Max-Plus Algebra and Applications to System Theory and Optimal Control

Max-Plus Working Group[1] *presented by J.P. Quadrat*[2]

[1] currently consisting of M.Akian, G. Cohen, S. Gaubert, J.P. Quadrat & M. Viot
[2] INRIA-Rocquencourt, B.P. 105, 78153 Le Chesnay Cedex, France

Proceedings of the International Congress of Mathematicians, Zurich, Switzerland, August 1994, published by Birkhaüser, pp. 1502–1511, 1995

In the modeling of human activities, in contrast to natural phenomena, quite frequently only the operations max (or min) and $+$ are needed. A typical example is the performance evaluation of synchronized processes such as those encountered in manufacturing (dynamic systems made up of storage and queuing networks). Another typical example is the computation of a path of maximum weight in a graph and more generally of the optimal control of dynamical systems. We give examples of such situations. The max-plus algebra which is a mathematical framework well suited to handle such situations. We present results on $i)$ linear algebra, $ii)$ system theory, $iii)$ duality between probability and optimization based on this algebra.

1 Max-Plus Linear Algebra

Definition 1. 1. A *abelian monoid* $\mathcal{K}$ is a set endowed with one operation $\oplus$ which is associative, commutative and has a *zero element* ε.

2. A *semiring* is an abelian monoid endowed with a second operation $\otimes$ which is associative and distributive with respect to $\oplus$ which has an *identity element* denoted e, with ε *absorbing* (that is $\varepsilon \otimes a = a \otimes \varepsilon = \varepsilon$).
3. A *dioid* is a semiring which is *idempotent* (that is $a \oplus a = a$, $\forall a \in \mathcal{K}$).
4. A *semifield* is a semiring having its second operation invertible on $\mathcal{K}_\star = \mathcal{K} \setminus \{\varepsilon\}$.
5. A semifield which is also a dioid is called an *idempotent semifield*.
6. We will say that these structures are *commutative* when the product is also commutative.
7. We call $\mathbb{R}_{\max}$ [resp. $\mathbb{R}_{\min}$]the set $\mathbb{R} \cup \{-\infty\}$ [resp. $\mathbb{R} \cup \{+\infty\}$] endowed with the two operations $\oplus = \max$ [resp $\oplus = \min$] and $\otimes = +$.
8. We call $\mathbb{R}_{\max}^{n\times n}$ and analogously $\mathbb{R}_{\min}^{n\times n}$ the set of $n \times n$ matrices with entries belonging to $\mathbb{R}_{\max}$ endowed with $\oplus$ denoting the $\max$ entry by entry and $\otimes$ defined by
$$[AB]_{ij} \overset{\text{def}}{=} [A \otimes B]_{ij} \overset{\text{def}}{=} \max_k [A_{ik} + B_{kj}] = \oplus_k A_{ik} \otimes B_{kj} \; .$$
9. We call $\mathcal{S}_{\max}$ [resp. $\mathcal{I}_{\max}$] the set of functions [resp. increasing functions], from $\mathbb{R}$ into $\mathbb{R}_{\max}$ endowed with $\oplus$ denoting the pointwise maximum and $\otimes$ the *sup-convolution* defined by
$$[f \otimes g](x) \overset{\text{def}}{=} [f \,\square\, g](x) \overset{\text{def}}{=} \sup_t [f(x-t) + g(t)] \; .$$

Analogously we define $\mathcal{S}_{\min}$ [resp. $\mathcal{I}_{\min}$]. The set $\mathcal{I}^d_{\min}$ is the restriction of $\mathcal{I}_{\min}$ to piecewise constant increasing functions with jumps at positive integer abscissas.

10. We call $\mathcal{C}_x$ [resp. $\mathcal{C}_v$] the set of lower [resp. upper] semicontinuous and proper (never equal to $-\infty$ [resp. ∞]) convex [concave] functions endowed with the $\oplus$ operator denoting the pointwise maximum [minimum] and $\otimes$ operator denoting the pointwise sum.
11. We call $\mathcal{C}_0$ the set of lower semicontinuous and proper strictly convex functions having 0 as infimum endowed with the $\otimes$ operator denoting the inf-convolution of two functions.

