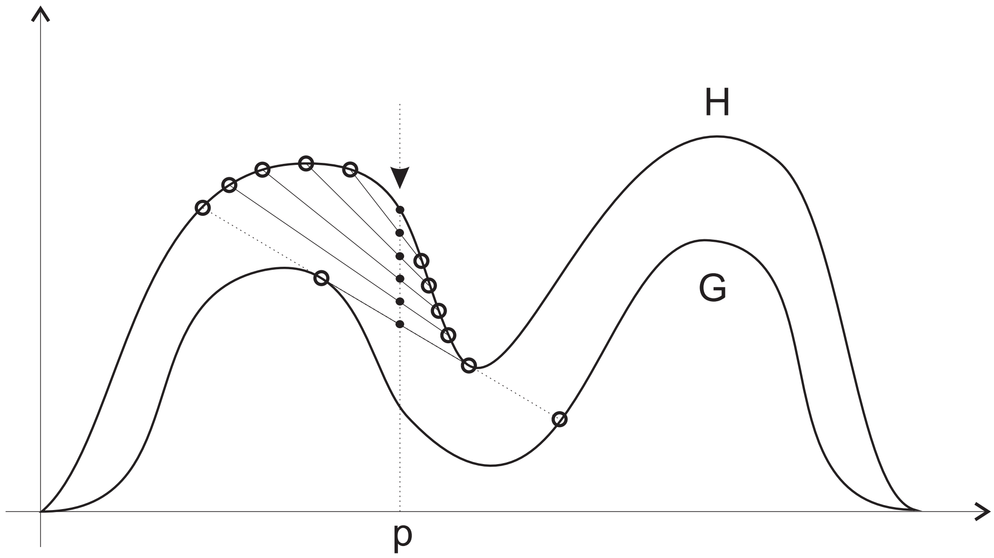

Figure 3. A dual (geometric/analytic) interpretation of the concave biconjugate function $G_H^{**}(p) = \inf_x \sup_{y \in \mathcal{A}_p^H(x)}[x(p-y)+G(y)]$ in the presence of H.

To derive the first identity in (4.8) take any $p \in (0,1)$ and set $c_* = G_H^{**}(p)$. We claim that $c_* \geq \hat{V}(p)$. Clearly, if $c_* = H(p)$ this is true, so let us suppose that $c_* < H(p)$. Then by definition of $G_H^{**}(p)$ if we take any $c \in (c_*, H(p))$ (close to c_*) we can find a slope x (depending on c) such that $\mathcal{A}_p^H(x) = [\ell_H^p(x,c), r_H^p(x,c)]$ is a nontrivial interval containing p. Consider a continuous function $F : [0,1] \to I\!R$ which is linear on $(\ell_H^p(x,c), r_H^p(x,c))$ and equals H on $[0, \ell_H^p(x,c)] \cup [r_H^p(x,c), 1]$. Note that $F(p) = c$ by definition of $\ell_H^p(x,c)$ and $r_H^p(x,c)$. Since F clearly belongs to $\texttt{Sup}[G,H]$ we see by definition of $\hat{V}$ that $\hat{V}(p) \leq F(p) = c$. Since $c \in (c_*, H(p))$ was arbitrary we can conclude that $\hat{V}(p) \leq c_*$ as claimed.

To see that $\hat{V}(p) = c_*$ let us assume that $\hat{V}(p) < c_*$. Then by definition of $\hat{V}$ there exists $F \in \texttt{Sup}[G,H]$ such that $F(p) < c_*$. Let $\ell_F = \sup\{\, y \in [0,p] : F(y) = H(y)\,\}$ and $r_F = \inf\{\, y \in [p,1] : F(y) = H(y)\,\}$. Since F is continuous it follows that $[\ell_F, r_F]$ is a nontrivial interval containing p. By definition of $\texttt{Sup}[G,H]$ we know that F is superharmonic on $[\ell_F, r_F]$ and hence concave on the same interval. Let s be a supporting line (tangent) for F at p. Set $\ell_s = \sup\{\, y \in [0,p] : s(y) = G(y)$ or $s(y) = H(y)\,\}$ and $r_s = \inf\{\, y \in [p,1] : s(y) = G(y)$ or $s(y) = H(y)\,\}$. Then by definition of $G_H^{**}(p)$ we know that either $s(\ell_s) = G(\ell_s)$ or $s(r_s) = G(r_s)$. Moreover, since $F(p) < c_*$ this is also true if we replace s by $s_\varepsilon := s+\varepsilon$ for $\varepsilon > 0$ sufficiently small. If $s_\varepsilon(\ell_{s_\varepsilon}) = G(\ell_{s_\varepsilon})$ then by definitions of $\texttt{Sup}[G,H]$ and s_ε we know that F is superharmonic on $[\ell_{s_\varepsilon}, p]$ and hence concave on the same interval. Since F is continuous and $F(0) = G(0)$ it follows that F must meet s_ε at some point in $[\ell_{s_\varepsilon}, p]$. This conclusion contradicts the fact that F is concave on $[\ell_{s_\varepsilon}, p]$. If $s_\varepsilon(r_{s_\varepsilon}) = G(r_{s_\varepsilon})$ then the same arguments can be applied to F on the interval $[p, r_{s_\varepsilon}]$ and this leads to a similar contradiction. In either case therefore we can conclude that $F(p) < c_*$ cannot be true and hence we must have $F(p) = c_*$ as claimed.

