On the Applications of Maslov Optimization Theory

P. Del Moral*, M. Doisy †

31/03/98

Introduction

Maslov Idempotent Measure Theory has recently emerged as a new branch of mathematics analysis for studying deterministic control problems, first order non linear partial differential equations such as Hamilton-Jacobi equations with discontinuous initial date and low lying eigen functions of Schrödinger operator.
During the decade its application area has grown establishing unexpected connections with number of other fields. Litvinov and Maslov's presentation of the so-called Correspondance Principle for Idempotent Calculus [28] includes a very precise and useful summary on this subject. The beginning 1990's several different approaches were suggested to construct an optimization theory at the same level of generality as probability and stochastic processes theory [4, ?, 34].
The key idea consists in replacing in the structural axioms of Probability Theory the role of the classical semi-ring of positive real numbers $(\mathbb{R}_+, +, \cdot, 0, 1)$ by the idempotent semi-ring

$$\mathbb{R}_{\max} = (\mathbb{R}_- \cup \{-\infty\}, \oplus, \odot, \mathbb{0}, \mathbb{1}) \stackrel{\text{def}}{=} (\mathbb{R}_- \cup \{-\infty\}, \max, +, -\infty, 0)$$

This new approach has the attribute of taking full advantage of the well-developed theory of Idempotent Analysis presented in [26, 27, 29, 30, 31, 32]. The main purpose of this article is to discuss some topological aspects of this (max,+)-version of Probability Theory and to present links between this theory and applications.

The paper is organized as follows. Before to present a formulation of Maslov optimization

*LSP-UMR C55830, CNRS, Bat.1R1, Univ. Paul Sabatier, 31062 Toulouse, delmoral@cict.fr.
†ENSEEIHT, 2 Rue Camichel, 31 071 Toulouse et LSP-UMR C55830, CNRS, doisy@len7.enseeiht.fr.

theory it is important to understand some of the basic principles of idempotent Analysis. In the first section we follow fairly closely [26]. We introduce some notations, definitions and properties concerning idempotent measures on general Hausdorff topological spaces and integration theory on locally compact and separable metric spaces.
In the second section we give a brief survey of the optimization theory developed in [13]. We also introduce some additional topological structure in the definition of an optimization or decision variable so that to present this material in an absolutely rigorous way. As we shall see this natural framework is dictated by the classical large deviation calculus as presented in [19] and [20]. The concepts of performance measure, optimization variables, independence and conditioning are the basic notions of our model. Using this framework we will see that the Bellman optimization principle is the (max,+)-version of the classical Markov causality principle. The optimization processes satisfying this principle will be called Bellman-Maslov processes in reference to optimal control theory.
The third and main part of this article is concerned with presenting some applications of this new branch of idempotent analysis.
Section 3.1 is devoted to the study of fixed points of non expansive transformations on non necessarily finite state spaces. These transformations were deeply studied by Kolokoltsov, Nussbaum and Maslov. In our framework we will see that such fixed points are the (max,+)-version of the invariant probability measures for the discrete time Bellman semi-group. We discuss the problem of the existence of invariant measures and we propose several explicit representations of such objects. We will also discuss the applications of this approach in the analysis of the long time behavior of generalized simulated annealing and genetic type algorithms.
In section 3.2 we present the (max,+)-version of the celebrated Dynkin's formula in stochastic process theory. We will see that the Hamiltonian operator is the (max,+)-version of the classical Kolmogorov differential operator of Stochastic Processes Calculus. In addition a stability criterion is proposed to study the qualitative properties of optimization processes on the basis of which the performance of a given admissible control subset can be easily calibrated.
The study of Maslov optimization processes lead to new advances in the understanding of optimal control problems, in particular, of its relations with the study of the solution of first order Hamilton-Jacobi differential equations with non necessarily continuous initial data or non necessarily smooth Hamiltonian. This part of the paper records works of Kolokoltsov and Maslov [26, 27]. The main contribution here is to extend their result to more general classes of Hamilton-Jacobi equations and to connect these new generalized solutions with the classical viscosity solutions (see [5, 10, 11] and references therein). What also makes our result interesting and new is that we propose a natural and general "probabilistic line of thought" to study these Hamilton-Jacobi equations.
Using (max,+)-semigroup technic we will prove that the so-called Maslov generalized so-

lution of a first order Hamilton-Jacobi differential equation coincide with the now classical viscosity solution. We will give a complete proof of this statement and explain its relevance to obtain explicit solutions such as the Oleinik-Lax formula and to study evolution equations arising in mathematical morphology.

In section 3.4 we study the limiting behavior of a class of Hamilton Jacobi equations. The weak law of large numbers and the central limit theorem presented in this section are the time continuous version of those presented in [34] and [13].

In the last subsection we use the preceding framework to make sense out of the evolution equations arising in Mathematical Morphology. This new application of idempotent analysis provides an alternative approach to the Beucher's morphological gradient or the sup-inf derivative notions presented in [8, 38].

1 Idempotent measures and Integration

The elements of idempotent analysis (analysis of functions with values in a general idempotent semi-ring) are developed in [21, 24, 26, 27, 28, 29, 30, 31, 32]. Let $\mathbb{R}_{\max} \stackrel{\text{def}}{=} [-\infty, +\infty[$ be the semi-ring of reals numbers endowed with the commutative semigroup laws $\oplus$, $\odot$, the neutrals elements $\mathbb{0}$, $\mathbb{1}$ and the exponential metric ρ such that

$$\begin{array}{lll} \mathbb{0} = -\infty, & a \oplus b = \sup(a, b), & \rho(a, b) = |e^a - e^b|, \\ \mathbb{1} = 0, & a \odot b = a + b, & \dfrac{a}{b}\circ = a - b \quad (b \neq \mathbb{0}). \end{array}$$

By $\rho(a, b) = \sup_{1 \leq i \leq n} |e^{a_i} - e^{b_i}|$ we also denote the exponential metric on $\mathbb{R}^n_{\max}$.

Let Ω be an Hausdorff topological space equipped with a topology τ (i.e. $\emptyset, \Omega \in \tau$ and, any union of sets of τ belongs to τ and, any finite intersections of elements of τ belongs to τ). Let $(S, \cup, \cap)$ be a semi-ring of sets in Ω.

Definition 1.1 *A Maslov idempotent measure μ on $(S, \cup, \cap)$ is a mapping from Ω in $\mathbb{R}_{\max}$ such that*

1. $\mu(\emptyset) = \mathbb{0}$
2. $\mu(A \cup B) = \mu(A) \oplus \mu(B)$ *for all sets.*

It is said to be bounded whenever $\mu(\Omega) < +\infty$.

Example 1 1. *For every function $f : \Omega \to \mathbb{R}_{\max}$, the mapping*

$$\mu_f : A \in S \mapsto \mu_f(A) = \sup_{x \in A} f(x) \tag{1}$$

is a Maslov idempotent measure.

2. *When $\Omega = \mathbb{R}$, the mapping μ given by*

$$\mu_f(A) = \begin{cases} \mathbb{1} & \textit{if } \mathbb{R} - A \textit{ countable} \\ \mathbb{0} & \textit{otherwise} \end{cases}$$

is a Maslov measure but cannot be represented in the form (2).

As pointed out in [29] these measures fail to be continuous on the empty set and they have a non unique extension to the σ-field spanned by S (i.e. the smallest σ-field that contains S) but there is a unique maximal extension to $\sigma(S)$. If S is taken as the set of all closed (or open) subsets of Ω the resulting σ-field is the Borel σ-field of Ω denoted by $\mathbf{B}(\Omega)$.

Example 2 *Assume that $f : \Omega \to \mathbb{R}_{\max}$ is an upper semicontinous function with compact level sets, that is*

$$\Omega(\alpha) = \{\omega \in \Omega \ : \ f(\omega) \geq \alpha\} \quad \text{are compact subsets of } \Omega$$

Then the mapping

$$\mu : A \in \mathbf{B}(\Omega) \mapsto \mu(A) = \sup_{\omega \in A} f(\omega) \tag{2}$$

is a Maslov idempotent measure on $\mathbf{B}(\Omega)$. Let $(\Omega', \mathbf{B}(\Omega'))$ be an auxiliary Hausdorff topological space and, let $g : \Omega \to \Omega'$ be a continuous function. As we shall see in the forthcoming development the covariant image $m = \mu \circ g^{-1}$ of μ under the transformation g is the Maslov idempotent measure given by

$$m(A) = \mu(g^{-1}(A)) = \sup_{\omega' \in A} h(\omega') \qquad \forall A \in \mathbf{B}(E)$$

where $h : \Omega' \to \mathbb{R}_{\max}$ is the u.s.c. function with compact level sets given by

$$h(\omega') = \sup\{f(\omega) | \omega \in \Omega \ : \ g(\omega) = \omega'\}$$

Let us specialize to the situation described in [26, 27] and [29, 30]. Namely, a more restrictive setting is to assume that Ω is a locally compact and separable space. In this situation Maslov has presented in [29] a new idempotent measure theory. Here some basic and well known results are briefly presented. Their proof can be found for instance in [26, 27, 29, 30]. Let us introduce some additional notations:

1. By $\mathcal{C}_{\mathbb{0}}(\Omega)$ we denote be the semi-module of continuous $\mathbb{R}_{\max}$-valued functions on Ω tending to $\mathbb{0}$ at infinity and, $\mathcal{C}_{\mathrm{K}}(\Omega)$ is the sub-semi-module of $\mathcal{C}_{\mathbb{0}}(\Omega)$ such that the support of any $f \in \mathcal{C}_{\mathrm{K}}(\Omega)$ is compact (i.e. $\mathrm{Supp} f \stackrel{\text{def}}{=} \overline{\{\omega \in \Omega \ : \ f(\omega) > \mathbb{0}\}}$ is compact).

2. We denote $\mathcal{C}'_{\mathbb{0}}(\Omega)$ (respectively $\mathcal{C}'_{\mathrm{K}}(\Omega)$) the semi-module of continuous and $\mathbb{R}_{\max}$-valued $(\oplus, \odot)$-linear forms on $\mathcal{C}_{\mathbb{0}}(\Omega)$ (respectively $\mathcal{C}_{\mathrm{K}}(\Omega)$).

3. By $I\!M(\Omega)$ we denote the semi-module of upper-semi continuous functions on Ω and, by $\mathrm{Lip}(\Omega)$ the semi-module of upper bounded and Lipschitz continuous functions $f : \Omega \to \mathbb{R}_{\max}$, i.e. upper bounded functions f such that

$$\rho(f(x), f(y)) \leq C \ d_{\Omega}(x, y) \qquad \forall x, y \in \Omega$$

The more general fact about the semi-modules $\mathcal{C}'_{\mathrm{K}}(\Omega)$ and $\mathcal{C}'_{\mathbb{0}}(\Omega)$ which will be helpful to have at our disposal is the following

Theorem 1.2 (Kolokoltsov-Maslov [26])

1. *The duality relation*

$$\begin{array}{rcl} \mathcal{C}_{\mathrm{K}}(\Omega) \times I\!M(\Omega) & \longrightarrow & \mathbb{R}_{\max} \\ (\varphi, f) & \mapsto & \displaystyle\int_{\Omega}^{\oplus} \varphi(\omega) \odot f(\omega) \otimes d\omega \stackrel{\mathrm{def}}{=} \sup_{\omega\in\Omega} \varphi(\omega) + f(\omega) \end{array}$$

   determines a representation of $\mathcal{C}'_{\mathrm{K}}(\Omega)$ *as* $I\!M(\Omega)$*.*

2. *To functionals* $m \in \mathcal{C}'_{\mathbb{0}}(\Omega)$ *there correspond bounded functions* f *such that*

$$m(\varphi) = \int_{\Omega}^{\oplus} \varphi(\omega) \odot f(\omega) \otimes d\omega \stackrel{\mathrm{def}}{=} m_f(\varphi)$$

   Moreover for any $m_{f_1}, m_{f_2} \in\in \mathcal{C}'_{\mathbb{0}}(\Omega)$ *we have*

$$m_{f_1} = m_{f_2} \iff f_1^\star = f_2^\star$$

   where $f^\star$ *is the upper semi-continuous envelope of* f*, i.e.*

$$f^\star(x) \stackrel{\mathrm{def}}{=} \sup\{\varphi(x) \mid \varphi \in \mathcal{C}_{\mathbb{0}}(\Omega) \ : \ f \leq \varphi\}$$

3. *For any* $m_1, m_2 \in\in \mathcal{C}'_{\mathbb{0}}(\Omega)$ *we have*

$$\forall \varphi \in Lip(\Omega) \qquad m_1(\varphi) = m_2(\varphi) \Longrightarrow m_1 = m_2$$

Many constructions and results can be borrowed to idempotent analysis from usual analysis and functional analysis, the interested reader is recommended to consult [28] and references therein.

2 Maslov Optimization Theory

A first study of the parallelism between Probability and Stochastic Processes on the one hand and Performance measures and Optimization Processes on the other hand can be founded in [13, 17, 14, 15] and [34].
But to present this material in an absolutely rigorous way some additional topological structure has to be added. These notions are strongly related to Large Deviations Calculus.
As far as applications are concerned, this branch of idempotent analysis has been motivated by control theoretic applications and non linear estimation problems.
The aim of this section is to present a brief survey of this duality and to introduce the basic concepts and the mathematical tools upon which the theory dwells.

2.1 Performance Spaces and Optimization Variables

As in probability theory, the basic notion is the (max,+)-version of the definition of a probability space.

Definition 2.1 *A performance space or an optimization basis is a triplet* $(\Omega, \mathbf{B}(\Omega), \mathbb{P})$ *were*

1. $(\Omega, \mathbf{B}(\Omega))$ *is an Hausdorff topological space equipped with the Borel* σ*-field* $\mathbf{B}(\Omega)$*.*
2. $\mathbb{P}$ *is a Maslov idempotent measure on* $(\Omega, \mathbf{B}(\Omega))$ *such that*

$$\forall A \in \mathbf{B}(\Omega) \qquad \mathbb{P}(A) = \sup_{\omega \in A} \mathbb{p}(\omega)$$

where $\mathbb{p} : \Omega \to \mathbb{R}_{\max}$ *is an upper semicontinous function such that for any* $\alpha \in \mathbb{R}$ *the level sets*

$$\Omega(\alpha, \mathbb{P}) = \{\omega \in \Omega \ : \ \mathbb{p}(\omega) \geq \alpha\}$$

are compact subsets of Ω
The measure $\mathbb{P}$ *is called a performance or cost measure on* $(\Omega, \mathbf{B}(\Omega))$ *and the upper semi-continuous function* $\mathbb{p}$ *is called the performance density of the measure* $\mathbb{P}$*.*

Remark 2.2:
The density $\mathbb{p}$ of a performance measure $\mathbb{P}$ on $(\Omega, \mathbf{B}(\Omega))$ is similar to the definition of a good rate function in classical large deviation calculus. One nice consequence of the present performance concept is that supremum over any closed subset of Ω are always achieved.
Let us recall some elementary properties of such functions.

Lemma 2.3 *Let* $(\Omega, \mathbf{B}(E), \mathbb{P})$ *be a performance space.*

1. *For any closed subset* $A \subset \Omega$ *there exists some* $\omega^\star \in A$ *such that*

$$\mathbb{p}(\omega^\star) = \sup_{\omega \in A} \mathbb{p}(\omega)$$

2. *If* $\{F_\epsilon \ : \ \epsilon > 0\}$ *is a nested family of closed sets (i.e.* $F_{\epsilon_1} \subset F_{\epsilon_2}$ *for any* $\epsilon_1 < \epsilon_2$ *) then*

$$\mathbb{P}(\cap_{\epsilon > 0} F_\epsilon) = \lim_{\epsilon \to 0} \mathbb{P}(F_\epsilon)$$

The proof of this lemma is standard, the interested reader is referred to lemma 2.1.2 Deuschel-Stroock [20] or lemma 4.1.6 in Dembo-Zeitouni [19].
The second basic notion is the (max,+)-version of an random variable.

Definition 2.4 *Let* $(\Omega, \mathbf{B}(\Omega), \mathbb{P})$ *be a performance space and let* $(E, \mathbf{B}(E))$ *be a Hausdorff topological space. A measurable mapping* $X : \Omega \to E$ *is an* E*-valued optimization variable provided its restrictions to the compact level sets* $\Omega(\alpha, \mathbb{P})$*,* $\alpha \in \mathbb{R}_-$*, are continuous.*

Let us give some comments on the above notions.
An optimization problem is described by an Hausdorff topological space $(\Omega, \mathbf{B}(\Omega))$ where Ω is the set of all possible control actions or policies and $\mathbf{B}(\Omega)$ its natural Borel σ-field. $X(\omega) \in E$ is the output of a given controlled system $X : \Omega \to E$ under the control $\omega \in \Omega$. The performance measure $\mathbb{P}$ given on $(\Omega, \mathbf{B}(\Omega))$ describes the quality of each control action. As in Probability Theory the minimal assumption on X is to be measurable. In this situation we can define the covariant image $\mathbb{P}^X = \mathbb{P} \circ X^{-1}$ of $\mathbb{P}$ under the transformation X. Namely,

$$\mathbb{P}^X(A) = \mathbb{P}(X^{-1}(A)) \qquad \forall A \in \mathbf{B}(E)$$

Note that

$$\mathbb{P}^X(A) = \sup_{x \in A} \mathbb{p}^X(x) \qquad \text{with} \quad \mathbb{p}^X(x) \stackrel{\text{def}}{=} \sup \{\mathbb{p}(\omega) \mid \omega \ : \ X(\omega) = x\}$$

Unfortunately the function $\mathbb{p}^X : E \to \mathbb{R}_{\max}$ is not necessarily u.s.c. and the compact level set condition is not always true. The first claim follows from the fact that a pointwise supremum of an arbitrary collection of u.s.c. functions is not necessarily u.s.c.
The continuity assumption on the optimization variable definition is a natural condition on X which ensures that $\mathbb{p}^X$ is a performance.

Our claim that this framework is a natural framework for formulating optimization problems will be amply justified by the results which will follows.

