Sci., 28:277–294, 1994.

214. L. E. Stern. Criteria of optimality in the infinite-time optimal control problem. *J. Optimiz. Theory Appl.*, 44(3):497–508, 1984.

215. S. Sternberg. Local contractions and a theorem of Poincaré. *Amer. J. Math.*, 79:809–824, 1957.

216. A. I. Subbotin. *Min-Max solutions and Hamilton–Jacobi Equation* [in Russian]. Nauka, Moscow, 1991.

217. E. Titchmarsh. *Eigenfunction Expansions Associated with Second Order Ordinary Differential Equations, Part 2.* Oxford Univ. Press., Oxford, 1962.

218. A. Truman and H. Z. Zhao. The stochastic Hamilton–Jacobi equation, stochastic heat equations, and Schrödinger equations. Preprint, University College of Swansea, 1994.

219. S. R. Varadan. Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math*, 19(3):231–281, 1966.

220. V. V. Voevodin. *Parallel Structures in Algorithms and Programs.* Otd. Vychisl. Mat. Akad. Nauk SSSR, Moscow, 1987.

221. V. V. Voevodin. *Mathematical Principles of Parallel Computations.* Moscow State University, Moscow, 1991.

222. C. C. von Weizgacker. Existence of optimal programs of accumulation for an infinite horizon. *Rev. Econom. Stud.*, 32:85–104, 1965.

223. E. Wagneur. Moduloids and pseudomodules 3. The lattice structure problems. In *Analysis and optimization of systems, Proceedings of the 9th Intern. Conf.*, number 144 in Lect. Notes in Contr. and Inform. Sci., pages 977–984, Antibes, June [illegible]. Springer-Verlag.

224. E. Wagneur. Moduloids and pseudomodules 1. Dimension theory. *Discrete Mathematics*, 98:57–73, 1991.

225. E. Wagneur. Subdirect sum decomposition of finite-dimensional pseudomodules. In *11th Intern. Confer. on Analysis and Optimization of Systems* [41], pages 322–[illegible].

226. E. Wagneur and C. Sriskandarajan. Optimal control of a class of DEDs: Flow-shop with state-dependent processing times. *Discrete Event Dynamical Systems: Theory and Applications*, 3:397–425, 1993.

227. X. P. Wang. Puits multiples pour l'operateur de Dirac. *Ann. Inst. H. Poincaré Sec. A*, 43(3):260–319, 1985.

228. M. Weitzman. Duality theory for infinite horizon convex models. *Management Sci.*, 19(7):783–789, 1973.

229. M. Wilkinson. Tunneling between tori in phase space. *J. Phys. D*, 21:341–354, 1986.

230. M. Wilkinson and J. H. Hanay. Multidimensional tunneling between excited states. *Phys. D*, 27:201–212, 1987.

231. E. Witten. Supersymmetry and Morse theory. *J. Diff. Geom.*, 17:661–692, 1982.

232. S. Y. Yakovenko. On the notion of infinite extremal in stationary problems of dynamic optimization. *Dokl. Akad. Nauk SSSR.*, 308(4):758–802, 1989.

233. S. Y. Yakovenko and L. A. Kontorer. Nonlinear semigroups and infinite horizon optimization. In *Idempotent Analysis* [165], pages 167–210.

234. U. Zimmermann. Duality for algebraic linear programming. *Linear Algebra and its Applications*, 32:9–31, 1980.

235. U. Zimmermann. Linear and combinatorial optimization in ordered algebraic structures. *Ann. Discrete Math*, 10:1–380, 1981.

Appendix. Maslov Optimization Theory. Optimality versus Randomness

by Pierre Del Moral

Abstract: We explore some of the fascinating links that exist between stochastic processes and optimization processes. We exhibit a common algebraic framework which yields the Markov generator and the Chapman–Kolmogorov equation on the one hand and the Hamiltonian and the Bellman optimization principles on the other hand, owing to the existence of algebraic morphisms between the two theories. The key point is in using the idempotent measure/integration theory due to Maslov. This makes the apparently nonlinear equation of optimization processes a linear operator in a suitable algebra. This correspondence principle for idempotent calculus was described by Litvinov and Maslov [33]. We derive some properties of optimization martingales, such as the (max,+)-version of the Doob up-down crossing lemma. This work offers an alternative to the classical representations of optimization problems and leads to new developments in the field of their qualitative study [39]. We code the information in optimization problems in a way compatible with that in which we shall treat conditional probability. We introduce some transformations between performance and probability measures, which clarify the relationships between optimization and estimation problems [19, 40]. Having in mind Monte-Carlo principles, we discuss a particle interpretation of nonlinear filtering and optimization. In that algorithm, the trajectories that are generated describe the entire probability/performance space in such a way that arbitrariness is excluded compared with procedures such as simulated annealing. We set the formulation of the whole problem and exhibit minimal hypotheses leading to a time-uniform convergence of our particle algorithms. For more detailed proofs and various extensions, see [13, 38, 39].

Introduction

Duality between control and estimation is a familiar concept in the theory of linear systems but rises a lot more problems in the nonlinear field. The well-known relationship between diffusion and optimal stochastic control with quadratic cost does not really pertain to optimization and heavily relies on quadratic properties. It has also been early recognized [7, 25] that the maximum likelihood estimation is related to an optimal control problem in which the cost functional is taken to be the likelihood density associated with the trajectories. But this does not explain the intriguing links that persist between optimization and mean-value stochastic evolution, although this kind of estimator is quite different from the former in the general case. As is shown in the sequel, the explanation is to be found at a higher level of abstraction.

Using Maslov idempotent measure/integration theory, we derive an optimization theory at the same level of generality as probability and the theory of stochastic processes. Most of these results are taken from the thesis [13], but they were presented for the first time in [15]. Also, preliminary work along the same lines appeared in [26]. The first section recalls the basic principles of Maslov idempotent measure theory, which will be used in the forthcoming development. In §2 we introduce a general performance theory at the same level of generality as in probability theory. The concepts of performance measures, optimization variables, and independence are the basic notions here. The concept of performance is the achievement of deductive reasoning in which we estimate the cost of some control event. The independence between optimization variables means that the performance of any variable is not affected by the performance of the other variables. As in probability theory, these concepts are fundamental; in fact, they justify the mathematical development of Maslov's performance theory as a separate discipline rather than merely a topic in idempotent analysis. In §3 we introduce an analog of Lebesgue spaces and Markov, Minkowski, and Hölder inequalities. Then we deal in §4 with several convergence modes of optimization variables. From the above concepts, some classical asymptotic theorems (e.g., the law of large numbers and the central limit theorem) can readily be derived as in [13, 44]. Another important step is to introduce sequences of optimization variables indexed by a subset of reals. This leads to a rich additional structure and deeper results. In §5 we provide a new way of studying optimization problems in terms of performance measure, without any of the usual geometric descriptions. We introduce the notion of a modal semimartingale and provide a general formulation for the evolution of the optimal process. We also state that the Markov causality principle in this theory is the same as the Bellman optimality principle. Moreover, just as in stochastic theory, the infinitesimal generator of the so-called Maslov processes is the Hamiltonian. We also derive some optimization martingale properties, such as the (max, +)-version of the Doob up-down crossing lemma and Dynkin's formula, which lead to new developments in the field of qualitative study of optimization processes. In §6 we briefly review some nice consequences of these results (see [13] and [18] for details). In particular, we demonstrate an explicit advantage of the performance function over some classes of controls. We finally describe the conditional performance evolution of optimal regulation problems as that of filtering theory in idempotent algebra.

Certain transformations between performance and probability measures, which clarify the relationships between optimization and estimation problems, are developed in §§7 and 8. These transformations lead to useful conclusions, because they make the links between the performance and the probability measure of an event obvious. One purpose of §8 is to further develop the Log-Exp transform so as to characterize the filtering problem associated with these optimization problems more completely. We show that there is a bijective correspondence between Maslov performance and optimization problems, on one hand, and Markov probabilities and filtering problems, on the other hand. This stochastic interpretation of optimization problems gives strong basis for transposing the recently developed particle procedures for nonlinear filtering to optimal regulation problems, because these procedures make the links between the performance and the probability measure of an event explicit.

Particle methods have been developed in physics since World War II, mainly for the need of fluid mechanics. They provide the most powerful approach to the numerical solution of infinite-dimensional problems where nonlinearities do not allow the use of analysis and nonstationarities do not allow the application of fixed discretization schemes. This is just the case in optimal nonlinear filtering and optimization problems, which have so far received little attention from numerical analysts. The pioneering work on filtering problems appeared in [48] as well as in [3, 32, 51]. Solving the fundamental equation of nonlinear filtering is an enormous task, since the equation is infinite-dimensional, nonlinear, and stochastic. As pointed out by Bucy in his early papers [4, 5, 6], the progress in nonlinear filtering lies in numerical studies of the so-called "realization problem," particularly with the help of parallel computing. As all particle methods, they are based on a Dirac comb representation of the performance/probability measure at stake, but the "teeth" of this comb depend on the flow of the system and its partial observations or the reference path, both in mass and position.

This method, which has revealed its efficiency in RADAR and G.P.S signal processing [11, 12, 15], can be used in general for nonlinear filtering and optimization problems for discrete-event dynamical systems such as the determination of some communication network or manufacturing system from some partial observations or from some reference values.

In §9 we briefly review some basic facts about Monte-Carlo principles and the forthcoming developments (see [19]). We show that these principles are a powerful tool to study the conditional mean of a random variable as well as the conditional optimal state of an optimization variable. The so-called particle interpretations of optimization and filtering problems are global methods of investigation. This section also contains the key result which allows one to transfer the particle procedure for nonlinear filtering to optimization problems. In §10 we introduce some recursive distributions, which can be used to explore the performance/probability space, and the weights used in the algorithm. Complementary to the exploration distribution, these weights are related to the likelihood of each exploration path among N particles. It is important to notice that they are time-degenerate, since individual paths have divergent likelihood. As we shall see, the degeneracy of Bayesian weights is eliminated by a regularization of the problem. Moreover, from a practical point of view, it is necessary to study the time asymptotic behavior.

Section 11 constitutes the last step on our way to particle procedures. It is centered around the time-uniform convergence of these approximations, and to make it as self-contained as possible, we compare the sufficient conditions which guarantee time-uniform convergence of the particle filter and the optimization particle estimate.

I wish to thank the Russian Academy of Science and in particular Victor P. Maslov, Grigori L. Litvinov, and Vladimir E. Nazaikinskii for their warm hospitality during my stay in Moscow.

1. Maslov's Integration Theory

In this section we introduce Maslov's integration theory and the semiring of reference [34] and embed these notions into a Lebesgue integration theory. We recall the one-to-one correspondence between the space of bounded Maslov measures and a space of continuous linear forms, which leads to the topology of wide convergence. The elements of idempotent analysis (analysis of functions with values in a general idempotent semiring) are developed in [21, 29, 30, 31, 34, 35, 36]. Let $\mathcal{A} \stackrel{\text{def}}{=} [-\infty, +\infty[$ be the semiring of real numbers endowed with the commutative semigroup laws $\oplus$ and $\odot$, the neutral elements $\mathbb{0}$ and $\mathbb{1}$, and the exponential metric ρ such that

$$\mathbb{0} = -\infty, \quad a \oplus b = \sup(a,b), \quad \rho(a,b) = |e^a - e^b|,$$
$$\mathbb{1} = 0, \quad a \odot b = a + b, \quad \frac{a}{b}^{\circ} = a - b \quad (b \neq \mathbb{0}).$$

By $\rho(a,b) = \sup_{1 \le i \le n} |e^{a_i} - e^{b_i}|$ we denote the exponential metric on $\mathcal{A}^n$. We specially indicate a notion that will be used throughout the study of the $(\max,+)$-version of the Lebesgue spaces. By $d(x,y) = \log \rho(x,y)$ we denote the logarithm of the exponential metric ρ. This mapping is characterized by the following properties:

$$d(x,y) = \mathbb{0} \iff x = y, \qquad d(x,y) = d(y,x),$$
$$d(x,y) \le \log(\exp d(x,z) \odot \exp d(z,y)).$$

More generally, if $\mathbf{L}$ is a $(\oplus,\odot)$-semimodule, then every mapping d of $\mathbf{L} \times \mathbf{L}$ into $\mathcal{A}$ satisfying these three conditions is called a $\oplus$-metric. Consider a Polish space (Ω, σ). Let us introduce some standard notation to be used throughout the paper. Let $(\mathcal{S}, \cup, \cap)$ be a semiring of sets in Ω. A Maslov

measure μ on $(\Omega, \mathcal{S})$ is a mapping of $\mathcal{S}$ into $\mathcal{A}$ such that $\mu(\emptyset) = \mathbb{0}$ and, for any disjoint sets $A, B \in \mathcal{S}$, $\mu(A \cup B) = \mu(A) \oplus \mu(B)$. It is said to be bounded whenever $\mu(\Omega) < +\infty$. These measures fail to be continuous on the empty set and they have a nonunique extension to the σ-algebra, but there is a unique maximal extension [13, 34]. *In order to construct a measure theory which permits obtaining limit results, we always assume the Maslov measure μ be defined and known on the σ-algebra.* Next, we consider the Lebesgue–Maslov integral of a measurable function with respect to a Maslov measure. We first introduce some notation. Let μ be a Maslov measure on (Ω, σ), and let $\mathcal{L}^0(\Omega, \sigma)$, (respectively, $\mathcal{E}(\Omega, \sigma)$) be the semiring of $\mathcal{A}$-valued measurable functions (respectively, step functions). As usual, we write

$$\int_\Omega^{\oplus} f_e \odot \mu = \int_\Omega^{\oplus} f_e(\omega) \odot \mu(d\omega) \stackrel{\text{def}}{=} \bigoplus_{d \in f_e(\Omega)} d \odot \mu(f_e^{-1}(\{d\})), \qquad f_e \in \mathcal{E}(\Omega, \sigma),$$

$$\int_\Omega^{\oplus} f \odot \mu = \sup \left\{ \int_\Omega^{\oplus} f_e \odot \mu : \mathbb{0} \le f_e \le f,\ f_e \in \mathcal{E}(\Omega, \sigma) \right\}, \qquad f \in \mathcal{L}^0(\Omega, \sigma).$$

In [13] we prove the $(\oplus,\odot)$-linearity of this integral and the transport measure theorem. The following facts will prove essential to our purpose, in the sense that they allow a tractable description of the topology of wide convergence for Maslov measures. Let $C_{\mathbb{0}}(\Omega, \mathcal{A})$ be the semiring of continuous $\mathcal{A}$-valued functions on Ω such that $\rho(f(x), \mathbb{0}) \le \varepsilon$ for any $\varepsilon > 0$ and for any x outside some compact set $K \subset \Omega$. Let $C'_{\mathbb{0}}(\Omega, \mathcal{A})$ be the semimodule of continuous $\mathcal{A}$-valued linear forms on $C_{\mathbb{0}}(\Omega, \mathcal{A})$. By $M(\Omega, \sigma)$ we denote the semiring of bounded Maslov measures μ on (Ω, σ) such that the mapping $f \mapsto \int^{\oplus} f \odot \mu$ belongs to $C'_{\mathbb{0}}(\Omega, \mathcal{A})$, and by $\mathrm{M}(\Omega, \sigma)$ the quotient semiring of $M(\Omega, \sigma)$ by the wide sense equivalence. Let $\mathbf{D}(\Omega, \sigma)$ be the subsemiring of $\mathrm{M}(\Omega, \sigma)$ spanned by the measures μ_f, where $f \in \mathcal{L}^0(\Omega, \sigma)$ is upper semicontinuous, defined for every $A \in \sigma$ by

$$\mu_f(A) \stackrel{\text{def}}{=} \sup_{\omega \in A} f(\omega) \stackrel{\text{def}}{=} \int_\Omega^{\oplus} \mathbb{1}_A(\omega) \odot f(\omega) \odot d\omega. \tag{1.1}$$

In that case, for every $\phi \in C'_{\mathbb{0}}(\Omega, \mathcal{A})$ we have $\int^{\oplus} \phi \odot \mu = \sup_{\omega \in \Omega}(\phi(\omega) + f(\omega))$. These measures μ_f are said to be absolutely continuous and f is called the density. In [13] we introduce the Stieltjes–Maslov measures and prove that the constant measure $dx(A) = \mathbb{1}$ is the $(\max,+)$-version of the Lebesgue measure. In [30], Maslov proves that $\mathrm{M}(\Omega, \sigma) = \mathbf{D}(\Omega, \sigma) = C'_{\mathbb{0}}(\Omega, \mathcal{A})$. When $\Omega = \mathbb{R}$, the Maslov measure μ defined by $\mu(A) = \mathbb{0}$ if A is countable and $\mu(A) = \mathbb{1}$, if $\mathbb{R} - A$ is countable, is not absolutely continuous, but its maximal extension is $\mu^* = \mathbb{1} \in \mathbf{M}(\Omega, \sigma)$. Finally, to complete this section we recall the notion of the convolution of Maslov measures.

Definition 1.1 Let $\mu_1, \mu_2 \in \mathbf{M}(\Omega, \mathcal{A})$. By $\mu_1 \circledast \mu_2$ we denote the unique element in $\mathbf{M}(\Omega, \mathcal{A})$ such that

$$\int_{\Omega\times\Omega}^{\oplus} f(x+y) \odot \mu_1(dx) \odot \mu_2(dy) = \int_{\Omega}^{\oplus} f(z) \odot (\mu_1 \circledast \mu_2)(dz) \tag{1.2}$$

for any $f \in \mathcal{C}_{\mathbb{0}}(\Omega, \mathcal{A})$

This law is compatible with the weak topology and commutative, associative, and distributive over the addition $\oplus$.

2. Performance Theory

All notation, assumptions, and results of §1 are in force. The concepts of performance measure and optimization variable are basic notions. Their analysis in the semiring $\mathcal{A}$ offers an alternative to the classical representations of optimization problems. The purpose of this section is to recall some of these axioms. The intuitive background of the concept of an optimization variable is as follows. Suppose that we are given an optimization problem described by a measurable space (Ω, σ), where Ω is the set of all possible controls and σ is the sigma-algebra of controls that are possible or interesting in the framework of the optimization problem. Now we are given a reference value $Y(\omega)$ associated with a control event ω. This value depends on ω. The assumption of measurability means that for every reference value there is a meaningful control event in the original space. For instance, there is in general less information in $Y(\omega)$ than in ω, a fact expressed by the condition that $\sigma(Y) \subset \sigma$. Let $\Omega = \{(i,j) : 1 \le i,j \le n\}$, $n \in \mathbb{N}$, and let the powerset of Ω serve as σ. Then every function on Ω is measurable. On the other hand, if $\sigma(Y)$ is the sigma-algebra generated by the sets $S_k = \{(i,j) : i+j = k\}$ with $1 \le k \le 2n$, then $Y((i,j)) = i+j$ is $\sigma(Y)$-measurable, but $X((i,j)) = i$ is not. The general conditional optimization problem will be studied in §5. Let (Ω, σ) and $(E, \mathcal{E})$ be two Polish spaces. A Maslov measure $\mathbb{P} \in \mathbf{M}(\Omega, \sigma)$ such that $\mathbb{P}(\Omega) = \mathbb{1}$ is called a *performance measure* and $(\Omega, \sigma, \mathbb{P})$ is called a *performance space*. As in probability theory, we assign performance to each event in Ω and define optimization variables whose domain consists of the elements of Ω. A measurable function X from (Ω, σ) into $(E, \mathcal{E})$ is called an E-valued optimization variable, and we denote its performance measure by $\mathbb{P}^X$ and its upper semicontinuous density by p^X. Whenever the Maslov integral exists, we write

$$\mathbb{E}(X) = \int_{\Omega}^{\oplus} X \odot \mathbb{P} = \int_{E}^{\oplus} x \odot \mathrm{p}^X(x) \odot dx, \tag{1.3}$$

$$\forall A \in \mathcal{E} \quad \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}) = \mathbb{P}^X(A)$$

$$= \int_{A}^{\oplus} \mathrm{p}^X(x) \odot dx = \mathbb{E}(\mathbb{1}_A(X)). \tag{1.4}$$

In our framework, the following implications hold:

$$\mathbb{P}(X \in \Omega - A) < \mathbb{1} \implies \mathbb{P}(X \in A) = \mathbb{1} \implies \mathrm{op}(X) \in A.$$

These facts show that it is important to control the performance of an event. To this end, we shall state the (max, +)-version of the Markov, Minkowski, and Hölder inequalities in §3.

Definition 1.2 Let X and Y be two $\mathcal{A}$-valued optimization variables. We say that X and Y are equal $\mathbb{P}$-almost everywhere ($\mathbb{P}$-a.e.) if

$$\mathbb{P}(\{\omega \in \Omega : X(\omega) \ne Y(\omega)\}) = \mathbb{0}.$$

Several other characterizations of this equivalence relation are as follows:

$$\begin{aligned} X = Y \quad \mathbb{P}\text{-a.e.} &\iff \forall \varepsilon > \mathbb{0} \quad \mathbb{P}\{\omega \in \Omega : d(X(\omega), Y(\omega)) \ge \varepsilon\} = \mathbb{0} \\ &\iff \exists A \in \sigma : \quad \mathbb{P}(\Omega - A) = \mathbb{0} \quad \text{and} \\ &\qquad\qquad \forall \omega \in A \quad X(\omega) = Y(\omega) \\ &\iff \forall A \in \sigma \quad \mathbb{E}(\mathbb{1}_A \odot X) = \mathbb{E}(\mathbb{1}_A \odot Y). \end{aligned}$$

By way of example, the last condition can be used to prove the uniqueness of the Maslov conditional expectation. Let $\mathbf{L}^0(\Omega, \sigma) = \mathcal{L}^0(\Omega, \sigma)/\mathbb{P}$-a.e. be the induced quotient semiring. Now we introduce the semiring of $\mathcal{A}$-valued integrable optimization variables by setting

$$d_1(X, Y) \stackrel{\text{def}}{=} \mathbb{E}(d(X, Y)),$$

$$\mathcal{L}^1(\Omega, \sigma, \mathbb{P}) \stackrel{\text{def}}{=} \{X \in \mathcal{L}^0(\Omega, \sigma) : d_1(X, \mathbb{0}) = \mathbb{E}(X) < +\infty\}.$$

In [14] we prove that $X = Y$ $\mathbb{P}$-a.e. if and only if $d_1(X, Y) = \mathbb{0}$ and d_1 is a $\oplus$-metric over $\mathbf{L}^1(\Omega, \sigma) = \mathcal{L}^1(\Omega, \sigma)/\mathbb{P}$-a.e. For instance, let X be a real optimization variable whose performance law is given by

$$\mathrm{p}(x) = -\frac{1}{2}\left(\frac{x-m}{a}\right)^2;$$

then

$$d_1(X, \mathbb{0}) = m \odot \left(\frac{a^2}{2}\right), \qquad d_1(X, m) = m \odot d_1\left(\frac{X}{m}, \mathbb{0}\right).$$

Moreover,

$$d_1\left(\frac{X}{m}, \mathbb{1}\right) = \sup_x \left(\log|e^x - 1| - \frac{1}{2}\left(\frac{x}{a}\right)^2\right) = \sup_{x>0} \log \Theta_a(x)$$

and for $x \geq 0$, we have

$$\Theta_a'(x) = 0 \iff e^x - 1 = (a^{-2}x - 1)^{-1}.$$

By standard numerical approximations,

$$d_1\left(\frac{X}{m}\circ, \mathbb{1}\right) = \log \Theta_a(x(a))$$

with $0 \leq x(a) \leq x'(a)$ and

$$x'(a) = (a^{-2}x'(a) - 1)^{-1} \quad (\iff a^{-2}x'(a)^2 - 1 = x'(a)).$$

Similarly, if $a < 2$, then $0 \leq x'(a) \leq x''(a)$, where

$$\frac{2}{a}(x''(a) - a) = x''(a) \quad \left(\iff x''(a) = \frac{2a}{2-a}\right),$$

and for $a < 2$, we have

$$0 \leq d_1(X, m) \leq m \odot d\left(\frac{2a}{2-a}, \mathbb{1}\right).$$

By the same argument, if

$$\mathbb{p}(x) = -\frac{1}{p}\left|\frac{x-m}{a}\right|^p.$$

for some $p \geq 2$, then we obtain, for $1/q + 1/p = 1$,

$$d_1(X, 0) = m \odot \left(\frac{a^q}{q}\right) \quad \text{and}$$

$$a < p \implies 0 \leq d_1(X, m) \leq m \odot d\left(\frac{pa}{p-a}, \mathbb{1}\right). \qquad (1.5)$$

Proposition 1.1 *Let $Q_m(a,p)$ be a real optimization variable whose performance is*

$$\mathbb{p}(x) = -\frac{1}{p}\left|\frac{x-m}{a}\right|^p \quad \textit{for some } p \geq 2,\ a > 0,\ m \in \mathbb{R}.$$

Then $\lim_{a\to 0} d_1(Q_m(a,p), m) = 0$, and for $1/q + 1/p = 1$ we have

$$d_1(Q_m(a,p), 0) = m \odot (a^q/q), \quad \textit{and}$$

$$a < p \implies 0 \leq d_1(Q_m(a,p), m) \leq m \odot d\left(\frac{pa}{p-a}, \mathbb{1}\right). \qquad (1.6)$$

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

Definition 1.3 Let $(X_i)_{i\in I}$ be a family (not necessarily finite) of E-valued optimization variables on the same performance space $(\Omega, \sigma, \mathbb{P})$. We say that they are $\mathbb{P}$-independent if for every finite subset $J = \{t_1, \ldots, t_n\} \subset I$, $n \geq 1$, one of the following equivalent assertions is satisfied:

1. $\forall i \in \{1, \ldots, n\}\ \forall A_i \in \sigma(X_{t_i})$, $\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \odot \cdots \odot \mathbb{P}(A_n)$.
2. $\mathbb{p}_J^X(x_1, \ldots, x_n) = \bigodot_{j=1}^n \mathbb{p}_{t_j}^X(x_j)$ where $\mathbb{p}_J^X$ (respectively, $\mathbb{p}_{t_j}^X$ $1 \leq j \leq n$) is the performance density of $X_J = (X_{t_1}, \ldots, X_{t_n})$ (respectively, X_{t_j}, $j \in [1, n]$).

