Min-Plus Probability Calculus

J.P. Quadrat *†

March 18, 1998

Abstract

We study here the duality appearing between probability calculus and optimization by substituting the semifield $(\mathbb{R} \cup \{+\infty\}, \min, +)$ for the standard semiring $(\mathbb{R}^+, +, \times)$.

1 Introduction

This lecture aid is not original, it is, mainly, a compilation and reorganization of some sections of the three papers [5, 48, 20] in which all the members of the Max-plus working group and in particular M. Viot and M. Akian have played a key role.

The min-plus probability calculus, called decision calculus, is obtained from the probability calculus by substituting the idempotent semifield $(\mathbb{R} \cup \{+\infty\}, \min, +)$ for the standard semiring $(\mathbb{R}^+, +, \times)$.

To the probability of an event corresponds the cost of a set of decisions. To random variables correspond decision variables. Almost all concepts and tools of probabilities have an analogue. First, we give the counterparts of characteristic functions, weak convergence, tightness and limit theorems.

The analogue of Markov chains are the so-called Bellman Chains. The asymptotic theorems for the Bellman chains are the general min-plus linear system ones. They can be seen in [10, 22, 23, 48] and will not discussed here. The min-plus product forms exist and correspond to computing geodesics on a $\mathbb{Z}$-module. In some cases, explicit formulae dual of the standard product forms give explicitly the minimal distance between two states. The general problem can be reduced to a standard flow problem.

*J.P. Quadrat : INRIA Domaine de Voluceau Rocquencourt, BP 105, 78153, Le Chesnay (France). Email : Jean-Pierre.Quadrat@inria.fr.

†This work has been partly supported by the ALAPEDES project of the European TMR programme.

The Cramer transform used in the large deviation literature is defined as the composition of the Laplace transform by the logarithm by the Fenchel transform. It transforms convolution into inf-convolution. Probabilistic results about processes with independent increments are then transformed into similar results on dynamic programming equations for systems with instantaneous costs which do not depend of the states. By this way we obtain explicit solutions of some Hamilton Jacobi Bellman equations (HJB) called Hopf formulae. This Cramer transform is well known in statistical mechanics. We illustrate, on a simple example, called min-plus perfect gaz, how the Cramer transform appears in the computing of the correpond-ing Gibbs distribution.

Bibliographic notes are given at the end of the paper.

2 Cost Measures and Decision Variables

2.1 Cost measures

Le us denote $\mathbb{R}_{\min}$ the idempotent semifield $(\mathbb{R} \cup \{+\infty\}, \min, +)$ and by extension the metric space $\mathbb{R} \cup \{+\infty\}$ endowed with the exponential distance $d(x, y) = |\exp(-x) - \exp(-y)|$. We start by defining cost measures which can be seen as normalized idempotent measures of Maslov in $\mathbb{R}_{\min}$ [41] .

We call a *decision space* the triplet $(U, \mathcal{U}, \mathbb{K})$ where U is a topological space, $\mathcal{U}$ the set of open sets of U and $\mathbb{K}$ a mapping from $\mathcal{U}$ to $\mathbb{R}_{\min}$ such that

1. $\mathbb{K}(U) = 0$,
2. $\mathbb{K}(\emptyset) = +\infty$,
3. $\mathbb{K}\left(\bigcup_n A_n\right) = \inf_n \mathbb{K}(A_n)$ for any $A_n \in \mathcal{U}$.

The mapping $\mathbb{K}$ is called a *cost measure*. A set of cost measures K is said *tight* if

$$\sup_{C \text{compact} \subset U} \inf_{\mathbb{K} \in K} \mathbb{K}(C^c) = +\infty \ .$$

A mapping $c : U \to \mathbb{R}_{\min}$ such that $\mathbb{K}(A) = \inf_{u \in A} c(u) \ \forall A \subset U$ is called a *cost density* of the cost measure $\mathbb{K}$.

The set $D_c \stackrel{\text{def}}{=} \{u \in U \mid c(u) \neq +\infty\}$ is called the *domain* of c.

Theorem 1. *Given a l.s.c. c with values in $\mathbb{R}_{\min}$ such that $\inf_u c(u) = 0$, the mapping $A \in \mathcal{U} \mapsto \mathbb{K}(A) = \inf_{u \in A} c(u)$ defines a cost measure on $(U, \mathcal{U})$. Conversely any cost measure defined on open sets of a second countable topological space*[1]

[1] i.e. a topological space with a countable basis of open sets.

admits a unique minimal extension $\mathbb{K}_*$ to $\mathcal{P}(U)$ (the set of subsets of U) having a density c which is a l.s.c. function on U satisfying $\inf_u c(u) = 0$.

Proof. This precise result is proved in Akian [1]. See also Maslov [41] and Del Moral [31] for the first part and Maslov and Kolokoltsov [39] for the second part. □

This theorem shows that on second countable spaces there is a bijection between l.s.c. functions and cost measures. In this paper, we will consider cost measures on $\mathbb{R}^n$, $\mathbb{R}^{\mathbb{N}}$, separable Banach spaces and reflexive Banach separable spaces with the weak topology which are all second countable topological spaces.

We will use very often the two following cost densities defined on $\mathbb{R}^n$ with $\|.\|$ the euclidian norm.

1. $\chi_m(x) \stackrel{\text{def}}{=} \begin{cases} +\infty & \text{for } x \neq m. \\ 0 & \text{for } x = m, \end{cases}$

2. $\mathcal{M}^p_{m,\sigma}(x) \stackrel{\text{def}}{=} \frac{1}{p}\|\sigma^{-1}(x-m)\|^p$ for $p \geq 1$ with $\mathcal{M}^p_{m,0} \stackrel{\text{def}}{=} \chi_m$.

By analogy with the conditional probability we define *conditional cost excess* to take the best decision in A knowing that it must be taken in B by

$$\mathbb{K}(A|B) \stackrel{\text{def}}{=} \mathbb{K}(A \cap B) - \mathbb{K}(B) \ .$$

2.2 Decision Variables

By analogy with random variables we define decision variables and related notions.

1. A *decision variable* X on $(U, \mathcal{U}, \mathbb{K})$ is a mapping from U to E (a second countable topological space). It induces a cost measure $\mathbb{K}_X$ on $(E, \mathcal{B})$ ($\mathcal{B}$ denotes the set of open sets of E) defined by $\mathbb{K}_X(A) = \mathbb{K}_*(X^{-1}(A))$ for all $A \in \mathcal{B}$. The cost measure $\mathbb{K}_X$ has a l.s.c. density denoted c_X . When $E = \mathbb{R}$, we call X a real decision variable; when $E = \mathbb{R}_{\min}$, we call it a *cost variable*.

2. Two decision variables X and Y are said *independent* when:
$$c_{X,Y}(x, y) = c_X(x) + c_Y(y).$$

3. The *conditional cost excess* of X knowing Y is defined by:
$$c_{X|Y}(x, y) \stackrel{\text{def}}{=} \mathbb{K}_*(X = x \mid Y = y) = c_{X,Y}(x, y) - c_Y(y).$$

4. The *optimum* of a decision variable is defined by

$$\mathbb{O}(X) \stackrel{\text{def}}{=} \arg\min_{x \in E} \operatorname{conv}(c_X)(x)$$

when the minimum exists, where conv denotes the l.s.c. convex hull and arg min the point where the minimum is reached. When a decision variable X with values in a linear space satisfies $\mathbb{O}(X) = 0$ we say that it is *centered*.

5. When the optimum of a decision variable X with values in $\mathbb{R}^n$ is unique and when near the optimum, we have

$$\operatorname{conv}(c_X)(x) = \frac{1}{p}\|\sigma^{-1}(x - \mathbb{O}(X))\|^p + o(\|x - \mathbb{O}(X)\|^p) \ ,$$

we say that X is of order $p$and we define its *sensitivity of order p* by $\mathbb{S}^p(X) \stackrel{\text{def}}{=} \sigma$. When $\mathbb{S}^p(X) = I$ (the identity matrix) we say that X is of *order p and normalized*.

6. The *value* of a cost variable X is $\mathbb{V}(X) \stackrel{\text{def}}{=} \inf_x(x + c_X(x))$, *the conditional value* is $\mathbb{V}(X \mid Y = y) \stackrel{\text{def}}{=} \inf_x(x + c_{X|Y}(x, y))$.