Proposition 2.5 *Let $(\Omega, \mathbf{B}(\Omega), \mathbb{P})$ be a performance space and let $(E, \mathbf{B}(E))$ be an Hausdorff topological space. Let $X : \Omega \to E$ be an optimization variable. Then*

1. *$(E, \mathbf{B}(E), \mathbb{P}^X)$ is a performance space and the density of $\mathbb{P}^X$ is the u.s.c. function $\mathbb{p}^X : E \to \mathbb{R}_{\max}$ given by*

$$\mathbb{p}^X(x) = \sup \{\mathbb{p}(\omega) \mid \omega \in \Omega \ : \ X(\omega) = x\} \qquad \forall x \in E \tag{3}$$

2. *Assume that E is a metric space and X is a continuous optimization variable and the function $\mathbb{p}^X$ is bounded. Then there exists a non increasing sequence of Lipschitz continuous functions $\mathbb{p}_n^X : E \to \mathbb{R}_{\max}$ such that*

$$\mathbb{p}^X(x) = \inf_{n \geq 1} \mathbb{p}_n^X(x)$$

Proof:
1) When the optimization variable is continuous this result is nothing else than the classical contraction principle of Large Deviation Theory. The extension to general optimization

variables taking values in a metric space can be found in the classical texts [19] theorem 4.2.23 pp. 117 or in [20] lemma 2.1.4 pp. 37. The general result is contained in [3].
2) We note that if $\mathbb{p}^X$ is bounded define $a = \inf_E \mathbb{p}^X \in \mathbb{R}_-$. Then for all$x \in E$

$$\mathbb{p}^X(x) = \sup\{\mathbb{p}(\omega) \mid \omega \ : \ X(\omega) = x \quad \text{and} \quad \mathbb{p}(\omega) \geq a\} \tag{4}$$

Now, let $\mathbb{p}^X_n$, $n \geq 1$, be the sequence of functions defined by

$$\mathbb{p}^X_n(x) = \sup\{\mathbb{p}(\omega) - n d_E(x, X(\omega)) \mid \omega \ : \ \mathbb{p}(\omega) \geq a\}$$

Using the standard inequality $\sup\{(1)\} - \sup\{(2)\} \leq \sup\{(1) - (2)\}$ one gets easily

$$\mathbb{p}^X_n(x) - \mathbb{p}^X_n(y) \leq n \sup\{d_E(y, X(\omega)) - d_E(y, X(\omega)) \mid \omega \ : \ \mathbb{p}(\omega) \geq a\}$$

for any $x, y \in E$. Using the triangle inequality one can check that

$$|\mathbb{p}^X_n(x) - \mathbb{p}^X_n(y)| \leq n \ d_E(x, y)$$

To complete the proof of 2) it suffices to show that

$$\mathbb{p}^X(x) = \inf_{n \geq 1} \mathbb{p}^X_n(x) \tag{5}$$

First for each $x \in E$ we note that

$$\inf_n \{-n d_E(x, X(\omega))\} = \delta_x \circ X(\omega)$$

where $\delta_\alpha(\mathbf{B}) = 0$ if $\alpha = \mathbf{B}$ and $-\infty$ otherwise. Therefore

$$\mathbb{p}^X(x) \quad = \sup_{\omega \in \Omega(\mathbb{P}, a)} (\mathbb{p}(\omega) + \delta_x(X(\omega))$$

This yields the inequality

$$\mathbb{p}^X \leq \inf_n \mathbb{p}^X_n \tag{6}$$

Finally, since the subsets $X^{-1}\{x\} \cap \Omega(a, \mathbb{P})$, $x \in E$, are compacts there exists a sequence $(\omega_n)_{n \geq 1}$, $\omega_n \in \Omega(a, \mathbb{P})$, such that

$$\mathbb{p}^X_n(x) = \mathbb{p}(\omega_n) - n \ d_E(x, X(\omega_n)) \tag{7}$$

As usually there exists a subsequence $(\omega_{n_k})_{k \geq 1}$, $\omega_{n_k} \in \Omega(a, \mathbb{P})$ such that

$$\lim_{k \to \infty} \omega_{n_k} \stackrel{\text{def}}{=} \omega \in \Omega(a, \mathbb{P})$$

Assume that $X(\omega) \neq x$. In this situation (7) implies that for any $b \in \mathbb{R}_-$ there exists some $k \geq 1$ such that

$$\mathbb{p}^X_{n_k}(x) \leq b$$

A contradiction with $\mathbb{p}^X_{n_k}(x) \geq a$ for all $k \geq 1$ and $x \in E$. This implies that $X(\omega) = x$ and in view of (7) it follows that

$$\inf_n \mathbb{p}^X_n(x) \leq \inf_k \mathbb{p}^X_{n_k}(x) \leq \lim_{k\to\infty} \mathbb{p}(\omega_{n_k}) \leq \mathbb{p}(\omega) \tag{8}$$

Combining (6) and (8) one gets the desired equality (5). ■

Remark 2.6:
It is important to note that the form of the density (3) is really justified only if the mapping satisfies the nearly continuity assumption of definition 2.4. In this situation the performance space $(E, \mathbf{B}(E), \mathbb{P}^X)$ is constructed in a completely rigorous way.

By analogy with the mean-value operator in Probability Theory we define the (max,+)-expectation of the optimization variable.

Definition 2.7 *Let X be an $\mathbb{R}$-valued optimization variable defined on a optimization basis $(\Omega, \mathbf{B}(\Omega), \mathbb{P})$. We denote by $\mathbb{E}(X)$ the (max,+)-expectation or mean valued of the optimization variable X the number*

$$\mathbb{E}(X) = \int_\Omega^\oplus X(\omega) \odot \mathbb{p}(\omega) \odot d\omega = \int_\mathbb{R}^\oplus x \odot \mathbb{p}^X(x) \odot dx$$

We list some basic properties of the (max,+)-expectation. In what follows (X, Y) is a pair optimization variable taking values in $\mathbb{R}^2$ and $a, b \in \mathbb{R}_{\max}$

$$\begin{aligned} \mathbb{E}(a \odot X \oplus b \oplus Y) &= a \odot \mathbb{E}(X) \oplus b \oplus \mathbb{E}(Y) \\ X \geq Y &\Rightarrow \mathbb{E}(X) \geq \mathbb{E}(Y) \\ Y = \mathbb{1}_A(X) &\Rightarrow \mathbb{E}(Y) = \mathbb{P}^X(A) = \int_A^\oplus \mathbb{p}^X(x) \odot dx \end{aligned}$$

where $\mathbb{1}_A(x) = \mathbb{1}(= 0)$ if $x \in A$ and $\mathbb{0}(= -\infty)$ otherwise.
When X is an optimization variable taking values in a locally compact and separable metric space E and for any test function $\varphi \in \mathcal{C}_K(E)$ the expectation of $\varphi(X)$ can be rewritten using the duality $\mathcal{C}'_K(E) = \mathbf{M}(E)$

$$\mathbb{E}(\varphi(X)) = \int_E^\oplus \varphi(x) \odot \mathbb{p}^X(x) \odot dx = < \varphi, \mathbb{P}^X >_{\mathcal{C}_K(E), \mathbf{M}(E)}$$

At this point it is already useful to give some examples of optimization variables and performance spaces which often appear in practical situations.

Example 3 *1.* $\Omega = \mathbb{R}^d$, $d \geq 1$, $\mathbb{p}(\omega) = -\frac{1}{p}\|\omega\|_p^p = -\frac{1}{p}\sum_{i=1}^d |w^i|^p$.

(a) The identity mapping $X : \Omega \to \mathbb{R}^d$ *given by* $X(\omega) = \omega$ *is an optimization variable with performance density*

$$\mathbb{p}^X(x) = -\frac{1}{p}\|x\|_p^p \qquad \forall x \in \mathbb{R}^d$$

(b) The projection on the first argument $X : \Omega \to \mathbb{R}^d$ *given by* $X(\omega) = \omega_1$ *is an optimization variable taking values in* $\mathbb{R}$ *with performance density*

$$\mathbb{p}^X(x) = -\frac{1}{p}|x|_p^p \qquad \forall x \in \mathbb{R}$$

2. $\Omega = \mathbb{R}^{d_1+d_2}$, $d_1, d_2 \geq 1$,

$$\mathbb{p}(\omega) = -\frac{1}{p_1}\|\pi_1(\omega)\|_{p_1}^{p_1} - \frac{1}{p_2}\|\pi_2(\omega) - h(\pi_1(\omega))\|_{p_2}^{p_2}$$

where $h : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$ *is a continuous function and*

$$\pi_1(\omega) \stackrel{\text{def}}{=} (\omega_1, \ldots, \omega_{p_1}) \qquad \textit{and} \qquad \pi_2(\omega) \stackrel{\text{def}}{=} (\omega_{p_1+1}, \ldots, \omega_{p_2})$$

In this situation the pair variable (X, Y) *given by*

$$\begin{array}{cccl} (X,Y) & \Omega & \to & \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \\ & \omega & \mapsto & (X(\omega), Y(\omega)) = (\pi_1(\omega), \pi_2(\omega)) \end{array}$$

is an optimization variable with density

$$\mathbb{p}^{(X,Y)}(x, y) = -\frac{1}{p_1}\|x\|_{p_1}^{p_1} - \frac{1}{p_2}\|y - h(x)\|_{p_2}^{p_2}$$

In addition one can easily check that the single mapping $X : \Omega \to \mathbb{R}^{d_1}$ *is an optimization variable with density*

$$\mathbb{p}^X(x) = -\frac{1}{p_1}\|x\|_{p_1}^{p_1}$$

3. Let $\Omega = H$ *be a separable Hilbert space with norm* $\|.\|$. *Let* Q *be a positive trace class operator in* Ω. *We may define on* $(H, \mathbf{B}(H))$ *a performance measure by letting*

$$\mathbb{p}(\omega) = \begin{cases} -\frac{1}{2}\|Q^{-\frac{1}{2}}\omega\|^2 & \textit{if} \quad x \in ImQ^{1/2} \\ -\infty & \textit{otherwise} \end{cases}$$

The interested reader is referred to Jabzick and Da prato [?] of other examples of performance measures on separable Hilbert spaces.

Another important example of optimization variable is the (max,+)-version of the classical Gaussian random variables in Probability theory. Since we will now be dealing with optimization variables defined on Banach space let us recall some basic properties. Let E be a separable Banach space. The Borel σ-field $\mathbf{B}(E)$ coincides with the smallest σ-field of subsets of E containing all subsets of the form

$$\{s \in E \ : \ f(x) \leq \alpha\} \qquad \alpha \in \mathbb{R} \qquad f \in E^\star$$

Let E be is a separable Hilbert space with inner product $< .,. >$, norm $\|\centerdot\|$ and equipped with the weak topology. In this situation the function

$$x \in E \to \|x\| = \sup_{h \in E} < h, x > \in \mathbb{R}_+$$

is lower semi continuous as a limit of the continuous functions $x \in E \to < h, x > \in \mathbb{R}$ and all subsets $\{x \in E \ : \ \|x\| \leq \alpha\}$, $\alpha \in \mathbb{R}$, are compact for the weak topology.

Definition 2.8
Let us assume that H is a separable Hilbert space with inner product $< .,. >$, norm $\|\centerdot\|$ and equipped with the weak topology. Let X be an optimization variable on a performance space $(\Omega, \mathbf{B}(\Omega), \mathbb{P})$ and taking values into H. It is called a quadratic optimization variable when the induced performance density p^X is given by

$$\mathrm{p}^X(x) = -\frac{1}{2}\|x\|^2$$

It is noteworthy that a linear transformation of a quadratic optimization variable is still quadratic.

Proposition 2.9 *Let X be a quadratic optimization variable taking values in a separable Hilbert space H with norm $\|\centerdot\|$ and equipped with the weak topology. Let E be an auxiliary separable Hilbert space and, let $T : H \to E$ be a bounded linear mapping such that $T(H) = E$.*
1) The mapping $Y = TX$ is a quadratic optimization variable and

$$\mathrm{p}^Y(y) = -\frac{1}{2}\|T^{-1}y\|^2$$

where T^{-1} is the pseudo-inverse of the transformation T.
2) Let $T^\star : E \to H$ be the adjoint of T. If $TT^\star$ is invertible then

$$\mathrm{p}^Y(y) = -\frac{1}{2}\|T^\star(TT^\star)^{-1}y\|^2$$

3) Assume that $E = \mathbb{R}^d$ with $d \geq 1$ and $T : H \to \mathbb{R}^d$ is given by $Tx = (T_1 x, \ldots, T_d x)$ for any $x \in H$ where $T_n : H \to \mathbb{R}$ are bounded linear transformations satisfying

$$T_m T_n^\star = \delta_m(n)\ Id \qquad \forall m, n \in \{1, \ldots, d\} \tag{9}$$

Then we have the formulae

$$\mathbb{p}^Y(y) = -\frac{1}{2} \sum_{m=1}^{d} \|T^\star y_m\|^2 \qquad \forall y = (y_1, \ldots, y_d) \in \mathbb{R}^d$$

Proof:
By the definition of the pseudo-inverse T^{-1}, if $y \in \mathrm{Im}T$ then $T^{-1}y$ is the element of $T^{-1}\{y\}$ of the minimal norm. This implies the first assertion.
Let us prove 2). Using 1) the second assertion is clear.
To prove the last part of the proposition it suffices to note that (9) implies that $TT^\star = Id$ and for any $y \in \mathrm{Im}T$

$$\begin{aligned} \|T^\star y\|^2 &= < T^\star y, T^\star y > \\ &= < \sum_{m=1}^{d} T_m^\star y_m, \sum_{m=1}^{d} T_m^\star y_m > = \sum_{m=1}^{d} < T_m^\star y_m, T_m^\star y_m > = \sum_{m=1}^{d} \|T_m^\star y_m\|^2 \end{aligned}$$

■

Next example initiate the study of optimization processes. Although these objects will be introduced in the end of this section this example shows that the above proposition gives a natural and simple way to calculate the densities of integral functions of a quadratic optimization variables.

Example 4 *Let U be a quadratic optimization variable with values in $H = L^2([0,1], \mathbb{R})$ (equipped with the weak topology) and with performance density*

$$\mathbb{p}^U(u) = -\frac{1}{2} \int_0^1 u_t^2\ dt$$

Denote by $X_t : H \to \mathbb{R}$, $t \in (0,1)$, the linear operator given by

$$X_t(u) = \int_0^t u_s\ ds$$

It is easy to see that $Im X_t = \mathbb{R}$ so that the adjoint operator $X_t^\star : \mathbb{R} \to H$ is defined by

$$X_t^\star(x)(s) = x\ 1_{[0,t]}(s) \qquad \forall s \in [0,1]$$

It is clear that $X_t X_t^\star = t\ Id$ *is invertible for any* $t \in (0,1)$ *and therefore, by proposition (2.9), the mapping* X_t *is an quadratic optimization variable and its performance is given by*

$$\mathbb{p}^X(x) = -\frac{1}{2}\int_0^t \left(\frac{x}{t}\right)^2 dt = -\frac{x^2}{2t} \qquad \forall t \in (0,1)$$

Let $t_1 < \ldots < t_n$, $n \geq 1$, *be a collection of times in* $(0,1)$ *and let* $\sigma = (t_1, \ldots, t_n)$ *the corresponding* n*-uple of times. We note* $X_\sigma : H \to \mathbb{R}^n$ *the linear operator given by*

$$X_\sigma(u) = (X_{t_1}(u), \ldots, X_{t_n}(u)) \qquad \forall u \in H$$

In view of proposition 2.9, X_σ *is a quadratic optimization variable taking values in* $\mathbb{R}^n$ *and for any* $x = (x_1, \ldots, x_n) \in Im X_\sigma$

$$\mathbb{p}^{X_\sigma}(x) = \sup\left\{\mathbb{p}^U(u) | u \in H \ :\ X_\sigma(u) = x\right\}$$

Note that $\mathbb{p}^{X_\sigma}$ *may also be written*

$$\mathbb{p}^{X_\sigma}(x) = \sup\left\{\mathbb{p}^U(u) | u \in H \ :\ R_\sigma(u) = \Delta_\sigma(x)\right\}$$

with

$$R_\sigma = (R_1, \ldots, R_n) \qquad R_i = \frac{X_i - X_{i-1}}{t_i - t_{i-1}} \qquad \forall 1 \leq i \leq n$$

with the convention $X_0 = t_0 = 0$ *and*

$$\Delta_\sigma(x) = \left(\frac{x_1}{t_1}, \frac{x_2 - x_1}{t_2 - t_1}, \ldots, \frac{x_n - x_{n-1}}{t_n - t_{n-1}}\right)$$

Note that the adjoint operators $R_i^\star$, $1 \leq i \leq n$, *are given by*

$$R_i^\star(z)(s) = \frac{z}{t_i - t_{i-1}}\ 1_{[t_{i-1}, t_i]}(s) \qquad \forall 1 \leq i \leq n$$

Clearly condition (9) holds and proposition 2.9 implies that for any $\sigma = (t_0, \ldots, t_n)$ *the optimization variable* X_σ *is quadratic and its performance is given by*

$$\mathbb{p}^{X_\sigma}(x) = -\frac{1}{2}\sum_{i=1}^n \frac{(x_i - x_{i-1})^2}{t_i - t_{i-1}}$$

with the convention $x_0 = t_0 = 0$.

2.2 Independence and conditioning

Our next aim is to transfer probabilistic axioms to optimization theory. The independence concept in such a framework is given by

Definition 2.10 *Let $(\Omega, \mathbf{B}(\Omega), \mathbb{P})$ be a performance space. Two subsets $A, B \subset \mathbf{B}(\Omega)$ are said to be independent if*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \odot \mathbb{P}(B)$$

Let $(E, \mathbf{B}(E))$ and $(F, \mathbf{B}(F))$ be two Hausdorff topological spaces and, let (X, Y) be a pair optimization variable taking values in $E \times F$. We assume that (X, Y) is defined on an optimization basis $(\Omega, \mathbf{B}(\Omega), \mathbb{P})$ and we denote by $\mathbb{p}^{X,Y}$ its performance density. We say that they are $\mathbb{P}$-independent if

$$\mathbb{p}^{X,Y}(x, y) = \mathbb{p}^{X}(x) \odot \mathbb{p}^{Y}(y)$$

for any $x, y \in E$.