As the opposite of the latter concept of independence, we introduce an extension of the Bayes formula to performance measures. The conditional performance of an event A, assuming an event B such that $\mathbb{P}(B) \neq 0$, and denoted by $\mathbb{P}(A/B)$, is, by definition, the ratio

$$\mathbb{P}(A/B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}\circ. \qquad (1.7)$$

The last conditions are the axioms of Maslov optimization theory. In the development of the theory, all conclusions are directly or indirectly based on these axioms alone. We now examine how the notions of conditional expectation and optimal value of a random variable are translated in such a framework.

Theorem, Definition 1 *Let X and Y be two E-valued optimization variables defined on the same performance space $(\Omega, \sigma, \mathbb{P})$.*

1. *Let $\sigma(X,Y)$ be the σ-algebra spanned by the optimization variables X and Y, and let $\sigma(Y)$ be the σ-algebra spanned by Y. For every function $\phi \in \mathbf{L}^1(\Omega, \sigma(X,Y), \mathbb{P})$, there exists a unique function in $\mathbf{L}^1(\Omega, \sigma(Y), \mathbb{P})$, called the conditional expectation of ϕ relative to $\sigma(Y)$ and denoted by $\mathbb{E}(\phi/\sigma(Y))$, such that*

$$\forall \psi \in \mathbf{L}^1(\Omega, \sigma(Y), \mathbb{P}) \quad \mathbb{E}(\phi \odot \psi) = \mathbb{E}(\mathbb{E}(\phi/\sigma(Y)) \odot \psi).$$

2. *The measurable function $\mathbb{p}^{X/Y}(x/y)$ defined as $\dfrac{\mathbb{p}^{X,Y}(x,y)}{\mathbb{p}^Y(y)}\circ$ if $\mathbb{p}^Y(y) > 0$ and equal to 0 otherwise is called the conditional performance density of X relative to Y. A conditional optimal state of X relative to Y, denoted by $\mathrm{op}(X/Y)$, is a (not necessarily unique) measurable function such that $\mathbb{p}^{X/Y}(\mathrm{op}(X/y)/y) = \mathbb{1}$.*

Proposition 1.2 *Let X, Y, and Z be three E-valued optimization variables on the same performance space $(\Omega, \sigma, \mathbb{P})$,*

$$\varphi, \psi \in \mathbf{L}^1(\Omega, \sigma(X,Y), \mathbb{P}), \quad \phi \in \mathbf{L}^1(\Omega, \sigma(X,Y,Z), \mathbb{P}),$$

and $a, b \in \mathcal{A}$. Then $\mathbb{P}$-a.e. we have

1. $\mathbb{E}(a \odot \varphi \oplus b \odot \psi/\sigma(Y)) = a \odot \mathbb{E}(\varphi/\sigma(Y)) \oplus b \odot \mathbb{E}(\psi/\sigma(Y))$, $\mathbb{E}(\phi) = \mathbb{E}(\phi/\{\emptyset, \Omega\})$;
2. $\mathbb{E}(\varphi/\sigma(Y)) = \varphi$ *if* $\varphi \in \mathbf{L}^1(\Omega, \sigma(Y))$, $\mathbb{E}(\mathbb{E}(\phi/\sigma(X,Y))/\sigma(Y)) = \mathbb{E}(\phi/\sigma(Y))$.

We shall always assume in the sequel that the conditional or nonconditional optimal states are well defined and unique.

Proposition 1.3 *Let X, Y, and Z be E-valued optimization variables defined on the same performance space $(\Omega, \sigma, \mathbb{P})$, and let φ be an upper semicontinuous function. Then*

$$\mathrm{op}(\mathrm{op}(X/Z, Y)/Y) = \mathrm{op}(X/Y), \qquad \mathrm{op}(\varphi(X)/Y) = \varphi(\mathrm{op}(X/Y)).$$

For instance, let F be a measurable function from $\mathbb{R}^n$ into $\mathbb{R}^n$, C a $n \times m$ matrix, and U and V two $\mathbb{P}$-independent optimization variables with $\mathbb{p}^U(u) \stackrel{\text{def}}{=} -\frac{1}{2}u'Q^{-1}u$ and $\mathbb{p}^V(v) \stackrel{\text{def}}{=} -\frac{1}{2}v'R^{-1}v$, where Q^{-1} and R^{-1} are, respectively, $n \times n$ and $m \times m$ symmetric positive definite matrices. Let $S^{-1} \stackrel{\text{def}}{=} Q^{-1} + C'R^{-1}C$ and $X_1 = F(X_0) + U$, $Y_1 = CX_1 + V$; then $\mathrm{op}(X_1/X_0) = F(X_0)$, $\mathrm{op}(Y_1/X_0) = CF(X_0)$, and

$$\mathrm{op}(X_1/X_0, Y_1) = F(X_0) + SC'R^{-1}(Y_1 - CF(X_0)).$$

If X and Y are two $\mathbb{P}$-independent E-valued optimization variables, then $\mathbb{p}^{X,Y}(x,y) = \mathbb{p}^X(x) \odot \mathbb{p}^Y(y)$, and the performance density of the sum $X+Y$ can be described by the $(\max,+)$-convolution

$$\mathbb{p}^{X+Y}(z) = \int_E^{\oplus} \mathbb{p}^X(z-y) \odot \mathbb{p}^Y(y) \odot dy \stackrel{\text{def}}{=} (\mathbb{p}^X \circledast \mathbb{p}^Y)(z).$$

Similarly, we claim that every $(\max,+)$-linear and continuous operator S on the class of upper semicontinuous functions that commutes with the delays U_x (defined by $U_x(f)(y) = f(y-x)$ for every upper semicontinuous function f) can be represented by a $(\max,+)$-convolution. Indeed, let h_n be a weak approximation of the Maslov–Dirac function ($\delta_0(x) = \mathbb{I}_0(x) = \mathbb{I}$ if $x = 0$ and $\delta_0(x) = \mathbb{0}$ otherwise). Let E be a normed space, and let $h_n(x) = -\frac{n}{2}\|x\|^2$. Then for every function $f \in \mathcal{C}_0$ we have:

$$\begin{aligned} S(f \circledast h_n) &= S\left(\int_E^{\oplus} U_y(f) \odot h_n(y) \odot dy\right) = \int_E^{\oplus} S(U_y(f)) \odot h_n(y) \odot dy \\ &= \int_E^{\oplus} f(y) \odot U_y(S(h_n)) \odot dy = \int_E^{\oplus} U_y(S(f)) \odot h_n(y) \odot dy \\ &= S(f) \circledast h_n = f \circledast S(h_n) \\ \Longrightarrow S(f) &= f \circledast \lim_{n \to +\infty} S(h_n) = f \circledast S(\delta_0). \end{aligned}$$

For instance, if $S(\delta_0)(x) = g(x)$, then $S(f)(x) = \sup_{y \in E}(g(x-y) + f(y))$. We now recall the definition of the *Fenchel transformation*. Let $\mathcal{S}_+$ be the class of proper upper semicontinuous concave functions from $\mathbb{R}$ into $\mathbb{R}$. The Fenchel transformation is the mapping $\mathcal{F}: \mathcal{S}_+ \to \mathcal{S}_+$, $f \mapsto \mathcal{F}(f)$ such that

$$-(\mathcal{F}f)(x^*) = \int_{\mathbb{R}}^{\oplus} x^*(x) \odot f(x) \odot dx. \tag{1.8}$$

Let τ_a be the translation on $\mathbb{R}$ associated with $a \in \mathbb{R}$. Then for every $x^* \in \mathbb{R}$ we have

$$(-\mathcal{F}(f \circ \tau_a))(x^*) = x^*(a) \odot (-\mathcal{F}(f))(x^*). \tag{1.9}$$

This property is characteristic of the Fourier transformation (Maslov [34]). The Fenchel transformation is then equal to the Fourier transformation in our setting.

Proposition 1.4 *Let $f, g \in \mathcal{S}_+$, and let X be an $\mathbb{R}$-valued optimization variable on $(\Omega, \sigma, \mathbb{P})$ whose optimal state $\mathrm{op}(X) \in \mathbb{R}$ is unique. Assume that $\mathcal{F}\mathbb{p}^X$ is twice continuously differentiable around 0 and*

$$(f \circledast g)(z) \stackrel{\text{def}}{=} \int^{\oplus} f(z-x) \odot g(x) \odot dx = \sup_{x \in \mathbb{R}}(f(z-x) + g(x)).$$

Then

1. $\mathcal{F} \circ \mathcal{F} = \mathcal{I}d$, *and* $\mathcal{F}(f \circledast g) = \mathcal{F}(f) \odot \mathcal{F}(g)$;
2. $\mathbb{E}(\omega X) = -(\mathcal{F}\mathbb{p}^X)(\omega)$, $-(\mathcal{F}\mathbb{p}^X)'(0) = \mathrm{op}(X)$, *and* $(\mathcal{F}\mathbb{p}^X)''(0) = ((\mathbb{p}^X)'')^{-1}(\mathrm{op}(X))$.

The following example is classical in convex analysis and will be essential in the study of convergence modes. If $a \in \mathbb{R}^+$, $m \in \mathbb{R}$, $a \neq 0$, $p > 2$, and $1/q + 1/p = 1$, then

$$-\mathcal{F}(f) = m + \frac{1}{p}|xa|^p \iff f = -\frac{1}{q}\left|\frac{x-m}{a}\right|^q. \tag{1.10}$$

3. Lebesgue–Maslov Semirings

Here we introduce an analog of Lebesgue spaces and Markov, Minkowski, and Hölder inequalities. Then we deal with several modes of convergence for optimization variables. Unfortunately, space limitations prevent us from coming into details (see [13, 14]). To focus on the main idea, the treatment will leave technical issues aside. In the sequel, all optimization variables are defined on the same performance space $(\Omega, \sigma, \mathbb{P})$. Most of the results are taken from [13], but they have been presented for the first time in [14] and [15]. As for random variables, we introduce the Kyn–Fan metric of optimization variables. For every $\mathcal{A}$-valued optimization variables X, Y we set

$$\begin{aligned}
\mathcal{K}(X,Y) &= \{(\varepsilon,\eta)\in(\mathbb{R}^+)^2 : \mathbb{P}\{\omega\in\Omega : d(X(\omega),Y(\omega)) > \log\eta\} \leq \log\varepsilon\},\\
\delta(X,Y) &= \inf\{\varepsilon+\eta : (\varepsilon,\eta)\in\mathcal{K}(X,Y)\},\\
\tilde{\delta}(f,g) &= 2\inf\{\varepsilon : (\varepsilon,\varepsilon)\in\mathcal{K}(X,Y)\},\\
e(X,Y) &= \log\delta(X,Y),\\
\tilde{e}(X,Y) &= \log\tilde{\delta}(X,Y).
\end{aligned}$$

Proposition 1.5 *δ and $\tilde{\delta}$ are metrics over $\mathbf{L}^0(\Omega,\sigma)$, and e and $\tilde{e}$ are $\oplus$-metrics over $\mathbf{L}^0(\Omega,\sigma)$ with*

$$e(X,Y) \leq \tilde{e}(X,Y) \leq c\odot e(X,Y) \qquad (c=\log 2) \tag{1.11}$$

One can also introduce the $\mathbf{L}^p$-semirings for $0<p\leq+\infty$. Let $X,Y\in\mathcal{L}^0(\Omega,\sigma)$. We write

$$\begin{aligned}
d_p(X,Y) &= \mathbb{E}(d(X,Y)^p)^{1/p},\\
\mathcal{L}^p(\Omega,\sigma,\mathbb{P}) &= \{X\in\mathcal{L}^0(\Omega,\sigma) : d_p(X,\mathbf{0}) < +\infty\},
\end{aligned} \tag{1.12}$$

where $a^p \stackrel{\text{def}}{=} pa$. For $p=+\infty$, we write

$$d_\infty(X,Y) = \inf\{m\geq\mathbf{0} : \mathbb{P}(\{\omega\in\Omega : d(X(\omega),Y(\omega))\geq m\}) = \mathbf{0}\}, \tag{1.13}$$
$$\mathcal{L}^\infty(\Omega,\sigma,\mathbb{P}) = \{X\in\mathcal{L}^0(\Omega,\sigma) : d_\infty(X,\mathbf{0}) < +\infty\}. \tag{1.14}$$

For every $0<p\leq+\infty$, we state [13, 14] that

$$X=Y \quad \mathbb{P}\text{–a.e.} \iff d_p(f,g)=\mathbf{0}.$$

If $\mathbf{L}^p(\Omega,\sigma,\mathbb{P}) = \mathcal{L}^p(\Omega,\sigma,\mathbb{P})/\mathbb{P}$-a.e., then, for every $0<p\leq+\infty$, d_p is a $\oplus$-metric over $\mathbf{L}^p$. Keeping in mind the notation of Proposition 1.1, we obtain

$$d_p(Q_m(a,r),m) = d_1(Q_m(ap^{1/r},r),m).$$

Moreover, if $X,Y\in L^\infty(\Omega,\sigma,\mathbb{P})$, then $d_p(X,Y)$ is an increasing sequence in $\mathcal{A}$ that converges to $d_\infty(X,Y)$ as p goes to $+\infty$. The following theorem gives an exhaustive list of properties that lead to useful conclusions, because they make the relationship between the latter $\oplus$-metrics and the Maslov expectation explicit. As usual, for any $a\in\mathcal{A}$ and $p>0$, we have $a^p=pa$ and $\mathbf{L}^p=\mathbf{L}^p(\Omega,\sigma,\mathbb{P})$.

Theorem 1.1 1. *For every $p>0$ and $X\in\mathbf{L}^p$, we have*

$$\mathbb{E}(X^p)^{1/p} \leq \inf\{m\geq\mathbf{0} : \mathbb{P}(\omega\in\Omega : X(\omega)\geq m) = \mathbf{0}\}.$$

2. **Markov Inequality:** *For any $X\in\mathbf{L}^p$, $p>0$, and $\varepsilon\geq\mathbf{0}$ we have*

$$\mathbb{P}(X\geq\varepsilon)\odot\varepsilon^p \leq \mathbb{E}(X^p).$$

3. *Let g be an increasing function from $\mathcal{A}$ into $\mathcal{A}$; then for any $a\geq\mathbf{0}$ and $X\in\mathbf{L}^0$, we have*

$$g(a)\odot\mathbb{P}(\{\omega\in\Omega : X(\omega)\geq a\}) \leq \mathbb{E}(g(X)) \leq d_\infty(g(X),\mathbf{0})\oplus g(a).$$

4. **Hölder Inequality:** *For any $0<p\leq q<\infty$ and $0<n<+\infty$ such that $1/p+1/q=1/n$, $X\in\mathbf{L}^p$, and $Y\in\mathbf{L}^q$, we have $X\odot Y\in\mathbf{L}^n$ and*

$$\mathbb{E}((X\odot Y)^n)^{1/n} \leq \mathbb{E}(X^p)^{1/p}\odot\mathbb{E}(Y^q)^{1/q}.$$

5. **Minkowski Inequality:** *For any $0<p<+\infty$ and $X,Y\in\mathbf{L}^p$, we have $X\oplus Y\in\mathbf{L}^p$, $X\odot Y\in\mathbf{L}^p$, and*

$$\mathbb{E}((X\oplus Y)^p)^{1/p} = \mathbb{E}(X^p)^{1/p}\oplus\mathbb{E}(Y^p)^{1/p},$$
$$\mathbb{E}((X\odot Y)^p)^{1/p} \leq c\odot(\mathbb{E}(X^p)^{1/p}\oplus\mathbb{E}(Y^p)^{1/p}),$$

where $c=\log 2$.

6. *For every $X,Y\in L^1$, we have*

$$d(\mathbb{E}(X),\mathbb{E}(Y)) \leq \mathbb{E}(d(X,Y)), \qquad \tilde{e}(X,Y) \leq c\odot\mathbb{E}(d(X,Y)),$$

where $c=\log 2$.

Consequently, for every p, $0\leq p\leq+\infty$, $\mathbf{L}^p$ is a semiring; in other words,

$$(\mathbf{L}^p(\Omega,\sigma,\mathbb{P}),\oplus,\mathbf{0}) \quad\text{and}\quad (\mathbf{L}^p(\Omega,\sigma,\mathbb{P}),\odot,\mathbb{1})$$

are two semigroups. In view of the results stated in Proposition 1.1, we readily obtain the following corollary.

Corollary 1.1 (*All notation of Proposition* 1.1 *is in force.*) *Let X^N be a sequence of real optimization variables whose performance densities satisfy*

$$p_N(x) \leq -\frac{1}{p}\left|\frac{x-m}{a_N}\right|^p$$

for some real numbers m, $p>0$, and a_N, $\lim_{N\to+\infty}a_N=0$. Then, for sufficiently large N and for every $\varepsilon>\mathbf{0}$, we have

$$\begin{aligned}
&\mathbb{P}(\{\omega\in\Omega : d(X^N(\omega),m)>\varepsilon\})\\
&\quad\leq \frac{\mathbb{E}(d(Q_m(a_N,p),m))}{\varepsilon}\circ \leq \frac{m}{\varepsilon}\circ\odot d\left(\frac{pa_N}{p-a_N},\mathbb{1}\right) \underset{N\to+\infty}{\longrightarrow} \mathbf{0},\\
&d(\mathbb{E}(X^N),m) \leq m\odot d\left(\frac{pa_N}{p-a_N},\mathbb{1}\right),\\
&\tilde{e}(X^N,m) \leq \log 2\odot d\left(\frac{p\,a_N}{p-a_N},\mathbb{1}\right).
\end{aligned}$$

4. Convergence Modes

One problem in performance theory is the determination of the asymptotic properties of optimization variables. In this section we focus on clarifying the underlying concepts. We start from a simple problem. Suppose that we wish to study the behavior of a sequence of performance convolutions $\mathbb{p}_1 \circledast \mathbb{p}_2 \circledast \cdots \circledast \mathbb{p}_n$. This problem is also related to asymptotic studies of the solution of some Bellman equation (see [2, 26, 44, 46, 47]). We claim that it is natural to analyze such equations by means of an appropriate sequence of optimization variables. Indeed, let $(X^n)_n$ be a sequence of independent optimization variables, and let $\mathbb{p}_n$ be the performance law of X^n. Let S_n be the sequence defined by $S_n = \sum_{i=1}^n X^i$, and let $\mathbb{p}^n$ be its performance law. Then

$$\mathbb{p}^n = \mathbb{p}_1 \circledast \mathbb{p}_2 \circledast \cdots \circledast \mathbb{p}_n.$$

For instance, let $X_k \stackrel{\text{def}}{=} \Delta X_0 + \sum_{l=1}^k \Delta X_l$, where the ΔX_l are $k+1$ independent optimization variables whose performance laws are defined by

$$\mathbb{p}^{\Delta X_l}(z) = -\frac{1}{q}\left|\frac{z - \mathrm{op}(\Delta X_l)}{\sigma_l}\right|^q, \qquad \sigma_l > 0,\ q \geq 2.$$

Then, using the properties of the Fenchel transform, for $1/p + 1/q = 1$ we obtain

$$\mathbb{p}^{X_k}(z) = \mathbb{p}^{X_0} \circledast \mathbb{p}^{\Delta X_1} \circledast \cdots \circledast \mathbb{p}^{\Delta X_k} \implies$$
$$\mathbb{p}^{X_k/(k+1)}(z) = -\frac{(k+1)^{q/p}}{q}\left|\frac{z - \Delta\overline{X}_k}{\overline{\sigma}_k}\right|^q = Q_{\Delta\overline{X}_k}\left(\frac{\overline{\sigma}_k}{(k+1)^{1/p}}, q\right),$$

where

$$\Delta\overline{X}_k = \frac{1}{k+1}\sum_{l=0}^k \mathrm{op}(\Delta X_l) = \mathrm{op}\left(\frac{1}{k+1}\sum_{l=0}^k \Delta X_l\right),$$
$$\overline{\sigma}_k = \left(\frac{1}{k+1}\sum_{l=0}^k \sigma_l^q\right)^{1/q}.$$

Consequently, if

$$\Delta\overline{X}_\infty \stackrel{\text{def}}{=} \lim_{k\to+\infty} \Delta\overline{X}_k < +\infty \quad\text{and}\quad \overline{\sigma}_\infty \stackrel{\text{def}}{=} \lim_{k\to+\infty} \overline{\sigma}_k < +\infty,$$

then $\frac{1}{k+1}X_k$ is weakly convergent to $\Delta\overline{X}_\infty$. Moreover, by virtue of Proposition 1.1, if $c = \log 2$ and $a_k = \frac{\overline{\sigma}_k}{(k+1)^{1/p}}$, then we have

$$\begin{aligned} d_1\left(\frac{X_k}{k+1}, \Delta\overline{X}_\infty\right) &\leq c \odot \left(d_1(\Delta\overline{X}_k, \Delta\overline{X}_\infty) \oplus d_1\left(\frac{X_k}{k+1}, \Delta\overline{X}_k\right)\right) \\ &\leq c \odot \left(d_1(\Delta\overline{X}_k, \Delta\overline{X}_\infty) \oplus \Delta\overline{X}_k \odot d_1\left(\frac{q a_k}{q - a_k}, \mathbb{1}\right)\right) \underset{k\to+\infty}{\longrightarrow} 0. \end{aligned}$$

According to Markov's inequality, it is easy to calibrate the performance of the events associated with the nonconvergence. For instance, for every $\varepsilon \geq 0$ we have

$$\mathbb{P}\left(\left\{\omega \in \Omega : d\left(\frac{X_k(\omega)}{k+1}, \Delta\overline{X}_\infty\right) \geq \varepsilon\right\}\right) \leq \frac{d_1\left(\frac{X_k}{k+1}, \Delta\overline{X}_\infty\right)}{\varepsilon} \underset{k\to+\infty}{\longrightarrow} 0.$$

The last assertion of the theorem shows that

$$\begin{aligned} &\inf\left\{\varepsilon \geq 0 : \mathbb{P}\left(\left\{\omega \in \Omega : d\left(\frac{X_k(\omega)}{k+1}, \Delta\overline{X}_\infty\right) \leq \varepsilon\right\}\right) \leq \varepsilon\right\} \\ &\quad \leq d_1\left(\frac{X_k}{k+1}, \Delta\overline{X}_\infty\right) \underset{k\to+\infty}{\longrightarrow} 0. \end{aligned}$$

These facts will be further developed in the end of the section. Next, we introduce various convergence modes involving sequences of optimization variables.

Definition 1.4 Let $(X_n)_{n\geq 1}$ be a sequence of $\mathcal{A}$-valued optimization variables, and let X be an $\mathcal{A}$-valued optimization variable.

1. **Uniform Convergence:** $\lim_{n\to+\infty} \sup_{\omega\in\Omega} d(X_n(\omega), X(\omega)) = 0$.
2. **$\mathbb{P}$-Convergence** (e-$\mathbb{P}$):

$$\forall\varepsilon > 0 \quad \lim_{n\to+\infty} \mathbb{P}(\{\omega \in \Omega : d(X_n(\omega), X(\omega)) \geq \varepsilon\}) = 0.$$

3. **$\mathbf{L}^\infty$-Convergence:** $\lim_{n\to+\infty} d_\infty(X_n, X) = 0$.
4. **$\mathbf{L}^p$-Convergence:** $(0 < p < +\infty)$ $\lim_{n\to+\infty} d_p(X_n, X) = 0$.
5. **$\mathbb{P}$-almost everywhere Convergence** ($\mathbb{P}$-a.e.):

$$\forall\varepsilon > 0 \quad \mathbb{P}\left(\left\{\omega \in \Omega : \limsup_{n\to+\infty} d(X_n(\omega), X(\omega)) \geq \varepsilon\right\}\right) = 0.$$

6. **Weak Convergence:** $\forall\phi \in \mathcal{C}_0(\Omega)$ $\lim_{n\to+\infty} \mathbb{E}(\phi(X_n)) = \mathbb{E}(\phi(X))$.

Next, we introduce the uniform integrability of a class of functions in $\mathbf{L}^1$.