This shows that the first identity in (4.8) holds. The second identity can be derived in exactly the same way (or follows by symmetry if we replace G and H by $-G$ and $-H$

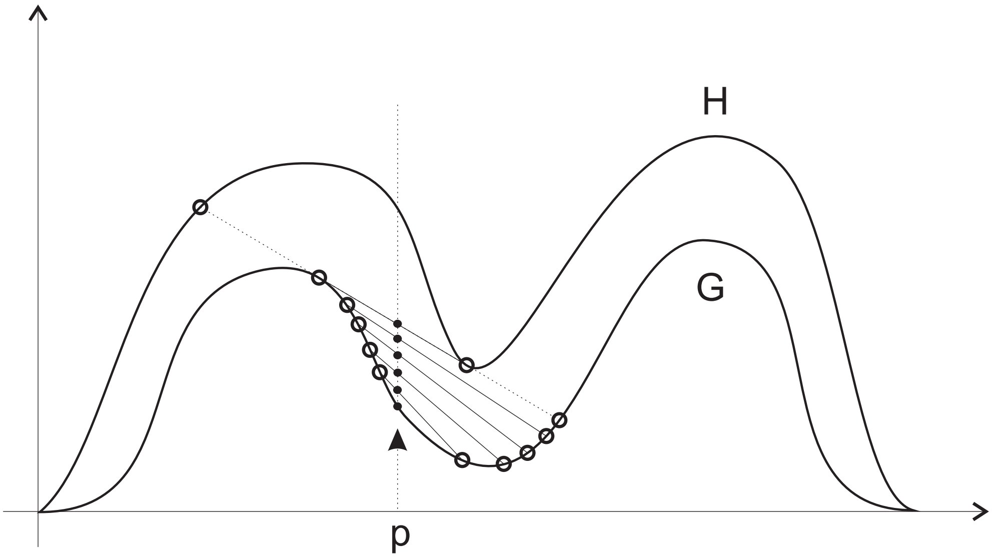

Figure 4. A dual (geometric/analytic) interpretation of the convex biconjugate function $H_{**}^{G}(p) = \sup_x \inf_{y \in \mathcal{A}_G^p(x)} [x(p-y)+H(y)]$ in the presence of G.

respectively). The duality relation (4.7) then follows by combining the identities (4.8) with the identities (2.4) as stated above. This completes the proof. $\square$

Remark 4.2. The duality relation (4.7) can also be restated by saying that the biconjugate Legendre transform of G in the presence of H coincides with the biconjugate Legendre transform of H in the presence of G. The joint value (4.7) is therefore referred to as the *biconjugate Legendre transform of G and H*. It is denoted by

$$p \mapsto \mathfrak{L}_G^H(p) \tag{4.9}$$

for $p \in [0,1]$. The proof above shows that the biconjugate Legendre transform (4.9) coincides with the value function of the optimal stopping game associated with G and H by means of standard Brownian motion in [0,1] absorbed at either 0 or 1.

Remark 4.3. It may be noted that certain elements in the statement and proof of the duality relation (4.7) are reminiscent of Fenchel's duality theorem [9] stating that the points having the minimal vertical separation between concave and convex functions are also the tangency points for the maximally separated parallel tangents (see [17] and [19]). The parallels between the two theorems appear to be both loose as well as indicative of deeper connections. Unlike Fenchel's duality theorem, however, the duality relation (4.7) applies to graphs of a very general nature and requires no assumption of convexity or concavity.