At the opposite of the latter concept of independence, we introduce the extension of Bayes formula to performance measure. The conditional performance of an event A, assuming an event B, and denoted by $\mathbb{P}(A|B)$ is by definition the ratio

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}\circ = \mathbb{P}(A \cap B) - \mathbb{P}(B)$$

with the convention $\mathbb{P}(A|B) = \mathbb{0}$ when $\mathbb{P}(B) = \mathbb{0}$. Continuing in the same vein, let $(E, \mathbf{B}(E))$ be a Hausdorff topological space and, let (X, Y) be a pair of E-valued optimization variables with density $\mathbb{p}^{(X,Y)}$. The function $\mathbb{p}^{X|Y}$ defined by

$$\mathbb{p}^{X|Y}(x|y) = \frac{\mathbb{p}^{(X,Y)}(x, y)}{\mathbb{p}^{Y}(y)}\circ = \mathbb{p}^{X,Y}(x, y) - \mathbb{p}^{Y}(y)$$

with the convention $\mathbb{p}^{X|Y}(x|y) = \mathbb{0}$ if $\mathbb{p}^{Y}(y) = \mathbb{0}$ is called the conditional performance density of X with respect to Y.
One can easily check that for any $y \in E$ such that $\mathbb{p}^{Y}(y) \neq \mathbb{0}$, the function $x \mapsto \mathbb{p}^{X|Y}(x|y)$ is a well defined performance density.
Let us give some comments on this definition. Using the framework of the previous development we discuss now the meaning of the conditional density $\mathbb{p}^{X|Y}$. Recall that the densities $\mathbb{p}^{(X,Y)}$ and $\mathbb{p}^{Y}$ may be written

$$\begin{aligned} \mathbb{p}^{(X,Y)}(x, y) &= \sup\{\omega \in \Omega \mid \omega \ : \ X(\omega) = x \ \text{ and } \ Y(\omega) = y\} \\ \mathbb{p}^{Y}(y) &= \sup\{\omega \in \Omega \mid \omega \ : \ Y(\omega) = y\} \end{aligned}$$

On the other hand

$$\mathbb{p}^Y(y) = \sup_{x\in E} \mathbb{p}^{(X,Y)}(x,y) \tag{10}$$

To see this claim we note that

$$\sup_{x\in E} \mathbb{p}^{(X,Y)}(x,y) = \sup_{x\in E} \sup_{\omega\in Y^{-1}(y)\cap X^{-1}(x)} \mathbb{p}(\omega) = \sup_{\omega\in Y^{-1}(y)} \mathbb{p}(\omega) = \mathbb{p}^Y(y) \tag{11}$$

This indicates that the conditional performance $\mathbb{p}^{X|Y}(\,\boldsymbol{.}\,|y)$ is indeed a normalized idempotent measure on the subsets $A\cap Y^{-1}(y)$, $A\in \mathbf{B}(\Omega)$.
Clearly we have also the following properties.

Proposition 2.11 *Let (X,Y,Z) be a triplet of optimization variable with performance density $\mathbb{p}^{(X,Y,Z)}$. The following properties hold.*

1. *For any z with $\mathbb{p}^Z(z)\neq \mathbb{0}$ we have*

$$\mathbb{p}^{X|Z}(x|z) = \int_E^{\oplus} \mathbb{p}^{X|Y,Z}(x|y,z)\odot \mathbb{p}^{Y|Z}(y|z)\odot dz$$

2. *For any x*

$$\mathbb{p}^X(x) = \int_E^{\oplus} \mathbb{p}^{X|Y}(x|y)\odot \mathbb{p}^Y(y)\odot dy$$

3. *For any (y,z) such that $\mathbb{p}^{Y,X}(y,z)\neq \mathbb{0}$*

$$\mathbb{p}^{X|Y,X}(x|y,z) = \delta_z(x)\odot \mathbb{p}^{X|Y}(x|y)$$

where $\delta_z(x)=0$ if $z=x$ and $-\infty$ otherwise.

2.3 Optimization processes

After the introduction of the basic axioms of performance measure analysis the next step consists in introducing optimization processes. As it will be obvious in the forthcoming development the objects on which the optimization processes are sought may vary considerably. It is therefore necessary to undertake the study in an abstract setting.
To achieve this program we need to introduce some additional notations. The basic definition of an optimization process with time space $\mathcal{T}$ ($\mathcal{T}=(s,t)\subset\mathbb{R}$ or $\mathbb{Z}$ with $s\leq t$) will consist of a sequence of optimization variables $\{X_t\ ;\ t\in\mathcal{T}\}$ taking values in a Polish space $(S,\mathbf{B}(S))$.

Definition 2.12 *An S-valued optimization (or decision) process with time space T is a system*

$$(\Omega, \mathbf{B}(\Omega), \mathbb{P}, X = (X_t)_{t\in T})$$

defined by : 1) A performance space $(\Omega, \mathbf{B}(\Omega), \mathbb{P})$ and 2) An optimization variable $X : \Omega \to \mathcal{E}(\mathcal{T}, S)$, where $\mathcal{E}(\mathcal{T}, S)$ is a metric space of mappings $x : \mathcal{T} \to S$ furnished with a topology such that the natural restriction mappings $\pi_{[s,t]} : \mathcal{E}(\mathcal{T}, S) \to \mathcal{E}([s,t], S)$, $s \le t$, and $\pi_{[s,t],[s',t']} : \mathcal{E}([s,t], S) \to \mathcal{E}([s',t'], S)$, $s \le s' \le t' \le t$ are continuous.

The optimization variable X_t is called the state of X at the time $t \in T$ and the curve $t \in \mathcal{T} \mapsto X_t(\omega) \in S$ is called the trajectory of the path $\omega \in \Omega$

<u>*Remark*</u> 2.13:
Before presenting some examples of optimization processes we remark that the above definition may cover continuous or discontinuous optimization processes. More precisely, let $\mathcal{T} = \mathbb{R}$ and $\mathcal{E}(\mathcal{T}, S) = \mathcal{D}(\mathcal{T}, S)$ be the space of right continuous paths which have a left limit. There exists a metric d on this space such that the natural restriction mappings are continuous. Moreover the relative topology which $\mathcal{D}(\mathcal{T}, S)$ inherits as a subset of the latter coincide with the uniform convergence on compact sets (see for instance lemma 5.4.29 in [20]).

When X is an optimization process on T the function $X_t : \Omega \to S$ is given by $X_t = \pi_t \circ X$ where π_t is the projection on the time $t \in T$. According to the definition of an optimization variable and π_t being continuous the mapping X_t satisfies the continuity assumption of definition 2.4. This implies that X_t is an optimization variable with performance density

$$\mathbb{p}_t^X(x) = \sup\{\mathbb{p}(\omega) \mid \omega \ : \ X_t(\omega) = x\}$$

More generally the composite mappings

$$X_{[s,t]} = \pi_{[s,t]} \circ X : \Omega \to \mathcal{E}([s,t], S)$$

can be regarded as optimization variables with performance density

$$\mathbb{p}_{[s,t]}^X(x) = \sup\{\mathbb{p}^X(z) \mid z \in \mathcal{E}(\mathcal{T}, S) \ : \ \pi_{[s,t]}(z) = x\}$$

Summarizing we have that

Proposition 2.14 *Let $(\Omega, \mathbf{B}(\Omega), \mathbb{P}, X = (X_t)_{t\in T})$ be an optimization process with time space T and taking values in a metric space (S, d_{S}). The following properties are satisfied*

1. *For any $s \le t$, $s, t \in \mathcal{T}$ the composite mapping $X_{[s,t]} = \pi_{[s,t]} \circ X : \Omega \to \mathcal{E}([s,t], S)$ is an optimization variable and its performance density is the upper continuous function $\mathbb{p}_{[s,t]}^X : \mathcal{E}([s,t], S) \to \mathbb{R}_{\max}$ (with compact level sets) given by*

$$\mathbb{p}_{[s,t]}^X(x) = \sup\{\mathbb{p}(\omega) \mid \omega \in \Omega \ : \ X(\omega) = x\} \qquad \forall x \in \mathcal{E}([s,t], S)$$

2. For any $s \leq s' \leq t' \leq t$ the performance densities $\mathbb{p}^X_{[s,t]}$ and $\mathbb{p}^X_{[s',t']}$ satisfy

$$\mathbb{p}^X_{[s',t']}(x_2) = \sup\{\mathbb{p}^X_{[s,t]}(x_1) \mid x_1 \in \mathcal{E}([s,t],S) \quad : \quad \pi_{[s,t],[s',t']}(x_1) = x_2\}$$

for any $x_2 \in \mathcal{E}([s',t'],S)$.

We now present a number of examples which can be handled in our framework.

Example 5 *Let us assume that $\mathcal{T} = [0,1]$, $S = \mathbb{R}^d$, $d \geq 1$, and*

$$\Omega = C_0([0,1],\mathbb{R}^d) = \{\omega \in \mathcal{C}([0,1],\mathbb{R}^d) \ : \ \omega(0) = 0\} \qquad \mathcal{E}(\mathcal{T},S) = \mathcal{C}([0,1],\mathbb{R}^d)$$

equipped with the uniform norm. For any $\omega \in \Omega$ we define $X : \Omega \to \mathcal{C}([0,1],\mathbb{R}^d)$ by

$$X_t(\omega) = \omega_t + \int_0^t f(X_s(\omega))\, ds \qquad \forall t \in T$$

where $f : \mathbb{R}^d \to \mathbb{R}^d$ satisfies the Lipschitz condition

$$\|f(x) - f(y)\| \leq C\, \|x - y\| \qquad C < \infty$$

One can check that X is a continuous mapping. Now, let $\mathbb{p} : \Omega \to \mathbb{R}_{\max}$ be defined by

$$\mathbb{p}(\omega) = \begin{cases} -\frac{1}{2}\int_0^1 \| \dot{\omega}_t^2 \|\, dt & \text{if} \quad \omega \in H^1([0,1],\mathbb{R}^d) \\ -\infty & \text{otherwise} \end{cases}$$

The function $\mathbb{p}$ is u.s.c. on $C_0([0,1],\mathbb{R}^d)$. On the other hand using the fact that bounded subsets of $H^1([0,1],\mathbb{R}^d$ are relatively compact in Ω . In addition the subsets

$$\{\omega \in \Omega \ : \ \mathbb{p}(\omega) \geq \alpha\} = \{\omega \in H^1([0,1],\mathbb{R}^d) \ : \ \|\omega\|_{H^1} \leq \sqrt{2|\alpha|}\} \qquad \alpha \in \mathbb{R}_-$$

are compacts. Hence, $\mathbb{p}$ is a well defined performance density on $(\Omega, \mathbf{B}(\Omega))$ (cf pp 12,13 in [20]).
Using the above observations one can check that X is a well defined optimization variable on $(\Omega, \mathbf{B}(\Omega))$ with performance density

$$\mathbb{p}^X(x) = -\frac{1}{2}\int_0^1 \| \dot{x}_t - f(x_t)\|^2\, dt$$

if the integral exists and $\mathbb{0} = -\infty$ otherwise.

Example 6 *Let us assume that* $\mathcal{T} = [0,1]$, $S = \mathbb{R}^d$, $d \geq 1$, $p \geq 1$ *and*

$$\Omega = L^p([0,1], \mathbb{R}^d) \qquad \mathcal{E}([0,1], \mathbb{R}^d) = \mathcal{C}([0,1], \mathbb{R}^d)$$

equipped respectively with the weak convergence topology and the uniform norm. For any $\omega \in \Omega$ *define* $X : \Omega \to \mathcal{C}([0,1], \mathbb{R}^d)$ *by setting*

$$\begin{cases} \dot{X}(\omega)_t &= A(X(\omega)_t) + B\,\omega_t \\ X(\omega)_0 &= x_0 \in \mathbb{R}^d \end{cases}$$

where $B \in \mathbb{R}^{d \times d}$ *and* $A : \mathbb{R}^d \to \mathbb{R}^d$ *is a bounded and uniformly Lipschitz mapping. Under our assumptions one can check that* X *is continuous.*
Now, let $\mathbb{p} : \Omega \to \mathbb{R}_{\max}$ *be defined as*

$$\mathbb{p}(\omega) = -\frac{1}{p}\, \|\omega\|_p^p$$

It is clear that $(\Omega, \mathbf{B}(\Omega), \mathbb{P})$ *is an optimization basis. Using proposition 2.5 we conclude that* X *is an optimization process with performance density*

$$\mathbb{p}^X(x) = \sup\{-\frac{1}{p}\, \|\omega\|_p^p \; ; \; \omega \in L^p([0,1], \mathbb{R}^d) \; : \; X(\omega) = x\} \qquad \forall x \in \mathcal{C}([0,1], \mathbb{R}^d)$$

Hence

$$\mathbb{p}^X(x) = -\frac{1}{p} \int_0^1 \; | \, \dot{x}_t - A(x_t)|^p \; dt$$

when the integral is defined and $\mathbb{0} = -\infty$ *otherwise.*

Using the same arguments, if the performance density $\mathbb{p} : \Omega \to \mathbb{R}_{\max}$ *is given by*

$$\mathbb{p}(\omega) = \int_0^1 \; L_t(X(\omega)_t, \omega_t) \; dt$$

for some $L_t : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_{\max}$ *then* X *is an optimization process with performance density*

$$\mathbb{p}^X(x) = \sup\{\int_0^1 \; L_t(X(\omega)_t, \omega_t) \; dt \; ; \; \omega \in \Omega \; : \; X(\omega) = x\} \qquad \forall x \in \mathcal{C}([0,1], \mathbb{R}^d)$$

Example 7 *As an example of discrete time oprimization process let us consider the case* $\mathcal{T} = \{0, \ldots, T\}$, $T \geq 1$ *and* $\Omega = U^T$ *where* (U, d_U) *is a metric space. For each* $1 \leq t \leq T$ *we denote by* $\mathbb{p}_t : U \to \mathbb{R}_{\max}$ *is a performance measure on* $(U, \mathbf{B}(U))$*. Assume that* U^T

is equipped with the metric $d(\omega, \omega') = \sup_{1 \le t \le T} d_U(\omega_t, \omega'_t)$ *and the performance densities* $\{\mathbb{p}_t \; ; \; 0 \le t \le T\}$ *are choosen so that the mapping*

$$\mathbb{p}(\omega) = \sum_{t=1}^{T} \mathbb{p}_t(\omega_t)$$

is a well defined performance density on $(\Omega, \mathbf{B}(\Omega))$*. Let* (S, d_S) *be another metric space. For any* $\omega \in \Omega$ *we define* $X(\omega) \in S^{T+1}$ *by setting*

$$\begin{cases} X(\omega)_t &= F(X(\omega)_{t-1}, \omega_t) \qquad 1 \le t \le T \\ X(\omega)_0 &= x_0 \in S \end{cases}$$

where $F : S \times U \to S$ *is a continuous function such that*

$$d_S(F(x,u), F(y,v)) \le d_S(x,z) + d_U(u,v)$$

Under our assumptions the mapping $X : U^T \to S^{T+1}$ *can be regarded as an optimization variable with performance density*

$$\mathbb{p}^X(z) = \sup \{\sum_{t=1}^{T} \mathbb{p}_t(\omega_t) \; ; \; \omega \in U^T \; : \; X(\omega) = z\}$$

It is interesting to note that $\mathbb{p}^X(z) = \sup_{\emptyset} = -\infty = \mathbb{0}$ *whenever* $z_0 \neq x_0$ *.*
Using lemma 2.3 for any $x_1, \ldots, x_T \in S$ *we can find some* $\omega(x) \in U^T$ *such that*

$$\mathbb{p}^X(x_0, x_1, \ldots, x_T) = \sum_{t=1}^{T} \mathbb{p}_t(\omega_t(x))$$

Finally, if we denote by $\mathbb{p}^X_{t|r,\ldots,s}$*,* $r \le s \le t$*, the conditional performance of* X_t *with respect to* $(X_r, \ldots, X_s)$ *one can check that*

$$\begin{aligned} \mathbb{p}^X_{t|0,\ldots,t-1}(x_t|x_0, \ldots, x_{t-1}) &= \sup \{\mathbb{p}_t(u) \; ; \; u \in U \; : \; F(x_{t-1}, u) = x_t\} \\ &= \mathbb{p}^X_{t|t-1}(x_t|x_{t-1}) \end{aligned}$$

and therefore the performance density of X *can be rewritten in the following form*

$$\mathbb{p}^X(x_0, x_1, \ldots, x_T) = \sum_{t=1}^{T} \mathbb{p}^X_{t|t-1}(x_t|x_{t-1})$$

2.4 Bellman-Maslov Processes

The essence of the Bellman-Hamilton-Jacobi Theory can be introduced in forward time with initial penalty. This shows the central role played by the concatenation semigroup of optimization transition performances in Bellman's optimality principle as probability transitions in Markov systems.

We state that this principle may be viewed as a basic definition of optimization processes like Markov's property rather than a conclusion. In other words, a Maslov process is an optimization process that satisfies the $(\max, +)$-version of Markov's causality principle. The reversal of time yields optimal control processes of regulation type. The groundwork for the theory of Markov stochastic processes was laid in 1906 by A. A. Markov. In his investigation of connected experiments, he formulated the principle that the "future" is independent of the "past" when the "present" is known. This principle is the causality principle of classical physics state carried over to stochastic dynamical systems. It specifies that the knowledge of the state of a system at a given time is sufficient to determine its state at any future time. The following concept is the extension of the Markov causality principle in the Maslov optimization framework.

Definition 2.15 *Let $(\Omega, \mathbf{B}(\Omega), \mathbb{P}, X = \{X_t\}_{t\in\mathcal{T}})$ be an S-valued optimization process with time space $\mathcal{T}$. It is called a Bellman-Maslov process whenever its future and its past are independent when its present is known. In other words, X is a Bellman-Maslov process if for every $r \leq s \leq t$ we have the relation*

$$\mathbb{p}^X_{[r,s],[s,t]|s}(x_1, x_2|x) = \mathbb{p}^X_{[r,s]|s}(x_1|x) \odot \mathbb{p}^X_{[s,t]|s}(x_2|x) \tag{12}$$

for any $x_1 \in \mathcal{E}([r,s], S)$, $x_2 \in \mathcal{E}([s,t], S)$ and $x \in S$; where $\mathbb{p}^X_{[r,s],[s,t]|s}$ denotes the conditional performance density of the pair $X_{[r,s],[s,t]} = (X_{[r,s]}, X_{[s,t]})$ with respect to X_s.

One consequence of (12) is that for any $s \leq t$ the marginal performance density of X at time t can be computed from the marginal of X at time s

$$\mathbb{p}^X_t(x) = \int_S^{\oplus} \mathbb{p}^X_{t|s}(x|z) \odot \mathbb{p}^X_s(z) \odot dz$$

In other words

$$\mathbb{p}^X_t(x) = \sup_{z\in S} \mathbb{p}^X_{t|s}(x|z) + \mathbb{p}^X_s(z) \qquad \text{(Bellman's Principle)}$$

This means that Bellman's causality principle is the (max,+)-version of the classical Chapman-Kolmogorov's equation for Markov stochastic processes.