Definition 1.5 Let $\mathcal{H} \subset L^1(\Omega, \sigma, \mathbb{P})$; then $\mathcal{H}$ is said to be *uniformly integrable* whenever for any $X \in \mathcal{H}$ the integrals

$$\int_{\{\omega\in\Omega: X(\omega)\geq c\}}^{\oplus} X \odot \mathbb{P} \tag{1.15}$$

uniformly converge to $\mathbf{0}$ as $c \geq 0$ tends to $+\infty$.

These classes can be characterized as follows.

Proposition 1.6 *Let $\mathcal{H} \subset L^1(\Omega, \sigma, \mathbb{P})$; then $\mathcal{H}$ is uniformly integrable if and only if the Maslov expectations $\mathbb{E}(X)$, $X \in \mathcal{H}$, are uniformly bounded and for every $\varepsilon > \mathbf{0}$ there exists a $\delta > \mathbf{0}$ such that*

$$\forall A \in \sigma \quad \mathbb{P}(A) \leq \delta \implies \int_A^{\oplus} X \odot \mathbb{P} \leq \varepsilon \quad \forall X \in \mathcal{H}.$$

Moreover, let G be a $\mathcal{A}$-valued function. If

$$\lim_{t\to+\infty} \frac{G(t)}{t}_{\circ} = +\infty \quad \text{and} \quad \sup_{X\in\mathcal{H}} \mathbb{E}(G(X)) < +\infty,$$

then $\mathcal{H}$ is uniformly integrable. Other topological results such as the dominated convergence theorem can be found in [13, 14, 15]. The following theorem gives an exhaustive list of comparisons between various convergence modes.

Theorem 1.2 *Let X and X_n be optimization variables defined on the same performance space $(\Omega, \sigma, \mathbb{P})$.*

1. *For $0 < p \leq q \leq +\infty$, the $\mathbf{L}^q$-convergence implies the $\mathbf{L}^p$-convergence and the e-$\mathbb{P}$-convergence. The e-$\mathbb{P}$-convergence implies the weak and the $\mathbb{P}$-a.e.-convergence.*
2. *If, for every $\nu > 0$, the sequence $d(X_{\nu+n}, X_n)$ e-$\mathbb{P}$-converges to $\mathbf{0}$, then X_n e-$\mathbb{P}$-converges.*
3. *X_n e-$\mathbb{P}$-converges to X if and only if $e(X_n, X)$ or $\mathbb{P}(\bigcup_{m=n}^{+\infty}\{d(X_m, X) \geq \varepsilon\})$ converges to $\mathbf{0}$ for every $\varepsilon > \mathbf{0}$. Moreover, if $\{X_n\} = \mathcal{H}$ is uniformly integrable, then X_n $\mathbf{L}^1$-converges.*

The reverse implications are not true (see [13, 14]). As was already mentioned, optimization problems involving independent optimization variables are useful in the time-discrete case; the continuous case will be dealt with later on. In other words, the useful case is typically a sequence $X^1, \ldots, X^N, \ldots$ of mutually independent optimization variables. Then, because of the independence property, the induced performance density is the convolution of the individual optimization variables. The key point to study the sums of these independent variables is the fact that the partial-sum performances $\mathbb{p}_N$ must be regular in the sense that there exists a $p > 0$ and some sequences $a_N > 0$ and m_N such that

$$\mathbb{p}_N(x) \leq -\frac{1}{p}\left|\frac{x - m_N}{a_N}\right|^p \quad \text{and}$$

$$\lim_{N\to+\infty} \frac{m_N}{N} = m, \quad \lim_{N\to+\infty} \frac{a_N}{N^{1/q}} < +\infty \qquad \left(\frac{1}{q} + \frac{1}{p} = 1\right).$$

In that case, the performance $\mathbb{p}^N$ of the normalized sums $S_N = \frac{1}{N}\sum_{i=1}^N X^i$ satisfies (see Proposition 1.1)

$$\mathbb{p}^N(x) = \mathbb{p}_N(Nx) \leq Q_{m_N/N}(a_N/N, p), \qquad \lim_{N\to+\infty} d_1(S_N, m) = \mathbf{0}.$$

According to Markov's inequality, for every $\varepsilon \geq \mathbf{0}$ we have

$$\mathbb{P}(\{\omega \in \Omega : d(S_N(\omega), m) \geq \varepsilon\}) \leq \frac{d_1(S_N, m)}{\varepsilon}_{\circ}.$$

More generally, for every $p > 0$, S_N $\mathbf{L}^p$-converges to m. By a dual argument applied to the Fenchel transform of the latter performances, these facts may be summarized as follows:

Theorem 1.3 *Let $(X^i)_{i\geq 1}$ be a sequence of real-valued optimization variables on $(\Omega, \sigma, \mathbb{P})$ whose optimal states are well defined. We assume they are $\mathbb{P}$-independent. Assume that*

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{i=1}^N \text{op}(X^i) = \lim_{N\to+\infty} \text{op}\left(\frac{1}{N}\sum_{i=1}^N X^i\right) \stackrel{\text{def}}{=} \overline{X} < +\infty$$

for every $i \geq 1$, $\mathcal{F}\mathbb{p}^{X^i} \in \mathcal{S}_+ \cap \mathcal{C}^2(\mathbb{R})$, and there exist two reals $a > 0$ and $r > 1$ such that for every $N \geq 1$, $\lambda \in [0, 1]$, and $x \in \mathbb{R}$ we have

$$\frac{1}{2}x\left(\frac{1}{N}\sum_{i=1}^N |(\mathcal{F}\mathbb{p}^{X^i})''|(\lambda x)\right)x \leq \frac{1}{r}(xax)^{r/2}.$$

Then for every $p > 0$ and $\delta > 0$ we have

$$\frac{1}{N}\sum_{i=1}^N X^i \xrightarrow[N\to+\infty]{\mathbf{L}^p} \overline{X}, \qquad \frac{1}{N^{\delta+1}}\sum_{i=1}^N X^i \xrightarrow[N\to+\infty]{\mathbf{L}^p} 0. \tag{1.16}$$

Let us give another equivalent statement of these conditions in terms of performance densities. If

$$\overline{X}^i = X^i - \text{op}(X^i) \quad \text{and} \quad \bigoplus_{i=1}^{+\infty} \mathbb{p}^{\overline{X}^i}(x) \leq -\frac{1}{r'}(xax)^{r'/2},$$

then for every $i \in \mathbb{N}$ we obtain (Eq. (1.10))

$$\mathcal{F}\mathbb{p}^{\overline{X}^i}(x^*) \geq -\frac{1}{r}\,|x^* a^{-1} x^*|^{r/2} \quad \text{with } \frac{1}{r}+\frac{1}{r'}=1.$$

By virtue of the properties of the Fenchel transformation (Proposition 1.4), for every $i \geq 1$ and $x^* \in R$ there exists a $\lambda^i(x^*) \in [0,1]$ such that

$$\frac{1}{N}\sum_{i=1}^{N}|\mathcal{F}(\mathbb{p}^{\overline{X}^i})|(x^*) = \frac{1}{2}\,x^*\frac{1}{N}\sum_{i=1}^{N}|\mathcal{F}\mathbb{p}^{\overline{X}^i}|''(\lambda^i(x^*)x^*)x^* \leq \frac{1}{r}\,|x^* a^{-1} x^*|^{r/2}.$$

Proof. For every $N \geq 1$, we set

$$\mathbb{p}^N \stackrel{\text{def}}{=} \left(\sum_{i=1}^{N} X^i\right) * \mathbb{P}, \qquad \mathbb{p}^{S_N} \stackrel{\text{def}}{=} \left(\frac{1}{N}\sum_{i=1}^{N} X^i\right) * \mathbb{P};$$

these are the performance measures of $\sum_N$ and S_N. The optimization variables X^i are $\mathbb{P}$-independent:

$$\mathbb{p}^N = \mathbb{p}^{X^1} \circledast \cdots \circledast \mathbb{p}^{X^N} \implies \mathcal{F}(\mathbb{p}^N) = \bigodot_{i=1}^{N}\mathcal{F}(\mathbb{p}^{X^i}).$$

By virtue of the properties of the Fenchel transformation, for every $i \geq 1$ and $x^* \in R$, there exists a $\lambda^i(x^*) \in [0,1]$ such that

$$\mathcal{F}(\mathbb{p}^{X^i})(x^*) = -\operatorname{op}(X^i) + \frac{1}{2}\,x^*(\mathcal{F}\mathbb{p}^{X^i})''(\lambda^i(x^*)x^*)x^*.$$

Then

$$\mathbb{p}^N(x) = \mathcal{F}\left(\bigodot_{i=1}^{N}\mathcal{F}(\mathbb{p}^{X^i})\right)(x) = -\sup_{x^*\in\mathbb{R}}\left(x^*(x) + \sum_{i=1}^{N}\mathcal{F}(\mathbb{p}^{X^i})(x^*)\right),$$

$$\mathbb{p}^{S_N}(x) = N\Bigg(-\sup_{x^*\in\mathbb{R}}\Bigg(x^*\left(x - \frac{1}{N}\sum_{i=1}^{N}\operatorname{op}(X^i)\right) - \frac{1}{2}\,x^*\left(\frac{1}{N}\sum_{i=1}^{N}(-\mathcal{F}\mathbb{p}^{X^i})''(\lambda^i(x^*)x^*)\right)x^*\Bigg)\Bigg).$$

By virtue of the second assumption, there exist two real numbers $r > 1$ and $a \geq 0$ such that

$$\mathbb{p}^{S_N}(x) \leq N\left(-\sup_{x^*\in\mathbb{R}}\left(x^*\left(x - \frac{1}{N}\sum_{i=1}^{N}\operatorname{op}(X^i)\right) - \frac{1}{r}(x^* a x^*)^{r/2}\right)\right).$$

For every $s > 1$, $m \in \mathbb{R}$ and $b > 0$, we introduce

$$Q_s(m,b)(x) \stackrel{\text{def}}{=} -\frac{1}{s}((x-m)b(x-m))^{s/2}.$$

According to the involution property of the Fenchel transformation, for $1/r + 1/s = 1$ we have

$$0 \leq \mathbb{p}^{S_N} \leq NQ_s\left(\frac{1}{N}\sum_{i=1}^{N}\operatorname{op}(X^i), a^{-1}\right) = Q_s\left(\frac{1}{N}\sum_{i=1}^{N}\operatorname{op}(X^i), a_N\right),$$

$$\lim_{N\to+\infty} a_N = \lim_{N\to+\infty}\frac{1}{aN^{2/s}} = 0.$$

Then (1.16) is an immediate consequence of Corollary 1.1. Q.E.D.

We now state a result pertaining to the estimation of the optimal state via the $\oplus$-sum of independent variables.

Proposition 1.7 *Let X be a real optimization variable whose performance density is equal to $\mathbb{I}$ at a single point $\operatorname{op}(X) \in \mathbb{R}$. Let $(X^i)_{1\leq i\leq N}$, $n \geq 1$, be N independent optimization variables that have the same performance law as X. Then*

$$\forall\varepsilon > 0 \quad \lim_{N\to+\infty}\mathbb{P}\left(d\left(\bigoplus_{i=1}^{N}\mathbb{p}^X(X^i), \mathbb{p}^X(\operatorname{op}(X))\right) > \varepsilon\right) = 0. \tag{1.17}$$

Assume that for every $\varepsilon > 0$ there exists an $\eta > 0$ such that $d(x, \operatorname{op}(X)) \leq \varepsilon$ whenever $d(\mathbb{p}^X(x), \mathbb{I}) \leq \eta$. In that case,

$$\forall\varepsilon > 0 \quad \lim_{N\to+\infty}\mathbb{P}(d(\operatorname{op}_N(X)), \operatorname{op}(X)) > \varepsilon) = 0,$$

$$\operatorname{op}_N(X) \stackrel{\text{def}}{=} \operatorname{Arg}\sup_{1\leq i\leq N}\mathbb{p}^X(X^i). \tag{1.18}$$

Proof. It suffices to notice that $\mathbb{P}(d(\mathbb{p}^X(X), \mathbb{p}^X(\operatorname{op}(X))) > \varepsilon) < \mathbb{I}$ and, by virtue of the independence of the X^i,

$$\mathbb{P}\left(d\left(\bigoplus_{i=1}^{N}\mathbb{p}^X(X^i), \mathbb{I}\right) > \varepsilon\right) = \mathbb{P}(\forall i \in \{1,\ldots,N\} : d(\mathbb{p}^X(X^i), \mathbb{I}) > \varepsilon)$$
$$= N\mathbb{P}(d(\mathbb{p}^X(X), \mathbb{I}) > \varepsilon). \quad \text{Q.E.D.}$$

5. Optimization Processes

After introducing the performance theory axioms, an important step consists in introducing optimization sequences indexed by a subset of reals. This leads to a lot of additional structures and therefore deeper results. It will prove rather essential to our purpose in that it allows a tractable description of optimization problems. We first introduce the tools to be used later. Let $(E, \mathcal{E})$ be a Polish space and I a subset of $\mathbb{R}^+$.

5.1. *Definitions*

Definition 1.6 A E-valued *optimization process* with time space I is a system $(\Omega, \sigma, \mathbb{P}, X = \{X_t\}_{t\in I})$ defined by:

1. An optimization basis $(\Omega, \sigma, \mathbb{P})$.
2. A family of E-valued optimization variables $(X = \{X_t\}_{t\in I})$ defined on $(\Omega, \sigma, \mathbb{P})$.

The optimization variable X_t is called the state of X at time t, and the curve $t \mapsto X_t(\omega)$ is called the trajectory or path of $\omega \in \Omega$.

Remark 1.1 Let Ω be the class of all measurable functions from $[0,T]$ into $\mathbb{R}^n$, $[0,T] \in \mathbb{R}$, endowed with the uniform topology. Let σ be the induced Borel σ-algebra.

Assume that $\mathbb{P}$ is defined for every $A \in \sigma$ by the formula

$$\mathbb{P}(A) = \sup\{\mathbb{p}(\omega) : \omega \in A\} \quad \text{with } \mathbb{p}(\omega) = -\frac{1}{2}\int_0^T \|\omega_\tau\|^2_{Q_\tau^{-1}}\, d\tau,$$

and let $(Q_\tau^{-1})_{\tau\in[0,T]}$ be a sequence of positive definite matrices. For every $t \in [0,T]$, $U_t(\omega) \overset{\text{def}}{=} \omega_t$ is an optimization variable whose density is given by

$$\mathbb{p}_t^U(u) = \sup\{\mathbb{p}(\omega) : \omega \in \Omega \text{ such that } \omega_t = u\}.$$

Consequently, $U = (U_t)_{t\in[0,T]}$ is an optimization process. Let $X = (X_t)_{t\in[0,T]}$ be defined by

$$\begin{cases} \dot{X}_t = f(X_t) + g(X_t)U_t, & t \in [0,T], \\ X_0 = x_0, & \text{(the initial condition)}, \end{cases}$$

where f and g satisfy the usual Lipschitz and boundedness conditions. Then X is a $\mathcal{C}([0,T],\mathbb{R}^{n_X})$-valued optimization variable whose density $\mathbb{p}^X$ is upper semicontinuous and is defined by

$$\mathbb{p}^X(x) = \frac{1}{2}\int_0^T \|\dot{x}_\tau - f(x_\tau)\|^2_{(g(x_\tau),Q_\tau g(x_\tau))^{-1}} d\tau$$

if x is absolutely continuous, and $\mathbf{0}$ otherwise. Consequently, X_t, $t \in [0,T]$ is an $\mathbb{R}^n$-valued optimization variable with continuous density

$$\mathbb{p}_t^X(x) = \sup\{\mathbb{p}^X(z) : z \in \mathcal{C}([0,T],\mathbb{R}^n) \text{ such that } z_t = x\}.$$

5.2. *Modal semimartingales*

We now introduce modal semimartingales, which are tools for the study of the general equation of the conditional optimal states with respect to a regulation reference as in filtering theory ([13]). We also give an example to suggest how

these results may be useful in analyzing optimization problems. Let X be a discrete-time real-valued optimization process defined on $(\Omega, \sigma, \mathbb{P})$, and let $\mathcal{F}_t^X$ be the σ-algebra spanned by the optimization variables $X_0, \ldots, X_t$. By $\mathbf{F}^X = (\mathcal{F}_i^X)_{t\geq 0}$ we denote the induced increasing filtration of σ. We always assume that the optimal conditional states are well-defined and unique.

Definition 1.7 A discrete-time real-valued optimization process A is said to be $\mathbf{F}^X$-predictable if

$$\forall t \geq 0, \quad \text{op}(A_t/X_{\underline{t-1}}) \overset{\text{def}}{=} \text{op}(A_t/\mathcal{F}_{t-1}^X) = A_t$$

with $X_{\underline{\tau}} \overset{\text{def}}{=} (X_0, X_1, \ldots, X_\tau)$, $\tau \geq 0$. A discrete-time real-valued optimization process M is said to be a modal $\mathbf{F}^X$-martingale if

$$\forall t \geq s \geq 0, \quad \text{op}(M_t/X_{\underline{s}}) \overset{\text{def}}{=} \text{op}(M_t/\mathcal{F}_s^X) = M_s.$$

Consider the optimization processes

$$A_k^X \overset{\text{def}}{=} \sum_{l=1}^k \text{op}(\Delta X_l/\mathcal{F}_{l-1}^X), \qquad X - X_0 - A^X = M^X.$$

By construction, A^X is an $\mathbf{F}^X$-predictable optimization process and M^X is a modal $\mathbf{F}^X$-martingale which is null at $k = 0$. In [13] one can find the following result, which gives an analog of the general filtering equation for the evolution of the conditional optimal process (see also [13]). The following theorem simplifies the evolution of the conditional optimal states.

Theorem 1.4 *Let X and Y be the optimization semimartingales defined by*

$$X = X_0 + A^X + M^X, \qquad Y = Y_0 + A^Y + M^Y,$$

where A^X and A^Y denote their $\mathbf{F}^{X,Y}$-predictable part and M^X and M^Y their modal $\mathbf{F}^{X,Y}$-martingale parts. For every $\mathbf{F}^{X,Y}$-measurable optimization process Z, we define its $\mathbf{F}^Y$-optional projection $\widehat{Z}$ for every $k \geq 0$ by $\widehat{Z}_k \overset{\text{def}}{=} \text{op}(Z_k/\mathcal{F}_k^Y)$. Then

$$\widehat{X} = \widehat{X}_0 + A^{\widehat{X}} + M^{\widehat{X}} \quad \textit{and} \quad Y = Y_0 + A^{\widehat{Y}} + M^{\widehat{X}},$$

where

1. *$A^{\widehat{X}}$ and $A^{\widehat{Y}}$ denote the $\mathbf{F}^Y$-predictable processes defined for every $k \geq 0$ by*

$$A^{\widehat{X}} \overset{\text{def}}{=} \sum_{l=1}^k \text{op}(\Delta A_l^X/\mathcal{F}_{l-1}^Y), \qquad A^{\widehat{Y}} \overset{\text{def}}{=} \sum_{l=1}^k \text{op}(\Delta A_l^Y/\mathcal{F}_{l-1}^Y).$$

2. *$M^{\widehat{X}}$ and $M^{\widehat{Y}}$ are two modal $\mathbf{F}^Y$-martingales.*

Remark 1.2 Let $X_0, U_1, \dots, U_T, V_0, \dots, V_T$ be a sequence of independent real optimization variables. By X and Y we denote the real optimization processes defined by the equations

$$\Delta X_l = f(X_{l-1}) + g(X_{l-1})U_l, \qquad X_0 \text{ is the initial condition,}$$
$$Y_l = CX_l + V_l \qquad (0 \le l \le T),$$

where f and g are continuous real functions and $C \in \mathbb{R}$. In this case, we have

$$\Delta \widehat{X}_l = f(\widehat{X}_{l-1}) + \Delta M_k^{\widehat{X}}, \qquad \Delta M_k^{\widehat{Y}} = \Delta Y_l - Cf(\widehat{X}_{l-1}).$$

If the space of modal $\mathbf{F}^Y$-martingales is spanned by the processes

$$G \cdot \left(Y - \sum_{l=0}^{\cdot} Cf(\widehat{X}_{l-1}) \right)_k \stackrel{\text{def}}{=} \sum_{l=0}^{k} G_l(\Delta Y_l - Cf(\widehat{X}_{l-1})),$$

where G is $\mathbf{F}^Y$-predictable, then there exists an $\mathbf{F}^Y$-predictable process G such that

$$\Delta \widehat{X}_l = f(\widehat{X}_{l-1}) + G_l(\Delta Y_l - Cf(\widehat{X}_{l-1})).$$

Whenever f is linear, g is constant, and the optimization variables U and V are quadratic, the optimization process G coincides with the Kalman gain. In that special case, this process can be calculated by solving a Riccati equation [28].

5.3. *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 time inversion 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 1.8 Let $(\Omega, \sigma, \mathbb{P}, X = \{X_t\}_{t \in I})$ be an E-valued optimization process. It is called a *Maslov process* whenever its future and its past are independent provided that its present is known. In other words, X is a Maslov process if for any subdivision $s_1 \le \dots \le s_m \le t \le t_1 \le \dots \le t_n$, $n, m \ge 1$, of I, any $\varphi \in \mathcal{L}^0(E^{m+1}, \bigotimes_{i=0}^m \mathcal{E})$, and any $\psi \in \mathcal{L}^0(E^{n+1}, \bigotimes_{i=0}^n \mathcal{E})$, we $\mathbb{P}$-a.e. have

$$\begin{aligned} &\mathbb{E}(\varphi(X_{s_1}, \dots, X_{s_m}, X_t) \odot \psi(X_t, X_{t_1}, \dots, X_{t_n})/\sigma(X_t)) \\ &= \mathbb{E}(\varphi(X_{s_1}, \dots, X_{s_m}, X_t)/\sigma(X_t)) \odot \mathbb{E}(\psi(X_t, X_{t_1}, \dots, X_{t_n})/\sigma(X_t)). \quad (1.19) \end{aligned}$$

It is stated in [13] that the Bellman optimality equation for free evolution problems is the same as the Chapman–Kolmogorov transition equation for the associated Maslov process.

Proposition 1.8 *Let $(\Omega, \sigma, \mathbb{P}, X = \{X_t\}_{t \in I})$ be an E-valued Maslov process. For every $0 \le r \le s \le t$, $r, s \in I$, we have*

$$\mathbb{p}^X_{t/r}(z/x) = \int_E^{\oplus} \mathbb{p}^X_{t/s}(z/y) \odot \mathbb{p}^X_{s/r}(y/x) \odot dy$$

(the Bellman optimality equation),

where $\mathbb{p}^X_{\tau_2/\tau_1} \stackrel{\text{def}}{=} \mathbb{p}^{X_{\tau_2}/X_{\tau_1}}$, $\tau_1 \le \tau_2$.

Remark 1.3 The optimization process X defined in Example 1.1 is a Maslov process.

More generally, let L be an upper semicontinuous function from $\mathbb{R}^n \times \mathbb{R}^n$ into $\mathcal{A}$, and let L_0 be an upper semicontinuous function from $\mathbb{R}^n$ into $\mathcal{A}$. For every $x \in \mathbb{R}^n$ we assume that $L(x, \cdot)$ and S are performance densities. Let $\Omega = \mathbb{R} \times (\mathbb{R}^n)^{[0,T]}$, $[0,T] \subset \mathbb{R}$, be endowed with the uniform topology, and let σ be the induced Borel σ-algebra. Let X_0 be an optimization variable and U an optimization process defined on $(\Omega, \sigma, \mathbb{P})$. Let $X = (X_t)_{t \in [0,T]}$ be the optimization process defined by $\dot{X}_t = F(X_t, U_t)$ (X_0 is the initial condition and $t \in [0,T]$), where F satisfies the usual Lipschitz and boundedness conditions. Whenever X is a $(\mathcal{C}([0,T], \mathbb{R}^n), \mathcal{F}^X)$-valued optimization variable whose density $\mathbb{p}^X$ is upper semicontinuous and defined by

$$\begin{aligned} \mathbb{p}^X(x) &\stackrel{\text{def}}{=} \sup\{\mathbb{p}^{X_0, U}(x_0, u)/(x_0, u) \in \Omega : X(x_0, u) = x\} \\ &= \begin{cases} \sup\left\{ L_0(z_0) + \int_0^T L(x_\tau, u_\tau)\, d\tau / (z_0, u) \in \Omega : \right. \\ \qquad \left. \dot{x} = F(x, u),\ x_0 = z_0 \right\} & \text{if } x \text{ is absolutely continuous,} \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

X is a Maslov process by virtue of the usual argument. In the sequel, by $\mathcal{F}^X_{t_1}$ we denote the sub-σ-algebra of $\mathcal{F}^X$ spanned by the optimization variables $(X_\tau)_{0 \le \tau \le t_1}$. We say that X is a *n-dimensional (F, L)-Maslov optimization process with respect to the filtration* $(\mathcal{F}^X_t)_{0 \le t \le T}$.