Clearly the algebraic structure $\mathbb{R}_{\max}$ *and* $\mathbb{R}_{\min}$ *are idempotent commutative semifields*, $\mathbb{R}^{n\times n}_{\max}$, $\mathbb{R}^{n\times n}_{\min}$, $\mathcal{S}_{\max}$, $\mathcal{S}_{\min}$, $\mathcal{I}_{\max}$, $\mathcal{I}_{\min}$, $\mathcal{I}^d_{\min}$ $\mathcal{C}_x$ and $\mathcal{C}_v$ are dioids, $\mathcal{C}_0$ is a commutative monoid. We will call all these vectorial structures based on $\mathbb{R}_{\max}$ or $\mathbb{R}_{\min}$ *max-plus algebras*. Working with these structures show that idempotency is as useful as the existence of a symmetric element in the simplification of formulas and therefore that these structures are very effective to make algebraic computations.

Application 2. 1. These mathematical structures introduce a linear algebra point of view to dynamic programming problems.
Given C in $\mathbb{R}^{n\times n}_{\min}$ we call *precedence graph* $\mathcal{G}(C)$ the graph having $i)$ n nodes, $ii)$ oriented arcs (i,j) of *weight* C_{ji} if $C_{ji}\neq\epsilon$ in the matrix C.
The min-plus linear dynamical system

$$X^{m+1}=C\otimes X^m,\ X^0_j=e,\ \text{for } j=i,\ X^0_j=\epsilon \text{ elsewhere}, \tag{1}$$

is a dynamic programming equation. The number X^m_j is equal to the *least weight* of all paths from i to j (the *weight of a path* is the sum of the weights its arcs) of *length* m (composed of m arcs).
The *minimal average weight by arc* of paths having their lengths going to infinity is obtained by computing λ solution of the spectral problem

$$\lambda\otimes X=C\otimes X\ .$$

The computation of the minimal weight of paths from i to a region described by $d\in\mathbb{R}^n_{\min}$ ($d_j=e$ if j belongs to the region $d_j=\epsilon$ elsewhere) is equal to X_i solution of

$$X=C\otimes X\oplus d\ .$$

2. The evaluation of some systems where synchronization between tasks appears (as in event graphs a subset of Petri nets) can be modeled linearly in $\mathbb{R}_{\max}$ or dually in $\mathbb{R}_{\min}$ by

$$X^{m+1}=F\otimes X^m\oplus G\otimes U^m,\ Y^{m+1}=H\otimes X^{m+1}\ . \tag{2}$$

In $\mathbb{R}_{\max}$, the number X^m_i has the interpretation of the earliest date of the m-th occurrence of the event i (for example the starting time of a task on a machine in manufacturing) has happened. The max operator models the fact that tasks can be performed as soon as all the preconditions are fulfilled. The vector U models the timing of the input preconditions. The vector Y denotes the timing of the outputs of the system.

In $\mathbb{R}_{\min}$ the number X_i^m has the interpretation of the maximum number of events of kind i that can occur before the date m. We can pass from (F, G, H) over $\mathbb{R}_{\max}$ to the one over $\mathbb{R}_{\min}$ by interchanging the role of the delays and the coefficients. (see [7] for more details).

3. Clearly it exists infinite dimensional and/or continuous time versions of the equation (1). For $c, \psi \in \mathcal{C}_0$ the problem

$$v_x^m = \min_u \left[\sum_{i=m}^{N-1} c(u^i) + \psi(x^N) \mid x^m = x \right], \; x^{i+1} = x^i - u^i \;,$$

may be called dynamic programming with independent instantaneous costs (c depends only on u and not on x). Clearly v satisfies the linear recurrence in $\mathcal{C}_0$

$$v^m = c \,\square\, v^{m+1}, \;\; v^N = \psi$$

To solve some of these applications we have to solve max-plus linear equations in $\mathbb{R}_{\max}^{n \times n}$ or $\mathbb{R}_{\min}^{n \times n}$. The general one can be written $A \otimes X \oplus b = C \otimes X \oplus d$. In this section we use three points of view (contraction, residuation, combinatorial) to study this kind of equations.

1.1 Spectral equations, Contraction and Residuation

As in the conventional algebra all the linear iterations are not contractions. We can characterize the contractions using the max-plus spectral theory. To simplify the discussion we give a simplified result under restrictive hypotheses on the connexity of the associated incidence graph. The general result will be found for example in [7].

Theorem 3. *1. If the graph $\mathcal{G}(C)$ associated with the matrix C has only a strongly connected component there exists a unique λ solution of $\lambda \otimes X = C \otimes X$. It has the graph interpretation*

$$\lambda = \max_\zeta \frac{|\zeta|_w}{|\zeta|_l} \,,$$

where $|\zeta|_w$ denotes the weight of the circuit ζ and $|\zeta|_l$ its length.