The duality relation (4.7) extends from (straight lines of) Brownian motion to (geodesics) of more general diffusion processes using known properties of the fundamental solutions (eigenvalues) to the killed generator equation. Focusing only on the case when the boundaries are absorbing and leaving other cases to similar arguments this can be done as follows.

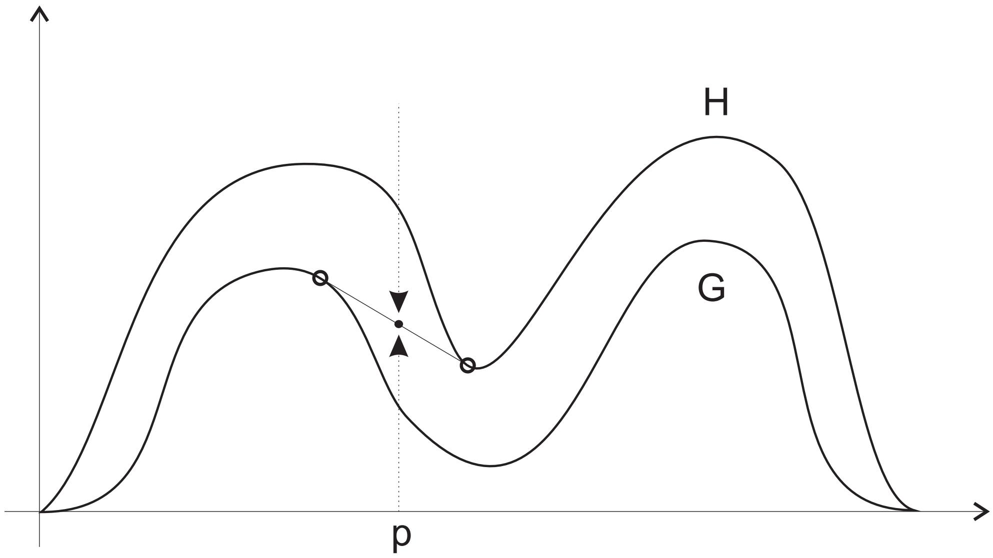

Figure 5. The duality principle for the Legendre transform stating that the concave biconjugate of G in the presence of H coincides with the convex biconjugate of H in the presence of G .

2. Let $X = (X_t)_{t \geq 0}$ be a regular diffusion process in $[0,1]$ absorbed at either 0 or 1 . Consider the optimal stopping game where the sup-player chooses a stopping time τ to maximise, and the inf-player chooses a stopping time σ to minimise, the expected payoff

$$\mathsf{M}_x^\lambda(\tau,\sigma) = \mathsf{E}_x\big[e^{-\lambda\tau}G(X_\tau)\,I(\tau<\sigma) + e^{-\lambda\sigma}H(X_\sigma)\,I(\sigma<\tau) + e^{-\lambda\tau}K(X_\tau)\,I(\tau=\sigma)\big] \tag{4.10}$$

for $x \in [0,1]$ where $G,\, H,\, K : [0,1] \rightarrow I\!\!R$ are continuous functions satisfying $G \leq K \leq H$ with $G(0) = H(0)$ and $G(1) = H(1)$. Since X is continuous we know by the results in Section 2 that the Stackelberg and Nash equilibria are satisfied so that the value of the game is unambiguously defined by

$$V(x) = \inf_\sigma \sup_\tau \, \mathsf{M}_x^\lambda(\tau,\sigma) = \sup_\tau \inf_\sigma \, \mathsf{M}_x^\lambda(\tau,\sigma) \tag{4.11}$$

for $x \in [0,1]$ with τ_{D_1} and σ_{D_2} being Nash optimal stopping times (for fuller details recall the text following (2.4) above). Let $I\!\!L_X$ be the infinitesimal generator of X , and let φ and ψ be continuous solutions to (3.20) on $[0,1]$ such that φ is increasing with $\varphi(0) > 0$ and ψ is decreasing with $\psi(1) > 0$. Recall that under regularity conditions we have that $I\!\!L_X$ is given by (3.21) above. Recall also that when $\lambda = 0$ we can take $\varphi = S$ and $\psi \equiv 1$ where S is the scale function of X . Define