7. The density cost of the sum Z of two independent variables X and Y is the *inf-convolution of their cost densities c_X and c_Y*, denoted $c_X \star c_Y$ defined by

$$c_Z(z) = \inf_{x,y} \left[c_X(x) + c_Y(y) \mid x + y = z\right] .$$

For a real decision variable X of cost $\mathcal{M}^p_{m,\sigma}$, $p > 1$, we have

$$\mathbb{O}(X) = m, \ \mathbb{S}^p(X) = \sigma, \ \mathbb{V}(X) = m - \frac{1}{p'}\sigma^{p'} .$$

2.3 Vector Spaces of Decision Variables

We can introduce vector spaces of decision variables which are the analogue of the standard $L^p(\Omega)$ spaces.

Theorem 2. *For $p > 0$, the numbers*

$$|X|_p \stackrel{\text{def}}{=} \inf\left\{\sigma \mid c_X(x) \geq \frac{1}{p}|(x - \mathbb{O}(X))/\sigma|^p\right\} \ \text{and} \ \|X\|_p \stackrel{\text{def}}{=} |X|_p + |\mathbb{O}(X)|$$

define respectively a seminorm and a norm on the vector space $\mathbb{L}^p$ of real decision variables having a unique optimum and such that $\|X\|_p$ is finite.

Proof. Let us denote $X' = X - \mathbb{O}(X)$ and $Y' = Y - \mathbb{O}(Y)$. We first remark that $\sigma > |X|_p$ implies

$$c_X(x) \geq \frac{1}{p}(|x - \mathbb{O}(X)|/\sigma)^p \;\forall x \Leftrightarrow \mathbb{V}(-\frac{1}{p}|X'/\sigma|^p) \geq 0 \; . \tag{1}$$

If there exists $\sigma > 0$ and $\mathbb{O}(X)$ such that (1) holds, then $c_X(x) > 0$ for any $x \neq \mathbb{O}(X)$ and $c_X(x)$ tends to 0 implies x tends to $\mathbb{O}(X)$ therefore $\mathbb{O}(X)$ is the unique optimum of X. Moreover $|X|_p$ is the smallest σ such that (1) holds.

If $X \in \mathbb{L}^p$, $\lambda \in \mathbb{R}$ and $\sigma > |X|_p$ we have

$$\mathbb{V}(-\frac{1}{p}|\lambda X'/\lambda\sigma|^p) = \mathbb{V}(-\frac{1}{p}|X'/\sigma|^p) \geq 0 \; ,$$

then $\lambda X \in \mathbb{L}^p$, $\mathbb{O}(\lambda X) = \lambda \mathbb{O}(X)$ and $|\lambda X|_p = |\lambda||X|_p$.

If X and $Y \in \mathbb{L}^p$, $\sigma > |X|_p$ and $\sigma' > |Y|_p$,

$$\mathbb{V}(-\frac{1}{p}(\max(|X'/\sigma|^p, |Y'/\sigma'|^p)) = \min(\mathbb{V}(-\frac{1}{p}|X'/\sigma|^p), \mathbb{V}(-\frac{1}{p}|Y'/\sigma'|^p)) \geq 0$$

and

$$\frac{|X'+Y'|}{\sigma+\sigma'} \leq \frac{\sigma}{\sigma+\sigma'}\frac{|X'|}{\sigma} + \frac{\sigma'}{\sigma+\sigma'}\frac{|Y'|}{\sigma'} \leq \max(\frac{|X'|}{\sigma}, \frac{|Y'|}{\sigma'}) \; ,$$

then

$$\mathbb{V}(-\frac{1}{p}(|X'+Y'|/(\sigma+\sigma'))^p) \geq 0 \; .$$

Therefore we have proved that $X + Y \in \mathbb{L}^p$ with $\mathbb{O}(X+Y) = \mathbb{O}(X) + \mathbb{O}(Y)$ and $|X+Y|_p \leq |X|_p + |Y|_p$.

Then $\mathbb{L}^p$ is a vector space, $|.|_p$ and $\|.\|_p$ are seminorms and $\mathbb{O}$ is a linear continuous operator from $\mathbb{L}^p$ to $\mathbb{R}$. Moreover, $\|X\|_p = 0$ implies $c_X = \chi$ thus $X = 0$ up to a set of infinite cost. $\square$

Theorem 3. *For two independent real decision variables X and Y and $k \in \mathbb{R}$ we have (as soon as the right and left hand sides exist)*

$$\mathbb{O}(X+Y) = \mathbb{O}(X) + \mathbb{O}(Y), \quad \mathbb{O}(kX) = k\mathbb{O}(X), \quad \mathbb{S}^p(kX) = |k|\mathbb{S}^p(X) \; ,$$

$$[\mathbb{S}^p(X+Y)]^{p'} = [\mathbb{S}^p(X)]^{p'} + [\mathbb{S}^p(Y)]^{p'}, \quad (|X+Y|_p)^{p'} \leq (|X|_p)^{p'} + (|Y|_p)^{p'} \; .$$

Proof. Let us prove only the last inequality. Consider X and Y in $\mathbb{L}^p$ and $\sigma > |X|_p$ and $\sigma' > |Y|_p$. Let us denote $\sigma'' = (\sigma^{p'} + \sigma'^{p'})^{1/p'}$, $X' = X - \mathbb{O}(X)$ and $Y' = Y - \mathbb{O}(Y)$. The Hölder inequality $a\alpha + b\beta \le (a^p + b^p)^{1/p}(\alpha^{p'} + \beta^{p'})^{1/p'}$ implies

$$(|X' + Y'|/\sigma'')^p \le |X'/\sigma|^p + |Y'/\sigma'|^p \ ,$$

then by the independency of X and Y we get

$$\mathbb{V}(-\frac{1}{p}(|X' + Y'|/\sigma'')^p) \ge 0 \ ,$$

and the inequality is proved. $\square$

Theorem 4 (Chebyshev). *For a decision variable belonging to $\mathbb{L}^p$ we have*

$$\mathbb{K}(|X - \mathbb{O}(X)| \ge a) \ge \frac{1}{p}(a/|X|_p)^p \ ,$$

$$\mathbb{K}(|X| \ge a) \ge \frac{1}{p}((a - \|X\|_p)^+/\|X\|_p)^p \ .$$

Proof. The first part is a straightforward consequence of the inequality $c_Y(y) \ge (|y|/|Y|_p)^p/p$ applied to centered decision variable Y.

The second part comes from the non increasing property of the function $x \in \mathbb{R}^+ \mapsto (a - x)^+/x$. $\square$

2.4 Characteristic Functions and Fenchel Transform

The role of the Laplace or Fourier transforms in probability calculus is played by the Fenchel transform in decision calculus.

Let $c \in \mathcal{C}_{\mathrm{X}}$, where $\mathcal{C}_{\mathrm{X}}$ denotes the set of functions from E (a reflexive Banach space with dual E') to $\mathbb{R}_{\min}$ convex, l.s.c. and proper[2]. Its *Fenchel transform* is the function from E' to $\mathbb{R}_{\min}$ defined by

$$\hat{c}(\theta) \stackrel{\text{def}}{=} [\mathcal{F}(c)](\theta) \stackrel{\text{def}}{=} \sup_x [\langle \theta, x \rangle - c(x)] \ .$$

Then the *characteristic function* of a decision variable is defined by $\mathbb{F}(X) \stackrel{\text{def}}{=} \mathcal{F}(c_X)$.

Important properties of the Fenchel transform are : for $f, g \in \mathcal{C}_{\mathrm{X}}$

[2] not always equal to $+\infty$

1. $\mathcal{F}(f) \in \mathcal{C}_X$,
2. $\mathcal{F}$ is an involution that is $\mathcal{F}(\mathcal{F}(f)) = f$,
3. $\mathcal{F}(f \star g) = \mathcal{F}(f) + \mathcal{F}(g)$,
4. $\mathcal{F}(f + g) = \mathcal{F}(f) \star \mathcal{F}(g)$.

Therefore, for two independent decision variables X and Y and $k \in \mathbb{R}$, we have

$$\mathbb{F}(X+Y) = \mathbb{F}(X) + \mathbb{F}(Y), \quad [\mathbb{F}(kX)](\theta) = [\mathbb{F}(X)](k\theta) \ .$$

Moreover, a decision variable with values in $\mathbb{R}^n$ is of order p if we have :

$$\mathbb{F}(X)(\theta) = \langle \mathbb{O}(X), \theta \rangle + \frac{1}{p'} \|\mathbb{S}^p(X)\theta\|^{p'} + o(\|\theta\|^{p'} \ .$$

2.5 Convergences of Decision Variables

The analogue of the topologies used in probability can be introduced. The standard limit theorems have a min-plus counterpart.