2. We denote C_λ the matrix defined by $C_\lambda \stackrel{\text{def}}{=} \lambda^{-1} \otimes C$, $C^* \stackrel{\text{def}}{=} E \oplus C \oplus C^2 \oplus \cdots \oplus C^{n-1}$ where E denotes the identity matrix and $C^+ \stackrel{\text{def}}{=} CC^*$. A column i of $[C_\lambda]^+$ such that $[C_\lambda]^+_{ii} = e$ is an eigen vector. In C_λ^+ it exists at least such a column.

3. There exists c such that for m large enough we have

$$C^{m+c} = \lambda^c C^m .$$

If $\mathcal{G}(C)$ has more than one strongly connected component, C may have more than one eigenvalue. The largest one is called the *spectral radius* of the matrix C and is denoted by $\rho(C)$.

Theorem 4. *The equation $\mu X = CX \oplus d$ has a least solution $X = [C_\mu]^* d_\mu$ when $\rho(C) \leq \mu$. The solution is unique when $\rho(C) < \mu$.*

The equation $Ax = d$ has not always a solution but its greatest subsolution can be computed explicitly

$$x = A\backslash d \stackrel{\text{def}}{=} \max\{x \mid Ax \leq d\} = \min_j (d_j - a_{j\cdot}) \ .$$

This computation, well known in residuation theory, defines a new binary operator $\backslash$ which can be seen as the dual operator of $\otimes$. The $\backslash$ is distributive with respect to $\wedge$ (defined as the $\min$ operator in the $\mathbb{R}_{\max}^{n\times n}$ context). With this two operators dual linear equations may be written.

Corollary 5. *The equation $\mu \setminus X = (C \setminus X) \wedge d$, has a solution as soon as $\mu \geq \rho(C)$. The largest X solution of this equation is*

$$X = [C_\mu]^* \setminus \mu d = \mu d \wedge (C_\mu \setminus \mu d) \wedge (C_\mu \setminus C_\mu \setminus \mu d) \wedge \cdots \ .$$

Application 6. In the event graphs framework described before this kind of equations appears when we compute the the latest date at which an event must occur if we want respect due times coded in d (see [7] for more details).

1.2 Symmetrization of the Max-Plus Algebra

Because every idempotent group is reduced to the zero element it is not possible to symmetrize the max operation. Nevertheless we can adapt the idea of the construction of $\mathbb{Z}$ from $\mathbb{N}$ to build an extension of $\mathbb{R}_{\max}$ such that the general linear scalar equation has always a solution.

Let us consider the set of pairs $\mathbb{R}_{\max}^2$ endowed with the natural idempotent semiring structure

$$(x', x'') \oplus (y', y'') = (x' \oplus y', x'' \oplus y'') \ ,$$
$$(x', x'') \otimes (y', y'') = (x'y' \oplus x''y'', x'y'' \oplus x''y') \ ,$$

with $(\varepsilon, \varepsilon)$ as the zero element and (e, ε) as the identity element and $\ominus(x', x'') \stackrel{\text{def}}{=} (x'', x')$.

Definition 7. Let $x = (x', x'')$ and $y = (y', y'')$. We say that x balances y (which is denoted $x \nabla y$) if $x' \oplus y'' = x'' \oplus y'$.

It is fundamental to notice that ∇ is *not* transitive and thus is not a congruence. However, we can introduce the congruence $\mathcal{R}$ on $\mathbb{R}_{\max}^2$ closely related to the balance relation:

$$(x', x'')\mathcal{R}(y', y'') \Leftrightarrow \begin{cases} x' \oplus y'' = x'' \oplus y' & \text{if } x' \neq x'', y' \neq y'' \\ (x', x'') = (y', y'') & \text{otherwise.} \end{cases} ,$$

We denote $\mathbb{S} \stackrel{\text{def}}{=} \mathbb{R}_{\max}^2 / \mathcal{R}$.

We distinguish three kinds of equivalence classes:

$\{(t, x'') \mid x'' < t\}$,	called positive elements,	represented by t;
$\{(x', t) \mid x' < t\}$,	called negative elements,	represented by $\ominus t$;
$\{(t, t)\}$,	called balanced elements,	represented by $t^\bullet$.