$$G_{\varphi,\psi} := \tfrac{G}{\psi} \circ \big(\tfrac{\varphi}{\psi}\big)^{-1} \quad \& \quad G_{\psi,\varphi} := \tfrac{G}{\varphi} \circ \big(-\tfrac{\psi}{\varphi}\big)^{-1} \tag{4.12}$$

$$H_{\varphi,\psi} := \tfrac{H}{\psi} \circ \big(\tfrac{\varphi}{\psi}\big)^{-1} \quad \& \quad H_{\psi,\varphi} := \tfrac{H}{\varphi} \circ \big(-\tfrac{\psi}{\varphi}\big)^{-1} \tag{4.13}$$

(recall Theorem 3.1 and Theorem 3.2 above).

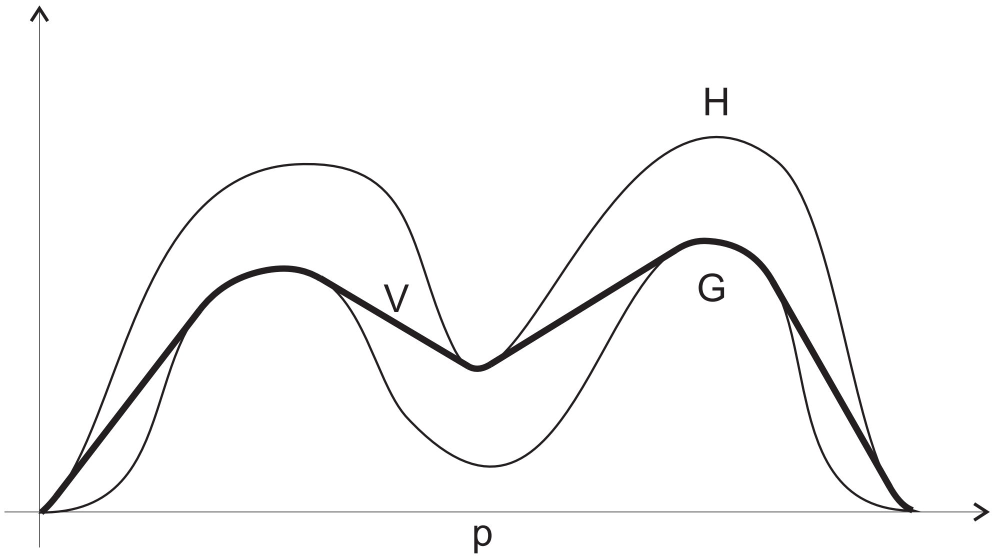

Figure 6. The duality principle for the Legendre transform yielding the shortest path between G and H by (i) depicting the semiharmonic characterisation of the value function and (ii) embodying the underlying Nash equilibrium.

Theorem 4.4. *Consider the optimal stopping game* (4.10)+(4.11), *and let* φ *and* ψ *be the solutions to* (3.20) *defined above. Then*

$$V(p) = \inf_x \sup_{y \in \mathcal{A}_p^{H_{\varphi,\psi}}(x)} \Big[x \big[\tfrac{\varphi}{\psi}(p) - y \big] + G_{\varphi,\psi}(y) \Big] \psi(p) \tag{4.14}$$

$$= \sup_x \inf_{y \in \mathcal{A}^p_{G_{\varphi,\psi}}(x)} \Big[x \big[\tfrac{\varphi}{\psi}(p) - y \big] + H_{\varphi,\psi}(y) \Big] \psi(p)$$

$$V(p) = \inf_x \sup_{y \in \mathcal{A}_p^{H_{\psi,\varphi}}(x)} \Big[x \big[\tfrac{\psi}{\varphi}(p) + y \big] + G_{\psi,\varphi}(y) \Big] \varphi(p) \tag{4.15}$$

$$= \sup_x \inf_{y \in \mathcal{A}^p_{G_{\psi,\varphi}}(x)} \Big[x \big[\tfrac{\psi}{\varphi}(p) + y \big] + H_{\psi,\varphi}(y) \Big] \varphi(p)$$

for any $p \in [0,1]$ *given and fixed.*

Proof. In parallel to (4.12) and (4.13) define

$$V_{\varphi,\psi} := \tfrac{V}{\psi} \circ \big(\tfrac{\varphi}{\psi} \big)^{-1} \quad \& \quad V_{\psi,\varphi} := \tfrac{V}{\varphi} \circ \big(-\tfrac{\psi}{\varphi} \big)^{-1} . \tag{4.16}$$