A *sequence of independent and identically costed (i.i.c.) real decision variables of cost c* on $(U, \mathcal{U}, \mathbb{K})$ is an application X from U to $\mathbb{R}^{\mathbb{N}}$ which induces the density cost

$$c_X(x) = \sum_{i=0}^{\infty} c(x_i), \quad \forall x = (x_0, x_1, \ldots) \in \mathbb{R}^{\mathbb{N}} \ .$$

The cost density is finite only on minimizing sequences of c, elsewhere it is equal to $+\infty$.

We have defined a decision sequence by its density and not by its value on the open sets of $\mathbb{R}^{\mathbb{N}}$ because the density always exists and can be defined easily.

In order to state limit theorems, we define several type of convergence of sequences of decision variables.

Definition 5. For the sequence of real decision variables $\{X_n, n \in \mathbb{N}\}$, cost measures $\mathbb{K}_n$ and c_n functions from U (a first countable topological space[3] to $\mathbb{R}_{\min}$ we say that :

1. $X_n \in \mathbb{L}^p$ *converges in p-norm* towards $X \in \mathbb{L}^p$ denoted $X_n \xrightarrow{\mathbb{L}^p} X$, if $\lim_n \|X_n - X\|_p = 0$;

[3]Each point admits a countable basis of neighbourhoods.

2. X_n *converges in cost* towards X, denoted $X_n \xrightarrow{\mathbb{K}} X$, if for all $\epsilon > 0$ we have $\lim_n \mathbb{K}\{u \mid |X_n(u) - X(u)| \geq \epsilon\} = +\infty$;

3. X_n *converges almost surely* towards X, denoted $X_n \xrightarrow{\text{a.s.}} X$, if we have $\mathbb{K}\{u \mid \lim_n X_n(u) \neq X(u)\} = +\infty$.

4. $\mathbb{K}_n$*converges weakly* towards $\mathbb{K}$, denoted $\mathbb{K}_n \stackrel{w}{\to} \mathbb{K}$, if for all f in $\mathcal{C}_b(E)$ [4] we have $\lim_n \mathbb{K}_n(f) = \mathbb{K}(f)$[5].

5. c_n converges in the epigraph sense (*epi-converges*) towards c, denoted $c_n \stackrel{epi}{\to} c$ if

$$\forall u, \quad \forall u_n \to u, \quad \liminf_n \; c_n(u_n) \geq c(u) \;, \tag{2}$$

$$\forall u, \quad \exists u_n \to u : \limsup_n \; c_n(u_n) \leq c(u) \;. \tag{3}$$

A sequence $\mathbb{K}_n$ of cost measures is said asymptoticaly tight if

$$\sup_{C \text{compact} \subset U} \liminf_n \mathbb{K}_n(C^c) = +\infty \;.$$

The different "weak" convergences have strong relations given in the following theorem.

Theorem 6. *Let $\mathbb{K}_n$, $\mathbb{K}$ be cost measures on a metric space U. Then the three following conditions are equivalent*

1. $\mathbb{K}_n \stackrel{w}{\to} \mathbb{K}$ *;*

2.

$$\liminf_n \mathbb{K}_n(F) \geq \mathbb{K}(F), \quad \forall F \textit{ closed} \;, \tag{4}$$

$$\limsup_n \mathbb{K}_n(G) \leq \mathbb{K}(G), \quad \forall G \textit{ open} \;; \tag{5}$$

3. $\lim_n \mathbb{K}_n(A) = \mathbb{K}(A)$ *for any set A such that* $\mathbb{K}(\mathring{A}) = \mathbb{K}(\bar{A})$.

---

[4] $\mathcal{C}_b(E)$ denotes the set of continuous and lower bounded functions from E to $\mathbb{R}_{\min}$.

[5] $\mathbb{K}(f) \stackrel{\text{def}}{=} \inf_u (f(u) + c(u))$ where c is the density of $\mathbb{K}$.

On asymptoticaly tight sequences $\mathbb{K}_n$ the weak convergence of $\mathbb{K}_n$ towards $\mathbb{K}$ is equivalent to (5) and

$$\liminf_n \mathbb{K}_n(C) \geq \mathbb{K}(C), \quad \forall C \text{ compact} . \tag{6}$$

On a first countable topological space, the epi convergence of l.s.c. cost densities is equivalent to (6) and (5).

Proof. See [5]. $\square$

In a locally compact space conditions (6) (5) are equivalent to the condition $\lim_n \mathbb{K}_n(f) = \mathbb{K}(f)$ for any continuous function with compact support. This is the definition of weak convergence used by Maslov and Samborski in [43]. These conditions are are also equivalent to the epigraph convergence of densities. This type of convergence does not insure that a weak-limit of cost measures is a cost measure (the infimum of the limit is not necessarily equal to zero).

Denoting by $\mathcal{K}(U)$ the set of cost measures on U (a metric space) endowed with the topology of the weak convergence, any tight set K of $\mathcal{K}(U)$ is relatively sequentially compact[6] .

These different kinds of convergence are connected in a nonstandard way.

Theorem 7. *We have the implications :*

1. *Convergence in p-norm implies convergence in cost but the converse is false.*

2. *Convergence in cost implies almost sure convergence and the converse is false.*

3. *For tight sequences, the convergence in cost implies the weak convergence.*

Proof. See Akian [2] for points 1 and 2 and 3 and Del Moral [31] for point 2. $\square$

We have the analogue of the law of large numbers.

Theorem 8. *Given a sequence $\{X_n,\ n \in \mathbb{N}\}$ of i.i.c. decision variables belonging to $\mathbb{L}^p,\ p \geq 1$, we have*

$$Y_N \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=0}^{N-1} X_n \to \mathbb{O}(X_0) ,$$

where the limit can be taken in the sense of almost sure, cost and p-norm convergence.

[6] that is any sequence of K contains a weakly convergent subsequence

Proof. We have only to estimate the convergence in p-norm. The result follows from simple computation of the p-seminorm of Y_N. Thanks to Theorem 3 we have $(|Y_N|_p)^{p'} \leq N(|X_0|_p)^{p'}/N^{p'}$ which tends to 0 as N tends to infinity. $\square$

We have the analogue of the central limit theorem.

Theorem 9. *Given an i.i.c. sequence $\{X_n, n \in \mathbb{N}\}$ centered of order p with l.s.c. convex cost, we have*

$$Z_N \stackrel{\text{def}}{=} \frac{1}{N^{1/p'}} \sum_{n=0}^{N-1} X_n \stackrel{w}{\to} \mathcal{M}^p_{0,\mathbb{S}^p(X_0)} .$$

Proof. We have $\lim_N [\mathbb{F}(Z_N)](\theta) = \frac{1}{p'} \|\mathbb{S}^p(X_0)\theta\|^{p'}$, where the convergence can be taken in the pointwise, uniform on any bounded set or epigraph sense. In order to obtain the weak convergence we have to prove the tightness of Z_N. But as the convergence is uniform on $B = \{\|\theta\| \leq 1\}$ we have for $N \geq N_0$, $\mathbb{F}(Z_N) \leq C$ on B where C is a constant. Therefore $c_{Z_N}(x) \geq \|x\| - C$ for $N \geq N_0$ and Z_N is asymptoticaly tight. $\square$

The central limit theorem may be generalized to the case of non convex cost densities.

We have the analogue of the large deviation theorem.

Theorem 10. *Given an i.i.c. sequence $\{X_n, n \in \mathbb{N}\}$ of tight cost density c, we have :*

$$1/n c_{(X_1+\cdots+X_n)/n} \stackrel{w}{\to} \hat{c} \ ,$$

where $\hat{c}$ denotes the convex hull of c.

Proof. Let us give only the ideas of the proof. The cost density of $X_1 + \cdots + X_n$ is $c^{\star n}$. Therefore we want $\lim_n 1/n c^{\star n}(x/n)$. But the epigraph of the convolution of two functions is equal to the sum of their epigraphs. Therefore the epigraph of $1/n c^{\star n}(x/n)$ is equal to the vectorial mean of n identical convex sets which are epigraphs of c. The limit (in the epigraph sense), when n goes to infinity, of the vectorial mean converges towards the convex hull of the epigraph of c [25]. See also [3]. $\square$

This last theorem is only a trivial case of law of large numbers for random sets studied in [25, 26]. The interpretation of this result as a min-plus large deviation theorem comes from by M. Akian.