When the time space $\mathcal{T} = \mathbb{N}$ an optimization process satisfying (12) will be called a Bellman-Maslov chain in reference to Markov chain theory. When the transitions performance density $\mathbb{p}^X_{t|s}$ only depends on $|t-s|$ the optimization process is said to be time homogeneous.

Example 8 *Let X be a discrete time optimization process taking values in a metric space S. Assume that its performance density $\mathbb{p}^X : \mathcal{E}(\mathbb{N}, S) \to \mathbb{R}_{\max}$ is given by*

$$\mathbb{p}^X(x) = c_0(x_0) + \sum_{t\geq 1} c_t(x_{t-1}, x_t)$$

where $C_0 : S \to \mathbb{R}_{\max}$ and $c_t(x, \centerdot) : S \to \mathbb{R}_{\max}$ are performance densities on S. By proposition 2.5 for any $t \geq 0$ the sequence $(X_0, \ldots, X_t)$ can be regarded as a S^{t+1}-valued optimization variable with performance density

$$\mathbb{p}^X_{0,\ldots,t}(x_0, \ldots, x_t) = c_0(x_0) + \sum_{s=1}^{t} c_s(x_{s-1}, x_s)$$

and therefore the conditional density of X_t with respect to $(X_0, \ldots, X_{t-1})$ is given by

$$\mathbb{p}^X_{t|0,\ldots,t-1}(x_t|x_0, \ldots, x_{t-1}) = c_t(x_{t-1}, x_t) = \mathbb{p}^X_{t|t-1}(x_t|x_{t-1})$$

This yields

$$\mathbb{p}^X_t(x) = \sup_{z\in S} \left(\mathbb{p}^X_{t|t-1}(x|z) + \mathbb{p}^X_{t-1}(z)\right) = \int_S^{\oplus} \mathbb{p}^X_{t|t-1}(x|z) + \mathbb{p}^X_{t-1}(z) \odot dz$$

Example 9 *Assume that Ω is a Polish space of controls on $(0,1)$ and taking values in a metric space U. To each $\omega \in \Omega$ we define $X : \omega \in \Omega \to X(\omega) \in \mathcal{C}((0,1), \mathbb{R}^d)$ by setting*

$$\begin{cases} \dot{X}(\omega)_t &= f(X_t(\omega), \omega_t) \\ X(\omega)_0 &= x_0 \in \mathbb{R}^d \end{cases} \tag{13}$$

where $f : \mathbb{R}^d \times U \to \mathbb{R}^d$ a a sufficiently regular mapping such that the mapping $X : \omega \in \Omega \to \mathcal{C}((0,1), S)$ is continuous (as usually the space $\mathcal{C}((0,1), \mathbb{R}^d)$ is furnished with the uniform topology). Finally suppose we are given a performance measure $\mathbb{P}$ on $\mathbf{B}(E)$ with a density of the form

$$\mathbb{p}(\omega) = \int_0^1 L_t(X_t(\omega), \omega_t)\, dt$$

for some $L_t : \mathbb{R}^d \times U \to \mathbb{R}_{\max}$. If the following condition is met

(H) $\forall x \in \mathbb{R}^d \quad \forall t \in (0,1) \quad \exists \omega \in \Omega$ *such that*

$$\int_t^1 L_t(X^x_s(\omega), \omega_s)\, ds = 0$$

where $\{X_s^x(\omega)\ :\ t \le s \le 1\}$ *is the solution of (13) such that* $X_t(\omega) = x$.

then $X_{[0,t]} = \pi_{[0,t]} \circ X$ *is an optimization variable with performance density*

$$\mathbb{p}^X_{[0,t]}(x) = \sup\{\int_0^t L_s(X_s(\omega), \omega_s)\, ds\ ;\ \omega \in \Omega\ :\ X_{[0,t]}(\omega) = x\}$$

Moreover if we denote by $\mathbb{p}^X_t$ *the performance density of* X_t *we have for any* $s \le t$

$$\mathbb{p}^X_t(x) = \sup_{z \in \mathbb{R}^d} \left(\mathbb{p}^X_{t|s}(x|z) + \mathbb{p}^X_s(z)\right) = \int_{\mathbb{R}^d}^{\oplus} \mathbb{p}^X_{t|s}(x|z) + \mathbb{p}^X_s(z) \odot dz$$

where

$$\mathbb{p}^X_{t|s}(x|z) = \sup\{\int_s^t L_r(X_r(\omega), \omega_r)\, dr\ ;\ \omega \in \Omega\ :\ X_s = z \text{ and } X_t(\omega) = x\}$$

3 Applications

3.1 Invariant Measures of Bellman-Maslov chains

Let $\{X_n\ ;\ n \geq 0\}$ be a time homogeneous Bellman-Maslov chain with transition performance densitie $\mathbb{P}(x/y)$ and initial performance measure μ. Throughout this section we assume that

- (E, d_{E}) is a compact metric space.
- $(x, y) \in E \times E \to \mathbb{P}(x/y) \in \mathbb{R}_{\max}$ is a continuous mapping such that

  $$\rho\left(\mathbb{P}(x/y), \mathbb{P}(x/z)\right) \leq C\ d_{\mathrm{E}}(y, z) \tag{14}$$

  for some finite constant which doesn't depend on $x, y, z \in E$.

Under these conditions the transition performance $\mathbb{P}(x/y)$ generates two optimization operators. One is acting on continuous functions $f \in \mathcal{C}(E)$ and the other one on performance densities $\mu : E \to \mathbb{R}_{\max}$

$$\mathbb{P}f(x) \;=\; \int_E^{\oplus} f(y) \odot \mathbb{P}(y/x) \odot dy \tag{15}$$

$$\mu\mathbb{P}(x) \;=\; \int_E^{\oplus} \mathbb{P}(x/y) \odot \mu(y) \odot dy \tag{16}$$

To see that these two operations are well defined we first notice that the mapping

$$(x, y) \in E \times E \to \mathbb{P}(x/y) \odot \mu(y) \in \mathbb{R}_{\max}$$

is a performance density and proposition 2.5 implies that the marginal $\mu\mathbb{P}$ defined in (16) is well defined performance density on E. On the other hand using the inequality (14) one concludes that

$$\rho\left(\mathbb{P}f(x), \mathbb{P}f(y)\right) \leq C\ d_{\mathrm{E}}(x, y)$$

and therefore $\mathbb{P}$ is a Feller transition in the sense that

$$\mathbb{P}\left(\mathcal{C}(E)\right) \subset \mathcal{C}(E)$$

and the operation (15) is well defined. We will denote $\mathbb{P}^n(x/y)$ the n-step transitions defined by the inductive formula

$$\mathbb{P}^n(x|z) = \int_E^{\oplus} \mathbb{P}(x/y) \odot \mathbb{P}^{n-1}(y|z) \odot dy \qquad \forall n \geq 1$$

The following result which is easy to prove is the (max,+)-version of the classical Feller Markov semi-groups in Stochastic Process Theory.

Proposition 3.1 *For any time $n \geq 1$ and for any test function $f \in \mathcal{C}(E)$ the function*

$$x \in E \longrightarrow \mathbb{P}^n f(x) = \mathbb{E}(f(X_n)|X_0 = x) = \int_E^{\oplus} f(y) \odot \mathbb{P}^n(y/x) \odot dy$$

is continuous and $\{\mathbb{P}^n \; ; \; n \geq 0\}$ is a monotone semigroup of $(\oplus, \odot)$-linear operators on $\mathcal{C}(E)$, that is

1. *$\mathbb{P}^0 = Id$ and for any $m, n \geq 0$ we have $\mathbb{P}^m \circ \mathbb{P}^n = \mathbb{P}^{m+n}$.*
2. *For any $f, g \in \mathcal{C}(E)$ such that $f \leq g$ we have $\mathbb{P}^n f \leq \mathbb{P}^n g$ for any $n \geq 0$.*
3. *For any $f, g \in \mathcal{C}(E)$ and $a, b \in \mathbb{R}_{\max}$ and $n \geq 0$*

$$\mathbb{P}^n(a \odot f \oplus b \odot g) = a \odot \mathbb{P}^n f \oplus b \odot \mathbb{P}^n g$$

The former semigroup arise in the analysis of optimal control problems with terminal penalties. In these settings the set of test functions $\mathcal{C}(E)$ represents the class of terminal penalties and for any penalty f the function $\mathbb{P}^n f$ is written in the more explicit form

$$\mathbb{P}^n f(x_0) = \sup \left\{ \sum_{m=1}^{n} \mathbb{P}(x_m/x_{m-1}) + f(x_n) \; ; \; x_1, \ldots, x_n \in E \right\}$$

One problem is to study the asymptotic behavior of the operators $\mathbb{P}^n$ as n tends to ∞. This problem is not trivial mainly because the transformation $\mathbb{P}$ is non expansive. More precisely if we note

$$\rho\left(\mathbb{P}f, \mathbb{P}g\right) = \sup_{x \in E} \rho\left(\mathbb{P}f(x), \mathbb{P}g(x)\right)$$

we have that

$$\rho\left(\mathbb{P}f, \mathbb{P}g\right) \quad \leq \quad \rho\left(f, g\right) \; \big(\sup_{x,y \in E} \exp \mathbb{P}(x/y)\big) = \rho\left(f, g\right)$$

The study of (max,+)-semi-groups have been started in [21, 24, 26, 27, 31], see also [28] for a complete bibliography on this subject. The most accurate works are those due to Kolokoltsov [25] who investigates the existence of eigenvalues and provide some applications to Mathematical Economy, Games Theory and Stochastic optimal control. The abstract spectral analysis of such transformations is described in full details in Lesin-Samborsky [**?**].

The main purpose of this section it to propose a probabilistic description of the fixed point of such transformations. In the study of asymptotic of Markov Processes invariant probability measures play a crucial role. The same notions are now introduced for performance measure.

Definition 3.2 *A performance measure π on $(E,\mathcal{E})$ with values in $\mathcal{A}$ is invariant for $\mathbb{P}(y/x)$ iff*

$$\forall y \in E \qquad \pi(y) = \int_E^{\oplus} \pi(x) \odot \mathbb{P}(y/x) \odot dx \ .$$

A performance measure π on $(E,\mathcal{E})$ with values in $\mathcal{A}$ is sub-invariant for $\mathbb{P}(y/x)$ iff

$$\forall y \in E \qquad \pi(y) \geq \int_E^{\oplus} \pi(x) \odot \mathbb{P}(y/x) \odot dx \ .$$

In the sequel the integral :

$$\int_E^{\oplus} \pi(x) \odot \mathbb{P}(y/x) \odot dx$$

will be noted $\pi \cdot \mathbb{P}(y)$. By linearity, if π_1 and π_2 are two invariant (resp. sub-invariant) performance measures and $\alpha\,,\, \beta \in \mathbb{A}$ then $\alpha \odot \pi_1 \oplus \beta \odot \pi_2$ is invariant (resp. sub-invariant). The subset of invariant (resp. sub-invariant) performance is a convex subset of $\mathbf{M}(E,\mathcal{E})$.
It is well known ([**?**]) that the *taboo* probabilities give an explicit construction of invariant measures. Then define

$$\tau_x = \inf\{n \geq 1 \ s.t. \ X_n = x\} \quad \forall x \in E \ .$$

and

$$_{x_0}\mathbb{P}^n(x/x_0) = \mathbb{P}(X_n = x\,,\, \tau_{x_0} \geq n \ /x_0)$$
$$(\forall x, x_0 \in E \ \ \forall n \in \mathbb{N}^* \ .)$$

For each x_0 in E let π_{x_0} defined by

$$\pi_{x_0}(x) = \oplus_{n\geq 1} \ {}_{x_0}\mathbb{P}^n(x/x_0) \quad \forall x \in E \ . \tag{17}$$

$\pi_{x_0}(\cdot)$ is a performance measure because

$$\begin{aligned} \oplus_{x\in E} \ \pi_{x_0}(x) &= \oplus_{n\geq 1} \left[\oplus_{x\in E} \ {}_{x_0}\mathbb{P}^n(x/x_0)\right] \\ &= \oplus_{n\geq 1} \ \mathbb{P}(\tau_{x_0} \geq n/x_0) \\ &= \mathbb{P}(\tau_{x_0} \geq 1/x_0) = 1\!\!1 \ . \end{aligned}$$

by definition of τ_{x_0} .

Proposition 3.3 *For all x_0 in E , π_{x_0} is a sub-invariant performance.*
If x_0 satisfies : $\mathbb{P}(\tau_{x_0} < \infty/x_0) = 1\!\!1$ then π_{x_0} is an invariant performance.

Proof:

$$\begin{aligned}\pi_{x_0} \cdot \mathbb{P}(x) &= \int_E^{\oplus} [\oplus_{n\geq 1} \ {}_{x_0}\mathbb{P}^n(y/x_0)] \odot \mathbb{P}(x/y) \odot dy \\ &= [\pi_{x_0}(x_0) \odot \mathbb{P}(x/x_0)] \oplus \\ &\int_{E\setminus\{x_0\}}^{\oplus} \oplus_{n\geq 1} \ {}_{x_0}\mathbb{P}^n(y/x_0) \odot \mathbb{P}(x/y) \odot dy\end{aligned}$$

By the Bellman Equation

$$\int_{E\setminus\{x_0\}}^{\oplus} \oplus_{n\geq 1} \ {}_{x_0}\mathbb{P}^n(y/x_0) \odot \mathbb{P}(x/y) \odot dy =_{x_0} \mathbb{P}^{n+1}(x/x_0) \ .$$

Then

$$\pi_{x_0} \cdot \mathbb{P}(x) = [\pi_{x_0}(x_0) \odot \mathbb{P}(x/x_0)] \oplus [\oplus_{m\geq 2} \ {}_{x_0}\mathbb{P}^m(x/x_0)] \ .$$

As $\mathbb{P}(x/x_0) = \ {}_{x_0}\mathbb{P}^1(x/x_0)$ and $\pi_{x_0}(x_0) \in [\mathbb{0}, \mathbb{1}]$ we have

$$\pi_{x_0}(x_0) \odot \mathbb{P}(x/x_0) \leq_{x_0} \mathbb{P}^1(x/x_0)$$

and then

$$\pi_{x_0} \cdot \mathbb{P}(x) \leq \oplus_{m\geq 1} \ {}_{x_0}\mathbb{P}^m(x/x_0) = \pi_{x_0}(x) \quad \forall x \in E \ .$$

π_{x_0} is a sub-invariant performance.
If moreover $\pi_{x_0}(x_0) = \mathbb{1}$ the inequality becomes an equality. But

$$\begin{aligned}\pi_{x_0}(x_0) &= \oplus_{n\geq 1} \ \mathbb{P}(\tau_{x_0} = n/x_0) \\ &= \mathbb{P}(\tau_{x_0} < \infty/x_0) \ .\end{aligned}$$

and the condition $\pi_{x_0}(x_0) = \mathbb{1}$ can be written
$\mathbb{P}(\tau_{x_0} < \infty/x_0) = \mathbb{1}$. ■

Thanks to (Max,+) calculus the condition
$\mathbb{P}(\tau_{x_0} < \infty/x_0) = \mathbb{1}$ has a more practical traduction.

Proposition 3.4 *For $x_0 \in E$ the two conditions are equivalent*

$$\begin{aligned} i) \ \ \mathbb{P}(\tau_{x_0} < \infty/x_0) &= \mathbb{1} && (18) \\ ii) \ \ \oplus_{n\geq 1} \ \mathbb{P}^n(x_0/x_0) &= \mathbb{1} \ . && (19)\end{aligned}$$

Proof:
Define

$$L(x_0, x_0) = \mathbb{P}(\tau_{x_0} < \infty/x_0) = \oplus_{n\geq 1} \ \mathbb{P}(\tau_{x_0} = n/x_0)$$

and

$$U(x_0, x_0) = \oplus_{n\geq 1} \mathbb{P}^n(x_0/x_0) \ .$$

Using the first entrance to x_0 and for $n \geq 2$

$$\mathbb{P}^n(x_0, x_0) = \mathbb{P}(\tau_{x_0} = n/x_0)\oplus$$

$$\oplus_{j=1}^{n-1} \mathbb{P}(\tau_{x_0} = j/x_0) \odot \mathbb{P}^{n-j}(x_0/x_0)$$

Summing on $n \in \mathbb{N}^*$ we have

$$U(x_0, x_0) = L(x_0, x_0) \oplus [L(x_0, x_0) \odot U(x_0, x_0)] \ .$$

The calculus in $[\mathbb{0}, \mathbb{1}]$ implies easily that

$$U(x_0, x_0) = \mathbb{1} \quad \text{iff} \quad L(x_0, x_0) = \mathbb{1} \ .$$

■

As in classical Markovian setting an other form of π_{x_0} can be obtained

$$\pi_{x_0} = \mathbb{E}_{x_0}[\oplus_{n=1}^{\tau_{x_0}} \mathbb{1}_{\{X_n = x\}}] \ \ \forall x \in E \ . \tag{20}$$

([?] p.235) where $\mathbb{E}_{x_0}$ is the integral under the distribution law of the process with the initial state x_0. At last let us mentioned a maximal property of the performance measure π_{x_0} .