The set of positive [resp. negative, resp. balanced] elements is denoted $\mathbb{S}^{\oplus}$ [resp. $\mathbb{S}^{\ominus}$, resp. $\mathbb{S}^{\bullet}$]. This yields the decomposition

$$\mathbb{S} = \mathbb{S}^{\oplus} \cup \mathbb{S}^{\ominus} \cup \mathbb{S}^{\bullet} \ .$$

We also denote $\mathbb{S}^{\vee} \overset{\text{def}}{=} \mathbb{S}^{\oplus} \cup \mathbb{S}^{\ominus}$ and $\mathbb{S}^{\vee}_{\star} = \mathbb{S}^{\vee} \setminus \{\varepsilon\}$.

If $x \,\nabla\, y$ and $x, y \in \mathbb{S}^{\vee}$, we have $x = y$. We call this result the *reduction of balances*.

We now consider a solution X, in $\mathbb{R}^n_{\max}$, of the equation $AX \oplus b = CX \oplus d$, then the definition of the balance relation implies that $(A \ominus C)X \oplus (b \ominus d) \,\nabla\, \varepsilon$. Conversely, assuming that X is a positive solution of $AX \oplus b \nabla CX \oplus d$, with $AX \oplus b$ and $CX \oplus d \in \mathbb{S}^{\oplus}$, using the reduction of balances we obtain that X is solution of $AX \oplus b = CX \oplus d$.

Theorem 8 (Cramer's rule). *Let $A \in \mathbb{S}^{n\times n}$, $b \in \mathbb{S}^n$, $|A|$ the determinant of the matrix A (defined by replacing $+$ by $\oplus$, $-$ by $\ominus$ and $\times$ by $\otimes$ in the conventional definition) and A_i the matrix obtained from A by replacing the i-th column by b, then if $|A| \in \mathbb{S}^{\vee}_{\star}$, and $|A_i| \in \mathbb{S}^{\vee}, \ \forall i = 1, \cdots, n$, then there exists a unique solution of $AX \,\nabla\, b$, belonging to $(\mathbb{S}^{\vee})^n$, which satisfies*

$$X_i = |A_i|/|A| \ .$$

2 Min-Plus Linear System Theory

System theory is concerned with the input (u)-output (y) relation of a dynamical system ($\mathcal{S}$) denoted $y = S(u)$ and by the improvement of this input-output relation (based on some engineering criterion) by altering the system through a feedback control law $u = F(y, v)$. Then the new input (v)-output (y) relation is defined implicitely by $y = S(F(y, v))$. Not surprisingly, system theory is well developed in the particular case of linear shift-invariant systems. Analogously, a min-plus version of this theory can also be developed. The typical application is the performance evaluation of systems which can be described in terms of event graphs.

2.1 Inf-convolution and Shift-Invariant Max-Plus Linear Systems

Definition 9.
1. A *signal* u is a mapping from $\mathbb{R}$ into $\mathbb{R}_{\min}$. The signals set, denoted $\mathcal{Y}$, is endowed with two operations, namely the pointwise minimum of signals denoted $\oplus$, and the addition of a constant to a signal denoted $\otimes$ which plays the role of the external product of a signal by a scalar.
2. A *system* is an operator $S : \mathcal{Y} \to \mathcal{Y}, u \mapsto y$. We call u (respectively y) the *input* (respectively *output*) of the system. We say that the system is *min-plus linear* when the corresponding operator is linear.
3. The set of linear systems is endowed with two internal and one external operations, namely
$i)$ *parallel composition* $S = S_1 \oplus S_2$ defined by pointwise minimum of output signals corresponding to the same input;
$ii)$ *series composition* $S = S_1 \otimes S_2$, or more briefly, $S_1 S_2$ defined by the composition of operators;
$iii)$ *amplification* $T = a \otimes S$, $a \in \mathbb{R}_{\min}$ defined by $T(k) = a \otimes S(k)$.

4. The improved input (v)-output (y) relation of a system S by a *linear feedback* $u = F(y) \oplus G(v)$ is obtained by solving the equation $y = S(F(y)) \oplus S(G(v))$ in y.
5. A linear system is called *shift-invariant* when it commutes with the shift operators on signals $(u(.) \mapsto u(.+k))$.