3 Bellman Chains

We denote $\mathcal{M}_p$ the dioid of (p, p)-matrices with entries in the semiring $\mathcal{K}$. The matrix product in $\mathcal{M}_p$ is

$$[AB]_{ij} \stackrel{\text{def}}{=} [A \otimes B]_{ij} \stackrel{\text{def}}{=} \min_k [A_{ik} + B_{kj}] \ .$$

All the entries of the zero matrix of $\mathcal{M}_p$ are $+\infty$. The diagonal entries of the identity matrix of $\mathcal{M}_p$ are 0, the other entries being $+\infty$.

With a matrix M in $\mathcal{M}_n(\mathcal{K})$, we associate a *precedence graph* $\mathcal{G}(M) = (\mathcal{N}, \mathcal{P})$ with nodes $\mathcal{N} = \{1, 2, \cdots, n\}$, and arcs $\mathcal{P} = \{xy \mid x, y \in \mathcal{N},\ \ M_{xy} \neq \varepsilon\}$. The number M_{xy}, when it is nonzero, is called the weight of the arc xy.

A path π, of length l, with origin x and end y, is an ordered set of nodes $\pi = \pi_0\pi_1 \cdots \pi_l$ with $\pi_0 = x$ and $\pi_l = y$, and $\pi_i\pi_{i+1} \in \mathcal{P}$ for all $i = 0, \cdots, l-1$. The couple $\pi_i\pi_{i+1}$ are called the arcs of π and the π_i its nodes. The *length* of the path π is denoted $|\pi|$. The couple xy of the *ends* of π is denoted $\langle\pi\rangle$. When the two ends of π are equal one says that π is a circuit. The *weight* of π, denoted $\pi(M)$, is the $\otimes$-product of the weights of its arcs. For example we have $xyz(M) = M_{xy} \otimes M_{yz}$.

The set of all paths with ends xy and length l is denoted $\mathcal{P}^l_{xy}$. The paths of length 0 are the nodes $\mathcal{P}^0 = \mathcal{N}$. Then, $\mathcal{P}^*_{xy}$ is the set of all paths with ends xy and $\mathcal{P}^*$ the set of all paths. We have :

$$\mathcal{P}^* \stackrel{\text{def}}{=} \bigcup_{l=0}^{\infty} \mathcal{P}^l \ .$$

For $\rho \subset \mathcal{P}$, $\langle\rho\rangle$ is the set of the ends of the paths of ρ. Then denoting $\mathcal{P}_N$ the set of arcs of the graph associated to the matrix N we have the following trivial accessibility results :

Proposition 11. *For $M \in \mathcal{M}_n$ we have :*

$$\mathcal{P}_{M^k} = \langle\mathcal{P}^k\rangle, \quad \mathcal{P}_{M^*} = \langle\mathcal{P}^*\rangle.$$

For $\rho \subset \mathcal{P}^*$ one define :

$$\rho(M) \stackrel{\text{def}}{=} \bigoplus_{\pi\in\rho} \pi(M) \ ,$$

which is the infimum of the weights of all the paths belonging to ρ.

We denote

$$M^* \stackrel{\text{def}}{=} \bigoplus_{i=0}^{\infty} M^i \ ,$$

which exists if we accept entries in $\overline{\mathbb{R}}_{\min}$. Then, we have the following interpretation of the matrix product in $\mathcal{M}_n$.

Proposition 12. *For $M \in \mathcal{M}_n$ we have*

$$\mathcal{P}^l_{xy}(M) = (M^l)_{xy}, \quad \mathcal{P}^*_{xy}(M) = (M^*)_{xy} . \tag{7}$$

The matrix M^* has no entries equal to $-\infty$ iff there is no circuits with negative weight in $\mathcal{P}$.

The min-plus Markov chain is called a Bellman chain, it is defined by a transition cost matrix $M \in \mathcal{M}_n$ satisfying $Me = e$ where e denotes the column of e of size n. Then, given an initial cost, which is a line vector c^0 satisfying $c^0 e = e$, we can define a cost c, on the set of paths $\mathcal{P}$, by $c(\pi) = c^0_{\pi_0}\pi(M)$ for all $\pi \in \mathcal{P}^l$ and $l \in \mathbb{N}$. Then, the analogue of the forward Kolmogorov equation is the forward Bellman equation $c^n = c^{n-1} \otimes M$, c^0 given. It gives the marginal cost, for the Bellman chain $X^n(\pi) \stackrel{\text{def}}{=} \pi_n$, to be in state (node) $x \in \mathcal{N}$ at time n.

If a transition cost matrix satisfies

$$M_{xy} = M_{yx} > 0, \; M_{xx} \geq 0, \; \forall y \neq x \in \mathcal{N} ,$$

then the matrix M^*_{xy} defines a metric. Indeed, we have $M^*_{xx} = 0$ and $M^*_{xy} \leq M^*_{xz} + M^*_{zy}$ by definition of the matrix product and the fact that $M^*M^* = M^*$. A path from x to y in $\mathcal{G}(M)$ achieving the minimal cost among the paths of any length is called a *geodesic* joining x to y. We will still call a geodesic an optimal path when the matrix M is nonsymmetric.

3.1 Min-plus closed Jackson service networks

A closed Jackson network of queues is a set of n customers and m services. The customers wait for services in queues attached to each service. The customers are served in the order of arrival. The service is random and markovian. In discrete time situation, a (m, m) transition probability matrix r is given. The entry r_{ij} is the probability that a customer, served at queue i, goes to queue j, if the queue i is not empty. If the queue is empty, this probability is 0. Such a system is a Markov chain with state space :

$$S^m_n \stackrel{\text{def}}{=} \{x \in \mathbb{N}^m : 1.x = n\} ,$$

where 1 denotes the vectors with all its entries equal to 1 with size adapted to the context (here m). It is clear that, if r is irreducible, the Markov chain describing the system is irreducible. Therefore it has a unique invariant measure p. This measure is explicitly computable :

$$p_x = k\theta_1^{x_1} \cdots \theta_m^{x_m} ,$$

with θ any solution of $\theta r = \theta$ and k a normalizing constant such that $p1 = 1$.

The best way to understand what is a min-plus closed Jackson service network is to consider the following problem. We consider a company renting cars. It has n cars and m parkings in which customers can rent cars. The customers can rent a car in a parking and leave the rented car in another parking. After some time the distribution of the cars in the parkings is not satisfactory and the company has to transport the cars to achieve a better distribution. Given r the (m, m) matrix of transportation cost from a parking to another, the problem is to determine the minimal cost of the transportation from a distribution $x = (x_1, \cdots, x_m)$ of the cars in the parking to another one $y = (y_1, \cdots, y_m)$ and to compute the best plan of transportation. Therefore the precise transportation problem is the following.

MIN-PLUS CLOSED JACKSON PROBLEM (TRANSPORTATION PROBLEM). *Given the (m, m) transition cost matrix r irreducible such that $r_{ij} > 0$ if $i \neq j = 1, \cdots, m$ and $r_{ii} = 0$ for all $i = 1, \cdots, m$, compute M^* for the the Bellman chain on S_n^m of transition cost M defined by $M_{x,T_{ij}(x)} \stackrel{\text{def}}{=} r_{ij}$ and*

$$T_{ij}(x_1, \cdots, x_m) \stackrel{\text{def}}{=} (x_1, \cdots, x_i - 1, \cdots, x_j + 1, \cdots, x_m) ,$$

for $i, j = 1, \cdots, m$.

The operator T_{ij} corresponds to the transportation of a car from the parking i to the parking j. We denote $\mathcal{T} \stackrel{\text{def}}{=} \{T_{ij},\ i, j = 1, \cdots, m\}$.

If $r_{ii} = e$ for all $i = 1, \cdots, m$ (the absence of transportation costs nothing) the previous problem corresponds to the computation of the largest invariant cost c satisfying $c = cM$, and $c_x = e$. Indeed, in this case the left eigen semimodule has as many independent generators as states[7]. Remarking that the diagonal entries of M^* are e, it is clear $M^*_{x.}M = M^*_{x.}$. Then, from the fact that $q = bM^*$ is the largest solution of $q = qM \oplus b$, we can prove that the searched extremal left eigenvector is $M^*_{x.}$.