Proposition 3.5 *If μ is an invariant performance such that $\mu(x_0) = \mathbb{1}$ then $\mu(x) \leq \pi_{x_0}(x) \quad \forall x \in E$.*

Proof:
We are going to prove by induction on n that

$$\mu(x) \leq \oplus_{k=1}^{n} {}_{x_0}\mathbb{P}^n(x/x_0) \ .$$

As μ is invariant

$$\begin{aligned} \mu(x) &= \int_E^{\oplus} \mu(y) \odot \mathbb{P}(x/y) \odot dy \\ &\leq \mu(x_0) \odot \mathbb{P}(x/x_0) \\ &\leq \mathbb{P}(x/x_0) \ \text{ if } \mu(x_0) = \mathbb{1} \\ &\leq {}_{x_0}\mathbb{P}^1(x/x_0) \end{aligned}$$

Then $\mu(x) \leq_{x_0} \mathbb{P}^1(x/x_0)$.
Suppose now that $\mu(x) \leq \oplus_{k=1}^{n} {}_{x_0}\mathbb{P}^k(x/x_0)$.

$$\begin{aligned}\mu(x) &= [\mu(x_0) \odot \mathbb{P}(x/x_0)] \oplus \\ &\quad \int_{E\setminus\{x_0\}}^{\oplus} \mu(y) \odot \mathbb{P}(x/y) \odot dy \\ &\leq {}_{x_0}\mathbb{P}^1(x/x_0) \oplus \\ &\quad \int_{E\setminus\{x_0\}}^{\oplus} \oplus_{k=1}^{n} {}_{x_0}\mathbb{P}^k(y/x_0) \odot \mathbb{P}(x/y) \odot dy \ .\end{aligned}$$

But

$$\int_{E\setminus\{x_0\}}^{\oplus} {}_{x_0}\mathbb{P}^k(y/x_0) \odot \mathbb{P}(x/y) \odot dy =_{x_0} \mathbb{P}^{k+1}(x/x_0) \ .$$

Then $\mu(x) \leq \oplus_{k=1}^{n+1} {}_{x_0}\mathbb{P}^k(x/x_0)$. Letting n tend to infinity : $\mu(x) \leq \pi_{x_0}(x) \quad \forall x \in E$. ∎

Suppose now that E is a *finite* set say $E = \{1, \cdots, n\}$. The invariance property of a performance π on E is here

$$\pi \cdot \mathbb{P}(i) = \oplus_{j=1}^{n} [\mathbb{P}(i/j) \odot \pi(j)] = \pi(i) \ \ \forall i \in E \ .$$

or in a more classical form

$$\pi(i) = \max_{j=1\cdots n} (\mathbb{P}(i/j) + \pi(j)) \ \ \forall i \in E \ .$$

By Proposition (3.4) if there exists $i_0 \in E$ s.t.
$\oplus_{n\geq 1} \mathbb{P}^n(i_0/i_0) = 1\!\!1$ then π_{i_0} as defined in (17) is an invariant measure. As E is a finite set this condition is equivalent to

$$\oplus_{i\in E} \oplus_{n\geq 1} \mathbb{P}^n(i/i) = 1\!\!1 \ .$$

or in usual algebra

$$\max \{ \sum_{i=1}^{n} \mathbb{P}(m_{i+1}/m_i) \} = 1\!\!1 \tag{21}$$

where the max is taken over all finite subsets $\{m_1, \cdots, m_n\}$ of E and where we let $m_{n+1} = m_1$.
Consider the *complete graph* G associated to $\{1, \cdots, n\}$ and define

- $\mathcal{C}$ the set of *circuits* of G
- $\mathcal{C}_e$ the set of *elementary circuits* of G *i.e.* no points is visited twice except the origin.

If $C \in \mathcal{C}$ and $[m_1, \cdots, m_k]$ are the vertex of C in the sense of the circuit, define the *performance cost* of C by

$$\mathbb{P}(C) = \odot_{i=1}^{k} \, \mathbb{P}(m_{i+1}/m_i) \quad (m_{k+1} = m_1) \ .$$

Then the condition (21) becomes

$$\oplus_{C \in \mathcal{C}} \ \mathbb{P}(C) = \mathbb{1} \ . \tag{22}$$

By induction on the length of C it is easy to express $\mathbb{P}(C)$ as an *ordinary* sum of performance of elementary circuits. But $\mathbb{P}$ taking its values in $[-\infty, 0]$, (22) is equivalent to

$$\oplus_{C \in \mathcal{C}_e} \mathbb{P}(C) = \mathbb{1} \ . \tag{23}$$

The set $\mathcal{C}_e$ is finite then (23) is

$$\exists C \in \mathcal{C}_e \ \text{ s.t. } \ \mathbb{P}(C) = \mathbb{1} \ .$$

or finally

$$\begin{aligned} \exists C &= [m_1, \cdots, m_k] \in \mathcal{C}_e \ \text{ s.t.} \\ \mathbb{P}(m_{i+1}/m_i) &= 0 \ \ \forall i = 1, \cdots, k \ \text{ and } \ m_{k+1} = m_1 \ . \end{aligned}$$

That is the existence of an elementary circuit with all null cost hedges.
In the Optimization setting, $\mathbb{P}(\cdot/j)$ is supposed to be a performance measure for all j *i.e.*

$$\oplus_{i=1}^{n} \ \mathbb{P}(i/j) = \mathbb{1} \qquad \forall j \in E$$

or in ordinary algebra

$$\forall j \in E \quad \exists n_j \in E \ \ s.t. \ \ \mathbb{P}(n_j/j) = 0 \ .$$

Then, *there always exists an elementary circuit with null cost* : indeed, starting from 1 find n_1 s.t.
$\mathbb{P}(n_1/1) = 0$, now starting from n_1 find n_2 s.t.
$\mathbb{P}(n_2/n_1) = 0$ and so on. After n times at last one point is visited twice and an elementary circuit with null cost was runned. The following Proposition resumed the results already obtained.

Proposition 3.6 *If for all j in E , $\mathbb{P}(\cdot/j)$ is a performance measure, there exists at last an elementary circuit, say $\mathcal{C}_0$, with null cost.*
For all i in $\mathcal{C}_0$ the measure π_i as defined in (17) is invariant.
For j in E , $\pi_i(j)$ can be expressed as :

$$\pi_i(j) = \max \ \{ \ \mathbb{P}(C) \ \ / \ \ C \in \mathcal{D}_{ij} \} \ .$$

where $\mathcal{D}_{ij}$ is the set of direct path from i to j i.e. with no circuit and without visiting i again.
For i and k in $\mathcal{C}_0$: $\pi_i = \pi_k$.

Proof:
The new form of π_i is clear : all circuit on the path from i to j can only reduce the total cost. For i and k in $\mathcal{C}_0$ the passage from i to k by $\mathcal{C}_0$ has a null cost, then it is easy to deduce that $\pi_i = \pi_k$ ■

As for Markov Processes one can ask about the convergence of $\mu \cdot \mathbb{P}^n$ for all initial performance μ and the link between invariant measures. A first approach in terms of operator is given by Nussbaum ([**?**]).
Define : $\mathbf{P} = \{p_{i\,j}\}$ with $p_{i\,j} = \mathbb{P}(i/j)\ \ \forall i, j \in E$ and let $\mathcal{T}_P$ be the operator :

$$\begin{array}{rcl} \mathcal{T}_P\ :\ (\bar{\mathbb{R}}_-)^n & \longrightarrow & (\bar{\mathbb{R}}_-)^n \\ x = (x_1, \cdots, x_n) & \longrightarrow & (\mathcal{T}_P(x))_i = \max\limits_{j=1\cdots n} (p_{i\,j} + x_j) \end{array}$$

Then an invariant performance π as defined in Definition (3.2) appears as a *fixed point* for $\mathcal{T}_P$. The existence of fixed points is of fundamental importance for Nussbaum's results. Thanks to Proposition (3.6) fixed points for $\mathcal{T}_P$ can be exhibit in Optimization setting and Nussbaum's conclusions hold for $\mathbb{P}$. For example his main result (Theorem 2.4) gives here
$\exists p \in \mathbb{N}$ and a map $\zeta\ \ (\bar{\mathbb{R}}_-)^n \longrightarrow (\bar{\mathbb{R}}_-)^n$
such that, for all initial performance measure μ

$$\exists k(\mu) \in \mathbb{N}\ \text{ s.t. }\ \mu \cdot \mathbb{P}^{kp} = \zeta(\mu)\ \ \forall k \geq k(\mu)$$

$\zeta(\mu)$ is an invariant performance for $\mathbb{P}^p$ and explicit formulas for the *cyclicity* integer p in terms of simple properties of $\mathbf{P}$ are also given.

An other approach developed in Baccelli et al. ([**?**]) is to study directly the asymptotic behavior of $\mathbf{P}^{\otimes n}$ to obtain an analogous to Perron-Frobenius theorem. Here the existence of a null cost circuit appears as an hypothesis (nevertheless the authors show how, a normalizing argument can lead to a general theory). The same cyclicity properties are obtained but expressed in (Max,+) calculus and the associated graph theory. For example the cyclicity integer p is related here to the length of elementary null costs circuits (see theorem 3.108 and 3.112 in ([**?**])).

The framework introduced in the preceding sections enable us to give a probabilistic description of an invariant performance measure with respect to some (max,+)-semigroup $\mathbb{P}(x/y)$. In the last part of this section we discuss the connections between these fixed points and the virtual energy function arising in the analysis of the long time behavior of a class of stochastic search algorithms such as the generalized simulated annealing (see for instance [37] and reference therein).
It is now convenient to introduce some additional definitions. Until the end of this section E is assumed to be a finite state space. What follows is an exposition of some basic definitions regarding the description of the virtual energy function which is due to Bott and

Mayberry [7] and also exposed in Freidlin-Wentzell [22].

Let $x \in E$, a graph consisting of arrows $y \to z$ with $(y, z) \in E - \{x\} \times E$ and $y \neq z$ is called an x-graph over E if it satisfies the following conditions

- every point $y \neq x$ is the initial point of exactly one arrow.
- there is no closed cycles in the graph.

We denote $G(x)$ the set of x-graphs.

Proposition 3.7 *If we denote by* $\nu : E \to \mathbb{R}_{\max}$ *the mapping defined by*

$$\nu(x) = \sup_{g \in G(x)} \sum_{(y \to z) \in g} \mathbb{P}(y/z)$$

then the normalized performance measure $\mu(x) = \nu(x) - \sup_{y \in E} \nu(y)$ *associated to* ν *is an invariant measure with respect to the transition* $\mathbb{P}(x/y)$, *i.e.* $\mu \mathbb{P} = \mu$.

Proof:
It clearly suffices to check that $\nu(x) = \sup_y \nu(y) + \mathbb{P}(y/x)$ for any $x \in E$.
Let $g^\star$ be an element of $G(x)$ such that

$$\nu(x) = \sum_{(y \to z) \in g^\star} \mathbb{P}(y/z)$$

Under our assumptions there exists a point $y(x) \in E$ such that $\mathbb{P}(y(x)/x) = 0$. By definition of $g^\star$ there also exists a path of a given length r

$$p_0 = y \to p_1 \to \ldots \to p_{r-1} \to p_r = x$$

such that $(p_{k-1} \to p_k) \in g^\star$ for any k. If we denote by $\widetilde{g}$ the graph obtained from $g^\star$ by replacing the arrow $(p_{r-1} \to x)$ by $(x \to y(x))$ we clearly have

$$\widetilde{g} \in G(p_{r-1}) \quad \text{and} \quad \sum_{(e_1 \to e_2) \in \widetilde{g}} \mathbb{P}(e_2/e_1) \;+\; \mathbb{P}(x/p_{r-1}) = \nu(x)$$

and therefore

$$\mathbb{P}(x/p_{r-1}) + \nu(p_{r-1}) \geq \nu(x)$$

The end the proof of the proposition we observe that for any $y \in E$

$$\mathbb{P}(x/y) + \nu(y) = \sup \left\{ \sum_{(e_1 \to e_2) \in g} \mathbb{P}(e_2/e_1) \right\} \leq \nu(x)$$

where the supremum runs over all graphs g having a cycle at point x. ■

In the study of simulated annealing algorithms the function μ is called the virtual energy associated to the virtual communication cost function $\mathbb{P}(x/y)$. In these settings the simulated annealing algorithm is a random search process $\{\xi_t\ :\ t \geq 0\}$ associated to a cooling schedule which ensures that ξ_t converges in probability as $t \to \infty$ to the set of global minima of the virtual energy function (the interested reader is referred to the classical texts on this subject and to [37, 33]) In [18] we discuss the applications of such algorithms in the study of the convergence of genetic-type algorithms. In this case the state space is a product space $E = S^N$ where S is an auxiliary finite set and N is the number of particles used in the genetic algorithm. Let us present a simple example of communication cost function $\mathbb{P}(x/z)$ which arise in this situation

$$\mathbb{P}(z/x) = \sup \left\{ -\sum_{i=1}^{N} d(x^i, y^i) - \sum_{i=1}^{N} (f(z^i) - \widehat{f}(y)) \right\} \tag{24}$$

where the supremum runs over all $y \in S^N$ such that

$$\left\{ z^i\ ;\ 1 \leq i \leq N \right\} \subset \left\{ y^i\ ;\ 1 \leq i \leq N \right\}$$

and

- $f : S \to \mathbb{R}_+$ is a given function.
- $d(u,v) = 0$ if $u = v$ and 1 otherwise and $\widehat{f}(y) = \min \{ f(y^i)\ ;\ 1 \leq i \leq N \}$

Notice that the first term in the right hand side of (24) is minimal when the particles $x = (x^1, \ldots, x^N)$ are kept unchanged so that $x = y$. the second term is null as soon as all particles y^i, $1 \leq i \leq N$, move to the best position so that $\{z^i\ ;\ 1 \leq i \leq N\} = \{y^i\ ;\ 1 \leq i \leq N\}$ and $f(z^i) = \widehat{f}(y)$ for any $1 \leq i \leq N$.
In these settings the crucial point is to prove that the particles will find the global minima of the fitness function $f : S \to \mathbb{R}_+$. In other words we have to check that

$$\left\{ x \in S^N\ ;\ \mu(x) = \min_E \mu \right\} \subset \left\{ u \in S\ ;\ f(u) = \min_S f \right\}^N \tag{25}$$

In [18] we propose a natural test set approach to study the global minima of the virtual energy. More precisely we give conditions for a given set $H \subset E$ to contain the global minima of μ and we apply this criterion to prove that (25) holds for a sufficiently large population size N.

3.2 Dynkin's Formula

In this section we discuss the (max,+)-version of Dynkin's formula. In this setup $(\Omega, \mathbf{B}(\Omega))$ denotes an Hausdorff topological space of controls $\omega = (\omega_t)_{t\in(0,T)}$, $T > 0$, taking values in a metric space U. To each control $\omega \in \Omega$ we associate a continuous path $X(\omega) \in \mathcal{C}((0,T), \mathbb{R}^d)$, $d \geq 1$, by setting

$$\left\{ \begin{array}{rcl} \dot{X}(\omega)_t & = & F(X_t(\omega), \omega_t) \\ X(\omega)_0 & = & x_0 \in \mathbb{R}^d \end{array} \right. \tag{26}$$

where $f : \mathbb{R}^d \times U \to \mathbb{R}^d$ a a sufficiently regular mapping such that the mapping

$$X : \omega \in \Omega \to X(\omega) \in \mathcal{C}([0,T], S)$$

is continuous (as usually the space $\mathcal{C}([0,T], \mathbb{R}^d)$ is furnished with the uniform topology) and therefore can be regarded as an optimization process.
We denote by $\mathbb{P}$ a performance measure on $(\Omega, \mathbf{B}(\Omega))$ with density

$$\mathbb{p}(\omega) = \int_0^T L(X_t(\omega), \omega_t)\, dt$$

where $L : \mathbb{R}^d \times U \to \mathbb{R}_{\max}$ is a sufficiently regular mapping such that $(\Omega, \mathbf{B}(\Omega), \mathbb{P})$ is a well defined optimization basis. We also assume that the following condition holds

(H0) For any $x \in \mathbb{R}^d$, $s \leq t$ there exists a control $\omega \in \Omega$ such that

$$\int_s^t L(X_r^{s,x}(\omega), \omega_r)\, dr = 0$$

where $\{X_r^{s,x}(\omega) \ : \ r \in [s,t]\}$ is the solution of (26) such that $X_s^{s,x}(\omega) = x$.

Under (H0) we have already shown that X is a well defined Bellman-Maslov process and the conditional performance of X_t given X_s, $s \leq t$, is given by the formula

$$\mathbb{p}^X_{t|s}(z|x) = \sup\left\{ \int_s^t L(X_r^{s,x}(\omega), \omega_r)\, dr \ ; \ \omega \in \Omega \ : \ X_t^{s,x}(\omega) = z \right\}$$

In the same way one can check that the conditional performance of $X_{[0,t]}$ given $X_{[0,s]}$ is

$$\mathbb{p}^X_{[0,t]|[0,s]}(z|x) = \sup\left\{ \int_s^t L(X_r^{(s,x_s)}(\omega), \omega_r)\, dr \ ; \ \omega \in \Omega \ : \ X_{[s,t]}^{(s,x_s)}(\omega) = z_{/[s,t]} \right\}$$

if $z_{/[0,s]} = x$ and $-\infty$ otherwise and therefore

$$\begin{array}{rcl} \mathbb{p}^X_{[s,t]|[0,s]}(y|x) & = & \sup\left\{ \int_s^t L(X_r^{(s,x_s)}(\omega), \omega_r)\, dr \ ; \ \omega \in \Omega \ : \ X_{[s,t]}^{(s,x_s)}(\omega) = y \right\} \\ & = & \mathbb{p}^X_{[s,t]|s}(y|x_s) \end{array}$$

Theorem 3.8 *Assume that (H0) holds and the test function $\varphi \in \mathcal{C}_b^1(\mathbb{R})$ satisfies the following condition*

(H1) For any $x, p \in \mathbb{R}^d$ there exists a control value $\omega(x,p) \in U$ such that

$$\omega(x,p) = Arg \sup_{u \in U} \left(p \bullet F(x,u) + L(x,u) \right)$$

and for any $s \in [0,T]$ and $x_s \in E$ the feedback control

$$t \in [s,T] \to \omega_t(X_t^{(s,x_s)}(\omega), D\varphi(X_t^{(s,x_s)}(\omega)))$$

is the restriction on $[s,T]$ of some control $\omega \in \Omega$.