Remark 1.4 (All previous notation is in force.) Let Ω be the product of $\mathbb{R}$ by the class of all measurable functions from $[0,T]$ into $\mathbb{R}^2$, $T>0$, endowed with the uniform topology, and let σ be the induced Borel σ-algebra. By X_0 we denote a real optimization variable and by U, V, two $\mathbb{R}^{[0,T]}$-valued optimization variables defined on $(\Omega,\sigma,\mathbb{P})$ and such that, with some abusive notation,

$$\mathbb{p}(x_0,u,v) \stackrel{\text{def}}{=} -\frac{1}{2}(x_0-\overline{x}_0)^2 - \frac{1}{2}\int_0^T u_\tau^2\, d\tau - \frac{1}{2}\int_0^T v_\tau^2\, d\tau.$$

The optimization process (X,Y) defined by $\dot{Y}=h(X)+V$, $\dot{X}=f(X)+g(X)U$ (X_0 is the initial condition), where f, g, and h satisfy the usual Lipschitz and boundedness conditions, is a two-dimensional (F,L)-Maslov optimization process with respect to the filtration $(\mathcal{F}_t^X)_{0\le t\le T}$ with $L_0(x_0)=-\frac{1}{2}(x_0-\overline{x}_0)^2$ and

$$F((x,y),(u,v)) = (f(x)+g(x)u, h(x)+v),$$
$$L((x,y),(u,v)) = -\frac{1}{2}u^2 - \frac{1}{2}(y-h(x))^2.$$

Moreover, it is easy to show that Y is also a real Maslov optimization process with respect to the filtration $(\mathcal{F}_t^Y)_{0\le t\le T}$. Nevertheless, X is not a Maslov optimization process with respect to the filtration $(\mathcal{F}Y_t)_{0\le t\le T}$, in the sense that for every $0\le s\le t$ we have $\mathbb{p}(x_t/y_{\underline{s}}) \ne \mathbb{p}(x_t/y_s)$. These facts will be further developed in §6.

5.4. *Optimization martingales*

This part constitutes the final step on our way to general theorems of Maslov optimization theory. It is devoted to two closely related kinds of results: one is the $(\max,+)$-version of Doob's inequality, and the second one is the $(\max,+)$-version of Doob's up-down crossing lemma. Let X be a discrete-time real-valued optimization process defined on $(\Omega,\sigma,\mathbb{P})$, and let $\mathcal{F}_t^X$ be the σ-algebra spanned by the optimization variables $X_0,\dots,X_t$. We denote by $\mathbf{F}^X=(\mathcal{F}_t^X)_{t\ge 0}$ the induced increasing filtration of σ. For simplicity, here we only consider real-valued optimization processes; this can readily be generalized (the details of the generalization are left to the reader). As in probability theory, processes of interest are the optimization martingales.

Definition 1.9 Let $(M_t)_{t\ge 0}$ be an optimization process defined on $(\Omega,\sigma,\mathbb{P})$ and adapted to an increasing sequence of σ-algebras $(\mathcal{F}_t^X)_{t\ge 0}$. The process $(M_t)_{t\ge 0}$ is called an optimization $\mathbf{F}^X$-martingale (resp., submartingale, supermartingale) if

1. Each optimization martingale M_t is integrable.
2. For every $0\le s\le t$, $\mathbb{E}(M_t/\mathcal{F}_s^X) = (\text{resp.}, \ge, \le) M_s$

By $\mathbb{L}(\mathbf{F}^X)$ we denote the $(\oplus,\odot)$-semimodule of optimization $\mathbf{F}^X$-martingales M with $\mathbb{E}(M_0)=\mathbb{1}$.

Remark 1.5 All notation and assumptions of Example 1.4 are in force. Let $\mathbb{P}_0$ be the new performance measure on $(\Omega,\mathcal{F}_T)$ defined for every $0\le t\le T$ by the formula

$$\frac{d\mathbb{P}}{d\mathbb{P}_0}{}^{\circ}/\mathcal{F}_t \stackrel{\text{def}}{=} Z_t = \int_0^t h(X_\tau)' R_\tau^{-1} Y_\tau\, d\tau - \frac{1}{2}\int_0^t \|h(X_\tau)\|^2_{R_\tau^{-1}}.$$

Under $\mathbb{P}_0$, X is unchanged and Y becomes a Maslov process, independent of X, with performance density

$$\mathbb{p}_0^Y(y) = -\frac{1}{2}\int_0^T \|y_\tau\|^2_{R_\tau^{-1}}\, d\tau.$$

Let $\mathbb{E}_0(\cdot)$ be the conditional expectation associated with the performance reference measure $\mathbb{P}_0$. We can readily establish that Z is a $\mathbb{P}_0$-optimization martingale such that $\mathbb{E}_0(Z_0)=\mathbb{1}$. Let $(\mathcal{F}_t^Y)_{t\in[0,T]}$ be the increasing filtration associated with the optimization process Y on $[0,T]$; in [13] we state the following analog of Kallianpur–Striebel formula:

For every $t\in[0,T]$ and $\varphi\in\mathbb{L}(\Omega,\mathcal{F}_t,\mathbb{P})$, $\mathbb{P}$-a.e. we have

$$\mathbb{E}(\varphi/\mathcal{F}_t^Y) = \frac{\mathbb{E}_0(\varphi\odot Z_t/\mathcal{F}_t^Y)}{\mathbb{E}_0(\mathbb{1}\odot Z_t/\mathcal{F}_t^Y)}{}^{\circ}.$$

Let us now recall the main properties of the optimization martingales such as the analog of the Doob up-down crossing lemma, which ensures the existence of the closure of the optimization supermartingale. The significance of these results will be clarified in §6.

Proposition 1.9 *Let M be an optimization $\mathbf{F}^X$-submartingale, $a\in\mathcal{A}$, and $T>0$. Then*

$$a\odot\mathbb{P}\Big(\Big\{\omega\in\Omega:\sup_{0\le t\le T} M_t(\omega)\ge a\Big\}\Big) \le \int^{\oplus}_{\{\omega\in\Omega:\sup_{0\le t\le T}M_t(\omega)\ge a\}} M_T\odot\mathbb{P} \le \mathbb{E}(M_T). \tag{1.20}$$

One can also combine Markov's inequality with the supermartingale property.

Proposition 1.10 *Let M be an optimization $\mathbf{F}^X$-supermartingale, and let $a\in\mathcal{A}$ and $T>0$. Then*

$$a\odot\mathbb{P}\Big(\Big\{\omega\in\Omega:\sup_{0\le t\le T} M_t(\omega)\ge a\Big\}\Big) \le \mathbb{E}(M_0). \tag{1.21}$$

We now introduce the Doob up-down crossing lemma in our framework (for details, the reader is referred to [13]).

Lemma 1.1 *Let M be an optimization $\mathbf{F}^X$-supermartingale. For every $0 \leq a \leq b < +\infty$, $b \neq a$, we have*

$$\forall p \geq 1 \qquad \mathbb{P}(\{\omega \in \Omega : U_M([a,b],\omega) \geq p\}) \leq \left(\frac{a}{b}\circ\right)^{p-1} \odot \frac{\mathbb{E}(M_0)}{b}\circ, \tag{1.22}$$

$$\mathbb{P}(\{\omega \in \Omega : D_M([a,b],\omega) \geq p\}) \leq \left(\frac{a}{b}\circ\right)^{p-1} \odot \frac{\mathbb{E}(M_0)}{a}\circ, \tag{1.23}$$

where

1. *For every $a, b \in \mathcal{A}$, $b \neq 0$, $\frac{a}{b}\circ \stackrel{\text{def}}{=} a - b$, $a^q \stackrel{\text{def}}{=} qa$*
2. *$U_M([a,b],\omega)$ (resp. $D_M([a,b],\omega)$) denotes the number of up-crossings (resp. down-crossings) of the paths $t \mapsto M_t(\omega)$ over $[a,b] \in \mathcal{A}$.*

In view of the previous lemma, for every optimization $\mathcal{F}^X$-supermartingale there exists an optimization variable M_∞ such that

$$\mathbb{P}\left(\limsup_{t\to+\infty} \rho(M_t, M_\infty) \geq \varepsilon\right) = 0 \qquad \forall \varepsilon > 0.$$

6. Applications

We are now in a position to expound most of the consequences of the above results for (F,L)-Maslov processes. We choose the shortest possible route, thus leaving apart a large number of interesting properties (see [13, 17, 18]). We also introduce the Hamiltonian associated with an (F,L)-Maslov process. This function is an essential tool to obtain the Kolmogorov operator and the associated Dynkin formula in such a framework. Let X be an $\mathbb{R}^n$-valued (F,L)-Maslov process defined on a given performance space $(\Omega, \sigma, \mathbb{P})$. We write, for every $\xi, x, u \in \mathbb{R}^n$,

$$H^X(\xi, x, u) \stackrel{\text{def}}{=} \xi' F(x,u) + L(x,u), \qquad H^X(\xi, x) \stackrel{\text{def}}{=} \sup_{u\in\mathbb{R}} H^X(\xi, x, u).$$

Theorem, Definition 2 *Let X be a (F,L)-Maslov process defined on the filtered performance basis $(\Omega, \mathcal{F}, \mathbf{F}^X, \mathbb{P})$, $\mathbf{F}^X \stackrel{\text{def}}{=} (\mathcal{F}_t^X)_{t\geq 0}$. For every continuously differentiable function ϕ, there exists an optimization martingale $M \in \mathbb{L}(\mathbf{F}^X)$ such that for every $t \geq 0$ we have*

$$\phi(X_t) = \phi(X_0) + \int_0^t H^X\left(X_\tau, \frac{d\phi}{dx}(X_\tau)\right) d\tau + M_t. \tag{1.24}$$

The operator

$$\mathcal{H}^X(\phi)(x) \stackrel{\text{def}}{=} H^X\left(x, \frac{d\phi}{dx}(x)\right)$$

is called the Hamilton–Jacobi operator associated with X. Furthermore,

$$\mathbb{E}\left(\phi(X_t) - \int_0^t (\mathcal{H}^X\phi)(X_\tau)\, d\tau / \mathcal{F}_0^X\right) = \phi(X_0) \quad \text{(Dynkin's formula)}. \tag{1.25}$$

Proof. Let $H^X(u/\xi, x) \stackrel{\text{def}}{=} H^X(\xi, x, u) - H^X(\xi, x)$. For every $t \geq s \geq 0$, one has $\mathbb{P}$-a.e.

$$\begin{aligned}\mathbb{E}(M_t/\mathcal{F}_s^X) &= M_s + \sup_{u\in\Omega}\left(\int_s^t \left(H\left(X_\tau, \frac{d\phi}{dx}(X_\tau), u_\tau\right)\right.\right.\\ &\qquad\left.\left. - H\left(X_\tau, \frac{d\phi}{dx}(X_\tau)\right)\right) d\tau\right)\\ &= M_s + \sup_{u\in\Omega}\left(\int_s^t H\left(u_\tau/X_\tau, \frac{d\phi}{dx}(X_\tau)\right) d\tau\right)\\ &= M_s. \qquad \text{Q.E.D.}\end{aligned}$$

Whenever ϕ is time dependent, we obviously obtain the same equation with the operator ∂_t. The following consequences are illustrations of the results stated in the previous section. These results lead to new developments in the field of qualitative studies of optimization processes mainly because they exhibit explicit bounds of the cost function over some classes of optimization variables. For this purpose, the (max, +)-version of Dynkin's formula is first required to construct optimization martingales. Let ϕ be a continuously differentiable function and $0 \leq s \leq t$. In view of the preceding theorem, Propositions 1.9 and 1.10, and Lemma 1.1, we obtain the following assertion.

1. If $\mathcal{H}^X(\phi) = 0$, then $\phi(X_t)$ is an optimization martingale, and this condition gives a mean to calculate conditional Maslov expectations (that is, $\mathbb{E}(\phi(X_t)/\mathcal{F}_s) = \phi(X_s)$).

In other words, if $\Omega_{s,x} = \{\omega \in \Omega : X_s(\omega) = x\}$, then

$$\sup_{(x_0,u)\in\Omega_{s,x}} \left(\phi(X_t) + \int_s^t L(x_\tau, u_\tau)\right) \doteq \phi(x).$$

2. If $\mathcal{H}^X(\phi) \geq 0$, then $\phi(X_t)$ is an optimization submartingale (that is, $\mathbb{E}(\phi(X_t)/\mathcal{F}_s) \geq \phi(X_s)$). By straightforward application of Proposition 1.9, we have, for every $a \in \mathcal{A}$,

$$\sup_{(x_0,u)\in\Omega_a} \left(L_0(x_0) + \int_0^T L(x_\tau, u_\tau)\right) \leq \frac{\mathbb{E}(\phi(X_T))}{a}\circ,$$

where $\Omega_a = \{\omega \in \Omega : \sup_{0\leq t\leq T} \phi(X_t(\omega)) \geq a\}$ and $\dot{x} = F(x,u)$; x_0 is the initial condition.

3. If $\mathcal{H}^X(\phi) \leq 0$, then $\phi(X_t)$ is an optimization supermartingale (that is, $\mathbb{E}(\phi(X_t)/\mathcal{F}_s) \leq \phi(X_s)$). By straightforward application of Proposition 1.9, we have, for every $x \in \mathbb{R}^n$ and $a \in \mathcal{A}$,

$$\sup_{(x_0,u)\in\Omega_{a,x}} \int_0^T L(x_\tau, u_\tau) \leq \frac{\phi(x)}{a}{}_\circ,$$

where $\Omega_{a,x} = \{\omega \in \Omega : X_0(\omega) = x,\ \sup_{0\leq t\leq T} \phi(X_t(\omega)) \geq a\}$.

Similarly, for every $p \geq 1$ and $a \leq b$, we have

$$\sup_{(x_0,u)\in\Omega_{[a,b],x}} \int_0^T L(x_\tau, u_\tau) \leq \left(\frac{a}{b}{}_\circ\right)^{p-1} \odot \frac{\phi(x)}{b}{}_\circ,$$

where $\Omega_{[a,b],x} = \{\omega \in \Omega : X_0(\omega) = x,\ M_{\phi(X)}([a,b],\omega) \geq p\}$.

It follows from the above that for every x there exists an optimization variable $\mathrm{op}(X_\infty/X_0) = x$ such that

$$\sup_{(x_0,u)\in\Omega_\infty^{\varepsilon,x}} \int_0^T L(x_\tau, u_\tau) = \mathbb{0}$$

for every $\varepsilon > 0$, where

$$\Omega_\infty^{\varepsilon,x} = \{\omega \in \Omega : X_0(\omega) = x,\ \limsup_{t\to+\infty} \rho(X_t(\omega), X_\infty(\omega)) \geq \varepsilon\}.$$

In other words,

$$\mathrm{op}(X_t/X_0 = x) \underset{t\to+\infty}{\longrightarrow} \mathrm{op}(X_\infty/X_0 = x).$$

4. Whenever $\mathcal{H}^X(\phi) \leq -b$, with $b > 0$, we obtain a time-explicit majorant

$$\sup_{(x_0,u)\in\Omega_{a,x}} \int_0^T L(x_\tau, u_\tau) \leq \frac{\phi(x)}{a \odot bT}{}_\circ \Big(\underset{T\to+\infty}{\longrightarrow} \mathbb{0}\Big),$$

where $\Omega_{a,x} = \{\omega \in \Omega : X_0(\omega) = x,\ \sup_{0\leq t\leq T} \phi(X_t(\omega)) \geq a\}$. In that special case, there exists some $T \gg 0$ such that

$$\mathbb{P}(\Omega - \Omega_{a,x}/X_0 = x) = \mathbb{1} \quad \text{and} \quad \sup_{0\leq t\leq T} \phi(\mathrm{op}(X_t/X_0 = x)) \leq a.$$

The last statement pertains only to free evolution problems, thus avoiding conditional optimization. Now we want to code the information in a different way, namely, the one that is compatible with the way we look at the conditional performance measure. Here we describe, with the aid of an example, a statement that captures the main idea, and we state the conditional performance evolution similar to that of filtering theory. All notation and assumptions of Example 1.4 are in force. We deal with quadratic assumptions

for the optimization variables X_0, U, and V, although these can readily be generalized. It is immediate to check that the Hamiltonian associated with the Maslov process X is given by the equation

$$H^X(x,\xi) = \xi f(x) + \frac{1}{2}(\xi g(x))^2 \qquad \forall \xi, x \in \mathbb{R}.$$

Therefore, for every continuously differentiable real function ϕ we have

$$\phi(X_t) - \phi(X_0) - \int_0^t \Big(\partial_x \phi\Big(X_\tau\Big) + \frac{1}{2}(\partial_x\phi(X_\tau)g(X_\tau))^2\Big) d\tau \in \mathbb{L}(\mathbf{F}^X).$$

One can also combine the latter with a reference optimization process Y, so that

$$\phi(X_t) - \phi(X_0) - \int_0^t \Big(\partial_x \phi(X_\tau) + \frac{1}{2}(\partial_x\phi(X_\tau)g(X_\tau))^2 - \frac{1}{2}(Y_\tau - h(X_\tau))^2\Big) d\tau \in \mathbb{L}(\mathbf{F}^{X,Y}).$$

Let us now examine how the notion of a conditional performance translates in these settings. With some obvious abusive notation, the performance density of $(X_t, Y_{\underline{t}})$, where $Y_{\underline{t}} \stackrel{\text{def}}{=} Y/[0,t]$, is given for every $x \in \mathbb{R}$ and $y_{\underline{t}} \in \mathbb{R}^{[0,t]}$ by the equation

$$\begin{aligned}\mathbb{p}_t(x, y_{\underline{t}}) &= \mathbb{P}(X_t = x \text{ and } \forall 0 \leq \tau \leq t,\ Y_\tau = y_\tau)\\ &= \sup_{(x_0,u):X_t(x_0,u)=x} \Big\{ -\frac{1}{2}(x_0 - \overline{x}_0)^2 - \frac{1}{2}\int_0^t u_\tau^2\, d\tau - \frac{1}{2}\int_0^t (y_\tau - h(x_\tau))^2\, d\tau \Big\},\end{aligned}$$

$$\mathbb{p}_t(x, y_{\underline{t}}) = \int_{\mathbb{R}}^{\oplus} \mathbb{p}_{t/s}(x, y_{\underline{t}}/z, y_{\underline{s}}) \odot \mathbb{p}_s(z, y_{\underline{s}}) \odot dz$$

with

$$\begin{aligned}\mathbb{p}_{t/s}(x, y_{\underline{t}}/z, y_{\underline{s}}) &\stackrel{\text{def}}{=} \frac{\mathbb{p}_{t,s}(x, z, y_{\underline{t}})}{\mathbb{p}_s(z, y_{\underline{s}})}{}_\circ\\ &= \sup_{(x_0,u):(X_s,X_t)(x_0,u)=(z,x)} -\frac{1}{2}\int_s^t u_\tau^2\, d\tau - \frac{1}{2}\int_s^t (y_\tau - h(x_\tau))^2\, d\tau.\end{aligned}$$

Consequently,

$$\begin{aligned}\mathbb{p}_t(x/y_{\underline{t}}) &= \frac{\mathbb{p}_t(x, y_{\underline{t}})}{\mathbb{p}(y_{\underline{t}})}{}_\circ \qquad \Big(\mathbb{p}(y_{\underline{t}}) \stackrel{\text{def}}{=} \int_{\mathbb{R}}^{\oplus} \mathbb{p}_t(x, y_{\underline{t}}) \odot dx\Big)\\ &= \int_{\mathbb{R}}^{\oplus} \frac{\mathbb{p}_{t/s}(x, y_{\underline{t}}/z, y_{\underline{s}})}{\mathbb{p}(y_{\underline{t}}/y_{\underline{s}})}{}_\circ \odot \mathbb{p}_s(z/y_{\underline{s}}) \odot dz\end{aligned}$$

(the Bellman optimality equation).

By the same line of argument as for free evolution problems, the conditional optimization process described by the conditional performance measure $\mathbb{P}(\cdot/Y)$ is $\mathbb{P}$-a.e. a Maslov process. Such processes are called *conditional Maslov processes* or *regulation processes*. Note the presence of an initial penalty, which makes this a maximum-likelihood type problem. For a regulation problem of control type, one has a terminal penalty, and the natural time is reversed. In the sequel, we mainly deal with optimization of the first type, i.e., forward type.

The Maslov measure $q_t(x) \stackrel{\text{def}}{=} p_t(x, y_{\underline{t}})$ is given by the "*unnormalized*" Hamilton–Jacobi equation

$$\partial_t q = (\mathcal{H}^X q) - \frac{1}{2}(y_t - h)^2, \qquad q_0(x) = -\frac{1}{2}(x - \overline{x}_0)^2.$$

Therefore, the conditional performance $p_t(x/Y_{\underline{t}})$ is given by the following Hamilton–Jacobi equation:

$$\begin{aligned} \partial_t p &= (\mathcal{H}^X p) + (h - \widehat{h}_t)(Y_t - \widehat{h}_t) - \frac{1}{2}(h - \widehat{h}_t)^2, \\ p_0(x) &= -\frac{1}{2}(x - \overline{x}_0)^2. \end{aligned} \tag{1.26}$$

Here $\widehat{X}_t \stackrel{\text{def}}{=} \text{op}(X_t/Y_{\underline{t}})$, $\widehat{h}_t = h(\widehat{X}_t)$. Suppose that $p_t(x/Y_{\underline{t}})$ is a smooth solution of the Hamilton-Jacobi equation (1.26) and the conditional optimal state $\widehat{X}$ is well defined. Then

$$\dot{\widehat{X}}_t = \hat{f}_t - \widetilde{P}(t)^{-1}\widehat{\partial_x h}_t(Y_t - \widehat{h}_t), \qquad \widehat{X}_0 = \overline{x}_0.$$

Here $\widetilde{P}(t) \stackrel{\text{def}}{=} (\partial^2_{xx} p_t(\widehat{X}_t/Y_{\underline{t}}))$, $\hat{f}_t = f(\widehat{X}_t)$, and $\widehat{\partial_x h}_t = \partial_x h_t(\widehat{X}_t)$.

For linear optimization processes, $\widetilde{P}$ is a solution of the usual Riccati equation. From a practical point of view, this conditional optimal state, as well as the conditional expectation induced by a given nonlinear filtering problem, requires infinite-dimensional computations.

7. Maslov and Markov Processes

The main purpose of this section is to show that Maslov optimization processes and Markov stochastic processes can be mapped into each other by various transformations. We introduce some transformations between performance and probability measures which make clear the relationships between optimization and estimation problems.

7.1. *Wentzell-Freidlin Transform*

This subsection deals with two closely related kinds of results. One is the integral extension of the formulas

$$a \oplus b = \lim_{\epsilon \to +\infty} \epsilon \log\left(e^{\frac{a}{\epsilon}} + e^{\frac{b}{\epsilon}}\right) \quad \text{and} \quad a \odot b = \epsilon \log\left(e^{\frac{a}{\epsilon}} . e^{\frac{b}{\epsilon}}\right) \qquad a, b \in \mathbb{R}.$$

The first place where this study appears is Pontryagin–Andronov–Vitt [43]; it is further developed in [23] by Wentzell and Freidlin and Hijab [25]. It is well know that the study of various limit theorems for random processes is motivated by dynamical systems subject to the effect of random perturbations sufficiently small compared with the deterministic constituents of the motion. In order to study the effect of perturbations on large time intervals, we must be able to estimate the probability of rare events. The so-called Wentzell-Freidlin transform provides a way of computing the probability of rare events in terms of performance measures. Roughly speaking, for some random variable sequence X^ϵ and some optimization variable X ranging in the same measurable space, we have

$$\forall A \in \mathcal{E} \qquad \epsilon \log P(X^\epsilon \in A) \stackrel{\epsilon \sim 0}{\approx} \mathbb{P}(X \in A). \tag{1.27}$$

This investigation includes also results like the law of large numbers and the central limit theorem. The second motivation is to introduce an asymptotic mapping between conditional Markov processes and conditional Maslov processes. In other words, our interest is in large deviation results for conditional measures and related asymptotics of the filtering equations. The topology used in this mapping is the Prokhorov topology of probabilities, which is equivalent to the weak topology. More precisely, we show that the conditional expectation weakly converges to the conditional optimal state. The Maslov optimization theory allows a very tractable description of these results. We give only a brief exposition and leave apart a large number of properties. We suppose that the reader is familiar with the basic facts about nonlinear filtering and, in particular, with the fundamentals of the so-called change-of-measure approach. We deal with the simplest case. First, we consider a complete probability basis $(\Omega, \mathcal{F}_T, P)$, $T > 0$, and an increasing filtration $(\mathcal{F}_t)_{t\in[0,T]}$ on which two independent and real Wiener processes W and V are defined as well as an independent real-valued random variable X_0^ϵ, whose probability distribution μ_0^ϵ is given by

$$\mu_0^\epsilon(dx) = C_\epsilon \exp\left(\frac{1}{\epsilon} S_0(x)\right) dx,$$

where C_ϵ is a positive normalization constant and S_0 a Lipschitz concave Maslov performance density such that $S_0\left(\overline{X}_0\right) = 0$ and $S_0(x) < 0$ whenever $x \neq \overline{X}_0$.