Theorem 10. *1. For a shift-invariant continuous*[3] *min-plus linear system S it exists $h : \mathbb{R} \mapsto \mathbb{R}_{\min}$ called* impulse response *such that*

$$y = h \otimes u \overset{\text{def}}{=} h \,\square\, u \;.$$

2. The set of impulse responses endowed with the pointwise minimum and the inf-convolution is the dioid $\mathcal{S}_{\min}$.

3. If f [resp. g] denotes the impulse response of the system SF [resp. SG], the impulse response h of a system S altered by the linear feedback $u = F(y) \oplus G(v)$ is solution of

$$h = f \otimes h \oplus g \;.$$

2.2 Fenchel Transform

The Fourier and Laplace transforms are important tools in automatic control and signal processing because the exponentials diagonalize all the convolution operators simultaneously and consequently the convolutions are converted into multiplications by the Fourier transform. Analogous tools exist in the framework of the min-plus algebra.

Definition 11. Let $c \in \mathcal{C}_{\mathrm{X}}$, its Fenchel transform is the function in $\mathcal{C}_{\mathrm{X}}$ defined by $\hat{c}(\theta) = [\mathcal{F}(c)](\theta) \overset{\text{def}}{=} \sup_x [\theta x - c(x)]$.

For example setting $l_a(x) = ax$ we have $[\mathcal{F}(l_a)](\theta) = \chi_a(\theta)$ with

$$\chi_a(\theta) = \begin{cases} +\infty & \text{for } \theta \neq a, \\ 0 & \text{for } \theta = a. \end{cases}$$

Theorem 12. *For $f, g \in \mathcal{C}_{\mathrm{X}}$ we have $i)$ $\mathcal{F}(f) \in \mathcal{C}_{\mathrm{X}}$, $ii)$ $\mathcal{F}$ is an involution that is $\mathcal{F}(\mathcal{F}(f)) = f$, $iii)$ $\mathcal{F}(f \,\square\, g) = \mathcal{F}(f) + \mathcal{F}(g)$, $iv)$ $\mathcal{F}(f+g) = \mathcal{F}(f) \,\square\, \mathcal{F}(g)$.*

Theorem 13. *The response to a conventional affine input (min-plus exponential) is a conventional affine output with the same slope. If $y = h \,\square\, u$ and $u = l_a$ we have*

$$y = l_a / [\mathcal{F}(h)](a).$$

Unfortunately, the class of min-plus linear combinations of affine functions is only the set of concave functions which is not sufficient to describe all the interesting inputs of min-plus linear systems.

2.3 Rational Systems

A general impulse responde is too complicated to be used in practise since it involves an infinite number of operations to be defined.

Definition 14. 1. An impulse response $h \in \mathcal{I}^d_{\min}$ is *rational* if it can be computed with a finite number of $\oplus$, $\otimes$ and $*$[4] operations, from the functions $a \otimes e$ ($a \in \mathbb{R}_{\min}$) and

[3] Linear also for infinite linear combinations.

[4] For an impulse response h we define the operator * by $h^* \overset{\text{def}}{=} e \oplus h \oplus h^2 \cdots$

$\chi_1 \otimes e$ where

$$e(t) \overset{\text{def}}{=} \begin{cases} e \text{ for } t \leq 0, \\ \epsilon \text{ for } t > 0. \end{cases}$$

2. It is called *realizable* if there exists (F, G, H) such that $h^m = FG^m H$. Then there exists X such that

$$X^{m+1} = F \otimes X^m \oplus G \otimes U^m, \quad Y^m = H \otimes X^m \ .$$

   The vector X is called the *state* of the realization.
3. The system is called *ultimatly periodic* if $h^{m+c} = c \times \lambda + h^m$, for m large enough.
4. The number λ is called *the ultimate slope* of h.

Theorem 15. *For SISO systems having an impulse response in $\mathcal{I}^d_{\min}$ the three notions of rationality, ultimate periodicity and realizability are equivalent.*

This theorem is a min-plus version of the Kleene Schutzenberger theorem. The realization of an impulse response with a vectorial state X of minimal dimension is an open problem in the discrete time case.

2.4 Feedback Stabilization

Feedback can be used to stabilize a system without slowing down its throughput (the ultimate slope of its impulse response).

Definition 16. 1. A realization of a rational system is *internally stable* if all the ultimate slopes of the impulse responses from any input to any state are the same.
2. A realization is *structurally controllable* if every state can be reached by a path from at least one input.
3. A realization is *structurally observable* if from every state there exists a path to at least one output.