3.2 Solution of the m-parkings transportation problem

Let us consider the m-parkings case. In this case a path $\pi \in \mathcal{P}$ is

$$xT^1(x)T^2 \circ T^1(x) \cdots y = T^l \circ T^{l-1} \circ \cdots T^1(x) ,$$

with $T^i \in \mathcal{T}$. Since the arcs $\mathcal{P}_r$ of r are $xT_{ij}(x)$ with $x \in \mathcal{N}$ and $T_{ij} \in \mathcal{T}$ we can code ($\simeq$) a path $\pi \in \mathcal{P}^*$ in a simpler way by the couple $\pi \simeq x\mu$ with $x \in \mathcal{N}$ a

[7]Let us recall that in the min-plus context the irreducibility of the transition matrix assures the uniqueness of the eigenvalue but not the uniqueness of the generators of the eigen semimodule see [10] Section 3.7.

node of $\mathcal{G}(M)$ and $\mu \in \mathcal{P}_r^*$ a path of $\mathcal{G}(r)$. Clearly we have :

$$\pi(M) = \mu(r),\ \forall \pi \simeq x\mu \in \mathcal{P}^* \ .$$

Remarking that the vector $T_{ij}(x) - x$ is independent of x let us call it γ_{ij} and denote $\Gamma = \{\gamma_{ij},\ i, j = 1, \cdots, m\}$. These vectors are not mutually independent indeed we have the relations :

$$\gamma_{ik} = \gamma_{ij} + \gamma_{jk},\ \forall i, j, k = 1, \cdots, m \ .$$

For a path $\mu \in \mathcal{P}_r^*$, the evaluation $\mu(\Gamma) \in \mathbb{Z}^m$ is obtained by using the morphism which to the concatenation associates the vectorial sum and to the letters associate the corresponding vectors of Γ. For example for the path $ijkl \in \mathcal{P}_r^*$ we have :

$$ijkl(\Gamma) = \gamma_{ij} + \gamma_{jk} + \gamma_{kl} \ .$$

Then, the constraint on the paths $\pi : \langle\pi\rangle = xy$ with $x, y \in \mathbb{Z}^m$ is equivalent to the constraint $\mu(\Gamma) = y - x$ for the path $\pi \simeq x\mu$.

The cost of a path $\mu(r)$ depends only of the number of times each arc appears in μ and not of the order of the arcs. Similarly the constraint $\mu(\Gamma) = y - x$ does not depend of the order of the arcs in the path μ, since the evaluation $\mu(\Gamma)$ corresponds to additions of vectors, and addition of vector is commutative. To take account of this symmetry of the problem we denote $\mathcal{P}_r^c$ the set of equivalent classes of paths (where two paths are equivalent if the arcs appear the same number of times). Therefore, for $\mu \in \mathcal{P}_r^c$ we can take the representative $\mu = \prod_{a \in \mathcal{P}_r} a^{n_a}$. For example the path $\mu = ijijk$ belongs to the class of $(ij)^2(ji)(jk)$.

It is clear that $\mu(r^*) \leq \mu(r)$ because $r_{ij} \geq r_{ij}^*$. Moreover for each μ it exists $\tilde{\mu}$ such that $\mu(r^*) = \tilde{\mu}(r)$ and $\mu(\Gamma) = \tilde{\mu}(\Gamma)$. The path $\tilde{\mu}$ is obtained by substituting the arcs ij of μ by paths μ_{ij} such that $\mu_{ij}(r_{ij}) = r_{ij}^*$. Inside S_n^m this substitution is always possible. This is not always possible on the boundary of S_n^m because the path $x\mu$ may leave S_n^m. To avoid this difficulty we suppose that the costs on the boundary arcs are not r_{ij} but r_{ij}^*.

We can summarize the previous considerations in the following proposition.

Proposition 13. *The optimal value of the transportation problem is :*

$$M_{xy}^* = \mathcal{P}_{xy}^*(M) = \Phi_{r^*}(y - x) \ ,$$

with

$$\Phi_{r^*}(z) \stackrel{\text{def}}{=} \bigoplus_{\substack{\mu \in \mathcal{P}_{r^*}^c \\ \mu(\Gamma) = z}} \mu(r^*) \ .$$

The mathematical program $\Phi_{r^*}(z)$ is a flow problem.

Proposition 14. *Denoting by $\mathcal{J}$ the incidence matrix nodes-arcs of the complete graph with m nodes we have :*

$$\Phi_{r^*}(z) = \inf_{\substack{\phi \geq 0 \\ \mathcal{J}\phi = z}} \phi . r^* ,$$

where $\phi . r = \sum_{i,j} r_{ij} \phi_{ij}$.

Corollary 15. *We have for all y and x satisfying $x_j = 0$ for $j \neq i$ and $x_i = n$*

$$M^*_{xy} = \bigotimes_{j, j \neq i} (r^*_{ij})^{y_j} ,$$

and for all x and y satisfying $y_j = 0$ for $j \neq i$ and $y_i = n$

$$M^*_{xy} = \bigotimes_{j, j \neq i} (r^*_{ji})^{x_j} .$$

Proof. In these two cases the flow problems are trivial. The nonnul components are respectively $\phi_{ij} = y_j$ and $\phi_{ji} = x_i$, for $j \neq i$. $\square$

This corollary gives the searched min-plus product form.

3.3 example

Let us consider the transportation system with 3 parkings and 6 cars, and transportation costs :

$$r = \begin{pmatrix} 0 & 1 & +\infty \\ +\infty & 0 & 1 \\ 1 & +\infty & 0 \end{pmatrix} = \begin{pmatrix} e & 1 & \epsilon \\ \epsilon & e & 1 \\ 1 & \epsilon & e \end{pmatrix} .$$

We have :

$$r^* = \begin{pmatrix} e & 1 & 2 \\ 2 & e & 1 \\ 1 & 2 & e \end{pmatrix} .$$

Let us suppose that $x = (0, 0, 6)$ and $y = (2, 3, 1)$, we can apply the corollary, we have :

$$M^*_{xy} = (r^*_{31})^2 (r^*_{32})^3 = 2 \times 1 + 3 \times 2 = 8 .$$

The Geodesic is given in Fig. 1.

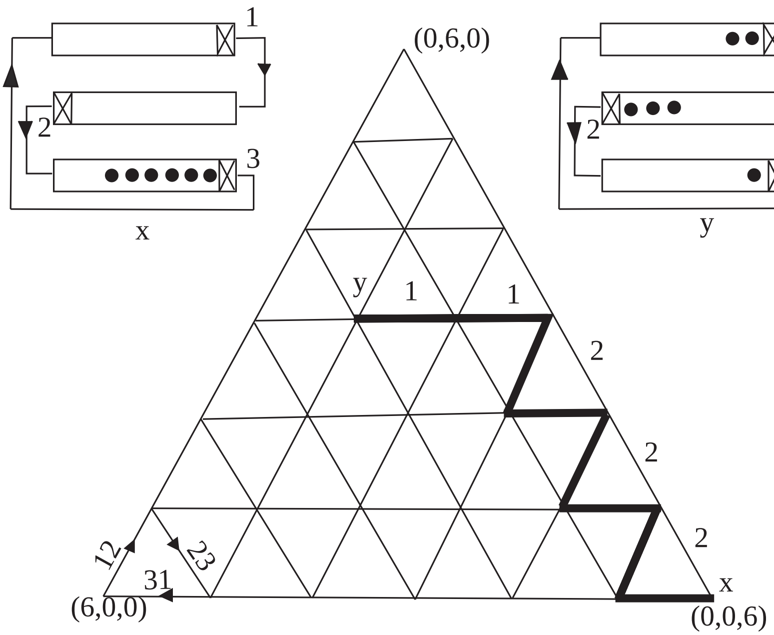

Figure 1: Transportation System (6 cars, 3 parkings)

4 Bellman Processes with independent Increments, Inf-Convolution and Cramer Transform

4.1 Bellman Processes

We can easily define continuous time decision processes which correspond to deterministic controlled processes. We discuss here only decision processes with continuous trajectories.