We choose two real functions f and g such that the following equations have a strong solution on $(\Omega, \mathcal{F}_T, P)$:

$$\begin{cases} dX_t^\epsilon = f(X_t^\epsilon)\,dt + \sqrt{\epsilon} dW_t \\ X^\epsilon(0) = X_0^\epsilon \end{cases}, \quad \text{and} \quad \begin{cases} dY_t^\epsilon = h(X_t^\epsilon)\,dt + \sqrt{\epsilon} dV_t \\ Y^\epsilon(0) = 0. \end{cases}$$

We set $\Omega_t \overset{def}{=} C([0,T],\mathbb{R})$, $\Omega_{0,t} \overset{def}{=} \{\eta \in \Omega_t \,:\, \eta(0) = 0\}$, and $\Omega_0 = \Omega_{0,T}$ and equip these spaces with the uniform topology. In the sequel, $\mathbf{B}(\mathcal{X})$ denotes the Borel sigma-algebra of a topological space $\mathcal{X}$. The following theorem gives a path integral representation for the conditional expectation, which is known as the Kallianpur–Striebel formula.

Theorem 1.5 *Let φ be a bounded measurable function from Ω_T into $\mathbb{R}$; then, P-a.e. we have*

$$E\left(\varphi(X^\epsilon) \big/ \mathcal{F}_T^{Y^\epsilon}\right) = \frac{\int \varphi(\theta)\, Z^\epsilon(\theta, Y^\epsilon)\, P^{X^\epsilon}(d\theta)}{\int Z^\epsilon(\theta, Y^\epsilon)\, P^{X^\epsilon}(d\theta)} = \frac{E_0\left(\varphi(X^\epsilon)\, Z^\epsilon(X^\epsilon, Y^\epsilon) / \mathcal{F}_T^{Y^\epsilon}\right)}{E_0\left(Z^\epsilon(X^\epsilon, Y^\epsilon) / \mathcal{F}_T^{Y^\epsilon}\right)},$$

where $\quad \epsilon \log Z^\epsilon(\theta, \eta) \overset{def}{=} \int_0^T h(\theta_\tau)\, d\eta_\tau - \frac{1}{2}\int_0^T h(\theta_\tau)^2\, d\tau$

and $E_0(\cdot)$ is the expectation associated with the probability measure P_0 defined by

$$\frac{dP}{dP_0}\Big|_{\mathcal{F}_T} = Z^\epsilon(X^\epsilon, Y^\epsilon).$$

Using the Itô formula of integration by parts, we obtain

$$h(X_t^\epsilon)\, dY_t^\epsilon = d(h(X_t^\epsilon) Y_t^\epsilon) - Y_t^\epsilon (\mathcal{L}^\epsilon h)(X_t^\epsilon)\, dt - Y_t^\epsilon \left(\frac{\delta h}{\delta x}\right)(X_t^\epsilon) \sqrt{\epsilon} dW_t,$$

where $\mathcal{L}^\epsilon$ is the Kolmogorov operator associated with the diffusion X^ϵ. Then

$$\epsilon \log Z^\epsilon(X^\epsilon, Y^\epsilon) = F^\epsilon(X^\epsilon, Y^\epsilon) - \int_0^T \sqrt{\epsilon} Y_t^\epsilon \left(\frac{\delta h}{\delta x}\right)(X_t^\epsilon)\, dW_t - \frac{1}{2}\int_0^T \left(Y_t^\epsilon \left(\frac{\delta h}{\delta x}\right)(X_t^\epsilon)\right)^2 dt$$

with

$$V^\epsilon(x,y) = \frac{1}{2}h^2(x) + y\left(f(x)\frac{\delta h}{\delta x}(x) + \frac{1}{2}\epsilon\frac{\delta^2 h}{\delta x^2}(x)\right) - \frac{1}{2}(y\frac{\delta h}{\delta x}(x))^2,$$

$$F^\epsilon(\theta, \eta) = h(\theta_T)\,\eta_T - \int_0^T V^\epsilon(\theta_t, \eta_t)\, dt.$$

Using Girsanov's theorem, we obtain:

$$E\left(\varphi(X^\epsilon) \big/ \mathcal{F}_T^{Y^\epsilon}\right) = \frac{\int_{\Omega_T} \varphi(\theta)\ exp\left(\frac{1}{\epsilon}F^\epsilon(\theta, Y^\epsilon)\right)\ \check{P}^\epsilon_{[Y^\epsilon]}(d\theta)}{\int_{\Omega_T} exp\left(\frac{1}{\epsilon}F^\epsilon(\theta, Y^\epsilon)\right)\ \check{P}^\epsilon_{[Y^\epsilon]}(d\theta)},$$

where $\check{P}^\epsilon_{[Y^\epsilon]}(d\theta)$ is the distribution on Ω_T of the diffusion

$$d\check{X}_t^\epsilon = \left(f\left(\check{X}_t^\epsilon\right) - Y_t^\epsilon \frac{\delta h}{\delta x}\left(\check{X}_t^\epsilon\right)\right) dt + \sqrt{\epsilon} dW_t \qquad \text{with} \quad \check{X}_0^\epsilon = X_0^\epsilon.$$

The term F^ϵ does not involve stochastic integration and thus is well defined for all $\eta \in \Omega_0$ rather than only on a set of ϵ-Wiener measure equal to one, so that the final estimates can be made uniform with respect to the parameter η over a compact subset in Ω_0. Futhermore, it continuously depends on $\eta \in \Omega_0$. These properties are inherited by the following measures ($A \in \mathbf{B}(\Omega_T)$, $B \in \mathbf{B}(\mathbb{R})$, $\eta \in \Omega_0$, $t \in [0,T]$):

$$\Sigma^\epsilon(A)(\eta) = \int_A e^{\frac{1}{\epsilon}F^\epsilon(\theta,\eta)} \check{P}^\epsilon_{[\eta]}(d\theta) \quad \sigma_t^\epsilon(B)(\eta) = \Sigma^\epsilon(\{\theta \in \Omega_T \,:\, \theta_t \in B\})(\eta),$$

$$\Pi^\epsilon(A)(\eta) = \frac{\Sigma^\epsilon(A)(\eta)}{\Sigma^\epsilon(\Omega_T)(\eta)} \quad \pi_t^\epsilon(B)(\eta) = \frac{\sigma_t^\epsilon(B)(\eta)}{\sigma_t^\epsilon(\mathbb{R})(\eta)}.$$

Let $(\Omega, \mathcal{F}_T, \mathbb{P})$ be a performance space equipped with a filtration $(\mathcal{F}_t)_{t\in[0,T]}$ on which two real $\mathbb{P}$-independent optimization processes U^X and U^Y are defined. We also define a real optimization variable X_0 with performance distribution $\mathbb{P}_0^X = S_0$. Assume that for every $(x,u,v) \in \mathbb{R} \times \Omega_T \times \Omega_T$ we have

$$\mathbb{P}^{X_0, U^X, U^Y}(x,u,v) = \begin{cases} S_0(x) - \frac{1}{2}\int_0^T \|u_t\|^2 dt - \frac{1}{2}\int_0^T \|v_t\|^2 dt \\ \quad \text{(whenever the integrals exist)} \\ 0 \quad \text{(otherwise)}. \end{cases}$$

We consider also the pair of real-valued optimization processes defined by

$$\dot{X}_t = f(X_t) + U_t^X, \quad X(0) = X_0 \quad \text{and} \quad \dot{Y}_t = h(X_t) + U_t^Y, \quad Y(0) = 0.$$

Let $\mathbb{P}^{X,Y}$ be the performance density on $(\Omega_T \times \Omega_0, \mathbf{B}(\Omega_T) \otimes \mathbf{B}(\Omega_0))$ of the pair of optimization variables (X,Y). For every $(x,y) \in \Omega_T \times \Omega_0$ we have

$$\mathbb{P}^{X,Y}(x,y) = S_0(x) - \frac{1}{2}\int_0^T \|\dot{x}_t - f(x_t)\|^2 dt - \frac{1}{2}\int_0^T \|\dot{y}_t - h(x_t)\|^2 dt$$

if x and y are absolutely continuous, and $\mathbb{P}^{X,Y}(x,y) = \mathbf{0}$ otherwise. One can check that $\mathbb{P}^{X,Y}$ is upper semicontinuous over $(\Omega_T \times \Omega_0)$ and

$$\mathbb{P}^X(x) = S_0(x_0) - \frac{1}{2}\int_0^T \|\dot{x}_t - f(x_t)\|^2 dt,$$

$$\mathbb{P}^Y(y) = \sup_{x\in\Omega^n x} \mathbb{P}^{X,Y}(x,y),$$

$$\mathbb{P}^{Y/X}(y/x) = \begin{cases} \frac{\mathbb{P}^{X,Y}(x,y)}{\mathbb{P}^X(x)}\circ = -\frac{1}{2}\int_0^T \|\dot{y}_t - h(x_t)\|^2\, dt & \text{if } \mathbb{P}^X(x) > 0\,, \\ 0 & \text{otherwise.} \end{cases}$$

Finally, for every bounded and measurable function φ from Ω_T into $\mathcal{A}$, one $\mathbb{P}$-*a.e.* has

$$\mathbb{E}\left(\varphi(X)/\mathcal{F}_T^Y\right) = \frac{\int^{\oplus} \varphi(\theta)\odot \mathbb{P}^{Y/X}(Y/\theta)\odot\mathbb{P}^X(\theta)\odot d\theta}{\int^{\oplus} \mathbb{P}^{Y/X}(Y/\theta)\odot\mathbb{P}^X(\theta)\odot d\theta}\circ$$
$$= \frac{\mathbb{E}_0\left(\varphi(X)\odot Z(X,Y)/\mathcal{F}_T^Y\right)}{\mathbb{E}_0\left(Z(X,Y)/\mathcal{F}_T^Y\right)}\circ,$$

where $\mathbb{E}_0(.)$ denotes the Maslov expectation associated with the performance measure on $(\Omega,\mathcal{F}_T)$ defined by

$$\frac{d\mathbb{P}}{d\mathbb{P}_0}^{\ominus}/_{\mathcal{F}_T} = Z(X,Y) = \int_0^T h(X_t)\,\dot{Y}_t dt - \frac{1}{2}\int_0^T h^2(X_t)\,dt.$$

By the same line of argument as before, we use these formulas to define several measures as follows ($A\in\mathbf{B}(\Omega_T)$, $B\in\mathbf{B}(\mathbb{R})$, $\eta\in\Omega_0$, and $t\in[0,T]$):

$$\Sigma(A)(\eta) = \int_A^{\oplus} Z(\theta,\eta)\odot\mathbb{P}^X(\theta)\odot(d\theta)\sigma_t(B)(\eta)$$
$$= \Sigma(\{\theta\in\Omega_T :\quad \theta_t\in B\})(\eta);$$
$$\Pi(A)(\eta) = \frac{\Sigma(A)(\eta)}{\Sigma(\Omega_T)(\eta)}\circ\ \pi_t(B)(\eta) = \frac{\sigma_t(B)(\eta)}{\sigma_t(\mathbb{R})(\eta)}\circ.$$

The classical theorems on large deviations can be stated as follows.

Theorem 1.6 *$\forall A\subset\Omega_T$, A closed, $\forall\mathcal{O}\subset\Omega_T$, $\mathcal{O}$ open, and $\forall x\in\mathbb{R}$, we have*

$$\limsup_{\epsilon\to 0}\epsilon\log P^{X^\epsilon/X_0^\epsilon}(A/x) \leq \mathbb{P}^{X/X_0}(A/x),$$
$$\liminf_{\epsilon\to 0}\epsilon\log P^{X^\epsilon/X_0^\epsilon}(\mathcal{O}/x) \geq \mathbb{P}^{X/X_0}(\mathcal{O}/x).$$

Theorem 1.7 *Suppose that $(P^\epsilon)_{\epsilon>0}$ is a sequence of probability measures over $(\Omega_T,\mathbf{B}(\Omega_T))$ and $\mathbb{P}$ is a Maslov performance measure over $(\Omega_T,\mathbf{B}(\Omega_T))$ such that for every open subset $\mathcal{O}$ and for every closed subset A in Ω_T we have*

$$\limsup_{\epsilon\to 0}\epsilon\log P^\epsilon(A)\leq\mathbb{P}(A),\qquad \liminf_{\epsilon\to 0}\epsilon\log P^\epsilon(\mathcal{O})\geq\mathbb{P}(\mathcal{O}).$$

If $(F^\epsilon)_{\epsilon>0}$ is a sequence of functions from Ω_T into $\mathbb{R}$ which uniformly converges to a function F as $\epsilon\to 0$, then

$$\limsup_{\epsilon\to 0}\epsilon\log\left(\int_A e^{\frac{1}{\epsilon}F^\epsilon(x)}P^\epsilon(dx)\right)\leq\int_A^{\oplus}F(x)\odot\mathbb{P}(dx), \tag{1.28}$$
$$\liminf_{\epsilon\to 0}\epsilon\log\left(\int_{\mathcal{O}} e^{\frac{1}{\epsilon}F^\epsilon(x)}P^\epsilon(dx)\right)\geq\int_{\mathcal{O}}^{\oplus}F(x)\odot\mathbb{P}(dx). \tag{1.29}$$

If we combine these theorems with the previous study, then

1. For every closed subset A and open subset $\mathcal{O}$ in Ω_T we have
$$\limsup_{\epsilon\to 0}\epsilon\log P^{X^\epsilon}(A)\leq\mathbb{P}^X(A),\qquad \liminf_{\epsilon\to 0}\epsilon\log P^{X^\epsilon}(\mathcal{O})\geq\mathbb{P}^X(\mathcal{O}).$$
2. For every closed subset A and open subset $\mathcal{O}$ in $\Omega_T\times\Omega_0$ we have
$$\limsup_{\epsilon\to 0}\epsilon\log P^{X^\epsilon,Y^\epsilon}(A)\leq\mathbb{P}^{X,Y}(A), \tag{1.30}$$
$$\liminf_{\epsilon\to 0}\epsilon\log P^{X^\epsilon,Y^\epsilon}(\mathcal{O})\geq\mathbb{P}^{X,Y}(\mathcal{O}). \tag{1.31}$$

According to the definition of the probability measure $\left(\breve{P}^\epsilon_{[\eta]}\right)_{\epsilon>0,\eta\in\Omega_0}$, we have the following statements:

1. For every closed subset A and open subset $\mathcal{O}$ in Ω_T, $\eta\in\Omega_0$,
$$\limsup_{\epsilon\to 0}\epsilon\log\breve{P}^\epsilon_{[\eta]}(A)\leq\mathbb{P}^{\breve{X}}_{[\eta]}(A),\qquad \liminf_{\epsilon\to 0}\epsilon\log\breve{P}^\epsilon_{[\eta]}(\mathcal{O})\geq\mathbb{P}^{\breve{X}}_{[\eta]}(\mathcal{O}),$$
where $\mathbb{P}^{\breve{X}}_{[\eta]}$ is the performance measure on Ω_T of the optimization process
$$\dot{\breve{X}}_t = f\left(\breve{X}_t\right)-\eta_t\frac{\delta h}{\delta x}\left(\breve{X}_t\right)+U_t^X\qquad\text{with}\quad \breve{X}_0 = X_0.$$
For every $\eta\in\Omega_0$, one can check that the sequence $(F^\epsilon(.,\eta))_{\epsilon>0}$ satisfies the condition of the theorem with
$$F(\theta,\eta)\overset{def}{=} h(\theta_T)\,\eta_T - \int_0^T\left(\frac{1}{2}h^2(\theta_t)+\eta_t f(\theta_t)\frac{\delta h}{\delta x}(\theta_t)-\frac{1}{2}\left(\eta_t\frac{\partial h}{\partial x}(\theta_t)\right)^2\right)dt.$$
2. For every closed subset A and open subset $\mathcal{O}$ in Ω_T and any $\eta\in\Omega_0$, we have
$$\limsup_{\epsilon\to 0}\epsilon\log\sum{}^\epsilon(A)\leq\sum(A)(\eta),\qquad \liminf_{\epsilon\to 0}\epsilon\log\sum{}^\epsilon(\mathcal{O})\geq\sum(\mathcal{O})(\eta).$$

Consequently, for every closed subset A and every open subset $\mathcal{O}$ in Ω_T and any $\eta\in\Omega_0$, we have

$$\limsup_{\epsilon\to 0}\epsilon\log\prod{}^\epsilon(A)\leq\prod(A)(\eta),\qquad \liminf_{\epsilon\to 0}\epsilon\log\prod{}^\epsilon(\mathcal{O})\geq\prod(\mathcal{O})(\eta).$$

In particular, for every closed subset A and every open subset $\mathcal{O}$ in $\mathbb{R}$ and any $\eta\in\Omega_0$, $t\in[0,T]$, we have

$$\limsup_{\epsilon\to 0}\epsilon\log\pi_t^\epsilon(A)\leq\pi(A)_t(\eta),\qquad \liminf_{\epsilon\to 0}\epsilon\log\pi_t^\epsilon(\mathcal{O})\geq\pi_t(\mathcal{O})(\eta).$$

Assume that for every $\eta \in \Omega_0$ and $t \in [0,T]$ there exists a unique conditional optimal state $op(X_t/\eta)$. In other words, for every $\gamma > 0$

$$\pi_t(\{x \in \mathbb{R} : \|x - op(X_t/\eta)\| \le \gamma\})(\eta) = \mathbb{1},$$

$$\pi_t(\{x \in \mathbb{R} : \|x - op(X_t/\eta)\| > \gamma\})(\eta) < \mathbb{1}.$$

Then, using the Prokhorov topology, we find that $\pi_t^\epsilon(\eta)$ weakly converges to $\delta_{op(X_t/\eta)}$ as ϵ tends to 0.

7.2. Log-Exp *Transform*

We briefly recall the Log-Exp transform (for details, see [13] and §8). This mapping leads to useful conclusions, because it makes the relationship between the performance and the probability measure of an event explicit. Let $\nu > 0$ and $d \ge 1$; let D_d^ν be the class of probability measures p on $\mathbb{R}^d$ such that

$$\int^{\oplus} \log(p(x)^\nu) \odot dx \stackrel{\text{def}}{=} \mathbb{N}_\nu(p) > 0,$$

and let $\mathbb{D}_d^\nu$ be the class of performance measures $\mathbb{p}$ on $\mathbb{R}^d$ such that

$$\int \exp\left(\frac{\mathbb{p}(x)}{\nu}\right) dx \stackrel{\text{def}}{=} N_\nu(\mathbb{p}) > 0.$$

We use the following conventions when discrete events are embedded in a continuous fashion:

$$\log\left(\sum_{n\ge 0} p_n \delta_{z_n}\right) = \bigoplus_{n \ge 0} \log(p_n) \odot \mathbb{1}_{z_n},$$

$$\exp\left(\bigoplus_{n\ge 0} \mathbb{p}_n \odot \mathbb{1}_{z_n}\right) = \sum_{n\ge 0} \exp(\mathbb{p}_n)\delta_{z_n}.$$

These spaces are in a one-to-one correspondence by the following transformations:

$$\mathbb{E}\mathrm{xp}_\nu(\mathbb{p}) \stackrel{\text{def}}{=} \frac{e^{\frac{1}{\nu}\mathbb{p}}}{N_\nu(\mathbb{p})}, \qquad \mathbb{L}\mathrm{og}_\nu(p) \stackrel{\text{def}}{=} \frac{\log p^\nu}{\mathbb{N}_\nu(p)^\circ}, \qquad \mathbb{E}\mathrm{xp}_\nu = \mathbb{L}\mathrm{og}_\nu^{-1}. \tag{1.32}$$

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be an optimization basis, and let $\mathbf{F} \stackrel{\text{def}}{=} (\mathcal{F}_k)_{k\ge 0}$ be an increasing filtration of $\mathcal{F}$ on which two independent real optimization processes U and V are defined ranging in $\mathbb{R}^{n_X}$ and $\mathbb{R}^{n_Y}$, respectively, $n_X, n_Y \ge 1$. Let us now define the following optimization processes on $(\Omega, \mathcal{F}, \mathbb{P})$ for every $k \ge 0$:

$$\mathcal{O}(X/Y): \quad X_k = F(X_{k-1}, U_k), \quad X_0 = U_0, \quad \text{and} \quad Y_k = H(X_k) + V_k,$$

where F is a measurable function from $\mathbb{R} \times \mathbb{R}$ into $\mathbb{R}$ and H is a measurable function from $\mathbb{R}$ into $\mathbb{R}$. From the preceding, we have

$$\mathbb{p}^{U_{\underline{k}}, Y_{\underline{k}}}(u,y) = \mathbb{p}^{U_{\underline{k}}}(u) \odot \mathbb{p}^{V_{\underline{k}}}(y - H(\phi(u))),$$

$$(U_{\underline{k}}, V_{\underline{k}}) \stackrel{\text{def}}{=} (U_0, U_1, \ldots, U_k, V_0, V_1, \ldots, V_k),$$

where $\phi(u) = X_{\underline{k}}$ is the state path associated with the value $U = u_{\underline{k}}$. It is now straightforward to apply the Log-Exp transform. For every $\nu > 0$ such that

$$\forall k \ge 0 \qquad N_\nu(k) \stackrel{\text{def}}{=} \int_{(\mathbb{R}\times\mathbb{R})^{[0,k]}} \exp\left(\frac{1}{\nu}\mathbb{p}^{U_{\underline{k}}, Y_{\underline{k}}}(u,y)\right) du\, dy > 0,$$

the measure

$$p^{W^\nu_{\underline{k}}, Y^\nu_{\underline{k}}} \stackrel{\text{def}}{=} \mathbb{E}\mathrm{xp}_\nu(\mathbb{p}^{U_{\underline{k}}, Y_{\underline{k}}}) = \frac{1}{N_\nu(k)} \exp(\mathbb{p}^{U_{\underline{k}}, Y_{\underline{k}}})$$

is the probability measure associated with the filtering problem $\mathcal{F}^\nu$ defined for every $k \ge 0$ by

$$\mathcal{F}^\nu(X/Y): \ X_k^\nu = F(X_{k-1}^\nu, W_k^\nu), \ X_0^\nu = W_0^\nu, \ \text{and} \ Y_k^\nu = H(X_k^\nu) + V_k^\nu,$$

where W^ν, V^ν are two P-independent stochastic processes with probability measures $\mathbb{E}\mathrm{xp}_\nu(\mathbb{p}^U)$ and $\mathbb{E}\mathrm{xp}_\nu(\mathbb{p}^V)$.

Example 1.1 Let $\tau \in [0,k]$ and $0 < \lambda < 1$; then

1. $\mathbb{p}^{U_\tau}(u) = -\frac{1}{2}u^2 \implies p^{W_\tau^\nu}(u) \stackrel{\text{def}}{=} \mathbb{E}\mathrm{xp}_\nu(\mathbb{p}^{U_\tau})(u) = \frac{1}{\sqrt{2\nu\pi}} e^{-\frac{1}{2\nu}u^2}$;
2. $\mathbb{p}^{U_\tau}(u) = \log\left(\frac{\lambda}{\lambda \oplus (1-\lambda)}\right) \odot \mathbb{1}_1 \oplus \log\left(\frac{1-\lambda}{\lambda \oplus (1-\lambda)}\right) \odot \mathbb{1}_0 \implies$

$$p^{W_\tau^\nu}(u) \stackrel{\text{def}}{=} \mathbb{E}\mathrm{xp}_\nu(\mathbb{p}^{U_\tau})(u) = \frac{\lambda^{1/\nu}}{\lambda^{1/\nu} + (1-\lambda)^{1/\nu}}\delta_1 + \left(1 - \frac{\lambda^{1/\nu}}{\lambda^{1/\nu} + (1-\lambda)^{1/\nu}}\right)\delta_0.$$

In other words, one may regard the regulation problem $\mathcal{O}(X/Y)$ as the maximum likelihood estimation problem associated with a filtering problem $\mathcal{F}^\nu$, $\nu > 0$. These facts will be further developed in §8.

7.3. *The Cramer Transform*

As is well known, the Cramer transform is defined by $\mathcal{C} \stackrel{\text{def}}{=} \mathcal{F} \circ \log \circ \mathcal{L}$, where $\mathcal{L}$ is the Laplace transform and $\mathcal{F}$ is the Fenchel transform. This transform maps the set of probability measures to the set of upper semicontinuous performance densities. It also converts probability convolutions into Maslov convolutions and the classical expectation of a random variable into the optimal state of the induced optimization variable. More details were developed in [26, 44, 15].