Then we have for any $t \in [0,T]$

$$\varphi(X_t) = \varphi(x_0) + \int_0^t H\left(X_s, D\varphi(X_s)\right) \, ds + M_t(X_{[0,t]})$$

where H is the Hamiltonian operator defined by

$$H(x,p) = \sup_{u \in U} \left(p \bullet F(x,u) + L(x,u) \right)$$

and $M_t : x \in \mathcal{C}([0,t], \mathbb{R}^d) \to M_t(x) \in \mathbb{R}_{\max}$ is a (max,+)-martingale in the sense that

$$\forall s \leq t \qquad \int^{\oplus} M_t(z) \odot \mathbb{P}^X_{[0,t]|[0,s]}(z|x) \odot dz = M_s(x)$$

In addition we have

$$\forall t \in [0,T] \qquad \int^{\oplus} M_t(z) \odot \mathbb{P}^X_{[0,t]}(z) \odot dz = 0$$

Proof:
We clearly have for any control $\omega \in \Omega$

$$\varphi(X_t(\omega)) = \varphi(x_0) + \int_0^t D\varphi(X_s(\omega)) \bullet F(X_s(\omega), \omega_s) \, ds$$

and therefore

$$M_t(X(\omega)) = \int_0^t \left[D\varphi(X_s(\omega)) \bullet F(X_s(\omega), \omega_s) - H\left(X_s(\omega), D\varphi(X_s(\omega))\right) \right] \, ds$$

This yields that for any $s \leq t$ the term

$$\int^{\oplus} \left(M_t(y) - M_s(x) \right) \odot \mathbb{P}^X_{[0,t]|[0,s]}(y|x) \odot dz$$

is equal to

$$\sup\left\{M_t(X^{s;x_s}(\omega)) - M_s(X^{(s,x_s)}(\omega)) + \int_s^t\ L(X_r^{(s,x_s)}(\omega),\omega_r)\ dr\ ;\ \omega\in\Omega\right\}$$

we get the desired result by taking the feedback control

$$r\in[s,T]\to\omega_r(X_r^{(s,x_s)}(\omega), D\varphi(X_r^{(s,x_s)}(\omega)))$$

■

Corollary 3.9 *Assume that (H0) holds and the test function $\varphi\in\mathcal{C}_b^1(\mathbb{R})$ satisfies (H1). Then for any $x\in E$*

$$\mathbb{E}\left(\varphi(X_t^x) - \int_0^t\ H\left(X_s^x, D\varphi(X_s^x)\right)\ ds\right) = \varphi(x) \qquad \text{(Dynkin's Formula)} \tag{27}$$

where $\{X_r^x(\omega)\ :\ r\in[0,t]\}$ is the solution of (26) such that $X_0^x(\omega)=x$.

Example 10 *The simplest case of optimization process in classical deterministic optimal control is defined as before with $d=1$, $\Omega=L^2([0,T];\mathbb{R})$ and*

$$L(x,u) = -\frac{u^2}{2} \qquad \text{and} \qquad F(x,u) = a\,x + b\,u \quad \text{with} \quad a,b\in\mathbb{R}$$

We immediately notice that the Hamiltonian associated to L is

$$H(x,p) = \ p\ a\ x + \frac{1}{2}\ |pb|^2$$

If $L^2([0,T];\mathbb{R})$ is equipped with the weak topology then $(\Omega,\mathbf{B}(\Omega),\mathbb{P})$ is a well defined optimization basis and

$$\mathbb{p}(\omega) = -\frac{1}{2}\ \int_0^T\ \omega_s^2\ ds$$

On the other hand the mapping $X:\omega\in L^2([0,T];\mathbb{R})\to X(\omega)\in\mathcal{C}([0,T],\mathbb{R})$ given by

$$\begin{cases} \dot{X}\ (\omega)_t &= a\ X_t(\omega) + b\ \omega_t \\ X(\omega)_0 &= x_0\in\mathbb{R}\end{cases}$$

is continuous and therefore $(\Omega,\mathbf{B}(\Omega),\mathbb{P},X)$ is a well defined Bellman-Maslov process with performance density $\mathbb{p}^X:\mathcal{C}([0,T],\mathbb{R})\to\mathbb{R}_{\max}$ given by the formula

$$\mathbb{p}^X(x) = -\frac{1}{2}\ \int_0^T\ (\dot{x}_s - a\ x_s)^2\ ds$$

if the integral exists and $-\infty$ otherwise. In this simple example it is not difficult to see that (H0) holds. One can also easily check that for any test function $\varphi \in \mathcal{C}_b^1(\mathbb{R})$ the optimal policy

$$t \in [s,T] \to \omega_t(X_t^{(s,x_s)}(\omega), \frac{\partial \varphi}{\partial x}(X_t^{(s,x_s)}(\omega))) = \frac{\partial \varphi}{\partial x}(X_t^{(s,x_s)}(\omega))\ b$$

belongs to $L^2([s,T];\mathbb{R})$. This implies that (H1) is satisfied and (27) holds with

$$H\left(\frac{\partial \varphi}{\partial x}(x), x\right) = a\ x\ \frac{\partial \varphi}{\partial x}(x) + \frac{1}{2}\ b^2\ \left(\frac{\partial \varphi}{\partial x}(x)\right)^2$$

Our next objective it to show how Dynkin's formula provides a very quick way to estimate the performance of a given set of admissible controls.

Corollary 3.10 *Assume that (H0) holds and the test function $\varphi \in \mathcal{C}_b^1(\mathbb{R})$ satisfies (H1) and*

$$H\left(x, D\varphi(x)\right) \leq -\ a \qquad \forall x \in A \tag{28}$$

for some $a > 0$ and $A \in \mathbf{B}(\mathbb{R}^d)$. Then, as soon as $x \in A$ we have

$$\mathbb{P}\left(\{\omega \in \Omega\ :\ \forall 0 \leq s \leq t \quad X_s^x(\omega) \in A\}\right) \leq \varphi(x) - \inf_A \varphi - a\ t \tag{29}$$

Proof:
To prove (29), note that, by (27) and (28)

$$\varphi(x) \geq \mathbb{E}\left(\left[\varphi(X_t^x) - \int_0^t\ H\left(X_s^x, D\varphi(X_s^x)\right)\ ds\right] \odot \mathbb{1}_{\{\forall s\in[0,t]\, X_s^x \in A\}}(X^x)\right)$$

Hence,

$$\varphi(x) \geq \inf_A \varphi + b\ t + \mathbb{P}\left(\{\omega \in \Omega\ :\ \forall 0 \leq s \leq t \quad X_s^x(\omega) \in A\}\right)$$

■

In other words if the test function $\varphi \in \mathcal{C}_b^1(\mathbb{R})$ satisfies (28) we have that

$$\sup\left\{\int_0^t\ L(X_s^x(\omega), \omega_s)\ ds\ ;\ \omega \in \Omega\ :\ \forall 0 \leq s \leq t \quad X_s^x(\omega) \in A\right\} < 0$$

as soon as $t > \frac{1}{a}\ (\varphi(x) - \inf_A \varphi)$ and therefore

$$\sup\left\{\int_0^t\ L(X_s^x(\omega), \omega_s)\ ds\ ;\ \omega \in \Omega\ :\ \exists 0 \leq s \leq t \quad X_s^x(\omega) \notin A\right\} = 0$$

Example 11 *The preceding example gives a clear application of the preceding corollary. Let us set*

$$\varphi(x) = \frac{x^2}{2} \qquad \text{and} \qquad A = [\epsilon, 1] \qquad \epsilon > 0$$

It is easily check that

$$\forall x \in [\epsilon, 1] \qquad H\left(x, D\varphi(x)\right) = (a + b^2/2)\ x^2 \leq \epsilon\ (a + b^2/2)$$

as soon as $a + b^2/2 < 0$. *This yields that*

$$\begin{aligned} \mathbb{P}\left(\forall 0 \leq s \leq t \quad \epsilon \leq X_s^x \leq 1\right) &\leq \frac{x^2}{2} - \inf_{[\epsilon,1]} \varphi - \epsilon\ |a + b^2/2|\ t \\ &\leq \frac{1}{2}(1 - \epsilon^2) - \epsilon\ |a + b^2/2|\ t \end{aligned}$$

and therefore for any $t > \dfrac{1}{\epsilon\ |2a + b^2|}$

$$\sup\left\{-\frac{1}{2}\int_0^t \omega_s^2\ ds \quad : \quad \forall 0 \leq s \leq t \quad \epsilon \leq X_s^x(\omega) \leq 1\right\} < -\frac{\epsilon^2}{2}$$

3.3 Generalized solution of Hamilton-Jacobi equations

Using Maslov's optimization framework we discuss now the solution of first order Hamilton-Jacobi equations. This part of the paper essentially records works of Kolokoltsov and Maslov [26]. In this paper the authors make use of idempotent analysis to propose a generalized solution of a class of Hamilton-Jacobi equations. The main contribution here is to extend their result to more general classes of Hamilton-Jacobi equations and to connect these new generalized solutions with the classical viscosity solutions (see [5, 10, 11] and references therein). As in [5] the main tools here are the so-called relaxed controls.
The class of differential equations studied in this paper have already been studied in details in several works (see for instance [5]). What makes our result interesting and new is that we propose a natural and general "probabilistic line of thought" to study these Hamilton-Jacobi equations. All the methods of proof we describe in this paper are built upon the application of the idempotent analysis framework developed in the preceding sections. This will provide a good illustration of the application of Maslov's optimization theory. A practical advantage of this approach is that it leads to a new understanding of such equations. For instance we will see that the Hamiltonian operator is the (max,+)-version of the Kolmogorov differential operator associated to a Markov diffusion process. On the other hand the viscosity solution of an Hamilton-Jacobi equation with discontinuous initial value coincide with the weak solution in idempotent analysis framework.
Let U be a compact metric space. By $\mathbf{M}_1(U)$ we denote the space of all probability measures on U furnished with the weak topology. Let us assume that

$$\Omega = L^\infty((0,1), \mathbf{M}_1(U))$$

equipped with the weak* topology. We recall that μ^n weak* converges to μ as $n \to \infty$ if and only if for any continuous function $\phi : (0,1) \times U \to \mathbb{R}$

$$\lim_{n\to\infty} \int_{(0,1)\times U} \phi(t,u)\mu_t^n(du)\ dt = \int_{(0,1)\times U} \phi(t,u)\mu_t(du)\ dt$$

For this topology $L^\infty((0,1), \mathbf{M}_1(U))$ is a compact space and the points of this set as usually called relaxed controls.
Let $\mathcal{C}((0,1), \mathbb{R}^d)$ be the space of all continuous mappings $d \geq 1$ furnished with the uniform norm. For any $x \in \mathbb{R}^d$ we define $X^x : \omega \in \Omega \to X^x(\omega) \in \mathcal{C}((0,1), \mathbb{R}^d)$ by setting

$$X_t^x(\omega) = x + \int_0^t F(X_s^x(\omega), \omega_s)\ ds \qquad F(x,\mu) = \int_U f(x,u)\ \mu(du) \tag{30}$$

where $f : \mathbb{R}^d \times U \to \mathbb{R}^d$ is a bounded continuous function so that

$$|f(x,u) - f(y,u)| \leq C\ |x-y|$$

In this framework the path $t \in (0,1) \to X_t^x(\omega) \in \mathbb{R}^d$ is called the relaxed path associated to ω with initial condition $x \in \mathbb{R}^d$. On can check that X^x is continuous and therefore it can be always regarded as an optimization variable taking values in $\mathcal{C}((0,1), \mathbb{R}^d)$.

Now, let $\mathbb{p}_x : \Omega \to \mathbb{R}_{\max}$ be defined as

$$\mathbb{p}_x(\omega) = \int_0^1 L(X_t^x(\omega), \omega_t)\, dt \qquad L(x,\mu) = \int_U l(x,u)\, \mu(du)$$

where $l : \mathbb{R}^d \times U \to \mathbb{R}_{\max}$ is a bounded continuous function so that

$$|l(x,u) - l(z,u)| \leq C\, |x - z|$$

We will always assume that condition (H0) page 34 holds and for any performance density μ on $\mathbb{R}^d$ the mapping

$$(x,\omega) \in \mathbb{R}^d \times \Omega \longrightarrow \mu(x) + \int_0^1 L(X_t^x(\omega), \omega_t)\, dt \in \mathbb{R}_{\max}$$

is a well defined performance density on $\mathbb{R}^d \times \Omega$. Under our assumptions we have that:

- For any $x \in \mathbb{R}^d$ there exists some $\omega_x \in \Omega$ such that $\mathbb{p}_x(\omega_x) = 0$ and therefore $\mathbb{P}_x(\Omega) = \mathbb{1}$
- For any $x \in \mathbb{R}^d$ $(\Omega, \mathbf{B}(\Omega), \mathbb{P}_x, X^x)$ is a well defined Bellman-Maslov process and the conditional density of X_t given X_s, $s \leq t$, is given by the formula

  $$\mathbb{p}_{t|s}^X(y|z) = \sup\left\{\int_s^t L(X_r^{s,z}(\omega), \omega_r)\, dr \;;\; \omega \in \Omega \;:\; X_t^{s,z}(\omega) = y\right\}$$

  where $\{X_r^{s,z}(\omega) \;:\; r \in [s,t]\}$ is the solution of (30) such that $X_s^{s,z}(\omega) = z$.
- For any performance density μ on $\mathbb{R}^d$ the mapping

  $$x \in \mathbb{R}^d \longrightarrow \int^{\oplus} \mathbb{p}_{t|0}(x|z) \odot \mu(z) \odot dz \in \mathbb{R}_{\max}$$

  is a performance measure on $\mathbb{R}^d$.

One clear consequence of proposition 2.5 is that the covariant measure $\mathbb{P}_x^X = \mathbb{P}_x \circ (X^x)^{-1}$ of $\mathbb{P}_x$ under the optimization variable X^x is a performance measure with density

$$\mathbb{p}_x^X(z) = \sup\{\int_0^1 L(X_t^x(\omega), \omega_t)\, dt \;;\; \omega \in L^\infty((0,1), \mathbf{M}_1(U)) \;:\; X^x(\omega) = z\}$$

By definition of the optimization process X we have that $\mathbb{p}_x^X(z) = -\infty = \mathbb{0}$ whenever $z_0 \neq x_0$. We will denote by $\mathbb{E}_x(\,.\,)$ the (max,+)-expectations with respect to the performance measure $\mathbb{P}_x$.

In order to carry our analysis to the next stage it is convenient to recall some classical results on relaxed controls. As usually we associate to each "classical" control $u(\centerdot) \in L^\infty((0,1),U)$ a relaxed control $t \to \delta_{u(t)} \in L^\infty((0,1),\mathbf{M}_1(U))$. Using this identification it is well known that $L^\infty((0,1),U)$ is a dense subset of $\Omega = L^\infty((0,1),\mathbf{M}_1(U))$.
Then, using the fact that the density $\mathbb{p}$ is continuous one concludes that

$$\forall x \in \mathbb{R}^d \;\; \forall A \in \mathbf{B}(\Omega) \qquad \begin{aligned} \mathbb{P}_x(A) &= \mathbb{P}_x(A \cap L^\infty((0,1),U)) \\ &= \sup\{\int_0^1 l(X_t^x(u),u_t)\,dt\ ;\ u \in A \cap L^\infty((0,1),U)\} \end{aligned} \tag{31}$$

Note that the above result remains valid if the density $\mathbb{p}$ is only upper-semi-continuous on Ω.

In section 3.2 we have presented the Hamilton-Jacobi operator associated to a time-homogeneous Bellman-Maslov process. We now present a new description of this operator which will be used in the sequel.

Definition-Proposition 3.11 *For any $x \in \mathbb{R}^d$ and $p \in \mathbb{R}^d$ we have*

$$\lim_{h \to 0} h^{-1}\, \mathbb{E}_x(\langle p, X_h^x - x\rangle) = H(x,p)$$

where $H : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_{\max}$ is the Hamiltonian function associated to l and given by

$$H(x,p) = \sup_{u \in U}(\langle p, f(x,u)\rangle + l(x,u))$$

Proof:
We first notice that

$$\begin{aligned} h^{-1}\, \mathbb{E}_x(\langle p, X_h^x - x\rangle) &= \sup\left\{\langle p, h^{-1}(X_h^x - x)\rangle + h^{-1}\int_0^h\ l(X_t^x(u),u_t)\,dt\right\} \\ &= \sup\left\{h^{-1}\int_0^h\ [\langle p, f(X_t^x,u_t)\rangle + l(X_t^x(u),u_t)]\ dt\right\} \end{aligned}$$

where the supremum run over all "classical" controls $u \in L^\infty((0,1),U)$ and the path $\{X_t^x(u)\ ;\ 0 \le t \le h\}$ is the solution of (30) associated to $u(\centerdot)$ and such that $X_0^x(u) = x$. Taking a constant control $u(\centerdot) = v \in U$ and letting $h \to 0$ we clearly have

$$\lim_{h \to 0} h^{-1}\, \mathbb{E}_x(\langle p, X_h^x - x\rangle) \ge \langle p, f(x,v)\rangle + l(x,v) \qquad \forall v \in U$$

Hence,

$$\lim_{h \to 0} h^{-1}\, \mathbb{E}_x(\langle p, X_h^x - x\rangle) \ge H(x,p)$$

On the other hand, the Lipschitz assumptions on the mappings l and f yield

$$\langle p, f(X_t^x, u_t)\rangle + l(X_t^x(u), u_t) \leq \langle p, f(x, u_t)\rangle + l(x, u_t) + \text{Cte } \|X_t^x(u) - x\|$$

This implies that

$$h^{-1} \int_0^h \left[\langle p, f(X_t^x, u_t)\rangle + l(X_t^x(u), u_t)\right] \, dt \leq H(x, p) + \text{Cte } \|f\| \, h$$

and therefore

$$\lim_{h \to 0} h^{-1} \, \mathbb{E}_x \left(\langle p, X_h^x - x\rangle\right) \leq H(x, p)$$

■

Let us recall the notion of continuous super-solution and sub-solution of the Hamilton-Jacobi equation

$$\frac{\partial \phi}{\partial t} = H(x, D_x \phi) \tag{32}$$

The main difficulty in the study of these equations comes from the fact that their solution $(t, x) \to \phi(t, x)$ may fail to be differentiable at each point (t, x).
It is therefore natural to introduce the notion of super-gradient $D^+\phi(t_0, x_0)$ and sub-gradient $D^-\phi(t_0, x_0)$ of a continuous function ϕ at (t_0, x_0)

$$\begin{aligned}
D^+\phi(t_0, x_0) &= \left\{(p, q) \in \mathbb{R} \times \mathbb{R}^d \, ; \, \limsup_{(t,x) \to (t_0, x_0)} \frac{\phi(t, x) - \phi(t_0, x_0) - p(t - t_0) - q.(x - x_0)}{(|t - t_0|^2 + |x - x_0|^2)^{1/2}} \leq 0\right\} \\
D^-\phi(t_0, x_0) &= \left\{(p, q) \in \mathbb{R} \times \mathbb{R}^d \, ; \, \liminf_{(t,x) \to (t_0, x_0)} \frac{\phi(t, x) - \phi(t_0, x_0) - p(t - t_0) - q.(x - x_0)}{(|t - t_0|^2 + |x - x_0|^2)^{1/2}} \geq 0\right\}
\end{aligned}$$

Definition 3.12 *A continuous function ϕ such that*

$$\forall (t_0, x_0) \quad \forall (p, q) \in D^+\phi(t_0, x_0) \qquad H(x_0, q) \geq p$$

is called a super-solution of the Hamilton-Jacobi equation (32). We say that ϕ is a sub-solution of (32) when

$$\forall (t_0, x_0) \quad \forall (p, q) \in D^-\phi(t_0, x_0) \qquad H(x_0, q) \leq p$$

ϕ is a solution of (32) if it is both a super-solution and sub-solution.