Theorem 17. *Any structurally controllable and observable realization can be made internally stable by a dynamic output feedback without changing the ultimate slope of the impulse response of the system.*

3 Bellman Processes

The functions stable by inf-convolution are known. They are the dynamic programming counterpart of the stable distributions of the probability calculus. They are the following functions

$$\mathcal{M}^p_{m,\sigma}(x) = \frac{1}{p}(|x-m|/\sigma)^p, \text{ with } \mathcal{M}^p_{m,0}(x) = \chi_m(x), \quad p \geq 1, m \in \mathbb{R}, \sigma \in \mathbb{R}^+ \ .$$

We have $\mathcal{M}^p_{m,\sigma} \, \square \, \mathcal{M}^p_{\bar{m},\bar{\sigma}} = \mathcal{M}^p_{m+\bar{m},[\sigma^{p'}+\bar{\sigma}^{p'}]^{1/p'}}$ with $1/p + 1/p' = 1$.

3.1 Cramer Transform

The Cramer transform ($\mathcal{C} \stackrel{\text{def}}{=} \mathcal{F} \circ \log \circ \mathcal{L}$, where $\mathcal{L}$ denotes the Laplace transform) maps probability measures to convex functions and transform convolutions into inf-convolutions:

$$\mathcal{C}(f * g) = \mathcal{C}(f) \,\square\, \mathcal{C}(g).$$

Therefore it converts the problem of adding independent random variables into a dynamic programming problem with independent costs. In Table 1 we give some properties of the Cramer transform. For a systematic study of the Cramer transform see Azencott [4].

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	$\hat{c}(\theta) = \log \int e^{\theta x} d\mu(x)$	$c(x) = \sup_\theta(\theta x - \hat{c}(\theta))$
$\mu \geq 0$	$\hat{c}$ convex l.s.c.	c convex l.s.c.
$m_0 \stackrel{\text{def}}{=} \int d\mu = 1$	$\hat{c}(0) = 0$	$\inf_x c(x) = 0$
$m_0 = 1,\ m \stackrel{\text{def}}{=} \int x d\mu$	$\hat{c}'(0) = m$	$c(m) = 0$
$m_0 = 1,\ m_2 \stackrel{\text{def}}{=} \int x^2 d\mu$	$\hat{c}''(0) = \sigma^2 \stackrel{\text{def}}{=} m_2 - m^2$	$c''(m) = 1/\sigma^2$
$m_0 = 1$ $\hat{c} = \lvert\sigma\theta\rvert^{p'}/p' + o(\lvert\theta\rvert^{p'})$	$\hat{c}^{(p')}(0^+) = \Gamma(p')\sigma^{p'}$	$c^{(p)}(0^+) = \Gamma(p)/\sigma^p$
$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(x-m)^2/\sigma^2}$	$m\theta + \frac{1}{2}(\sigma\theta)^2$	$\mathcal{M}^2_{m,\sigma}$
stable distrib. Feller [9]	$m\theta + \frac{1}{p'}\lvert\sigma\theta\rvert^{p'}$	$\mathcal{M}^p_{m,\sigma}$ with $p > 1,\ 1/p + 1/p' = 1$

3.2 Decision Space, Decision Variables

These remarks suggest the existence of a formalism anologous to probability calculus adapted to optimization. We start by defining cost measures which can be viewed as the normalized idempotent measures of Maslov [12].

Definition 18. 1. We call *decision space* the triplet $(U, \mathcal{U}, \mathbb{K})$ where U is a topological space, $\mathcal{U}$ is the set of the open subsets of U and $\mathbb{K}$ a map from $\mathcal{U}$ into $\overline{\mathbb{R}}^+$[5] such that $i)\ \mathbb{K}(U) = 0$, $ii)\ \mathbb{K}(\emptyset) = +\infty$, $iii)\ \mathbb{K}(\bigcup_n A_n) = \inf_n \mathbb{K}(A_n)$ for any $A_n \in \mathcal{U}$.

2. The map $\mathbb{K}$ is called a *cost measure*.
3. A map $c : u \in U \mapsto c(u) \in \overline{\mathbb{R}}^+$ such that $\mathbb{K}(A) = \inf_{u \in A} c(u),\ \forall A \in \mathcal{U}$ is called a *cost density* of the cost measure $\mathbb{K}$.