8. Nonlinear Filtering and Deterministic Optimization

In [13, 14, 15, 20] and §5, we state that the Bellman optimality principle may be viewed as a basic definition of optimization processes like Markov property rather than a deductive conclusion. In forward time, the so-called Maslov optimization processes and Markov stochastic processes can be mapped into each other via various transforms. The Log-Exp transform is a powerful tool to study the stochastic interpretation of Maslov performance. A simple example of this mapping gives some details of how this transform provides essential insight into analyzing optimization problems similarly to nonlinear filtering. Let $T > 0$. Let X be an $\mathbb{R}^{T+1}$-valued optimization variable defined on some performance space $(\Omega, \mathcal{F}, \mathbb{P})$ with

$$\mathbb{p}^X(x) = \sum_{n=0}^{T} \mathbb{p}_n(x_n/x_{n-1}) \qquad (\mathbb{p}_0(x_0/x_{-1}) \stackrel{\text{def}}{=} \mathbb{p}(x_0)),$$

where for every $x \in \mathbb{R}$ the functions $\mathbb{p}_n(\,\cdot\,/x)$ are upper semicontinuous performance densities. Then one can readily check that X is a Maslov optimization process and its transition performances are given by $\mathbb{p}_n(x_n/x_{n-1})$. For such Maslov processes, we can formulate general conditions of integrability type which guarantee the existence of an associated Markov stochastic process. If

$$\int \exp(\mathbb{p}(x_n/x_{n-1}))\, dx_n < +\infty \quad \text{and} \quad \int \exp(\mathbb{p}^X(x))\, dx < +\infty,$$

then

$$p^{X^e}(x) \stackrel{\text{def}}{=} \mathbb{E}\text{xp}(\mathbb{p}^X)(x) = \frac{\exp \mathbb{p}^X(x)}{\int \exp \mathbb{p}^X(x)\, dx} = \prod_{n=0}^{T} p_n(x_n/x_{n-1}),$$
$$p_n(x_n/x_{n-1}) = \mathbb{E}\text{xp}(\mathbb{p}_n(\,\cdot\,/x_{n-1}))(x_n),$$

is the probability density of some $\mathbb{R}$-valued Markov stochastic process X^e defined on a suitable probability space $(\Omega, \mathcal{F}, P)$.

Similarly, let U be an $\mathbb{R}^T$-valued optimization variable defined on some performance space $(\Omega, \mathcal{F}, \mathbb{P})$ with

$$\mathbb{p}^U(u) = \sum_{n=1}^{T} L(\phi(u)_{n-1}, u_n),$$

where $\phi(u)_n$ denotes the n-time value of the path solution $\phi(u)$ of some recursive system $x_n = f(x_{n-1}, u_n)$ with fixed initial condition x_0. Then

$$p^{U^e}(u) \stackrel{\text{def}}{=} \mathbb{E}\text{xp}(\mathbb{p}^U)(u) = \frac{\exp(\mathbb{p}^U(u))}{\int \exp(\mathbb{p}^U(u))\, du} = \prod_{n=1}^{T} \frac{\exp(L(\phi(u)_{n-1}, u_n))}{\int \exp(L(\phi(u)_{n-1}, u_n))\, du_n}.$$

This mapping permits us to give a direct stochastic interpretation of Maslov performances. We develop this mapping of optimization problems into filtering problems. To simplify the notation, to any real-valued optimization problem on $[0,T]$, denoted by $\mathcal{O}(\mathcal{X}/\mathcal{Y})$ and defined on a performance space $(\Omega, \mathcal{F}, \mathbb{P})$, we assign a real-valued nonlinear filtering problems on $[0,T]$, denoted by $\mathcal{F}(\mathcal{X}/\mathcal{Y})$ and defined on some convenient probability space $(\Omega, \mathcal{F}, P)$ as follows:

$\mathcal{O}(\mathcal{X}/\mathcal{Y})$	$\mathcal{F}(\mathcal{X}/\mathcal{Y})$
$\begin{cases} X_n = \phi(X_{n-1}, U_n), \quad X_0 = U_0, \\ Y_n = H(X_n) + V_n; \end{cases}$	$\begin{cases} X^e_n = \phi(X^e_{n-1}, U^e_n), \quad X^e_0 = U^e_0, \\ Y^e_n = H(X^e_n) + V^e_n; \end{cases}$
U and V are two $\mathbb{P}$-independent optimization processes	U^e and V^e are two P-independent stochastic processes
$(\mathbb{L}\text{og}\, p^{U^e,V^e} = \mathbb{p}^{U,V})$	$(\mathbb{E}\text{xp}\, \mathbb{p}^{U,V} = p^{U^e,V^e})$

where ϕ and H are two measurable real functions. When carefully examining these problems, several comments are in order. With some obvious abusive notation, we have:

1) $\mathbb{p}(u) = \bigodot_{n=0}^{T} \mathbb{p}(u_n)$ and $p(u^e) = \prod_{n=0}^{T} p(u^e_n)$;
2) $\mathbb{p}(u,v) = \mathbb{p}(u) \odot \mathbb{p}(v)$ and $p(u^e, v^e) = p(u^e)p(v^e)$;
3) $\mathbb{p}(u,y) = \mathbb{p}(y/u) \odot \mathbb{p}(u)$ and $p(u^e, y^e) = p(y^e/u^e)p(u^e)$;
4) $\mathbb{p}(y) = \int^{\oplus} \mathbb{p}(y/u) \odot \mathbb{p}(u) \odot du$ and $p(y^e) = \int p(y^e/u^e)p(u^e)\, du^e$;
5) $\mathbb{p}(u/y) = \frac{\mathbb{p}(y/u)}{\mathbb{p}(y)} \odot \mathbb{p}(u)$ and $p(u^e/y^e) = \frac{p(y^e/u^e)}{p(y^e)} p(u^e)$;
6) $\mathbb{p}(y/u) = \mathbb{p}^V(y - H(\phi(u)))$ and $p(y^e/u^e) = p^{V^e}(y^e - H(\phi(u^e)))$

In this sense, the random variable V^e and the optimization variable V completely describe the Bayesian factor, which produces the conditional probability or the performance measure. The following examples suggest how these results can be useful in analyzing some optimization problems. Let $H(x) = x$, $n \in [0, T]$, $0 < \lambda < 1$, and $c \in \mathbb{R}$. Then

1. $\mathbb{p}(v_n) = -\frac{1}{2} v_n^2 \implies \mathbb{p}^V(c - H(\phi(u))) = -\frac{1}{2} \sum_{n=0}^{T} (c - \phi_\tau(u))^2$;
2. $\mathbb{p}(u_n) = -\frac{1}{2} u_n^2 \implies p(u_n^e) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(u_n^e)^2}$;
3. Poisson processes are also realistic models for a large class of point processes: photon count, electron emission, telephone calls, data communications, departure, waiting, servicing, etc.

Assume that U^e is a Poisson counting process with nonhomogeneous intensity function λ; its sample function is given for every piecewise constant path u^e such that $\Delta u_n^e \in \{0, 1\}$ by

$$p(u^e) = \exp\left(- \int_0^T \lambda_\tau \, d\tau + \int_0^T \log(\lambda_\tau) \, du_\tau^e \right),$$

$$\mathbb{p}(u) = \int_0^T \log(\lambda_\tau) \, du_\tau.$$

4. $$\mathbb{p}(u_n) = \log\left(\frac{\lambda}{\lambda \oplus (1-\lambda)}\right) \odot \mathbb{I}_1(u_n) \oplus \log\left(\frac{1-\lambda}{\lambda \oplus (1-\lambda)}\right) \odot \mathbb{I}_0(u_n)$$

$$\implies p(u_n^e) = \frac{\lambda}{\lambda + (1-\lambda)} \delta_1(u_n^e) + \left(1 - \frac{\lambda}{\lambda + (1-\lambda)}\right) \delta_0(u_n^e);$$

5. Initial constraint: $\mathbb{p}(u_0) = \mathbb{I}_{x_0} \implies p(u_0^e) = \delta_{x_0}(u_0^e)$;
6. Final constraint: $\mathbb{p}(v_T) = \mathbb{I}_0(v_T) \implies \mathbb{p}^{V_T^e}(c - H(\phi_T(u))) = \mathbb{I}_c(\phi_T(u))$ $\implies p(v_T^e) = \delta_0(v_T^e)$.

Clearly, these results can be generalized to the vector case. As a matter of fact, the differential equations introduced in $\mathcal{F}(\mathcal{X}/\mathcal{Y})$ or $\mathcal{O}(\mathcal{X}/\mathcal{Y})$ are usually constructed using the conventional addition and multiplication. We end this section by recalling that *nonlinear filtering and optimization may be useful in optimization and control of communication networks and manufacturing systems.* This class of systems, often referred to as discrete-event dynamical systems contains man-made systems that involve a finite amount of resources (machines, communications, ...) shared by several users. The time behavior of such systems might be described in terms of $(\oplus, \odot)$-differential equations as a physical phenomenon (the reader is referred to [1], [8] for $(\oplus, \odot)$-linear systems). In such areas, *the functions ϕ and H are constructed using the operations min and max and the conventional addition and multiplication.* The state variable X_n denotes the earliest epoch at which the nodes or machines become active for the nth time, U_n is the epoch at which the resources become available for the nth time, and Y_n is the epoch at which the final products of the system are delivered to the outer world for the nth time. In the stochastic case, the input sequences U could be exponentially distributed with a parameter which depends on the places of the event graph (nonlinear filtering). In the deterministic case, the inputs could be performed according to a given cost or performance function (deterministic optimization). We should note in passing that some stochastic event graphs (in which holding, firing, and lag times are random variables) and queuing systems might be modeled in the same fashion. Unfortunately, we do not have enough room here to carry out a thorough study on the modeling of such systems; the reader is referred to the book [1] for $(\oplus, \odot)$-linear systems.

9. Monte-Carlo Principles

This section constitutes the second step on our way to particle methods for filtering and optimization problems. We first briefly review some basic facts about Monte-Carlo principles and show that these principles can be used to study the mean of a random variable as well as the optimal state of an optimization variable. The concept of probability is the achievement of deductive reasoning in which we estimate the chances of some event realization. When the event is associated with a random error of some approximation, this measure evaluates the chances to get such a precision. In what follows, particle algorithms will be studied in that way. The independence between random variables means that the realization of some variable is not altered by the realizations of the others. This concept is fundamental; in fact, it justifies the mathematical development of probability not merely as a topic in measure theory, but as a separate discipline. The significance of independence arises in the context of repeated trials.

1) Mean Estimation: Let X be a real random variable X^i, and let $(X^i)_{i \geq 1}$ be a sequence of independent random variables with the same probability law as X and defined on the same probability space (Ω, σ, P); then for every $N \geq 1$ we have

$$E((E_N(X) - E(X))^2) = \frac{1}{N} E((X - E(X))^2),$$

where $E_N(X) = \frac{1}{N} \sum_{i=1}^N X^i$. Applying Chebyshev's inequality, for every $\varepsilon > 0$ we obtain

$$P(|E_N(X) - E(X)| > \varepsilon) \leq \frac{1}{N\varepsilon^2} E((X - E(X))^2).$$

In other words, if $E((X - E(X))^2) < +\infty$, then $E_N(X) \xrightarrow[N \to +\infty]{\mathbf{L}^0} E(X)$

2) Optimal State Estimation: Let X be a real optimization variable defined on a performance space $(\Omega, \sigma, \mathbb{P})$, let $\mathbb{p}^X$ be its performance function, and let $\mathrm{op}(X)$ be its unique optimal state. Assume that $\mathbb{p}^X$ is regular in the following sense: for every $\varepsilon > 0$, there exists an $\eta > 0$ such that

$$\rho(\mathbb{p}^X(x), \mathbb{1}) = |\exp \mathbb{p}^X(x) - \exp \mathbb{p}^X(\mathrm{op}(X))| \leq \eta \implies |x - \mathrm{op}(X)| \leq \varepsilon. \tag{1.33}$$

One can also formulate general conditions of second-derivative type which guarantee this kind of regularity [13]. In the sequel, X^e *will stand for some random variable on a probability space* (Ω, σ, P). Then for every $\varepsilon > 0$ we have

$$P(|\mathrm{op}_N(X) - \mathrm{op}(X)| > \varepsilon) \leq (1 - P(|X^e - \mathrm{op}(X)| \leq \varepsilon))^N,$$

where

$$\mathrm{op}_N(X) = \mathrm{Arg} \sup_{x \in \Omega_N} \mathbb{p}^X(x) \stackrel{\mathrm{def}}{=} A \bigoplus_{i=1}^N \mathbb{p}^X(X^i) = A \left\langle \bigoplus_{i=1}^N \mathbb{1}_{X^i}, \mathbb{p}^X \right\rangle,$$

$\Omega_N = \{X^1, \ldots, X^N\}$, and $(X^i)_{i \geq 1}$ is a sequence of independent random variables with the same probability law as X^e.

In other words,

$$\text{if } P(|X^e - \mathrm{op}(X)| \leq \varepsilon) > 0 \text{ for every } \varepsilon > 0, \text{ then } \mathrm{op}_N(X) \xrightarrow[N \to +\infty]{\mathbf{L}^0} \mathrm{op}(X).$$

Finally, let us note that the random variable X^e need not depend on the performance $\mathbb{p}^X$. When the probability law of X^e is given by $\mathbb{E}\mathrm{xp}(\mathbb{p}^X)$, then the condition $P(|X^e - \mathrm{op}(X)| \leq \varepsilon) > 0$ is clearly satisfied for every $\varepsilon > 0$. We continue our investigation of particle methods for nonlinear filtering and optimization problems. In the sequel, for every sequence u of real numbers and $n \in [0, T]$ we write

$$u_{\underline{n}} = (u_0, \ldots, u_n), \qquad \|u\|_2^2 = \sum_{n=0}^T u_n^2, \qquad \|u_{\underline{n}}\|_2^2 = \sum_{m=0}^n u_m^2.$$

Using the preceding and the Bayes formula, we shall derive an $\mathbf{L}^0$-approximation of the conditional expectation, as well as conditional optimal control, starting with an example. By the same line of argument as before, let $T > 0$, and let U, V be two $\mathbb{P}$-independent $\mathbb{R}^T$-valued optimization variables defined on a performance space $(\Omega, \sigma, \mathbb{P})$ with $\mathbb{p}(u, v) = -\frac{1}{2}\|u\|_2^2 - \frac{1}{2}\|v\|_2^2$.

Let X and Y be the real-valued optimization processes defined on $(\Omega, \sigma, \mathbb{P})$ by the dynamical systems

$$\mathcal{O}(\mathcal{X}/\mathcal{Y}): \quad X_n = \phi(X_{n-1}, U_n), \quad X_0 = U_0, \quad Y_n = H(X_n) + V_n. \tag{1.34}$$

Then, by obvious considerations,

$$\mathbb{p}(u, y) = \mathbb{p}(y/u) \odot \mathbb{p}(u) = -\frac{1}{2}\|u\|_2^2 - \frac{1}{2}\|y - H(\phi(u))\|_2^2,$$

where $\phi_n(u)$ is the nth value of the path controlled by u. In particular, if $H(x) = x$ and $y_n = c$ is a fixed constant, then the performance function $\mathbb{p}(u/c) = -\frac{1}{2}\|c - \phi(u)\|_2^2 - \frac{1}{2}\|u\|_2^2$ is clearly associated with the minimum energy regulation problem with reference value c.

The filtering problem associated with (1.34) is then defined on some probability space (Ω, σ, P) by

$$\mathcal{F}(\mathcal{X}/\mathcal{Y}): \quad X_n^e = \phi(X_{n-1}^e, U_n^e), \quad X_0^e = U_0^e, \quad Y_n^e = H(X_n^e) + V_n^e, \tag{1.35}$$

where U^e and V^e are two sequences of P-independent Gaussian random variables with zero mean and unit variance. In this case, the probability density is given by

$$p(u^e, y^e) = p(y^e/u^e)p(u^e) = \prod_{n=0}^T \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_n^e - H(\phi_n(u^e)))^2} \prod_{n=0}^T \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u_n^2}.$$

As is well known, H can always be chosen as a linear function, through a suitable state-space basis, so that the conditions for $\mathbf{L}^0$*-convergences have the simplest form.* By the same lines of argument as before, with some obvious abusive notation, if $(U^i)_{i \geq 1}$ is a sequence of independent random variables with the same probability law as U^e and defined on the same probability space (Ω, σ, P), then for every $N \geq 1$, $1 \leq n \leq T$, and $\varepsilon > 0$, one has:

1) Conditional Expectation Estimate:

$$P(|E_N(\phi_n(U^e)/Y_{\underline{n}}^e) - E(\phi_n(U^e)/Y_{\underline{n}}^e)| > \varepsilon) \leq \frac{C_T}{N\varepsilon^2} E((\phi_n(U^1) - E(\phi_n(U^e)/Y_{\underline{n}}^e))^2).$$

Here

$$E_N(\phi_n(U^e)/Y_{\underline{n}}^e) = \sum_{i=1}^N \frac{p(Y_{\underline{n}}^e/U_{\underline{n}}^i)}{\sum_{j=1}^N p(Y_{\underline{n}}^e/U_{\underline{n}}^j)} \phi_n(U^i)$$

and $C_T > 0$. In other words, if the right-hand side of the last inequality is finite, then

$$\frac{\int \phi_n(u) p(Y_{\underline{n}}^e/u_{\underline{n}}) \frac{1}{N}\sum_{i=1}^N \delta_{U_{\underline{n}}^i}(du_{\underline{n}})}{\int 1\, p(Y_{\underline{n}}^e/u_{\underline{n}}) \frac{1}{N}\sum_{j=1}^N \delta_{U_{\underline{n}}^j}(du_{\underline{n}})} \xrightarrow[N \to +\infty]{\mathbf{L}^0} \frac{\int \phi_n(u) p(Y_{\underline{n}}^e/u_{\underline{n}})\, dp(u_{\underline{n}})}{\int 1\, p(Y_{\underline{n}}^e/u_{\underline{n}})\, dp(u_{\underline{n}})}$$

$$= E_N(\phi_n(U^e)/Y_{\underline{n}}^e) \qquad\qquad = E(\phi_n(U^e)/Y_{\underline{n}}^e).$$

2) Conditional Optimization Estimate: Let Y be a reference value for which the conditional performance $\mathbb{p}(u_{\underline{n}}/Y_{\underline{n}})$ satisfies the following regularity conditions (1.33):

1) $\forall \varepsilon > 0 \quad \exists \eta > 0 : \quad \forall 0 \le n \le T$
$$\rho(\mathbb{p}(u_{\underline{n}}/Y_{\underline{n}}), \mathbb{I}) \le \eta \implies \|u_{\underline{n}} - \mathrm{op}(U_{\underline{n}}/Y_{\underline{n}})\|_2 \le \varepsilon,$$
2) $\forall \varepsilon > 0 \quad \exists \eta > 0 : \quad \forall 0 \le n \le T$
$$\|u_{\underline{n}} - \mathrm{op}(U_{\underline{n}}/Y_{\underline{n}})\|_2 \le \eta \implies \rho(\mathbb{p}(u_{\underline{n}}/Y_{\underline{n}}), \mathbb{I}) \le \varepsilon.$$

After some algebraic manipulations, one can prove that for every $\varepsilon > 0$ there exists an $\eta > 0$ such that

$$\begin{aligned} P(\| \mathrm{op}_N(U_{\underline{n}}/Y_{\underline{n}}) - \mathrm{op}(U_{\underline{n}}/Y_{\underline{n}})\|_2 > \varepsilon) \\ \le (1 - P(\|U^1_{\underline{n}} - \mathrm{op}(U_{\underline{n}}/Y_{\underline{n}})\|_2 \le \eta))^N, \end{aligned}$$

where $\Omega_N = \{U^1_{\underline{n}}, \dots, U^N_{\underline{n}}\}$ and

$$\begin{aligned} \mathrm{op}_N(U_{\underline{n}}/Y_{\underline{n}}) &= \mathrm{Arg} \sup_{u \in \Omega_N} \mathbb{p}(u, Y_{\underline{n}}) \stackrel{\text{def}}{=} A \bigoplus_{i=1}^N \mathbb{p}(U^i_{\underline{n}}, Y_{\underline{n}}) \\ &= A \Big\langle \bigoplus_{i=1}^N \mathbb{I}_{U^i_{\underline{n}}}, \mathbb{p}^{U_{\underline{n}}, Y_{\underline{n}}} \Big\rangle. \end{aligned}$$

In other words, if $P(\|U^1_{\underline{n}} - \mathrm{op}(U_{\underline{n}}/Y_{\underline{n}})\|_2 \le \varepsilon) > 0$ for every $\varepsilon > 0$, then

$$\mathrm{op}_N(U_{\underline{n}}/Y_{\underline{n}}) \xrightarrow[N \to +\infty]{\mathbf{L}^0} \mathrm{op}(U_{\underline{n}}/Y_{\underline{n}}).$$

Example 1.2 Whenever $\mathbb{p}(u,v) = -\frac{1}{2}\|u\|_2^2 - \frac{1}{2}\|v\|_2^2$, we can readily check that the second regularity condition is satisfied for $\mathbb{p}(Y_{\underline{n}}) < +\infty$. Indeed, in this case we have

$$\begin{aligned} \rho(\mathbb{p}(u_{\underline{n}}/Y_{\underline{n}}), \mathbb{I}) &\le \big| \|u_{\underline{n}}\|_2^2 - \| \mathrm{op}(U_{\underline{n}}/Y_{\underline{n}})\|_2^2 + \| \mathrm{op}(Y_{\underline{n}} - H(\phi(u_{\underline{n}})))\|_2^2 \\ &\quad - \| \mathrm{op}(Y_{\underline{n}} - H(\phi(\mathrm{op}(U_{\underline{n}}/Y_{\underline{n}}))))\|_2^2 \big| \\ &\le \|u_{\underline{n}} - \mathrm{op}(U_{\underline{n}}/Y_{\underline{n}})\|_2^2 \\ &\quad + 2\|u_{\underline{n}} - \mathrm{op}(U_{\underline{n}}/Y_{\underline{n}})\|_2 \sup(1, \mathbb{p}(Y_{\underline{n}})). \end{aligned}$$

10. Particle Interpretations

The particle algorithms are based on a Dirac comb which depends on the flow of the system and its partial observations or reference values, both in mass and position. In order to obtain time-uniform convergence, it is necessary to introduce convenient exploring distributions and to regularize their complementary weights. The significance of these facts will be given in the

forthcoming development. *For filtering problems*, these exploring distributions are clearly dictated by the *Bayes principles*. *For optimization problems*, we consider the associated estimation problems. In other words, in that case we introduce such exploring distribution by an original method, based on the *use of the* Log-Exp *transform*. As was shown in §8, this transform explicitly characterizes the filtering problem associated with this optimization problem. The stochastic interpretation allows us to introduce, as well as for filtering problems, an exploring distribution that depends on the reference values. Hence, it suffices to start from the conditional probability function of some filtering problem $\mathcal{F}(\mathcal{X}/\mathcal{Y})$. For convenience, we keep, as a point of reference, the terminology and assumptions of §8 for the descriptions of the nonlinear filtering and optimization problems $\mathcal{F}(\mathcal{X}/\mathcal{Y})$ and $\mathcal{O}(\mathcal{X}/\mathcal{Y})$. Moreover, in order to clarify the notation, the symbol $(\cdot)^e$ will be omitted and the random variables V will be centered Gaussian variables with zero mean and variance function R. All stochastic processes defined in what follows are assumed to be carried by some probability space. Before starting the discussion, we give some consequences of the Bayes formula in our setting:

1. $p(u_{\underline{n}}, y_{\underline{n}}) = p(y_{\underline{n}}/u_{\underline{n}}) p(u_{\underline{n}})$.
2. $p(u_{\underline{n}}) = \prod_{m=0}^n p(u_m)$ and $p(y_{\underline{n}}/u_{\underline{n}}) = \prod_{m=0}^n p(y_m/u_{\underline{m}})$.
3. $$\begin{aligned} p(y_m/u_{\underline{m}}) &= p^{V_m}(y_m - H(\phi_m(u))) \\ &= \frac{1}{\sqrt{2\pi |R_m|}} \exp\Big(-\frac{1}{2R_m} (y_m - H(\phi_m(u)))^2 \Big). \end{aligned}$$
4. $$\begin{aligned} p(u_{\underline{n}}, y_{\underline{n}}) &= \prod_{m=0}^n p(y_m/u_{\underline{m}}) p(u_m) \\ &= \prod_{m=0}^n p(y_m/u_{\underline{m-1}}) \frac{p(y_m/u_{\underline{m}})}{p(y_m/u_{\underline{m-1}})} p(u_m) \\ &= \prod_{m=0}^n p(y_m/u_{\underline{m-1}}) \prod_{m=0}^n p(u_m/u_{\underline{m-1}}, y_m). \end{aligned}$$
5. $p(y_m/u_{\underline{m-1}}) = \int p(y_m/u_{\underline{m-1}}, u_m)\, dp(u_m) = \int p(y_m/u_{\underline{m}})\, dp(u_m)$.

We are now in a position to describe some time-recursive exploring distributions, which will be used in forward time. Using the above, we emphasize that these distributions are exhibited by some natural change of probability functions. The detailed assumptions under which these convergences are uniform in time will be studied in the last section.