Definition 16. 1. A *continuous time Bellman process* X_t with continuous trajectories is a decision variable with values in $\mathcal{C}(\mathbb{R}^+)$ [8] having the cost density

$$c_X(x(\cdot)) \stackrel{\text{def}}{=} \phi(x(0)) + \int_0^{\infty} c(t, x(t), x'(t))dt \ ,$$

with $c(t, \cdot, \cdot)$ a family of transition costs (that is a function c from $\mathbb{R}^3$ to $\mathbb{R}_{\min}$ such that $\inf_y c(t, x, y) = 0, \ \forall t, x$) and ϕ a cost density on $\mathbb{R}$. When the integral is not defined the cost is by definition equal to $+\infty$.

[8] $\mathcal{C}(\mathbb{R}^+)$ denotes the set of continuous functions from $\mathbb{R}^+$ to $\mathbb{R}$.

2. The Bellman process is said *homogeneous* if c does not depend on time t.

3. The Bellman process is said *with independent increments* if c does not depend on state x. Moreover if this process is homogeneous, c is reduced to the cost density of a decision variable.

4. The *p-Brownian decision process*, denoted by B_t^p, is the process with independent increments and transition cost density $c(t, x, y) = \frac{1}{p}|y|^p$ for all x.

As in the discrete time case, the marginal cost to be in state x at time t can be computed recursively using a forward Bellman equation.

Theorem 17. *The marginal cost* $v(t, x) \stackrel{\text{def}}{=} \mathbb{K}(X_t = x)$ *is given by the Bellman equation:*

$$\partial_t v + \hat{c}(\partial_x v) = 0, \quad v(0, x) = \phi(x) \ , \tag{8}$$

where $\hat{c}$ *means here* $[\hat{c}(\partial_x v)](t, x) \stackrel{\text{def}}{=} \sup_y [y \partial_x v(t, x) - c(t, x, y)]$.

For the Brownian decision process B_t^p *starting from* 0*, the marginal cost to be at time* t *in state* x *satisfies the Bellman equation*

$$\partial_t v + (1/p')|\partial_x v|^{p'} = 0, \quad v(0, \cdot) = \chi \ .$$

Its solution can be computed explicitly, it is $v(t, x) = \mathcal{M}^p_{0, t^{1/p'}}(x)$ *therefore we have*

$$\mathbb{V}[f(B_t^p)] = \inf_x \left[f(x) + \frac{x^p}{pt^{\frac{p}{p'}}} \right] . \tag{9}$$

4.2 Cramer Transform

Definition 18. The Cramer transform $\mathcal{C}$ is a function from $\mathcal{M}$, the set of positive measures on $E = \mathbb{R}^n$, to $\mathcal{C}_{\mathrm{X}}$ defined by $\mathcal{C} \stackrel{\text{def}}{=} \mathcal{F} \circ \log \circ \mathcal{L}$, where $\mathcal{L}$ denotes the Laplace transform[9].

From the definition and the properties of the Laplace and Fenchel transforms the following result is clear.

Theorem 19. *For* $\mu, \nu \in \mathcal{M}$ *we have* $\mathcal{C}(\mu * \nu) = \mathcal{C}(\mu) \star \mathcal{C}(\nu)$.

[9] $\mu \mapsto \int_E e^{\langle \theta, x \rangle} \mu(dx)$.

Table 1: Properties of the Cramer transform.

$\mathcal{M}$	$\log(\mathcal{L}(\mathcal{M})) = \mathcal{F}(\mathcal{C}(\mathcal{M}))$	$\mathcal{C}(\mathcal{M})$
μ	$\hat{c}_\mu(\theta) = \log \int e^{\theta x} d\mu(x)$	$c_\mu(x) = \sup_\theta(\theta x - \hat{c}(\theta))$
0	$-\infty$	$+\infty$
δ_a	θa	χ_a
$\lambda e^{-\lambda x - H(x)}$	$H(\lambda - \theta) + \log(\lambda/(\lambda - \theta))$	$H(x) + \lambda x - 1 - \log(\lambda x)$
$p\delta_0 + (1-p)\delta_1$	$\log(p + (1-p)e^\theta)$	$x \log(\frac{x}{1-p})$ $+(1-x)\log(\frac{1-x}{p})$ $+H(x) + H(1-x)$
stable distrib.	$m\theta + \frac{1}{p'}\lvert\sigma\theta\rvert^{p'} + H(\theta)$ $1 < p' < 2$	$c(x) = \mathcal{M}^p_{m,\sigma}\ , x \geq m$ $c(x) = 0,\ x < m,$ $1/p + 1/p' = 1$
Gauss distrib.	$m\theta + \frac{1}{2}\lvert\sigma\theta\rvert^2$	$\mathcal{M}^2_{m,\sigma}$
$\mu * \nu$	$\hat{c}_\mu + \hat{c}_\nu$	$c_\mu \star c_\nu$
$k\mu$	$\log(k) + \hat{c}$	$c - \log(k)$
$\mu \geq 0$	$\hat{c}$ convex l.s.c.	c convex l.s.c.
$m_0 \overset{\text{def}}{=} \int \mu$	$\hat{c}(0) = \log(m_0)$	$\inf_x c(x) = -\log(m_0)$
$m_0 = 1$	$\hat{c}(0) = 0$	$\inf_x c(x) = 0$
$S_\mu \overset{\text{def}}{=} \overline{\text{cvx}(\text{supp}(\mu))}$ $m_0 = 1$	$\hat{c}$ strictly convex in $D_{\hat{c}}$ $\hat{c}$ is C^∞ in $\overset{\circ}{D}_{\hat{c}}$	$\overset{\circ}{D}_c = \overset{\circ}{S}_\mu$ c is C^1 in $\overset{\circ}{D}_c$
$m_0 = 1,\ m \overset{\text{def}}{=} \int x\mu$	$\hat{c}'(0) = m$	$c(m) = 0$
$m_0 = 1,\ m_2 \overset{\text{def}}{=} \int x^2\mu$	$\hat{c}''(0) = \sigma^2 \overset{\text{def}}{=} m_2 - m^2$	$c''(m) = 1/\sigma^2$
$m_0 = 1,\ 1 < p' < 2$ $\hat{c} = \lvert\sigma\theta\rvert^{p'}/p' + o(\lvert\theta\rvert^{p'})$ $+H(\theta)$	$\hat{c}^{(p')}(0^+) = \Gamma(p')\sigma^{p'}$	$c^{(p)}(0^+) = \Gamma(p)/\sigma^p$

The Cramer transform changes the convolutions into inf-convolutions and consequently independent random variables into independent decision variables. In Table 1 we summarize the main properties and examples concerning the Cramer transform when $E = \mathbb{R}$. The difficult results of this table can be found in Azencott [8]. In this table we have denoted

$$H(x) \overset{\text{def}}{=} \begin{cases} 0 & \text{for } x \geq 0, \\ +\infty & \text{elsewhere.} \end{cases}$$

Let us give an example of utilization of these results in the domain of partial differential equations (PDE). Processes with independent increments are transformed in decision processes with independent increments. This implies that a generator

$\hat{c}(-\partial_x)$ of a stochastic process is transformed in the generator of the corresponding decision process $v \mapsto -\hat{c}(\partial_x v)$.

Theorem 20. *The Cramer transform v of the solution r of the PDE on $E = \mathbb{R}$*

$$-\partial_t r + [\hat{c}(-\partial_x)](r) = 0,\ r(0,.) = \delta\ ,$$

(with $\hat{c} \in \mathcal{C}_X$) satisfies the HJB equation

$$\partial_t v + \hat{c}(\partial_x v) = 0,\ v(0,.) = \chi\ . \tag{10}$$

This last equation is the forward HJB equation of the control problem of dynamic $x' = u$, instantaneous cost $c(u)$ and initial cost χ.

Remark 21. First let us remark that $\hat{c}$ is convex l.s.c.and not necessarily polynomial which means that fractional derivatives may appear in the PDE .