Theorem 3.13 *For any $\varphi \in \mathrm{Lip}(\mathbb{R}^d)$ the mapping*

$$\phi(t, x) = \mathbb{P}_t \varphi(x) \stackrel{\text{def}}{=} \mathbb{E}(\varphi(X_t^x))$$

belongs to $\mathrm{Lip}((0,1)\times\mathbb{R}^d)$ *and it is the unique bounded and uniformly continuous solution of the Hamilton-Jacobi equation*

$$\left\{\begin{array}{rcl}\dfrac{\partial\phi}{\partial t} & = & H(x, D_x\phi)\\ \phi(0,\centerdot) & = & \varphi(\centerdot)\end{array}\right. \tag{33}$$

In addition $\{\mathbb{P}_t\ ;\ t\geq 0\}$ *is a monotone semi-group of* $(\oplus,\odot)$*-linear operators on* $\mathrm{Lip}(\mathbb{R}^d)$*:*

1. $\forall 0\leq s\leq t\leq 1\quad \mathbb{P}_t\circ\mathbb{P}_s=\mathbb{P}_{t+s}\qquad \mathbb{P}_0=Id$
2. $\varphi_1\leq\varphi_2\Longrightarrow \mathbb{P}_t\varphi_1\leq\mathbb{P}_t\varphi_2\quad \forall t\in[0,1]$
3. *For any* $a_1,a_2\in\mathbb{R}_{\max}$ *and* $\varphi_1,\varphi_2\in\mathrm{Lip}(\mathbb{R}^d)$ *and* $t\in[0,1]$

$$\mathbb{P}_t\left(a_1\odot\varphi_1\oplus a_2\odot\varphi_2\right)=a_1\odot\mathbb{P}_t\left(\varphi_1\right)\oplus a_2\odot\mathbb{P}_t\left(\varphi_2\right)$$

Proof:
The proof of the theorem is essentially contained in [5]. Since the initial condition is Lipschitz the uniqueness of the solution in the class of bounded and uniformly continuous functions is a well known result. We now show that the tools presented before induce a simple proof of this theorem.
To prove that $\phi\in\mathrm{Lip}((0,1)\times\mathbb{R}^d)$ we use the fact that for any $\omega,\omega'\in\Omega$ and $x,y\in\mathbb{R}^d$

$$|X_t^x(\omega)-X_t^y(\omega')|\leq\ \mathrm{Cte(t)}\ |x-y| \tag{34}$$

The proof of this inequality is standard and it is a consequence of Gronwall's lemma. Using the above observation (31) and the classical inequality

$$\sup\{\ (1)\ \}-\sup\{\ (2)\ \}\leq\sup\{\ (1)-(2)\ \}$$

one obtain easily for any $s\leq t$

$$\mathbb{P}_t(\varphi)(x)-\mathbb{P}_s(\varphi)(y)\leq\sup\left\{\int_0^t\ l(X_r^x(u),u_r)\ dr-\int_0^s\ l(X_r^y(u),u_r)\ dr+\varphi(X_t^x(u))-\varphi(X_s^y(u))\right\}$$

where the supremum runs over all $u\in L^\infty((0,1),U)$. Under our Lipschitz assumptions on l and φ this implies that

$$\mathbb{P}_t(\varphi)(x)-\mathbb{P}_s(\varphi)(y)\leq\ \mathrm{Cte}\ \left(|t-s|+\int_0^s\ \|X_r^x(u)-X_r^y(u)\|\ dr+\|X_t^x(u)-X_t^y(u)\|\right)$$

Combining the above inequality with (34) we conclude that

$$\mathbb{P}_t(\varphi)(x)-\mathbb{P}_s(\varphi)(y)\leq\ \mathrm{Cte}\ (|t-s|+|x-y|)$$

This proves that $\phi \in \mathrm{Lip}((0,1)\times\mathbb{R}^d)$.
To prove that $\{\mathbb{P}_t \,;\, t \geq 0\}$ is a monotone semi-group of $(\oplus,\odot)$-linear operators on $\mathrm{Lip}(\mathbb{R}^d)$ it suffices to note that

$$\begin{aligned}
\mathbb{P}_{t+s}(\varphi)(x) &= \mathbb{E}(\varphi(X^x_{t+s})) \\
&= \mathbb{E}\left(\mathbb{E}(\varphi(X^x_{t+s})|X^x_s)\right) = \mathbb{E}\left(\mathbb{P}_t(X^x_s)\right) = \mathbb{P}_s\left(\mathbb{P}_t\varphi(x)\right)
\end{aligned}$$

To end the proof of the theorem it remains to check that $\phi(t,x) = \mathbb{P}_t(\varphi)(x)$ is solution of (33). First we note that

$$(p,q) \in D^+\phi(t,x) \Rightarrow \phi(t-h, X^x_h) - \phi(t,x) \leq -p\ h + q_\bullet(X^x_h - x) + o(h)$$

and

$$(p',q') \in D^-\phi(t,x) \Rightarrow \phi(t-h, X^x_h) - \phi(t,x) \geq -p'\ h + q'_\bullet(X^x_h - x) + o(h)$$

On the other hand we have

$$\phi(t,x) = \mathbb{E}\left(\mathbb{E}\left(\varphi(X^x_t)|X^x_h\right)\right) = \mathbb{E}\left(\phi(t-h, X^x_h)\right)$$

Hence,

$$\mathbb{E}\left(\phi(t-h, X^x_h) - \phi(t,x)\right) = 0$$

By definition of the super-gradient $(p,q) \in D^+$ and sub-gradient $(p',q') \in D^-\phi(t,x)$ one gets

$$\begin{aligned}
h^{-1}\ \mathbb{E}_x\left(\langle q, X^x_h - x\rangle\right) &\geq p + O(h) \\
h^{-1}\ \mathbb{E}_x\left(\langle q', X^x_h - x\rangle\right) &\leq p' + O(h)
\end{aligned}$$

Letting $h \to 0$ and using proposition 3.11 one concludes that

$$H(x,q) \geq p \qquad H(x,q') \leq p'$$

and the proof of the theorem is completed. ∎

By Rademacher's theorem every Lipschitz function is differentiable at almost all points (t,x). In view of the proof of the theorem, if ϕ is differentiable at (t,x) then we have

$$h^{-1}\ \mathbb{E}_x\left(\langle D_x\phi(t,x), X^x_h - x\rangle\right) = \frac{\partial\phi}{\partial t}(t,x) + O(h)$$

Letting $h \to 0$ it follows that $(t,x) \to \phi(t,x) = \mathbb{P}_t\varphi(x)$ satisfies (32) for almost all (t,x).

In order to point out the connections between these results and the classical viscosity solution we recall that the super-gradient and the sub-gradient of a continuous function ϕ can also be defined as follows

$$\begin{aligned} D^+\phi(t,x) &= \left\{ D\psi(t,x)\ ;\ \psi \in \mathcal{C}^1((0,1)\times\mathbb{R}^d) \ :\ (t,x) \text{ is a local maxima of } \phi-\psi \right\} \\ D^-\phi(t,x) &= \left\{ D\psi(t,x)\ ;\ \psi \in \mathcal{C}^1((0,1)\times\mathbb{R}^d) \ :\ (t,x) \text{ is a local minima of } \phi-\psi \right\} \end{aligned}$$

As a clear consequence of the above observation, ϕ is a super-solution of (32) if and only if for any (t,x) and $\psi \in \mathcal{C}^1$ such that (t,x) is a local maxima of $\phi - \psi$

$$H(x, D_x\psi(t,x)) \geq \frac{\partial \psi}{\partial t}(t,x)$$

If $\psi \in \mathcal{C}^1$ and (t,x) be a local maximum of $\phi-\psi$ then there exists an open ball $B((t,x),r)$, $r>0$, such that

$$\phi(s,y) - \psi(s,y) \leq \phi(t,x) - \psi(t,x) \qquad \forall (s,y) \in B((t,x),r)$$

Since

$$\psi(s,y) = \psi(t,x) + \frac{\partial \psi}{\partial t}(t,x)\,(s-t) + D_x\psi(t,x)\,(y-x) + o(|t-s|+|y-x|)$$

letting $s = t-h$ and $y = X_h^x(u)$, $u \in L^\infty$ this implies that

$$\phi(t-h, X_h^x(u)) - \phi(t,x) \leq \frac{\partial \psi}{\partial t}(t,x)\,(-h) + D_x\psi(t,x)\,(X_h^x(u) - x) + oh)$$

for sufficiently small $h>0$. Using the same line of arguments as before one conclude that

$$H(x, D_x\psi(t,x)) \geq \frac{\partial \psi}{\partial t}(t,x)$$

If the initial value condition of an Hamilton-Jacobi equation is not sufficiently regular theorem 3.13 is not applicable. More precisely solutions of Hamilton-Jacobi equations with discontinuous initial value are not always continuous and so we should not be able to apply the appoach presented above.

In order to carry our analysis to the next stage we note that the transition performance densities

$$\mathbb{P}^X_{t|0}(z|x) = \sup\left\{ \int_s^t\ l(X_r^x(u), u_r)\ dr\ ;\ u \in L^\infty\ :\ X_t^x(u) = z \right\}$$

generates two optimization operations. The first one acting on Lipschitz continuous function φ and the second one on performance densities μ

$$\begin{aligned}\mathbb{P}_t(\varphi)(x) &= \int^{\oplus} \varphi(z) \odot \mathbb{p}^X_{t|0}(z|x) \odot dz \\ (\mu\mathbb{P}_t)(z) &= \int^{\oplus} \mathbb{p}^X_{t|0}(z|x) \odot \mu(x) \odot dz\end{aligned}$$

We have already proved that the first operation gives a complete answer to the solution of the Hamilton-Jacobi equation (33) with initial value φ. The second operation can also be rewritten using the duality relation

$$< \mu, \mathbb{P}_t\varphi > = < \mu\mathbb{P}_t, \varphi > \qquad \forall \varphi \in \mathrm{Lip}(\mathbb{R}^d)$$

Now let us assume that $\mu \in \mathrm{Lip}(\mathbb{R}^r)$. In this situation we have

$$\mu\mathbb{P}_t(x) = \sup\left\{\mu(z) + \int_0^t l(X^z_s(u), u_s)\, ds \;;\; u \in L^\infty \;:\; X^z_t(u) = x\right\}$$

and therefore

$$\mu\mathbb{P}_t(x) = \sup\left\{\int_0^t l(\widetilde{X}^x_s(u), u_s)\, ds + \mu(\widetilde{X}^x_t(u)) \;;\; u \in L^\infty\right\}$$

where

$$\forall 0 \leq s \leq t \qquad \dot{\widetilde{X}}^x_s(u) = -f(\widetilde{X}^x_s(u), u_s) \qquad \widetilde{X}^x_0(u) = x$$

Theorem 3.13 clearly implies that $m(t,x) = \mu\mathbb{P}_t(x)$ is the unique bounded and uniformly continuous solution of the dual Hamilton-Jacobi equation

$$\left\{\begin{aligned}\frac{\partial m}{\partial t} &= H^{(\star)}(x, D_x m) \\ m(0, \cdot) &= \mu(\cdot)\end{aligned}\right. \tag{35}$$

where

$$H^{(\star)}(x,p) = \sup_{u \in U} (< p, -f(x,u) > + l(x,u)) = H(x, -p)$$

Definition 3.14 *For any performance density μ on $\mathbb{R}^d$ a mapping $(t,x) \to m_t(x) \in \mathbb{R}_{\max}$ such that for any $t \in (0,1)$ m_t is a performance density on $\mathbb{R}^d$ and $m_0(x) = \mu(x)$ and*

$$\forall \varphi \in \mathrm{Lip}(\mathbb{R}^d) \qquad < \varphi, m_t > = < \mu, \mathbb{P}_t\varphi >$$

is called a weak solution of the dual Hamilton-Jacobi equation (35).

The above definition and theorem 3.13 and theorem 1.2 lead to the following

Theorem 3.15 *For any performance density $\mu : \mathbb{R}^d \to \mathbb{R}_{\max}$ the mapping*

$$m(t,x) = \mu \mathbb{P}_t$$

is a performance density and it is the unique weak solution of the dual Hamilton-Jacobi equation

$$\begin{cases} \dfrac{\partial m}{\partial t} &= H^{(\star)}(x, D_x m) \\ m(0, \cdot) &= \mu(\cdot) \end{cases} \tag{36}$$

In addition $\{\mathbb{P}_t \ ; \ t \geq 0\}$ is a monotone semi-group of $(\oplus, \odot)$-linear operators on the set of performance densities:

1. $\forall 0 \leq s \leq t \leq 1 \quad \mathbb{P}_t \circ \mathbb{P}_s = \mathbb{P}_{t+s} \qquad \mathbb{P}_0 = Id$
2. $\mu_1 \leq \mu_2 \Longrightarrow \mu_1 \mathbb{P}_t \leq \mu_2 \mathbb{P}_t \quad \forall t \in [0,1]$
3. *For any $a_1, a_2 \in \mathbb{R}_{\max}$ such that $a_1 \oplus a_2 = \mathbb{1}$ and μ_1, μ_2 and $t \in [0,1]$*

$$(a_1 \odot \mu_1 \oplus a_2 \odot \mu_2)\, \mathbb{P}_t = a_1 \odot \mu_1 \mathbb{P}_t \oplus a_2 \odot \mu_2 \mathbb{P}_t$$

Our next objective is to connect the notion weak solution defined above with the classical viscosity solution of (36). We recall that a locally bounded upper-semi-continuous function $m(t,x)$ is a viscosity solution of (36) whenever

1. m is a super-solution of (36) i.e. for any (t,x) and $\psi \in \mathcal{C}^1$ such that (t,x) is a local maxima of $m - \psi$
$$H^{(\star)}(x, D_x \psi(t,x)) \geq \frac{\partial \psi}{\partial t}(t,x)$$
2. m is a sub-solution of (36) i.e. for any (t,x) and $\psi \in \mathcal{C}^1$ such that (t,x) is a local minima of $m - \psi$
$$H^{(\star)}(x, D_x \psi(t,x)) \leq \frac{\partial \psi}{\partial t}(t,x)$$

We will use the following conditions on the Hamiltonian function

(H) H is a convex function on $\mathbb{R}^d \times \mathbb{R}^d$ and a uniformly continuous function on $\mathbb{R}^d \times \overline{B}(0,R)$ for any $R > 0$ such that

$$\left| \frac{\partial H}{\partial x}(x,p) \right| \leq C\,(1 + |p|) \qquad \left| \frac{\partial H}{\partial p}(x,p) \right| \leq C$$

for some finite constant C.
Under these assumptions for any bounded performance density μ there exists a unique upper-semi-continuous solution of (36) with initial value μ (see for instance theorem 5.15 in [5]).
Let us check that in this case $(t,x) \to m(t,x) = \mu\mathbb{P}_t(x)$ is the desired solution.
Since μ is a bounded performance density there exists a non increasing sequence of uniformly bounded and Lipschitz functions $\{\mu^{(n)} \; ; \; n \geq 1\}$ such that

$$\mu = \inf_{n\geq 1} \mu^{(n)}$$

Letting $m^{(n)}(t,x) = \mu^{(n)}\mathbb{P}_t(x)$ we clearly have

$$m(t,x) = \mu\mathbb{P}_t(x) \leq \inf_{n\geq 1} m^{(n)}(t,x)$$

Now, let $\{\omega^n \; ; \; n \geq 1\}$ be a sequence of relaxed controls such that

$$m^{(n)}(t,x) = \int_0^t \; L(\widetilde{X}_s^x(\omega^n), \omega_s^n) \; ds + \mu^{(n)}(\widetilde{X}_t^x(\omega^n))$$

Under our assumptions there exists a subsequence $\{\omega^{n_k} \; ; \; k \geq 1\}$ such that ω^{n_k} converges to some relaxed control ω and $\widetilde{X}^x(\omega^{n_k})$ converges to $\widetilde{X}^x(\omega)$ and

$$\int_0^t \; L(\widetilde{X}_s^x(\omega^{n_k}), \omega_s^{n_k}) \; ds \longrightarrow \int_0^t \; L(\widetilde{X}_s^x(\omega), \omega_s) \; ds$$

as $k \to \infty$. This yields

$$\inf_{n\geq 1} m^{(n)}(t,x) \leq \int_0^t \; L(\widetilde{X}_s^x(\omega), \omega_s) \; ds + \limsup_{k\to\infty} \mu^{n_k}(\widetilde{X}_t^x(\omega^{n_k}))$$

Using the fact that $\{\mu_{n_k} \; ; \; k \geq 1\}$ is a non increasing sequence of uniformly bounded and continuous functions and μ is upper-semi-continuous one conclude that

$$\inf_{n\geq 1} m^{(n)}(t,x) \leq \int_0^t \; L(\widetilde{X}_s^x(\omega), \omega_s) \; ds + \mu(\widetilde{X}_t^x(\omega))$$

It is now not difficult to check that

$$m(t,x) = \mu\mathbb{P}_t(x) = \inf_{n\geq 1} m^{(n)}(t,x) = \inf_{n\geq 1} (\mu^{(n)}\mathbb{P}_t(x))$$

Since each $m^{(n)}$ is solution of the Hamilton-Jacobi equation (36), m is again solution of (36) (see for instance theorems 5.13 and 5.15 in [5]).

3.4 Limit Theorems

Our aim is now to study some limit theorems relating to the limiting behavior of the solution of the Hamilton-Jacobi equation (36) as $t \to \infty$. We will only consider a very special class of equations, the extension to more general situations is still an open subject of investigation. The weak law of large numbers and the central limit theorem presented hereafter are the time continuous version of those presented in [34] and [13].
In a first stage we will show how the preceding results apply to obtain the so-called Oleinik-Lax explicit formula. Then we use this formula to study the limiting behavior of these explicit solutions as $t \to \infty$.