A priori exploration: The following change of probability is a simple consequence of the Bayes formula:

$$p(u_{\underline{n}}, y_{\underline{n}}) = Z^0_{\underline{n}}(u,y) p_0(u_{\underline{n}}, y_{\underline{n}}) = Z^0_n(u,y) p_0(u_{\underline{n}}/y_{\underline{n}}) p_0(y_{\underline{n}}),$$

with

$$\begin{cases} p_0(y_{\underline{n}}) = G_n(y), & p_0(u_{\underline{n}}, y_{\underline{n}}) = p_0(u_{\underline{n}}) = p(u_{\underline{n}}), \\ G_n(y) = g_n(y) G_{n-1}(y), & g_n(y) = p^{V_n}(y_n), \\ Z_n^0(u,y) = z_n^0(u,y) Z_{n-1}^0(u,y), & z_n^0(u,y) g_n(y) = p(y_n/u_{\underline{n}}). \end{cases}$$

Let $\check{U}_{\underline{n}}$ be a *generic* explorating stochastic process P-independent of U and V with distribution $p(u_{\underline{n}})$, and let $(U^i)_{i\geq 1}$ be a sequence on independent copies of $\check{U}$. By the same line of argument as before, we obtain:

$$E_N(\phi_n(U)/Y_{\underline{n}}) \xrightarrow[N\to+\infty]{\mathbf{L}^0} E(\phi_n(U)/Y_{\underline{n}}),$$

$$\mathrm{op}_N(U_{\underline{n}}/Y_{\underline{n}}) \xrightarrow[N\to+\infty]{\mathbf{L}^0} \mathrm{op}(U_{\underline{n}}/Y_{\underline{n}}),$$

$$E_N(\phi_n(U)/Y_{\underline{n}}) = \sum_{i=1}^N \frac{Z_n^0(U^i,Y)}{\sum_{j=1}^N Z_n^0(U^j,Y)} \phi_n(U^i), \tag{1.36}$$

$$\mathrm{op}_N(U_{\underline{n}}/Y_{\underline{n}}) = A \bigoplus_{i=1}^N \mathbb{p}(U_{\underline{n}}^i, Y_{\underline{n}}).$$

Moreover, it is obvious that P-a.e. $E(\phi_n(\check{U})/\check{Y}_{\underline{n}})(Y_{\underline{n}}) = E(\phi_n(U)/Y_{\underline{n}})(Y_{\underline{n}})$ where $\check{Y}$ is the observation process of $\check{U}$ defined similarly to Y with (U,V) replaced by $(\check{U},\check{V})$ and $\check{V}$ is P-independent of $\check{U}$ with the same law as V. The significance of this obvious remark will be clarified later. In addition to the exploration processes U^i, the weights $Z_n^0(U^i,Y)$ and $\mathbb{p}(U_{\underline{n}}^i, Y_{\underline{n}})$ are related to the likelihood of U^i. It is important to notice that they are time-degenerate. For instance, if $\phi(x,u) = u$, $h(x) = x$, and U is a centered Gaussian process with zero mean and unit variance, then we have, for every $i \neq j$ and $q \in]0, 1/\sqrt{5}[$,

$$\sup_{n\geq 0} E\left(\left|\frac{Z_n^0(U^i,Y)}{Z_n^0(U^j,Y)}\right|^q\right) = +\infty = \sup_{n\geq 0} E\left(\left|\frac{\mathbb{p}(U_{\underline{n}}^i, Y_{\underline{n}})}{\mathbb{p}(U_{\underline{n}}^j, Y_{\underline{n}})}\circ\right|^2\right).$$

The degeneracy of these weights is eliminated by using a regularization of the problem. This regularization is clearly dictated by the form of these weights. To this purpose, we denote by $\mathcal{R}$ the space of functions $\alpha: \mathbb{N}^2 \to [0,1]$ such that, for every $m, u \in \mathbb{N}$, $\alpha_m(0) = 1$ and $u \mapsto \alpha_m(u)$ and $m \mapsto \alpha_m(u)$ are non-increasing. We endow this space with the pointwise convergence topology and define the α-regularized weights $Z_n^\alpha(u,y)$ and $\mathbb{p}^\alpha(u,y)$ as follows:

$$\log Z_n^\alpha(u,y) = \sum_{m=0}^n \alpha_m(n-m) \log z_m^0(u,y),$$

$$\mathbb{p}^\alpha(u_{\underline{n}}, y_{\underline{n}}) = \sum_{m=0}^n \alpha_m(n-m) \mathbb{p}(y_m, u_m/u_{\underline{m-1}}),$$

where

$$\mathbb{p}(y_m, u_m/u_{\underline{m-1}}) = \mathbb{p}(y_m/u_{\underline{m}}) \odot \mathbb{p}(u_m).$$

By $\mathrm{op}^\alpha(U_{\underline{n}}/Y_{\underline{n}})$ we denote the conditional optimal estimate associated with this α-regularized performance.

Then, recalling the previous example if $\alpha_l(u) = 1_{[0,T]}(u)$, we obtain

$$\sup_{n\geq 0} E\left(\left|\frac{Z_n^\alpha(U^i,Y)}{Z_n^\alpha(U^j,Y)}\right|^q\right) = \left(\frac{2}{\sqrt{1-5q^2}}\right)^T < +\infty,$$

$$\sup_{n\geq 0} E\left(\left|\frac{\mathbb{p}^\alpha(U_{\underline{n}}^i, Y_{\underline{n}})}{\mathbb{p}^\alpha(U_{\underline{n}}^j, Y_{\underline{n}})}\circ\right|^2\right) = 8T < +\infty.$$

For convenience, we now state that this α-regularization corresponds to an α-regularization of the observation process $\check{Y}$ or an α-regularization of the optimization problem $\mathcal{O}(\mathcal{X}/\mathcal{Y})$. In [13], we prove that P-a.e.

$$E^\alpha(\phi_n(U)/Y_{\underline{n}}) \stackrel{\text{def}}{=} E(\phi_n(\check{U})/\check{Y}_{\underline{n}}(\alpha,n))(Y_{\underline{n}}) = \frac{\int \phi_n(u) Z_n^\alpha(u, Y_{\underline{n}})\, dp(u_{\underline{n}})}{\int 1\, Z_n^\alpha(u, Y_{\underline{n}})\, dp(u_{\underline{n}})}, \tag{1.37}$$

where $\check{Y}_{\underline{n}}(\alpha,n)$ is the observation process of $\check{U}$ defined similarly to $\check{Y}$ with $\check{V}$ replaced by a Gaussian process P independent of $\check{U}$ and with zero mean and variance function $R(\alpha,n)$ defined for every $0 \leq m \leq n$ by

$$R(\alpha,n)_m^{-1} = \alpha_m(n-m) R_m^{-1}.$$

On the other hand, let $(U(\alpha,n), V(\alpha,n))$ be two $\mathbb{P}$-independent optimization processes carried by some performance space $(\Omega, \sigma, \mathbb{P})$ whose performance function is given by

$$\mathbb{p}^\alpha(u_{\underline{n}}, v_{\underline{n}}) = \bigodot_{m=0}^n \alpha_m(n-m) \mathbb{p}(u_m) \odot \bigodot_{m=0}^n \alpha_m(n-m) \mathbb{p}(v_m).$$

Then $\mathbb{p}^\alpha(u_{\underline{n}}, y_{\underline{n}})$ is the performance function of the optimization problem $\mathcal{O}^\alpha(\mathcal{X}/\mathcal{Y})$ defined similarly to $\mathcal{O}(\mathcal{X}/\mathcal{Y})$ with the optimization processes (U,V) replaced by $(U(\alpha,n), V(\alpha n))$. Finally, by the same line of argument as before, we note that once the α-regularization has been performed, we have

$$E_N^\alpha(\phi_n(U)/Y_{\underline{n}}) \xrightarrow[N\to+\infty]{\mathbf{L}^0} E^\alpha(\phi_n(U)/Y_{\underline{n}}),$$

$$\mathrm{op}_N^\alpha(U_{\underline{n}}/Y_{\underline{n}}) \xrightarrow[N\to+\infty]{\mathbf{L}^0} \mathrm{op}^\alpha(U_{\underline{n}}/Y_{\underline{n}}),$$

$$E_N^\alpha(\phi_n(U)/Y_{\underline{n}}) = \sum_{i=1}^N \frac{Z_n^\alpha(U^i,Y)}{\sum_{j=1}^N Z_n^\alpha(U^j,Y)} \phi_n(U^i), \tag{1.38}$$

$$\mathrm{op}_N^\alpha(U_{\underline{n}}/Y_{\underline{n}}) = A \bigoplus_{i=1}^N \mathbb{p}^\alpha(U_{\underline{n}}^i, Y_{\underline{n}}).$$

We have described an a priori exploration of the probability/performance space. In order to obtain a time-uniform convergence in general (e.g., for unstable systems), an important step consists in introducing recursive exploration distributions, which depend on the observation process (for filtering problems) or on the reference process (for deterministic optimization problems). Indeed, for obvious reasons, the quality of the particle estimates is greatly improved if the exploration distribution depends on the observed or the reference process.

Conditional exploration: The following changes of probabilities are simple consequences of the Bayes formula and the preceding remarks:

$$p(u_{\underline{n}}, y_{\underline{n}}) = Z_n^1(u,y) p_1(u_{\underline{n}}, y_{\underline{n}}) = Z_n^1(u,y) p_1(u_{\underline{n}}/y_{\underline{n}})\, p_1(y_{\underline{n}}),$$

where

$$\begin{cases} p_1(y_{\underline{n}}) = G_n(y), & p_1(u_{\underline{n}}/y_{\underline{n}}) = p(u_n/u_{\underline{n-1}}, y_{\underline{n}}) p_1(u_{\underline{n-1}}/y_{\underline{n-1}}), \\ G_n(y) = g_n(y) G_{n-1}(y), & g_n(y) = p^{V_n}(y_n), \\ Z_n^1(u,y) = z_n^1(u,y) Z_{n-1}^1(u,y), & z_n^1(u,y) g_n(y) = p(y_n/u_{\underline{n-1}}). \end{cases}$$

Let $\check{U}_{\underline{n}}$ be a *generic* explorating stochastic process P-independent of U and V with Y-conditional distribution $p_1(u_{\underline{n}}/y_{\underline{n}})$, and let $(U^i)_{i\geq 1}$ be a sequence of Y-conditionally independent copies of $\check{U}$. By the same line of argument as before (see Eq. (1.36)), we obtain

$$E_N(\phi_n(U)/Y_{\underline{n}}) \xrightarrow[N\to+\infty]{\mathbf{L}^0} E(\phi_n(U)/Y_{\underline{n}}),$$

$$\mathrm{op}_N(U_{\underline{n}}/Y_{\underline{n}}) \xrightarrow[N\to+\infty]{\mathbf{L}^0} \mathrm{op}(U_{\underline{n}}/Y_{\underline{n}}),$$

$$E_N(\phi_n(U)/Y_{\underline{n}}) = \sum_{i=1}^N \frac{Z_n^1(U^i,Y)}{\sum_{j=1}^N Z_n^1(U^j,Y)}\, \phi_n(U^i),$$

$$\mathrm{op}_N(U_{\underline{n}}/Y_{\underline{n}}) = A \bigoplus_{i=1}^N \mathbb{p}(U_{\underline{n}}^i, Y_{\underline{n}}).$$

Let $\check{Y}$ be the observation process of $\check{U}$ defined by $\check{Y}_n = H(\phi(\check{X}_{n-1}, \widetilde{U}_n)) + \check{V}_n$, where $\check{X}_{n-1} = \phi_{n-1}(\check{U})$ and the processes $(\widetilde{U}, \check{V})$ are P-independent of $\check{U}$ and have the same law as (U,V). Let us make several comments:

1. With some obvious abusive notation, the definition of $\check{Y}$ yields

$$p(\check{y}_{\underline{n}}/\check{u}_{\underline{n}}) = p(\check{y}_n/\check{u}_{\underline{n}}) p(\check{y}_{\underline{n-1}}/\check{u}_{\underline{n-1}}),$$

$$p(\check{y}_n/\check{u}_{\underline{n}}) = \int p^{Y_n/U_{\underline{n}}}(\check{y}_n/\check{u}_{\underline{n-1}}, u_n)\, dp(u_n) = p^{Y_n/U_{\underline{n-1}}}(\check{y}_n/\check{u}_{\underline{n-1}}),$$

$$p(\check{u}_{\underline{n}}, \check{y}_{\underline{n}}/y_{\underline{n}}) = p(\check{y}_{\underline{n}}/\check{u}_{\underline{n}}, y_{\underline{n}}) p(\check{u}_{\underline{n}}/y_{\underline{n}}) = p(\check{y}_{\underline{n}}/\check{u}_{\underline{n}}) p(\check{u}_{\underline{n}}/y_{\underline{n}})$$
$$\Longrightarrow E(\phi_n(\check{U})/\check{Y}_{\underline{n}}, Y_{\underline{n}})(Y_{\underline{n}}) = E(\phi_n(U)/Y_{\underline{n}})(Y_{\underline{n}}).$$

2. In the following example, the nonlinear structure of the problem in hand can be directly exploited. If $\phi(x,u) = F(x) + u$, $H(x) = C \cdot x$, and U is a discrete-time Gaussian process with zero mean and with variance function Q, then

$$p(u_n/u_{\underline{n-1}}, y_{\underline{n}}) = \frac{1}{\sqrt{2\pi|S_n|}} \exp\left(-\frac{1}{2|S_n|}(u_n - S_n C R_n^{-1}(y_n - CF(\phi_{n-1}(u))))^2 \right),$$

where $S_n^{-1} = Q_n^{-1} + CR_n^{-1}C$.

3. The conditional exploration transitions may depend on several observation values. We claim that, by using a suitable state space basis, this case may be reduced to the following. Let σ_n an increasing sequence in $\mathbb{N}$. We set

$$\sigma_{-1} = -1, \quad \overline{\sigma}_n =]\sigma_{n-1}, \sigma_n] \cap \mathbb{N}, \quad u_{\overline{\sigma}_n} = (u_{\sigma_m})_{m\in\overline{\sigma}_n}, \quad \text{and}$$
$$u_{\overline{\sigma}_{\underline{n}}} = (u_{\overline{\sigma}_0}, \ldots, u_{\overline{\sigma}_n}). \tag{1.39}$$

Next, Eq. (1.40) leads to similar changes of sample functions:

$$p(u_{\overline{\sigma}_{\underline{n}}}, y_{\overline{\sigma}_{\underline{n}}}) = p(y_{\overline{\sigma}_n}/u_{\overline{\sigma}_{\underline{n-1}}}) p(u_{\overline{\sigma}_n}/u_{\overline{\sigma}_{\underline{n-1}}}, y_{\overline{\sigma}_n}) p(u_{\overline{\sigma}_{\underline{n-1}}}, y_{\overline{\sigma}_{\underline{n-1}}}). \tag{1.40}$$

In that case, any exploration transition $p(u_{\overline{\sigma}_n}/u_{\overline{\sigma}_{\underline{n-1}}}, y_{\overline{\sigma}_n})$ depends on $|\Delta\sigma_n| = \sigma_n - \sigma_{n-1}$ observation values.

Probability space discretization. For the explicit determination of the conditional exploration transitions, we generally need another particle approximation scheme. To this end, we form a *stochastic tree* that represents all a priori possible transitions. For every $n \geq 0$ and $M \geq 1$, let $\chi = \cup_{n\geq 0}\chi_n$ be the *stochastic tree* defined by

$$\chi_n = \{(u_0^{i_0}, \ldots, u_n^{i_0,\ldots,i_n}),\ i_k \in \{1,\ldots,M\}\},$$
$$\chi_{n/m} = \{u_n^{\varrho(m),i_n},\ i_n \in \{1,\ldots,M\}\} \quad \forall m \in \chi_{n-1},$$

where for each n, $(u_n^{i_0,\ldots,i_n})_{i_0,\ldots,i_n}$ is a sequence of independent random variables with the same law as U_n, and for every $m = (u_0^{i_0}, \ldots, u_{n-1}^{i_0,\ldots,i_{n-1}}) \in \chi_{n-1}$, we have $\varrho(m) = (i_0, \ldots, i_{n-1})$. In order to clarify the presentation, we still use U to denote random variables whose χ-conditional distribution is given by $\frac{1}{M^{n+1}} \sum_{m\in\chi_n} \delta_m$. In other words, in this case we have

$$p(u_{\underline{n}}) = \frac{1}{M^{n+1}} \sum_{m\in\chi_n} \delta_m \quad \text{and}$$
$$p(u_n/u_{\underline{n-1}}) = \sum_{m\in\chi_{n-1}} \left(\frac{1}{M} \sum_{b\in\chi_{n/m}} \delta_b \right) 1_m(u_{\underline{n-1}}). \tag{1.41}$$

This particle approximation scheme is a time-uniform approximation to the perturbation distribution in the sense that for every continuous function

$$\phi: \mathbb{R}^{n+1} \to \mathbb{R}$$

we have, if $\overline{\phi} = E(\phi(U_{\underline{n}}))$,

$$\left\| \frac{1}{M^{n+1}} \sum_{m \in \chi_n} \phi(m) - \overline{\phi} \right\|_{\mathbf{L}^2}^2 = \frac{1}{M} \sum_{p=0}^{n} \left(1 - \frac{1}{M}\right)^{n-p} \frac{1}{M^p} \Gamma_p^n$$
$$\leq \frac{1}{M} \|\phi(m) - \overline{\phi}\|_{\mathbf{L}^2}^2,$$

with

$$\Gamma_p^n = \sum_{1 \leq t_0 \leq \cdots \leq t_p \leq n} \|\widehat{\phi}^{t_0,\ldots,t_p} - \overline{\phi}\|_{\mathbf{L}^2}^2,$$
$$\widehat{\phi}^{t_0,\ldots,t_p} = E(\phi(U_{\underline{n}})/U_{t_0}, \ldots, U_{t_p}).$$

The last inequality has been obtained by n applications of the Cauchy–Schwartz inequality. This particle approximation scheme will prove essential to our purpose in the sense that it allows for an explicit description of the conditional exploration transitions. The following formula is a direct consequence of the Bayes formula and Eq. (1.41):

$$p(u_n/u_{\underline{n-1}}, y_n) = \frac{p(y_n/u_{\underline{n-1}}, u_n)}{\int p(y_n/u_{\underline{n-1}}, u_n)\, dp(u_n/u_{\underline{n-1}})} p(u_n/u_{\underline{n-1}}) \implies$$
$$p(u_n/u_{\underline{n-1}}, y_n) = \sum_{m \in \chi_{n-1}} \left(\sum_{b \in \chi_{n/m}} \frac{p(y_n/u_{\underline{n-1}}, b)}{\sum_{c \in \chi_{n/m}} p(y_n/u_{\underline{n-1}}, c)} \delta_b(u_n) \right) 1_m(u_{\underline{n-1}}).$$

The degeneracy of these weights can be eliminated by using the same regularization as before. Here the α-regularized weights $\mathbb{p}^\alpha(u, y)$ are again defined as

$$\mathbb{p}^\alpha(u_{\underline{n}}, y_{\underline{n}}) = \sum_{m=0}^{n} \alpha_m(n-m) \mathbb{p}(y_m, u_m/u_{\underline{m-1}})$$
$$= \sum_{m=0}^{n} \alpha_m(n-m)(\mathbb{p}(y_m/u_{\underline{m}}) \odot \mathbb{p}(u_m)),$$

and $\mathrm{op}^\alpha(U_{\underline{n}}/Y_{\underline{n}})$ denotes the conditional optimal estimate associated with this α-regularized performance. On the other hand, the α-regularization of the weights $Z^1(u, y)$ is introduced as the α-regularization of the observation process $\check{Y}$. In other words, P-a.e. we have

$$E^\alpha(\phi_n(U)/Y_{\underline{n}}) \stackrel{\text{def}}{=} E(\phi_n(\check{U})/\check{Y}_{\underline{n}}(\alpha, n), Y_n)$$
$$(Y_{\underline{n}}, Y_{\underline{n}}) = \frac{\int \phi_n(\check{u}) Z_n^\alpha(\check{u}, Y_{\underline{n}})\, dp(\check{u}_{\underline{n}}/y_{\underline{n}})}{\int 1\, Z_n^\alpha(\check{u}, Y_{\underline{n}})\, dp(\check{u}_{\underline{n}}/y_{\underline{n}})},$$

where $\check{Y}_{\underline{n}}(\alpha, n)$ is the observation process of $\check{U}$ defined in the same way as $\check{Y}$ but with $\check{V}$ replaced by a Gaussian process, P-independent of $\check{U}$, with zero mean and variance function $R(\alpha, n)$ defined for every $0 \leq m \leq n$ by

$$R(\alpha, n)_m^{-1} = \alpha_m(n-m) R_m^{-1}.$$

For the *window regularization*, that is, for $\alpha = 1_{[0,t]}$, $t > 0$, we obtain

$$\log Z_n^\alpha(u, y) = \sum_{m=n-t}^{n} \log z_m^1(u, y).$$

Finally, once an α-regularization has been selected, we obtain $\mathbf{L}^0$-convergence results similar to those stated in Eqs. (1.38).

Sampling/Resampling Exploration. To conclude this section, we propose a strategy to accelerate the exploration of the performance/probability space. The following algorithms are an extension of the well-known sampling/resampling (S/R) principles introduced in [11, 22, 49], and more recently in [13, 24, 45, 50]. Other interesting particle schemes based on birth and death principles have been introduced in [27] and [42]. The sampling/resampling approach differs from the others in the way it stores and updates the information that is accumulated through the resampling of the positions. The basic idea is to iteratively build up a pure Dirac comb approximation (with out weights) of the conditional sample functions $p(u_{\underline{n}}/y_{\underline{n}})$, that is, to construct discrete-time stochastic processes $\widehat{U}_{\underline{n}}^n = (\widehat{U}_0^n, \ldots, \widehat{U}_n^n)$ such that, in a sense to be defined,

$$\frac{1}{N} \sum_{i=1}^{N} \delta_{\widehat{U}_{\underline{n}}^{n,i}} \underset{N \to +\infty}{\longrightarrow} p(u_{\underline{n}}/y_{\underline{n}}),$$

where $\widehat{U}_{\underline{n}}^{n,i}$ is a sequence of independent processes having the same law as $\widehat{U}_{\underline{n}}^n$. The symbol $\widehat{(\cdot)}^n$ means that the conditional sample function of the process depends on the observation path $y_{\underline{n}}$. Before starting the discussion, we recall that if $p(a, b)$ is the distribution of some random variable (A, B), then, in some sense,

$$\frac{1}{N} \sum_{i=1}^{N} \delta_{(A^i, B^i)} \underset{N \to +\infty}{\longrightarrow} p(a, b), \tag{1.42}$$

where (A^i, B^i) is a sequence of independent random variables having the same law as (A, B).

We initialize the sampling/resampling algorithm by first introducing a sequence of independent stochastic processes $U_{\underline{n}}^i = (U_0^i, \ldots, U_n^i)$ having the same law as $U_{\underline{n}}$. By the same line of argument as before, in some sense, we have

1) $p_0(u_0) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} \delta_{U_0^i} \underset{N \to +\infty}{\longrightarrow} p(u_0);$

2) $$p_0(u_0/y_0) \stackrel{\text{def}}{=} \frac{p(y_0/u_0)}{\int p(y_0/u_0)\,dp_0(u_0)}\,p_0(u_0) \underset{N\to+\infty}{\longrightarrow} p(u_0/y_0). \qquad (1.43)$$

By $\widehat{U}_0^0$ we denote the random variable whose law is given by $p_0(u_0/y_0)$ and by $\widehat{U}_0^{0,i}$ a sequence of independent variables with the same law as $\widehat{U}_0^0$. The last remark (1.42) and the approximation (1.43) yield

1) $$p_1(u_{\underline{1}}/y_0) \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^N \delta_{\widehat{U}_{\underline{1}}^{0,i}} \underset{N\to+\infty}{\longrightarrow} p(u_1)p(u_0/y_0) = p(u_{\underline{1}}/y_0)$$
(the first S/R update)

with $\widehat{U}_{\underline{1}}^{0,i} \stackrel{\text{def}}{=} (\widehat{U}_0^{0,i}, U_1^i)$;

2) $$p_1(u_{\underline{1}}/y_{\underline{1}}) \stackrel{\text{def}}{=} \frac{p(y_1/u_{\underline{1}})}{\int p(y_1/u_{\underline{1}})dp_1(u_{\underline{1}}/y_0)}\,p_1(u_{\underline{1}}/y_0) \underset{N\to+\infty}{\longrightarrow} p(u_0/y_0)$$
(the first Bayesian Correction).

Now by $\widehat{U}_{\underline{1}}^1 = (\widehat{U}_0^1, \widehat{U}_1^1)$ we denote a variable with distribution $p_1(u_{\underline{1}}/y_{\underline{1}})$. By the same line of argument, if $\widehat{U}_{\underline{1}}^{1,i}$ is a sequence of independent variables with the same law as $\widehat{U}_{\underline{1}}^1$, then we obtain:

1) $$p_2(u_{\underline{2}}/y_{\underline{1}}) \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^N \delta_{\widehat{U}_{\underline{2}}^{1,i}} \underset{N\to+\infty}{\longrightarrow} p(u_2)p(u_{\underline{1}}/y_{\underline{1}}) = p(u_{\underline{2}}/y_{\underline{1}})$$
(the second S/R update)

with $\widehat{U}_{\underline{2}}^{1,i} \stackrel{\text{def}}{=} (\widehat{U}_{\underline{1}}^{1,i}, U_2^i)$;

2) $$p_2(u_{\underline{2}}/y_{\underline{2}}) \stackrel{\text{def}}{=} \frac{p(y_2/u_{\underline{2}})}{\int p(y_2/u_{\underline{2}})\,dp_2(u_{\underline{2}}/y_{\underline{1}})}\,p_2(u_{\underline{2}}/y_{\underline{1}}) \underset{N\to+\infty}{\longrightarrow} p(u_{\underline{2}}/y_{\underline{2}})$$
(the second Bayesian Correction).