Then the arguments used in the proof of Theorem 3.2 combined with the arguments used in the proof of Theorem 4.1 show that the duality relation (4.7) leads to

$$V_{\varphi,\psi} = \big(G_{\varphi,\psi} \big)^{**}_{H_{\varphi,\psi}} = \big(H_{\varphi,\psi} \big)^{G_{\varphi,\psi}}_{**} \tag{4.17}$$

$$V_{\psi,\varphi} = \big(G_{\psi,\varphi} \big)^{**}_{H_{\psi,\varphi}} = \big(H_{\psi,\varphi} \big)^{G_{\psi,\varphi}}_{**} . \tag{4.18}$$

Substituting $p = (\varphi/\psi)^{-1}(q)$ and $p = (-\psi/\varphi)^{-1}(q)$ we see that (4.17) and (4.18) reduce to (4.14) and (4.15) respectively. Note that y in (4.15) can be taken with the positive sign since the infimum and supremum are taken over all $x \in I\!R$ (i.e. both positive and negative). This completes the proof. $\square$

5. Shortest path

Let $G : [0,1] \to I\!R$ and $H : [0,1] \to I\!R$ be continuous functions satisfying $G \le H$ with $G(0) = H(0)$ and $G(1) = H(1)$, let $\mathfrak{L}_G^H$ be the biconjugate Legendre transform of G and H defined in Remark 4.2 above, and consider the Euclidean distance in $I\!R^2$ to measure length.

Theorem 5.1. *The graph of the mapping*

$$p \mapsto \mathfrak{L}_G^H(p) \tag{5.1}$$

represents the shortest path from $(0, G(0)) = (0, H(0))$ *to* $(1, G(1)) = (1, H(1))$ *between the graphs of* G *and* H *when* p *runs from* 0 *to* 1.

Proof. We show that no continuous path between the graphs of G and H can be shorter. For this, take any continuous function $F : [0,1] \to I\!R$ satisfying $G \le F \le H$ on $[0,1]$ and suppose that $F(p) \ne \mathfrak{L}_G^H(p)$ for some $p \in (0,1)$. Consider first the case where $F(p) > \mathfrak{L}_G^H(p)$. Then the duality relation (4.7) and the definition of $G_H^{**}(p)$ yield the existence of $x \in I\!R$ such that the straight line $y \mapsto x(y-p)+F(p)$ meets the graph of H before the graph of G when y runs from p both backwards to 0 and forwards to 1. Moreover, since $F(p)$ is strictly larger than $\mathfrak{L}_G^H(p)$ this is also true if we replace $F(p)$ above with $F_\varepsilon(p) := F(p)-\varepsilon$ for $\varepsilon > 0$ sufficiently small. In other words, the the straight line $y \mapsto x(y-p)+F_\varepsilon(p)$ meets the graph of H before the graph of G when y runs from p both backwards to 0 and forwards to 1. Since $F \in [G, H]$ on $[0,1]$ it follows that the graph of $y \mapsto F(y)$ must meet the straight line $y \mapsto x(y-p)+F_\varepsilon(p)$ (for the first time) at some $(y_0, z_0) \in [0,p) \times I\!R$ when y runs from p backwards to 0, and likewise the graph of $y \mapsto F(y)$ must meet the same straight line $y \mapsto x(y-p)+F_\varepsilon(p)$ (for the first time) at some $(y_1, z_1) \in (p,1] \times I\!R$ when y runs from p forwards to 1. Since the straight line $y \mapsto x(y-p)+F_\varepsilon(p)$ represents the shortest path from (y_0, z_0) to (y_1, z_1) (relative to the Euclidean distance in $I\!R^2$), and $F(p)$ is strictly larger than $F_\varepsilon(p)$ by construction, we see that the graph of $y \mapsto F(y)$ defines a strictly longer path on $[y_0, y_1]$. This shows that the graph of F cannot represent the shortest path between the graphs of G and H on $[0,1]$ whenever $F(p) > \mathfrak{L}_G^H(p)$. The case $F(p) < \mathfrak{L}_G^H(p)$ can be ruled out in exactly the same way using the duality relation (4.7) and the definition of $H_{**}^G(p)$ instead. In either case therefore it follows that the graph of F cannot represent the shortest path from $(0, G(0)) = (0, H(0))$ to $(1, G(1)) = (0, H(0))$ between the graphs of G and H unless $F = \mathfrak{L}_G^H$ as claimed. This completes the proof. $\square$