Let us suppose that the Lagrangian function $l(x, u)$ doesn't depend on the first parameter and it is a convex function on U. For any $x \in \mathbb{R}^d$ let X^x be the Bellman-Maslov process simply defined by the equation

$$X^x(u)_t = x + \int_0^t u_s \, ds \qquad u \in L^\infty((0,1); U)$$

where U is a compact subset of $\mathbb{R}^d$. We first note that in this simple example the Hamiltonian function $H(x, p)$ does'nt depend on the first parameter and it is the convex function given by

$$H(p) = \sup_{u \in U} (< p, u > + l(u)) = (-l)^\star(p)$$

where $(\,.\,)^\star$ stands for the Fenchel transform of a convex function. By the involutive property of the Fenchel transform we also have

$$H^\star = (-l)^{\star\star} = -l$$

Then, for any $\varphi \in \mathrm{Lip}((0,1) \times \mathbb{R}^d)$ the function

$$\phi(t, x) = \mathbb{P}_t \varphi(x)$$

can be rewritten in the form

$$\phi(t, x) = \sup \left\{ - \int_0^t H^\star(u_s) \, ds + \varphi(X_t^x(u)) \; ; \; u \in L^\infty((0,1); U) \right\}$$

and hence

$$\phi(t, x) = \sup_{z \in \mathbb{R}^d} \left(\varphi(z) + \mathbb{p}_{t|0}^X(z|x) \right)$$

where

$$\mathbb{p}_{t|0}^X(z|x) = \sup \left\{ - \int_0^t H^\star(u_s) \, ds \; ; \; u \in L^\infty((0,1); U) \; : \; X_t^x(u) = z \right\}$$

If $z \notin x + t\,U$ we clearly have $\mathbb{p}^X_{t|0}(z|x) = -\infty$. Now let us suppose that $z \in x + t\,U$. Letting

$$u(s) = \frac{z-x}{t} \qquad \forall 0 \le s \le t$$

we obtain $X^x_t(u) = z$ and therefore

$$\mathbb{p}^X_{t|0}(z|x) \ge -t\ H^\star\left(\frac{z-x}{t}\right) \tag{37}$$

On the other hand, using the fact that $H^\star$ is convex

$$\int_0^t H^\star(u_s)\ ds \ge t\ H^\star\left(\frac{1}{t}\int_0^t u_s\,ds\right) \ge t\ H^\star\left(\frac{z-x}{t}\right)$$

as soon as $u(\boldsymbol{.})$ is chosen so that $X^x_t(u) = z$. This together with (37) gives

$$\mathbb{p}^X_{t|0}(z|x) = \begin{cases} -t\ H^\star\left(\dfrac{z-x}{t}\right) & \text{if } z \in x + t\,U \\ -\infty & \text{otherwise} \end{cases}$$

one concludes that

$$\mathbb{P}_t\varphi(x) = \sup_{z\in x+t\,U}\left(\varphi(z) - t\ H^\star\left(\frac{z-x}{t}\right)\right) \qquad \text{(Oleinik-Lax Formula)}$$

and for any performance density μ on $\mathbb{R}^d$

$$\mu\mathbb{P}_t(z) = \sup_{x\,:\,z-x\in t\,U}\left(\mu(x) - t\ H^\star\left(\frac{z-x}{t}\right)\right)$$

Let us remark that we can replace in the above construction the time space $(0,1)$ by the interval $(0,T)$, $T>0$. Moreover if we choose the initial value $x=0$ and we define

$$X_t(u) = \int_0^t u_s\ ds$$

then the performance density μ_t of X_t is given by

$$\mu_t(x) = \begin{cases} -t\ H^\star\left(\dfrac{x}{t}\right) & \text{if } x \in t\,U \\ -\infty & \text{otherwise} \end{cases} \tag{38}$$

Proposition 3.16 (Weak Law of Large Numbers)
If $H^\star : U \to \mathbb{R}^+$ is a convex function and has a unique minimum $X^\star$ such that $H^\star(X^\star) = 0$ then

$$\frac{X_t}{t} \xrightarrow[t\to\infty]{w} X^\star$$

in the sense that the performance measures of X_t/t weakly converges to $\delta_{X^\star}$ as $t\to\infty$

Proof:
Let m_t be the performance density of the optimization variable X_t/t. In view of (38) we have that

$$m_t(x) = \begin{cases} -t\ H^\star(x) & \text{if } x \in U \\ -\infty & \text{otherwise} \end{cases}$$

For any bounded and continuous test function φ on $\mathbb{R}^d$ we clearly have

$$< \varphi, m_t >= \sup_{x \in U} (\varphi(x) - t\ H^\star(x)) \xrightarrow[t \to \infty]{\text{w}} \varphi(X^\star) =< \varphi, \delta_{X^\star} >$$

■

Proposition 3.17 (Central Limit Theorem)
Suppose the assumptions of proposition 3.16 hold with $X^\star = 0$ and $H^\star$ is a twice differentiable function on U such that

$$(H^\star)''(0) > 0 \qquad \text{and} \qquad \|H^\star(u) - H^\star(v)\| \leq C\ |u - v|^\alpha$$

for some $C < \infty$ and $\alpha > 0$. Then we have

$$\frac{X_t}{\sqrt{t}} \xrightarrow[t \to \infty]{w} Z$$

where Z is a quadratic optimization variable with performance density

$$\nu(z) = -\frac{1}{2}\, z'\ (H^\star)''(0)\, z$$

Proof:
Let m_t be the performance density of the optimization variable $X_t/\sqrt{t}$. In view of (38) we have that

$$m_t(x) = \begin{cases} -t\ H^\star\left(x/\sqrt{t}\right) & \text{if } x/\sqrt{t} \in U \\ -\infty & \text{otherwise} \end{cases}$$

For any bounded and continuous test function φ with compact support K on $\mathbb{R}^d$ we clearly have

$$< \varphi, m_t >= \sup_{x \in K} \left(\varphi(x) - t\ H^\star\left(x/\sqrt{t}\right)\right)$$

Under our assumptions for any x there exists a constant $\lambda \in [0,1]$ such that

$$H^\star\left(x/\sqrt{t}\right) = \frac{1}{2t}\, x' H^\star\left(\lambda\, x/\sqrt{t}\right)\, x$$

This yields that

$$< \varphi, m_t >= \sup_{x \in K}\left(\varphi(x) - \frac{1}{2}\, x' H^\star\left(\lambda\, x/\sqrt{t}\right)\, x\right)$$

and therefore

$$| < \varphi, m_t > - < \varphi, \nu > | \leq \sup_{x \in K} -\frac{1}{2}\, x' \| H^\star\left(\lambda\, x/\sqrt{t}\right) - H^\star(0) \|\, x \leq \mathrm{Cte}\; t^{-\alpha/2}$$

the end of the proof is now straighforward.

∎

3.5 Applications to Mathematical Morphology

Our aim is now to discuss the connections between Maslov's optimization theory and Mathematical Morphology. More precisely we will now show how the preceding framework provides a natural analytic basis for studying the evolution of the so-called morphological dilations. These transformations model the rate of change of a given signal under a continuous multiscale dilation (see [8, 23]). Mathematical Morphology is a particular discipline in Image processing. In this framework the image is model as a function $\mu : \mathbb{R}^d \to \mathbb{R}$. The multiscale morphological dilation of the signal μ is defined as a space-scale function which represents at each scale $t \in \mathbb{R}_+$ the transformation of the signal by a given structuring function $g : U \to \mathbb{R}$ were U is a compact subset of $\mathbb{R}^d$.
This transformation is defined as follows

$$D_t(\mu)(x) = \sup_{z\,:\,x-z \in tU}\left(\mu(z) - t\ g\left(\frac{x-z}{t}\right)\right) \tag{39}$$

When $g(U) = \{0\}$ the structuring function is said to be flat and the dilation transformation takes the form

$$D_t(\mu)(x) = \sup_{z \in tU} \mu(x-z)$$

In this particular situation the transformation is called a multiscale dilation of the signal μ by the structuring element U.

Let us assume that μ is a performance density on $\mathbb{R}^d$ and $g : U \to \mathbb{R}_+$ is a convex structuring function. Using the preceding results on conclude that the mapping $((t,x) \to D_t(\mu)(x) = \mu\mathbb{P}_t(x)$ is a generalized solution of the Hamilton-Jacobi equation

$$\left\{ \begin{array}{rcl} \dfrac{\partial m}{\partial t} & = & g^\star(-D_x m) \\ m(0,\centerdot) & = & \mu \end{array} \right. \tag{40}$$

Workers on image processing have attempted to make sense out of the scale evolution associated with the continuous-space mutliscale morphological dilation (39). We note that the idempotent analysis framework is a rather natural way to think about such transformations and caracterizes their time evolution in terms of Hamilton-Jacobi equations.
This new application of idempotent analysis also provides an alternative approach to the Beucher's morphological gradient or the sup-inf derivative notions presented in [8]. Moreover, in contrast to the classical works on the subject we do not need to assume that the compact set U has smooth boundaries or that the structuring function g is strickly convex and/or have an inverse gradient everywhere [38].

To make the above ideas firm let us present some structuring functions arising in this framework.

Example 12 *If $g(U) = \{0\}$ and $U = \{u \in \mathbb{R}^d \ : \ \|u\|_\alpha \leq 1\}$ for some $\alpha \in [1,\infty]$ then using the classical formula*

$$\sup_{\|u\|_\alpha \leq 1} < q, u >= \|q\|_\beta \qquad \frac{1}{\beta} + \frac{1}{\alpha} = 1$$

we arrive at $g^\star(q) = \|q\|_\beta$ and therefore the equation (40) has the form

$$\frac{\partial m}{\partial t} = \|D_x m\|_\beta$$

For instance we have

1. *If $\alpha = 1, \beta = \infty$ then (40) takes the form*

$$\frac{\partial m}{\partial t} = \bigoplus_{1 \leq i \leq d} \left| \frac{\partial m}{\partial x_i} \right|$$

2. *If $\alpha = \infty, \beta = 1$ then (40) takes the form*

$$\frac{\partial m}{\partial t} = \sum_{1 \leq i \leq d} \left| \frac{\partial m}{\partial x_i} \right|$$

3. If $\beta \geq 1$ then (40) takes the form

$$\left(\frac{\partial m}{\partial t}\right)^{\beta} = \sum_{1 \leq i \leq d} \left|\frac{\partial m}{\partial x_i}\right|^{\beta}$$

Example 13 *If the structuring function g is given by*

$$g : u \in U \to g(u) = g(0)\,(1 - \|u\|_{\alpha}^{\alpha})^{1/\alpha}$$

where $g(0) \geq 0$ and $U = \{u \in \mathbb{R}^d \ : \ \|u\|_{\alpha} \leq 1\}$ for some $\alpha \in [1, \infty]$ then in the same way as before we arrive at

$$g^{\star}(q) = \|(g(0), q)\|_{\beta} = \left(g(0)^{\beta} + \|q\|_{\beta}^{\beta}\right)^{1/\beta} \qquad \frac{1}{\beta} + \frac{1}{\alpha} = 1$$

and therefore the equation (40) has the form

$$\frac{\partial m}{\partial t} = \|(g(0), D_x m)\|_{\beta}$$

For instance we have

1. If $\alpha = 1, \beta = \infty$ then (40) takes the form

$$\frac{\partial m}{\partial t} = g(0) \oplus \bigoplus_{1 \leq i \leq d} \left|\frac{\partial m}{\partial x_i}\right|$$

2. If $\alpha = \infty, \beta = 1$ then (40) takes the form

$$\frac{\partial m}{\partial t} = g(0) + \sum_{1 \leq i \leq d} \left|\frac{\partial m}{\partial x_i}\right|$$

3. If $\beta \geq 1$ then (40) takes the form

$$\left(\frac{\partial m}{\partial t}\right)^{\beta} = g(0)^{\beta} + \sum_{1 \leq i \leq d} \left|\frac{\partial m}{\partial x_i}\right|^{\beta}$$

References

[1] M. Akian, J.P. Quadrat, M. Viot, *Bellman Processes* $11^{\oplus eme}$ Conférence Internationale sur l'Analyse et l'Optimisation des Systèmes. Ecole des Mine s Sophia-Antipolis, France, 15-16-17 Juin 1994. Lecture Note in Control and Information Sciences 199. Springer-Verlag.

[2] M. Akian, J.P. Quadrat, M. Viot, *Duality between Probability and optimization*, Idempotency, Cambridge University Press, Editor J. Gunawardena, Publications of the Newton Institute, 1998.

[3] R. Azencott, *Grandres déviations et applications*, Ecoles d'Eté de Probabilités de Saint Flour VIII-1978", Edited by P.L. Hennequin. Lecture Notes in Mathematics 774, Springer, Berlin, pp. 1-176, 1980.

[4] F. Bacelli, G. Cohen, G. J. Olsder, J. P. Quadrat, *Synchronization and Linearity, An Algebra for Discrete Event Systems*, John Wiley and Sons, 1992

[5] G. Barles, *Solutions de viscosité des équations de Hamilton-Jacobi*, Mathématiques et Applications, Vol 17, SMAI, Springer Verlag, 1994.

[6] R. E. Bellman, S. E. Dreyfus, *Dynamic Programing and Applications*, Dunod, Paris, 1965.

[7] R. Bott, J.P. Mayberry, *Matrices and trees*, Economics Activity Analysis, Wiley, New York, 1954.

[8] R.W. Brockett, P. Maragos, *Evolution Equations for Continuous-Scale Morphological Filtering*, IEEE Trans. on Signal Processing, Vol. 42, No 12, December 1994.

[9] Cox H., *Estimation of state variables via dynamic programming*, Proc. 1964 Joint Automatic Control Conf., pp. 376–381, Stanford, California.

[10] M.G. Crandall, H. Ishii, P.L. Lions, *User's guide to viscosity solutions of second order partial differential equations*, Bull. Amer. Soc. 27, 1992, PP. 1-67.

[11] M.G. Crandall, P.L. Lions, *Viscosity solutions of Hamilton-Jacobi equations*, Trans. Amer. Math. Soc. 277, 1983, PP. 1-42.

[12] R. Cuninghame-Green, *Minimax Algebra*, Lectures Notes in Economics and Mathematical Systems, Springer-Verlag, 166, 1979.

[13] P. Del Moral, *Maslov Optimization Theory: Optimality versus Randomness.* in *Idempotency Analysis and its Applications*, Eds V. Kolokoltsov and Victor P. Maslov, (1997), Kluwer Publishers.

[14] P. Del Moral, *On the stability of Maslov optimization processes* 34th Conference on Decision and control, December 1995.

[15] P. Del Moral *Maslov optimization theory, topological aspects* Idempotency, Cambridge University Press, Editor J. Gunawardena, Publications of the Newton Institute, 1998.

[16] P. Del Moral, G. Salut, *Random Particle Methods in (max,+)-optimization problems.* Idempotency, Cambridge University Press, Editor J. Gunawardena, Publications of the Newton Institute, 1998.

[17] P. Del Moral, J. C. Noyer, G. Salut, *Maslov Optimization Theory*: *Stochastic Interpretation, Particle Resolution*, $11^{\oplus eme}$ Conférence Internationale sur l'Analyse et l'Optimisation des Systèmes. Ecole des Mine s Sophia-Antipolis, France, 15-16-17 Juin 1994. Lecture Note in Control and Information Sciences 199. Springer-Verlag.

[18] P. Del Moral, L. Miclo, *On the Convergence and the Applications of the Generalized Simulated Annealing*, Publications du Laboratoire de Statistiques et Probabilités, Université Paul Sabatier 1996.

[19] A. Dembo, O. Zeitouni, *Large deviations techniques and Applications.* Jones and Bartlett, 1993.

[20] J.D. Deuschel, D.W. Stroock, *Larges deviations. Academic Press.* 1989.

[21] P. I. Dudnikov, S. N. Samborskii, *Endomorphisms of semimodules over semirings with an idempotent operation*, Preprint N^o 87-48, Inst. Mat. Akad. Nauk Ukrain., Kiev, 1987 (Russian).

[22] M.I. Freidlin, A.D. Wentzell, *Random Perturbations of Dynamical Systems*, Springer Verlag, 1984.

[23] H.J.A. Heijmans, *The Algebraic Basis of Mathematical Morphology. I. Dilations and Erosions*, Computer Vision, Graphics, and Image Processing, 50, 245-295 (1990).

[24] V. N. Kolokoltsov, *Semiring analogs of linear equivalent spaces*, Abstr. Conf. on Optimal Control, Geometry, Kemerovo, 1988, p. 30 (Russian).

[25] V. N. Kolokoltsov Proceeedings of the Idempotency workshop, Hewlett Packard Laboratories, Bristol 3-7 october 1994.

[26] V. N. Kolokoltsov, V. Maslov, *Idempotent calculus as the apparatus of optimization theory.* I. Funktsional. Anal. i Prilozhen. 23 (1989), no. 1, 1-14; II, 23, no. 4, 53-62; English transl. in Functional Anal. Appl. 23 (1989).

[27] V. N. Kolokoltsov, V. P. Maslov, *The general form of endomorphisms in the space of continuous functions with values in a numerical semiring with idempotent addition*, Dokl. Akad. Nauk SSSR 295 (1987), N^o 2, 283-287; English transl. Soviet Math. Dokl. 36 (1988), N^o 1, 55-59.

[28] G. Litvinov, V.P. Maslov *Correspondence principle for idempotent calculus and some computer applications.* Institut des Hautes Etudes Scientifiques, IHES-M-95-33.

[29] V. Maslov, *Méthodes opératorielles*, Edition Mir, 1987.

[30] V. Maslov, *Quasilinear systems which are linear in some semimoduli* Congrès international sur les problèmes hyperboliques, 13-17 Janvier 1986, Saint Etienne, France.

[31] V. Maslov, S. N. Samborskii, *Idempotent Analysis*, Advances in Soviet Mathematics, 13, Amer. Math Soc., Providence (1992).

[32] V. Maslov, S. N. Samborskii, *Stationary Hamilton–Jacobi and Bellman equations*, Advances in Soviet Mathematics 13 AMS (1992) 119-133.

[33] L. Miclo, *Recuit simulé sans potentiel sur un ensemble fini*, in Séminaire de Probabilités, Lecture Notes in Mathematics 1526, 47-60, Springer Verlag, 1992.

[34] J. P. Quadrat, *Théorèmes asymptotiques en programmation dynamique*, Comptes Rendus à l'Académie des Sciences, 311: 745-748, 1990.

[35] R. T. Rockafellar, *Convex Analysis*, Princeton University, 1970.

[36] R. T. Rockafellar, *Convex functions and duality in optimization problems and dynamics*, Mathematical Systems Theory and Economics, I. H.W. Kuhn, G.P. Szergo editors, Lecture Notes in ORME, Springer-Verlag, 1969.

[37] A. Trouvé, *Parallélisation massive du recuit simulé*, Thèse de Doctorat, Université Paris XI, Janvier 1993.

[38] R. Van der Boomgaard, *Mathematical Morphology : Extensions towards computer vision*, Ph. D. Thesis, Univ. of Amsterdam, 1992.