Proof. The Laplace transform of r denoted q satisfies:

$$-\partial_t q(t,\theta) + \hat{c}(\theta)q(t,\theta) = 0,\ q(0,.) = 1\ .$$

Therefore $w = \log(q)$ satisfies:

$$-\partial_t w(t,\theta) + \hat{c}(\theta) = 0,\ w(0,.) = 0\ , \tag{11}$$

which can be easily integrated. As soon as $\hat{c}$ is l.s.c and convex w is l.s.c and convex and can be considered as the Fenchel transform of a function v. The function v satisfies a PDE which can be easily computed. Indeed we have:

$$w(t,\theta) = \sup_x(\theta x - v(t,x)) \Longrightarrow \begin{cases} \theta = \partial_x v\ , \\ \partial_t w = -\partial_t v\ . \end{cases}$$

Therefore v satisfies equation (10). This equation is the forward HJB equation of the control problem with dynamic $x' = u$, instantaneous cost $c(u)$ and initial cost χ because $\hat{c}$ is the Fenchel transform of c and the HJB equation of this control problem is

$$-\partial_t v + \min_u\{-u\partial_x v + c(u)\} = 0,\ v(0,.) = \chi\ .$$

□

If $\hat{c}$ is independent of time the optimal trajectories are straight lines and $v(x) = tc(x/t)$. This can be obtained by using (11).

Solution of linear PDE with constant coefficients can be computed explicitely by Fourier transform. The previous theorem shows that that non linear convex first order PDE with constant coefficients are isomorphic to linear PDE with constant coefficients and therefore can be computed explicitely. Such explicit solutions of HJB equation are known as Hopf formulas [16]. Let us develop the computations on a non trivial example.

Example 22. Let us consider the HJB equation

$$\partial_t v + \frac{1}{2}(\partial_x v)^2 + \frac{2}{3}(|\partial_x v|)^{\frac{3}{2}} = 0,\ v(0,.) = \chi \ .$$

From (11) we deduce that :

$$w(t,\theta) = t(\frac{1}{2}\theta^2 + \frac{2}{3}|\theta|^{\frac{3}{2}}) \ ,$$

therefore using the fact that the Fenchel transform of a sum is an inf-convolution we obtain:

$$v(t,x) = \frac{x^2}{2t} \star \frac{|x|^3}{3t^2} \ .$$

We can verify on this explicit formula a continuous time version of the central limit theorem. Using the scaling $x = yt^{2/3}$, we have

$$\lim_{t\to+\infty} v(t, yt^{2/3}) = y^3/3 \ ,$$

since the shape around zero of the corresponding instantaneous cost $c(u) = (u^2/2) \star (|u|^3/3)$ is $|u|^3/3$. Indeed a simple computation shows that $c(u)$ is obtained from

$$\begin{cases} c = 1/2y^4 + 1/3|y|^3 \ , \\ \quad u = |y|y + y \ , \end{cases}$$

by elimination of y. This system may be also considered as a parametrical definition of $c(u)$.

4.3 Min-Plus Perfect Gas

Let us show how the Cramer transform appears in statistical mechanics. Here a mechanical system means min-plus linear finite state system. More precisely, we consider the analogue of a system of independent particles (perfect gas) by building a large min-plus system composed of independent min-plus subsystems. Following standard methods of statistical mechanics, we compute the Gibbs distribution of the min-plus subsystems. In this computation appears naturally the Cramer transform.

The tensor product of two min-plus rectangular matrices A and B is the min-plus tensor of order 4 denoted $C = A \odot B$ with entries $C_{jj'ii'} = A_{ji} \otimes B_{j'i'} = A_{ji} + B_{j'i'}$. On the set of such tensors, we define the product $[C \otimes D]_{ii'kk'} = \bigoplus_{jj'} C_{ii'jj'} \otimes D_{jj'kk'}$.

Proposition 23. *Given a set of m min-plus matrices* $A_i \in \mathcal{M}_{n_i}$ *such that* $\mathcal{G}(A_i)$ *are irreducible, denoting* λ_i *their eigenvalues and* e_i *the identity matrix of dimension* n_i*, we have*

$$(\odot_i A_i)(\odot_i X_i) = (\otimes_i \lambda_i)(\odot_i X_i) \ , \tag{12}$$

$$\bigoplus_i \left[(\odot_{k=1}^{i-1} e_k) \odot A_i \odot (\odot_{k=i+1}^{m} e_k) \right] (\odot_i X_i) = (\oplus_i \lambda_i)(\odot_i X_i) \ ,$$

for all eigenvectors $(X_i)_{i=1,n}$ *of* $(A_i)_{i=1,n}$.

Let us consider a system composed of N independent subsystems (particles) of k different kinds defined by their min-plus matrices A_i, $i = 1, \cdots, k$, which are supposed to be irreducible with eigenvalues λ_i.

The repartition $(N_i, i = 1, \cdots, k)$ (with $\sum_i N_i = N$) of the N subsystems among the k possibilities defines the probability

$$p = (p_i \stackrel{\text{def}}{=} N_i/N, i = 1, \cdots, k) \ .$$

The number of possible ways to achieve a given distribution p is

$$M \stackrel{\text{def}}{=} N!/(N_1!N_2! \cdots N_k!) \ .$$

Using the Stirling formula, we have

$$S \stackrel{\text{def}}{=} (\log M)/N \sim -\sum_{i=1}^{k} p_i \log p_i, \text{ when } N \to +\infty \ .$$

This gives the asymptotics (with respect to N) of the probability to observe the empirical distribution p in a sample, of size N, drawn with the uniform law on $(1, \cdots, k)$.

Let us suppose that we observe the eigenvalue E of the complete system (the total energy of the complete system in the mechanical analogy). Thanks to (12), it is given by:

$$E = \bigotimes_{i=1}^{k} (\lambda_i)^{N_i} \ .$$

that is

$$\sum_i p_i \lambda_i = U \stackrel{\text{def}}{=} E/N \ . \tag{13}$$

Then, in a standard way, the *Gibbs distribution* is defined as the one maximizing S among all the distributions satisfying the constraint (13).

Theorem 24. *The Gibbs distribution is given by*

$$p_i(\theta) = \frac{e^{\theta\lambda_i}}{\sum_j e^{\theta\lambda_j}} \ , \tag{14}$$

where θ achieves the optimum in

$$c_\mu(U) = \max_\theta[\theta U - \log \mathbb{E}(e^{\theta\lambda})] \ .$$

where λ is a random variable with uniform law (μ) on $\{\lambda_i\}$.

Proof. The function $p \mapsto -S(p)$ is convex. Therefore we have to minimize a convex function subject to linear constraints. Let us introduce the Lagrangian

$$L(\theta, \mu, p) = \sum_i (p_i \log p_i) + \mu \left(1 - \sum_i p_i\right) + \theta \left(U - \sum_i p_i\lambda_i\right) \ .$$

The saddle point $(\theta, \mu, p)^*$ realizing $\max_\theta \max_\mu \min_p L(\theta, \mu, p)$ gives the Gibbs distribution. First solving $\max_\mu \min_p L(\theta, \mu, p)$ we obtain (14).

To compute θ as a function of U we have to maximize the Lagrangian with respect to θ, that is

$$\max_\theta \left[\theta U - \log\left(\sum_i e^{\theta\lambda_i}\right)\right] \ ,$$

which can be written as $\max_\theta[\theta U - \log \mathbb{E}(e^{\theta\lambda})] - \log k$, if λ is a random variable with uniform law on $(\lambda_i)_{i=1,\cdots,k}$. □

5 Notes and Comments.

Bellman [18] was aware of the interest of the Fenchel transform (which he calls max transform) for the analytic study of the dynamic programming equations. The bicontinuity of the Fenchel transform has been well studied in convex analysis [38, 7, 6].

Maslov has started the study of idempotent integration in [41]. He has been followed in particular by [39, 42, 32, 31, 4, 1, 2]. In [43] idempotent Sobolev spaces have been introduced as a way to study HJB equation as a linear object. In this paper the min-plus weak convergence has been also introduced but for compact support test functions. This weak convergence is used in [33] for the approximation of HJB equations. In [47] and [10] the law of large numbers and the central limit theorem for decision variables has been given in the particular case $p = 2$. In two independent works [32, 31] and [17] the study of decision variables has been started. The second work has been continued in [4]. A lot of results announced in [4] are proved in [1] and [2]. Many of the missing proofs of results given here can be found in [5]. The large deviation result is a known result in convex analysis [25, 26] and the bibliography of these papers. The min-plus product form result presented here comes from [20].

The Cramer transform is an important tool in large deviations literature [8, 37, 50, 34]. In [32, 10, 4] Cramer transform has been used in the min-plus context. In [34] the connection between large deviations and statistical mechanics is done. See also [48] for discussion of links existing between mechanics, thermodynamics statistical mechanics. In this paper min-plus ergodic theorems are also presented. Limit theorems of probabilities and decision theory can be put in a same framework see [3].