[5] $\overline{\mathbb{R}}^+ \stackrel{\text{def}}{=} \mathbb{R}^+ \cup \{+\infty\}$

4. The *conditional cost excess* to take the best decision in A knowing that it must be taken in B is
$$\mathbb{K}(A|B) \stackrel{\text{def}}{=} \mathbb{K}(A \cap B) - \mathbb{K}(B) \ .$$

Theorem 19. *Given a l.s.c. positive real valued function* c *such that* $\inf_u c(u) = 0$*, the expression* $\mathbb{K}(A) = \inf_{u \in A} c(u)$ *for all* $A \in \mathcal{U}$ *defines a cost measure. Conversely any cost measure defined on the open subsets of a Polish space admits a unique minimal extension* $\mathbb{K}_*$ *to* $\mathcal{P}(U)$ *(the set of the parts of* U*) having a density* c[6] *which is a l.s.c. function on* U *satisfying* $\inf_u c(u) = 0$*.*

This precise result is proved in Akian [1].

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

Definition 20. 1. A *decision variable* X on $(U, \mathcal{U}, \mathbb{K})$ is a mapping from U into E a topological space. It induces $\mathbb{K}_X$ a cost measure on $(E, \mathcal{B})$ ($\mathcal{B}$ denotes the set of open sets of E) defined by $\mathbb{K}_X(A) = \mathbb{K}_*(X^{-1}(A)), \ \forall A \in \mathcal{B}$. The cost measure $\mathbb{K}_X$ has a l.s.c. density denoted c_X.

2. When $E = \mathbb{R}$ [resp. $\mathbb{R}^n$, resp. $\mathbb{R}_{\min}$] with the topology induced by the absolute value [resp. the euclidian distance, resp. $d(x, y) = |e^{-x} - e^{-y}|$] then X is called a *real [resp. vectorial, resp. cost]* decision variable.
3. Two decision variables X and Y are said *independent* when
$$c_{X,Y}(x, y) = c_X(x) + c_Y(y).$$
4. The *optimum* of a real decision variable is defined by $\mathbb{O}(X) \stackrel{\text{def}}{=} \arg\min_x c_X(x)$ when the minimum exists. When a decision variable X satisfies $\mathbb{O}(X) = 0$, we say that it is *centered*.
5. When the optimum of a real decision variable X is unique and when near the optimum, we have
$$c_X(x) = \frac{1}{p} \left| \frac{x - \mathbb{O}(X)}{\sigma} \right|^p + o(|x - \mathbb{O}(X)|^p) \ ,$$
we define the *sensitivity of order* p of $\mathbb{K}$ by $\sigma^p(X) \stackrel{\text{def}}{=} \sigma$. When a decision variable satisfies $\sigma^p(X) = 1$, we say that it is *of order p and normalized*.
6. 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 set of decision variables having a unique optimum such that $\|X\|_p$ is finite. The corresponding set of decision variables is called $\mathbb{D}^p$. The space $\mathbb{D}^p$ is a conventional vector space and $\mathbb{O}$ is a linear operator on $\mathbb{D}^p$.
7. The *characteristic function* of a real decision variable is $\mathbb{F}(X) \stackrel{\text{def}}{=} \mathcal{F}(c_X)$ (clearly $\mathbb{F}$ characterizes only decision variables with cost in $\mathcal{C}_{\mathrm{X}}$).

---
[6] We extend the previous definition to a general subset of U.

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

Theorem 21. *If the cost density of a decision variable is convex, admits a unique minimum and is of order* p*, we have*[7]*:*

$$\mathbb{F}(X)'(0) = \mathbb{O}(X), \;\; [\mathbb{F}(X - \mathbb{O}(X))]^{(p')}(0) = \Gamma(p')[\sigma^p(X)]^{p'}, \; \textit{with } 1/p + 1/p' = 1 \; .$$

Theorem 22. *For two independent decision variables* X *and* Y *of order* p *and* $k \in \mathbb{R}$ *we have*

$$c_{X+Y} = c_X \,\square\, c_Y, \;\; \mathbb{F}(X+Y) = \mathbb{F}(X) + \mathbb{F}(Y), \;\; [\mathbb{F}(kX)](\theta) = [\mathbb{F}(X)](k\theta) \; ,$$

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

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

3.3 Limit Theorems for Decision Variables

We now study the behavior of normalized sums of real decision variables. They correspond to asymptotic theorems (when the number of steps goes to infinity) for dynamic programming. We have first to define convergence of sequences of decision variables. We have defined counterparts of each of the four classical kinds of convergence used in probability in previous papers (see [3]). Let us recall the definition of the two most important ones.