Then by $\widehat{U}_{\underline{2}}^2 = (\widehat{U}_0^2, \widehat{U}_1^2, \widehat{U}_2^2)$ we denote a variable with distribution $p_2(u_{\underline{2}}/y_{\underline{2}})$.
For every $n \geq 1$, the S/R process $\widehat{U}_{\underline{n}}^n$ is defined recursively as follows:
Let

$$p_{n-1}(u_{\underline{n-1}}/y_{\underline{n-1}}) \underset{N\to+\infty}{\longrightarrow} p(u_{\underline{n-1}}/y_{\underline{n-1}})$$

be the sampling function of the process $\widehat{U}_{\underline{n-1}}^{n-1}$; then

1) $$p_n(u_{\underline{n}}/y_{\underline{n-1}}) \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^N \delta_{\widehat{U}_{\underline{n}}^{n-1,i}} \underset{N\to+\infty}{\longrightarrow} p(u_n)p(u_{\underline{n-1}}/y_{\underline{n-1}}) = p(u_{\underline{n}}/y_{\underline{n-1}}) \quad \textit{(the nth S/R update)}$$

with $\widehat{U}_{\underline{n}}^{n-1,i} \stackrel{\text{def}}{=} (\widehat{U}_{\underline{n-1}}^{n-1,i}, U_n^i)$.

2) $$p_n(u_{\underline{n}}/y_{\underline{n}}) \stackrel{\text{def}}{=} \frac{p(y_n/u_{\underline{n}})}{\int p(y_n/u_{\underline{n}})\,dp_n(u_{\underline{n}}/y_{\underline{n-1}})}\,p_n(u_{\underline{n}}/y_{\underline{n-1}}) \underset{N\to+\infty}{\longrightarrow} p(u_{\underline{n}}/y_{\underline{n}}) \quad \textit{(the nth-Bayesian Correction)}$$

Then by $\widehat{U}_{\underline{n}}^n$ we denote the stochastic process whose sampling function is $p_n(u_{\underline{n}}/y_{\underline{n}})$. These S/R principles can be summarized as follows:

$$\widehat{U}_{\underline{n}}^{n-1} \;\overrightarrow{N \text{ a priori samplings}}\; \widehat{U}_{\underline{n}}^{n-1} \;\overrightarrow{N \text{ conditional resamplings}}\; \widehat{U}_{\underline{n}}^n.$$

When we use the previous approximations, we have, in a sense to be defined,

$$\frac{1}{N}\sum_{i=1}^N \delta_{\widehat{U}_{\underline{n}}^{n,i}} \underset{N\to+\infty}{\longrightarrow} p(u_{\underline{n}}/y_{\underline{n}}),$$

$$E_N(\phi_n(U)/Y_{\underline{n}}) \stackrel{\text{def}}{=} \frac{1}{N}\sum_{i=1}^N \phi_n(\widehat{U}_{\underline{n}}^{n,i}) \underset{N\to+\infty}{\longrightarrow} E(\phi_n(U)/Y_{\underline{n}}).$$

The analysis of this convergence necessarily involves the study of all approximations that lead to $p_n(u_{\underline{n}}/y_{\underline{n}})$. One open problem is to find sufficient conditions for these S/R particle schemes to L^0-converge uniformly in time to the conditional expectation. Some local proofs can be found in [13]. As usual, these S/R particle schemes are also applicable to *deterministic optimization problems*. Indeed, if Y is a reference path for which the conditional performance $\mathbb{p}(u_{\underline{n}}/Y_{\underline{n}})$ satisfies the regularity conditions stated in (1.33), then for every $\varepsilon > 0$ there exists an $\eta > 0$ such that

$$P(\|\operatorname{op}_N(U_{\underline{n}}/Y_{\underline{n}}) - \operatorname{op}(U_{\underline{n}}/Y_{\underline{n}})\|_2 > \varepsilon) \leq (1 - P(\|\widehat{U}_{\underline{n}}^n - \operatorname{op}(U_{\underline{n}}/Y_{\underline{n}})\|_2 \leq \eta))^N,$$

where

$$\operatorname{op}_N(U_{\underline{n}}/Y_{\underline{n}}) = A \bigoplus_{i=1}^N \mathbb{p}(\widehat{U}_{\underline{n}}^{n,i}, Y_{\underline{n}}).$$

In other words,

$$\forall \varepsilon > 0 \quad P(\|\widehat{U}_{\underline{n}}^n - \operatorname{op}(U_{\underline{n}}/Y_{\underline{n}})\|_2 \leq \varepsilon) > 0$$
$$\Longrightarrow \operatorname{op}_N(U_{\underline{n}}/Y_{\underline{n}}) \underset{N\to+\infty}{\overset{L^0}{\longrightarrow}} \operatorname{op}(U_{\underline{n}}/Y_{\underline{n}}).$$

More generally, the resampling updates may be given by some timing sequence schedule t_n. The recent literature ([50], [24], [11], and [13]) describes several different schemes for choosing $\Delta t_n = t_n - t_{n-1}$, none of which, in our opinion, is completely reliable. It is our opinion that the choice of the control parameters Δt_n and the assignment of the schedule must require physical insight and/or trial-and-error experiments. To clarify the presentation, we have restricted the study to the case $\Delta t_n = 1$. In fact, a suitable state-space augmentation allows one to reduce the more general case to the one considered above.

11. Convergence

The aim of this last section is to give sufficient conditions for the time-uniform convergence of our particle schemes. Our results indicate that once the filtering or optimization problem satisfies some natural *stochastic detectability* (in reference to the linear terminology) and *continuity* assumptions, $\mathbf{L}^0$-convergence is uniform in time. Let α^r be an increasing sequence of regularizations in $\mathcal{R}$ which converges pointwise to 1 as $r \to +\infty$ (for example, $\alpha^r_m(t) = 1_{[0,r]}(t)$, $\alpha^r_m(t) = (1-r^{-1})^t, \ldots$). For convenience, we use the following notation:

1. $\check{U}$ is one of the generic exploration processes defined in the previous section.

2. The index $(\cdot)^{\alpha^r}$ will be replaced by $(\cdot)^r$, and we write $\|u_{\underline{n}}\|^2_{2,r} = \sum_{m=0}^n \alpha^r_m(n-m)u^2_m$.

3. For every function $\theta: \mathbb{N} \to \mathbb{R}^+$ we set $\theta^\star = \sup_n \theta_n$. If A is a discrete-time stochastic process and A^N is a sequence of discrete-time stochastic processes, then we define the $\mathbf{L}^p$-time uniform convergences as follows.

$$\text{For } p=0:\quad A^N \xrightarrow[N\to+\infty]{\mathbf{L}^{0,\star}} A \iff \forall \varepsilon > 0 \quad P(|A^N_{\cdot} - A_{\cdot}| > \varepsilon)^\star \underset{N\to+\infty}{\longrightarrow} 0.$$

$$\text{For } p \geq 1:\quad A^N \xrightarrow[N\to+\infty]{\mathbf{L}^{p,\star}} A \iff$$

$$\|A^N_{\cdot} - A_{\cdot}\|^\star_p = E((A^N_{\cdot} - A_{\cdot})^p)^{p^{-1},\star} \underset{N\to+\infty}{\longrightarrow} 0.$$

In the sequel, we suppose that the conditional expectation satisfies the natural asymptotic condition $\|\phi_{\cdot}(U) - E(\phi_{\cdot}(U)/Y_{\underline{\cdot}})\|^\star_2 < +\infty$ and, for every $r \geq 1$, $(\sum_{m=0}^{\cdot} \alpha^r_m(\cdot - m))^\star < +\infty$.

Theorem 1.8 ([13]) *Assume that the following conditions are satisfied:*

1. *Stochastic detectability:* $\|\phi_{\cdot}(\check{U}) - E(\phi_{\cdot}(U)/Y_{\underline{\cdot}})\|^\star_2 < +\infty$.

2. *Continuity:* $E^r(\phi_{\cdot}(U)/Y_{\underline{\cdot}}) \xrightarrow[r\to+\infty]{\mathbf{L}^{0,\star}} E(\phi_{\cdot}(U)/Y_{\underline{\cdot}})$.

Then there exists an increasing parameter sequence $r(N)$ such that

$$E^{r(N)}_N(\phi_{\cdot}(U)/Y_{\underline{\cdot}}) \xrightarrow[N\to+\infty]{\mathbf{L}^{0,\star}} E(\phi_{\cdot}(U)/Y_{\underline{\cdot}}). \tag{1.44}$$

Corollary 1.2 *Let $\alpha^r = 1_{[0,r]}$, and assume that the following conditions are satisfied:*

1. *Stochastic detectability:* $\|\phi_{\cdot}(\check{U}) - E(\phi_{\cdot}(U)/Y_{\underline{\cdot}})\|^\star_2 < +\infty$.
2. *Asymptotic independence:*

$$E(\phi_n(U)/\phi_{n-r}(U), Y_{n-r}, \ldots, Y_n) \xrightarrow[r\to+\infty]{\mathbf{L}^{0,\star}} E(\phi_n(U)/Y_{n-r}, \ldots, Y_n).$$

Then there exists an increasing sequence of parameters $r(N)$ such that the $\mathbf{L}^0$-time uniform convergence (1.44) holds.

Suppose that Y is a reference value such that the conditional performances $\mathbb{P}^r(u_{\underline{n}}/Y_{\underline{n}})$ satisfy the following regularity conditions (1.33):

1) $\forall \varepsilon > 0 \quad \exists \eta > 0: \quad \forall n \geq 0 \quad \rho(\mathbb{P}^r(u_{\underline{n}}/Y_{\underline{n}}), \mathbb{I}) \leq \eta$
$$\Longrightarrow \|u_{\underline{n}} - \mathrm{op}(U_{\underline{n}}/Y_{\underline{n}})\|_{2,r} \leq \varepsilon,$$

2) $\forall \varepsilon > 0 \quad \exists \eta > 0: \quad \forall n \geq 0 \quad \|u_{\underline{n}} - \mathrm{op}(U_{\underline{n}}/Y_{\underline{n}})\|_{2,r} \leq \eta$
$$\Longrightarrow \rho(\mathbb{P}^r(u_{\underline{n}}/Y_{\underline{n}}), \mathbb{I}) \leq \varepsilon.$$

Let $\check{U}$ be one of the generic exploration processes defined in the previous section or one of the R/S-exploration processes.

Theorem 1.9 *Assume that the following conditions are satisfied:*

1. *Stochastic detectability: for every $r \geq 1$ and $\varepsilon > 0$,*
$$P(\|\check{U}_{\underline{\cdot}} - \mathrm{op}^r(U_{\underline{\cdot}}/Y_{\underline{\cdot}})\|_{2,r} > \varepsilon)^\star < 1.$$
2. *Continuity:* $\lim_{r\to+\infty} P(\|\mathrm{op}^r(U_{\underline{\cdot}}/Y_{\underline{\cdot}}) - \mathrm{op}(U_{\underline{\cdot}}/Y_{\underline{\cdot}})\|_2 > \varepsilon)^\star = 0$.

Then there exists an increasing sequence of parameters $r(N)$ such that

$$\lim_{N\to+\infty} P(\|\mathrm{op}^{r(N)}_N(U_{\underline{\cdot}}/Y_{\underline{\cdot}}) - \mathrm{op}(U_{\underline{\cdot}}/Y_{\underline{\cdot}})\|_{2,r(N)} > \varepsilon)^\star = 0. \tag{1.45}$$

For detailed proofs and for the optimization of the parameters $r(N)$ for each finite number of particles, see [13].

Conclusions

In this paper we have introduced Maslov optimization theory as a natural normed and idempotent semiring-valued measure theory at the same level of generality as that of probability and stochastic process theory. This work offers an alternative to classical geometric descriptions of optimization problems and leads to new developments in the qualitative studies of optimization processes. From a practical point of view, this parallelism between probability theory and optimization theory allows us to apply the recently developed particle methods [13, 15] to optimization problems. This paper is only concerned with deterministic processes. Stochastic optimization problems can be studied along the same lines, but the stochastic Bellman optimality equation is no longer $(\max,+)$-linear. For linear systems, the stochastic performance evolution may be described by the Maslov $\circledast$ or the classical $\star$ convolution of measures. Let $\mathbb{p}$ be a real performance measure, let p be a real-valued probability density, and let ϕ be a mapping from $\mathbb{R}$ into $\mathcal{A}$. Let U_k be a sequence of independent optimization variables with the same performance $\mathbb{p}$, and let W_k be a sequence of random variables with the same distribution p. Consider the simplest stochastic optimization problem (with W as the stochastic disturbance)

$$X_k = X_{k-1} - U_k - W_k, \qquad 0 \le k \le T.$$

We want to select an adapted strategy, still denoted by U, which optimizes the performance function

$$\sup_{U_0,\dots,U_T} E\Big(\sum_{l=0}^{T-1} \mathbb{p}(U_l) + \phi(X_T)\Big),$$

where $E(\cdot)$ denotes the usual probabilistic expectation over W. Then the induced performance function defined by the equation

$$\overline{\mathbb{p}}_k(x) = \sup_{U_k,\dots,U_T} E\Big(\sum_{l=k}^{T-1} \mathbb{p}(U_l) + \phi(X_T)/X_k\Big)(x), \quad \overline{\mathbb{p}}_T = \phi,$$

satisfies the stochastic optimality equation:

$$\overline{\mathbb{p}}_k(x) = \int_{\mathbb{R}}^{\oplus} \mathbb{p}(u) \odot E(\overline{\mathbb{p}}_{k+1}(x-u-W)) \odot du$$
$$= \int_{\mathbb{R}}^{\oplus} \mathbb{p}(x-z) \odot E(\overline{\mathbb{p}}_{k+1}(z-W)) \odot dz.$$

This equation can be written using the classical and $(\max,+)$-convolutions of measures in a nondistributive way. Indeed, with the change of the time index $n = T-k$, we obtain

$$\overline{\mathbb{p}}_n = \mathbb{p} \circledast (\overline{\mathbb{p}}_{n-1} \star p) \quad \text{and} \quad \overline{\mathbb{p}}_n = \mathbb{p}^n \circledast \phi \star p^n,$$

where for each $2 \le n \le T$ we have

$$\mathbb{p}^n \circledast \phi \star p^n = \mathbb{p} \circledast (\mathbb{p}^{n-1} \circledast \phi \star p^{n-1}) \star p, \qquad \mathbb{p} \circledast \phi \star p = \mathbb{p} \circledast (\phi \star p).$$

In the general case, however, there is no escape from the fact that using both algebras makes the problem nonlinear in any of them. In that respect, [9] simply shifts the difficulty to the suitable choice of the Lagrange multipliers (the equations defining these multipliers are again nonlinear). Although there has been a large amount of theoretical work in the field of nonlinear filtering and dynamic optimization in the last thirty years since their inception, little attention has been paid to actual realization of nonlinear estimates for real problems, which remains a challenge for the computation of global estimates. Another contribution of this paper is to introduce some particle principles that fully exploit the structure and the nonlinearities of the systems, and we give conditions for the $\mathbf{L}^0$-time uniform convergence of these schemes in terms of *stochastic detectability and observability*. This paper is also a milestone in showing that the original particle algorithms developed for nonlinear filtering can be used to solve optimization problems. It is our opinion that particle principles are more likely to be used in practice than linearization or fixed grid schemes. Furthermore, in practical situations, the structure and nonlinearity of the problem in hand should be fully exploited. The results on sufficient conditions for the particle procedures to converge uniformly in time are natural. Despite many successful applications of these techniques, important questions remain to be answered. We should mention that

1. A gap in the theory is the lack of simpler conditions implying detectability and regularity.
2. There are a number of questions regarding the convergence of the particle approximations based on sampling and resampling principles. As we alluded to earlier, the complete treatment of the S/R principles is a very complicated and sophisticated subject. All we have attempted to do here is to introduce natural modeling and to guide the reader to some starting points in the literature.
3. Detectability is usually connected with the positive Hessian of some value function. There remain unresolved issues in our investigations for nonlinear filtering problems.

The main point of particle resolution relies on the sampling of a large number of particle paths. Progress also lies in numerical studies particularly with the help of parallel computing. The field of Monte–Carlo simulations has long been a rendez-vous point for practitioners, algorithm makers and theorists. There are reasons to hope that this will continue in the future to benefit the development of the field of nonlinear filtering and optimization problems.

Laboratoire d'Analyse et d'Architecture des Systèmes (LAAS) du CNRS
7, Avenue du Colonel Roche 31077 Toulouse Cedex, France
Current address: Laboratoire de Statistiques et Probabilites
URA 745-BAT 1R1
118 Route de Narbonne, 31062 Toulouse Cedex, France
E-mail: delmoral@cict.fr

References

1. F. Bacelli, G. Cohen, G. J. Olsder, and J. P. Quadrat. *Synchronization and Linearity, An Algebra for Discrete Event Systems*. John Wiley and Sons, 1992.
2. R. E. Bellman and S. E. Dreyfus. *Dynamic Programming and Applications*. Dunod, Paris, 1965.
3. R. S. Bucy and P. D. Joseph. *Filtering for Stochastic Processes with Applications to Guidance*. John Wiley, 1968.
4. R. S. Bucy. Realization of nonlinear filters. *Proc. 2nd Symposium on Nonlinear Estimation Theory and Its Applications*, San Diego, 1971, pages 51–58.
5. R. S. Bucy. Nonlinear filtering. *Proc. of 1975 IEEE-USSR Joint Workshop on Information Theory*, December 15-19, 1975, pages 20–25.
6. R. S. Bucy, M. J. Merritt, and D. S. Miller. Hybrid Computer Synthesis of Optimal Discrete Nonlinear Filters. *Proc. 2nd Symposium on Nonlinear Estimation Theory and Its Applications*, San Diego, 1971, pages 59–87.
7. Cox H.. Estimation of state variables via dynamic programming. *Proc. 1964 Joint Automatic Control Conf.*, Stanford, California, pages 376–381.
8. R. Cuninghame-Green. Minimax Algebra. *Lect. Notes in Economics and Mathematical Systems*, 166, Springer-Verlag, 1979.
9. M. H. A. Davis and G. Burnstein. A deterministic approach to stochastic optimal control with application to anticipative control. *Stochastics*, 40:203–256, 1992.
10. P. Del Moral, T. Huillet, G. Rigal, and G. Salut. Optimal versus random processes: the nonlinear case. Rapport LAAS N° 91131, Avril 1991.
11. P. Del Moral, G. Rigal, and G. Salut. Filtrage non linéaire non-Gaussien appliqué au recalage de plates-formes inertielles-Mise en équations spécifiques. Contrat DIGILOG-S.T.C.A.N., N° A.91.77.013, Rapport N° 1, Septembre 1991.
12. P. Del Moral, J. C. Noyer, G. Rigal, and G. Salut. Traitement non-linéaire du signal par réseau particulaire: application Radar. 14ième Colloque GRETSI, Juan les Pins, 13-16 Septembre 1993.
13. P. Del Moral. Résolution particulaire des problèmes d'estimation et d'optimisation non linéaires. Thèse Université Paul Sabatier, Toulouse, 1994.
14. P. Del Moral. Maslov Optimization theory: Topological Aspects. april 1994, LAAS-CNRS report N° 94 110.
15. P. Del Moral, G. Rigal, and G. Salut. Estimation et commande optimale non linéaire: Méthodes de résolutions particulaires. Rapport final de contrat DIGILOG-DRET N° 89.34.553.00.470.75.01, Octobre 1992.
16. P. Del Moral and G. Salut. Maslov Optimization Theory: Optimality Versus Randomness. LAAS report N° 94211 (34 p), May 1994.
17. P. Del Moral. An introduction to the theory of optimization processes. LAAS report N° 94127 (27 p), April 1994.
18. P. Del Moral. Optimization processes: Hamilton–Lyapunov stability. LAAS report N° 94207 (16 p), May 1994.
19. P. Del Moral, J. C. Noyer, and G. Salut. Maslov Optimization Theory: Stochastic Interpretation, Particle Resolution. 11eme *Conférence Internationale sur l'Analyse et l'Optimisation des Systèmes*. Ecole des Mines Sophia-Antipolis, France, 15-16-17 Juin 1994. *Lect. Notes Contr. Inform. Sci.*, 199, Springer-Verlag, 1995.
20. P. Del Moral and G. Salut. Maslov Optimization theory. *Russian Journal of Mathematical Physics*, 1997 (to appear).
21. P. I. Dudnikov and S. N. Samborskii. Endomorphisms of semimodules over semirings with an idempotent operation [in Russian]. Preprint N° 87-48, Inst. Mat. Akad. Nauk Ukrain., Kiev, 1987.
22. B. Efron. *The Bootstrap, Jackknife and Other Resampling Plans*. Philadelphia: Society of Industrial and Applied Mathematics, 1982.
23. M. I. Freidlin and A. D. Wentzell. *Random Perturbations of Dynamical Systems*. Springer Verlag, 1979.

24. N. J. Gordon, D. J. Salmond, and A. F. M. Smith. *Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation*. IEE, 1993
25. O. Hijab. Minimum energy estimation. Doctoral Dissertation, University of California, Berkeley, 1980.
26. T. Huillet and G. Salut. Stochastic processes and optimal processes. LAAS-CNRS report N° 89025, January 1989.
27. T. Huillet and G. Salut. Interprétation des équations du filtrage non-linéaire. Séances du GdR Automatique du C.N.R.S. (Pôle non-linéaire), Paris, 8 Novembre 1989.
28. A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970.
29. V. N. Kolokoltsov. Semiring analogs of linear equivalent spaces [in Russian]. Abstr. Conf. on Optimal Control, Geometry, Kemerovo, 1988, p. 30.
30. V. N. Kolokoltsov and V. Maslov. Idempotent calculus as the apparatus of optimization theory. *Funktsional. Anal. i Prilozhen.*, 23(1):1–14, 23(4)53–62 (1989). Engl. transl. in *Functional Anal. Appl.*, 23 (1989).
31. V. N. Kolokoltsov and 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(2):283–287 (1987). Engl. transl. in *Sov. Math. Dokl.*, 36(1):55–59 (1988).
32. H. J. Kushner. On the differential equations satisfied by conditional probability densities of Markov processes. *SIAM J. Control*, 2:106–119, 1964.
33. G. L. Litvinov and V. P. Maslov. Correspondence principle for idempotent calculus and some computer applications. Preprint IHES-M-95-33.
34. V. P. Maslov. *Méthodes opératorielles* [in French]. Mir Publishers, 1987.
35. V. P. Maslov. Quasilinear systems which are linear in some semimoduli. *Congrès international sur les problèmes hyperboliques*, 13-17 Janvier 1986, Saint Etienne, France.
36. V. P. Maslov and S. N. Samborskii. Idempotent Analysis. *Advances in Soviet Mathematics*, 13, AMS, Providence (1992).
37. V. P. Maslov and S. N. Samborskii. Stationary Hamilton–Jacobi and Bellman equations. *Advances in Soviet Mathematics*, 13, AMS, Providence (1992).
38. P. Del Moral. Nonlinear filtering using random particles. *Probability Theory*, 40(4), 1995 (to appear).
39. P. Del Moral. On the stability of Maslov optimization processes. *34th Conference on Decision and Control*, December 1995.
40. P. Del Moral. Maslov optimization theory, topological aspects. *Proceeedings of the Idempotency Workshop*, Hewlett Packard Laboratories, Bristol 3-7 october 1994.
41. P. Del Moral and G. Salut. Particle interpretation of nonlinear filtering and optimization. *Russian J. Math. Phys.*, 1995 (to appear).
42. P. Muller. Monte Carlo integration in general dynamic models. *Contemp. Math.*, 115:145–163, 1991.
43. L. S. Pontryagin, A. A. Andronov, and A. A. Vitt. Statistical approach to dynamical systems. *Zh. Eksper. Teor. Fiz.*, 3 (1933)
44. J. P. Quadrat. Théorèmes asymptotiques en programmation dynamique. *Comptes Rendus à l'Académie des Sciences*, 311:745–748, 1990.
45. G. Rigal. Filtrage non linéaire, résolution particulaire et applications au traitement du signal. Thesis dissertation, University Paul Sabatier, Toulouse, July 1993.
46. R. T. Rockafellar. *Convex Analysis*. Princeton University, 1970.
47. R. T. Rockafellar. Convex functions and duality in optimization problems and dynamics. In *Mathematical Systems Theory and Economics*, I. H.W. Kuhn and G.P. Szergo, editors, Lecture Notes in ORME, Springer-Verlag, 1969.
48. R. L. Stratonovitch. Conditional Markov processes. *Theory of probability and its applications*, 5(2):156–178, 1960.
49. D. B. Rubin. *Using the SIR Algorithm to Simulate Posterior Distributions in Bayesian Statistics*. J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, editors, Cambride, MA, Oxford University Press, 1988, pages 395–402.

50. A. F. M. Smith and A. E. Gelfand. Bayesian Statistics without tears: A sampling-resampling perspective. *The American Statistician*, 46(2), May 1992.
51. R. L. Stratonovitch. *Conditional Markov Processes and Their Applications to the Theory of Optimal Control*. Elsevier, 1968.

Index