Remark 5.2. An interesting (and computationally elegant) feature of the algorithm for the shortest path obtained in this way is that its implementation has a *local character* in the sense that it is applicable at any point in the domain with no reference to calculations made earlier or elsewhere. In essence this is a consequence of the fact revealed by the duality principle that finding the *shortest path* between the graphs of functions is *equivalent* to establishing a *Nash* *equilibrium.* The result of Theorem 5.1 extends from (straight lines of) Brownian motion to (geodesics of) more general diffusion processes using the methodology described above.

Acknowledgements. The author gratefully acknowledges financial support from (i) the Centre for the Study of Finance and Insurance, Osaka University, Japan and (ii) the Department of Mathematical Sciences & Quantitative Finance Research Centre, University of Technology Sydney, Australia where the present research was initiated (February 2010) and concluded (June 2010) respectively. The author is grateful to Professor H. Nagai at the former institution and to Professor A. A. Novikov at the latter institution for the kind hospitality and insightful discussions. The author is indebted to Professor S. Pickenhain for the valuable comments on the origins of duality in optimal control (especially [10]) made during the 5th Workshop on Nonlinear PDEs and Financial Mathematics, University of Leipzig, Germany (March 2010).

References

[1] Bismut, J. M. (1973). Conjugate convex functions in optimal stochastic control. *J. Math. Anal. Appl.* 44 (384-404).

[2] Borodin, A. N. *and* Salminen, P. (2002). *Handbook of Brownian motion: Facts and Formulae.* Birkhäuser.

[3] Dynkin, E. B. (1963). The optimum choice of the instant for stopping a Markov process. *Soviet Math. Dokl.* 4 (627-629).

[4] Dynkin, E. B. (1965). *Markov Processes.* Springer-Verlag.

[5] Dynkin, E. B. (1969). Game variant of a problem of optimal stopping. *Soviet Math. Dokl.* 10 (16-19).

[6] Dynkin, E. B. *and* Yushkevich, A. A. (1969). *Markov processes: Theorems and Problems.* Plenum Press.

[7] Ekström, E. *and* Peskir, G. (2008). Optimal stopping games for Markov processes. *SIAM J. Control Optim.* 47 (684-702).

[8] Fenchel, W. (1949). On conjugate convex functions. *Canadian J. Math.* 1 (73-77).

[9] Fenchel, W. (1953). *Convex Cones, Sets and Functions.* Princeton Univ. Press.

[10] Friedrichs, K. (1929). Ein Verfahren der Variationsrechnung das Minimum eines Integrals als das Maximum eines anderen Ausdruckes darzustellen. *Nachr. Göttingen* (13-20).

[11] Itô, K. *and* McKean, H. P. Jr. (1974). *Diffusion Processes and Their Sample Paths.* Springer-Verlag.

[12] Mandelbrojt, S. (1939). Sur les fonctions convexes. *C. R. Acad. Sci. Paris* 209 (977-978).

[13] Neumann, J. von (1928). Zur Theorie der Gesellsehaftsspiele. *Math. Ann.* 100 (295-320).

[14] Peskir, G. (2007). Principle of smooth fit and diffusions with angles. *Stochastics* 79 (293-302).

[15] Peskir, G. (2008). Optimal stopping games and Nash equilibrium. *Theory Probab. Appl.* 53 (558-571).

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

[17] Rockafellar, R. T. (1966). Extension of Fenchel's duality theorem for convex functions. *Duke Math. J.* 33 (81-89).

[18] Rockafellar, R. T. (1970). Conjugate convex functions in optimal control and the calculus of variations. *J. Math. Anal. Appl.* 32 (174-222).

[19] Rockafellar, R. T. (1970). *Convex Analysis.* Princeton Univ. Press.

[20] Samee, F. (2010). On the principle of smooth fit for killed diffusions. *Electron. Commun. Probab.* 15 (89-98).

[21] Sion, M. (1958). On general minimax theorems. *Pacific J. Math.* 8 (171-176).

[22] Snell, J. L. (1952). Applications of martingale system theorems. *Trans. Amer. Math. Soc.* 73 (293-312).

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