Some aspects of [51, 11] are strongly related to the morphism between probability and decision calculus in particular the morphism between LQG and LEG problems and the link with H_∞ problem.

References

[1] Akian, M.: Densities of idempotent measures and large deviations. INRIA Report **2534** (1995). 2.1, 5, 5

[2] Akian, M.: Theory of cost measures: convergence of decision variables. INRIA Report **2611** (1995). 6, 5, 5

[3] Akian, M. : From probabilities to cost measures via large deviation, Rap. INRIA, in preparation. 6, 5

[4] Akian, M. , Quadrat, J.P. and Viot, M. : Bellman Processes L.N. in control and Inf. Sciences N. 199, Ed. G. Cohen, J.P. Quadrat, Springer Verlag (1994). 5, 5, 5, 5

[5] Akian, M. , Quadrat, J.P. and Viot, M. : Duality between probability and optimization In J. Gunawardena (Editor), “Idempotency”, Cambridge University Press (1998). 1, 2.5, 5

[6] Attouch, H: Variational convergence for functions and operators. Pitman (1984). 5

[7] Attouch, H., Wets, R.J.B.: Isometries for the Legendre-Fenchel transform. Transactions of the American Mathematical Society **296**, No. 1 (1986) 33–60. 5

[8] Azencott, R., Guivarc’h, Y., Gundy, R.F.: Ecole d’été de Saint Flour 8. Lect. Notes in Math., Springer-Verlag, Berlin (1978). 4.2, 5

[9] Baskett, F., Chandy, F., Muntz, M. and Palacios, R. : “Open, closed and mixed networks of queues with different classes of customers”, JACM 22 (1975).

[10] Baccelli, F., Cohen, G., Olsder, G.J., Quadrat, J.P.: Synchronization and linearity: an algebra for discrete event systems. John Wiley and Sons, New York (1992). 1, 7, 5, 5

[11] Basar, T., Bernhard, P.: H_∞ optimal control and relaxed minimax design problems. Birkhaüser (1991). 5

[12] Bertsimas, D. and Tsitsiklis, J. N. : “Linear Optimization”, Athena Scientific, Belmont Mass. (1997).

[13] Courtois, P. J. : Decomposability Queuing and Computer System Applications, Academic Press, New York (1977).

[14] Van Dijk, N. M. : Queuing networks and Product forms : a system approach, Wiley (1993).

[15] Gondran, M. and Minoux, M. : Graphs and Algorithms, Wiley, (1984).

[16] Bardi, M., Evans, L.C. On Hopf’s formulas for solutions of Hamilton-Jacobi equations. Nonlinear Analysis, Theory, Methods & Applications, Vol.8., No. 11 (1984) 1373–1381. 4.2

[17] Bellalouna, F.: Un point de vue linéaire sur la programmation dynamique. Détection de ruptures dans le cadre des problèmes de fiabilité. Thesis dissertation, University of Paris-IX Dauphine (1992). 5

[18] Bellman, R., Karush, W.: Mathematical programming and the maximum transform. SIAM Journal of Applied Mathematics **10** (1962). 5

[19] Billingsley, P.: Convergence of probability measures. John Wiley & Sons, New York (1968).

[20] Fall, O. and Quadrat, J.P. : About min-plus product forms submitted to CARI’98, Dakar, Octobre 98. 1, 5

[21] Fayolle, G. and Lasgouttes, J.M. : ‘Asymptotics and scalings for Large Product-Form Networks via the Central Limit Theorem, Markov Processes and Related Fields, V.2, N.2, p.317–349 (1996).

[22] Gaubert, S. : Théorie des systèmes linéaires dans les dioides, Thesis dissertation, École des Mines de Paris, (1992). 1

[23] Gaubert, S. and Plus, M.: Methods and applications of (max,+) linear algebra, in R. Reischuk and M. Morvan, editors, *STACS’97*, number 1200 in LNCS, Lübeck, Springer, (1997). 1

[24] Gunawardena, J. (Editor) : Idempotency, Cambridge University Press (1998).

[25] Hess, C. : Quelques théorèmes limites pour des ensembles aléatoires bornés ou non, séminaire d’analyse convexe, Montpellier 1984, exposé n. 12. 6, 6, 5

[26] Hess, C. : On the distribution of random sets and the multivalued strong law of large numbers in the Wijsman topology, Cahier CEREMADE n. 9250, Paris Dauphine 1992. 6, 5

[27] Kelly, F. P. : Reversibility and Stochastic Networks, Wiley, (1979).

[28] Kemeny, J. G. and Snell,J. L. : Finite Markov Chains, Van Nostrand, Princeton, N.J. (1967).

[29] Kolokoltsov, V. N. and Maslov, V. P. : Idempotent Analysis and Its Applications, Kluwer (1997).

[30] Pap, E. : Null-Additive Set Functions, Mathematics and Its applications 337, Kluwer academic Publishers, Dordrecht (1995).

[31] Del Moral, P.: Résolution particulaire des problèmes d'estimation et d'optimisation non-linéaires. Thesis dissertation, Toulouse, France (1994). 2.1, 6, 5, 5

[32] Del Moral, P., Thuillet, T., Rigal, G., Salut, G.: Optimal versus random processes: the non-linear case. LAAS Report, Toulouse, France (1990). 5, 5, 5

[33] Dudnikov, P.I., Samborski, S.: Networks methods for endomorphisms of semi-modules over min-plus Algebras, in 11th Inter. Conf. on Analysis and Optimization of Systems. L.N. in Control and Inf. Sc. No.199, Springer Verlag (1994). 5

[34] Ellis, R. S.: Entropy, large deviations, and statistical mechanics. Springer Verlag, New York (1985). 5, 5

[35] Feller, W.: An introduction to probability theory and its applications, John Wiley and Sons, New York (1966).

[36] Fenchel, W.: On the conjugate convex functions. Canadian Journal of Mathematics **1** (1949) 73–77.

[37] Freidlin, M.I., Wentzell, A.D.: Random perturbations of dynamical systems. Springer-Verlag, Berlin (1984). 5

[38] Joly, J.L.: Une famile de toplogies sur l'ensemble des fonctions convexes pour lesquelles la polarité est bicontinue. J. Math. pures et appl. **52** (1973) 421–441. 5

[39] Kolokoltsov, V. N. and Maslov, V. P.: The general form of the endomorphisms in the space of continuous functions with values in a numerical commutative semiring (with the operation $\oplus$=max). Soviet Math. Dokl. Vol. 36, No. 1 (1988) 55–59. 2.1, 5

[40] Landau, L. and Lifchitz, E. : Physique Statistique, Editions MIR, Moscou (1967).

[41] Maslov, V.: Méthodes opératorielles. Éditions MIR, Moscou (1987). 2.1, 2.1, 5

[42] Maslov, V., Samborski, S.N.: Idempotent analysis. Advances In Soviet Mathematics **13**, Amer. Math. Soc., Providence (1992). 5

[43] Maslov, V., Samborski, S.N.: Stationary Hamilton-Jacobi and Bellman equations (existence and uniqueness of solutions), in Idempotent analysis. Advances In Soviet Mathematics **13**, Amer. Math. Soc., Providence (1992). 2.5, 5

[44] Petersen, K. : Ergodic Theory, Cambridge University Press (1983).

[45] Pulhaskii, A. : Large deviations of semimartingales via convergence of the predictable characteristics, Stochastic and stochastic reports, N. 49, p.27–85 (1994).

[46] Pulhaskii, A. : Large deviation analysis of the single server queue, Queuing Systems, No.21, p.5–66, (1995).

[47] Quadrat, J.P.: Théorèmes asymptotiques en programmation dynamique. Note CRAS **311** Paris (1990) 745–748. 5

[48] Quadrat, J.P. and Plus, M. : Min-Plus linearity and statistical mechanics, Markov Processes and Related Fields Feb. 98. 1, 1, 5

[49] Rockafellar, R.T.: Convex analysis. Princeton University Press Princeton, N.J. (1970).

[50] Varadhan, S.R.S.: Large deviations and applications. CBMS-NSF Regional Conference Series in Applied Mathematics No. 46, SIAM Philadelphia, Penn. (1984). 5

[51] Whittle, P.: Risk sensitive optimal control. John Wiley and Sons, New York (1990). 5

[52] P. Whittle : ‘’ Systems in Stochastics Equilibrium”, Wiley, New York (1986).