Definition 23. For the decision variable sequence $\{X^m, m \in \mathbb{N}\}$ we say that

1. X^m*weakly converges* towards X, denoted $X^m \stackrel{w}{\to} X$, if for all f in $\mathcal{C}_b(E)$ (where $\mathcal{C}_b(E)$ denotes the set of uniformly continuous and lower bounded functions on E into $\mathbb{R}_{\min}$), $\lim_m \mathbb{M}[f(X^m)] = \mathbb{M}[f(X)]$, with $\mathbb{M}(f(X)) \stackrel{\text{def}}{=} \inf_x (f(x) + c_X(x))$.
2. $X^m \in \mathbb{D}^p$ *converges in p-sensitivity* towards $X \in \mathbb{D}^p$, denoted $X^m \stackrel{\mathbb{D}^p}{\longrightarrow} X$, if $\lim_m \|X^m - X\|_p = 0$.

Theorem 24. *Convergence in sensitivity implies convergence and the converse is false.*

The proof is given in Akian [2].

We have the analogue of the law of large numbers and the central limit theorem.

Theorem 25 (large numbers and central limit). *Given a sequence* $\{X^m, \; m \in \mathbb{N}\}$ *of independent identically costed (i.i.c.) real decision variables belonging to* $\mathbb{D}^p, \; p \geq 1$*, we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{m=0}^{N-1} X^m = \mathbb{O}(X^0) \; ,$$

where the limit is taken in the sense of p-sensitivity convergence.

[7] Γ denotes the classical Gamma function.

Moreover if $\{X^m, m \in \mathbb{N}\}$ *is centered and of order p we have*

$$^8\text{weak}^* \lim_N \frac{1}{N^{1/p'}} \sum_{m=0}^{N-1} X^m = X,\ \textit{with}\ 1/p + 1/p' = 1\ ,$$

where X *is a decision variable with cost equal to* $\mathcal{M}^p_{0,\sigma^p(X^0)}$.

The analogues of Markov chains, continuous time Markov processes, Brownian and diffusion processes have also been given in [3].

References

1. Akian, M.: Idempotent integration and cost measures. INRIA Report (1994), to appear.
2. Akian, M.: Theory of cost measures: convergence of decision variables. INRIA Report (1994), to appear.
3. Akian, M., Quadrat, J-P, Viot, M.,: Bellman Processes, Proceedings of the 11th International Conference on Analysis and Optimization of Systems, LNCIS, Springer-Verlag, (June 1994).
4. Azencott, R., Guivarc'h, Y., Gundy, R.F.: École d'été de Saint Flour 8. Lect. Notes in Math., Springer-Verlag, Berlin (1978).
5. Bellalouna, F.: Processus de décision min-markovien. Thesis dissertation, University of Paris-Dauphine (1992).
6. Bellman, R., Karush, W.: Mathematical programming and the maximum transform. SIAM Journal of Applied Mathematics **10** (1962).
7. 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).
8. Del Moral, P.: Résolution particulaire des problèmes d'estimation et d'optimisation nonlinéaires, thèse Toulouse, France (Juin 1994).
9. Feller, W.: An Introduction to Probability Theory and its Applications. John Wiley and Sons, New York (1966).
10. Freidlin, M.I., Wentzell, A.D.: Random Perturbations of Dynamical Systems. Springer-Verlag, Berlin (1979).
11. Gaubert, S.:, Théorie des systèmes linéaires dans les dioides, thèse École des Mines de Paris, (Juillet 1992).
12. Maslov, V.: Méthodes Opératorielles. Éditions MIR, Moscou (1987).
13. Maslov, V., Samborski, S.N.: Idempotent Analysis. Advances In Soviet Mathematics **13** Amer. Math. Soc., Providence, (1992).
14. Quadrat, J.P.: Théorèmes asymptotiques en programmation dynamique. Note CRAS Paris **311** (1990) 745–748.
15. Rockafellar, R.T.: Convex Analysis. Princeton University Press Princeton, N.J. (1970).
16. Varadhan, S.R.S.: Large deviations and applications. CBMS-NSF Regional Conference Series in Applied Mathematics, N. 46, SIAM Philadelphia, Penn., (1984).
17. Whittle, P.: Risk Sensitive Optimal Control. John Wiley and Sons, New York (1990).

[8] The weak* convergence corresponds to the restriction of test functions to the conventional linear ones in the definition of the weak convergence.