Appears in *Markov Processes and Related Fields*, Vol. 3, N. 4, pp 565–587 (1997).

MIN-PLUS LINEARITY AND STATISTICAL MECHANICS

J.P. QUADRAT AND MAX-PLUS WORKING GROUP

ABSTRACT. We revisit some results obtained recently in min-plus algebra following the ideas of statistical mechanics. Computation of geodesics in a graph can be done by min-plus matrix products. A min-plus matrix is seen as a kind of finite states mechanical system. The energy of this system is the eigenvalue of its min-plus matrix. The graph interpretation of the eigenvalue may be seen as a kind of Mariotte law. The Cramer transform is introduced by statistics on populations of independent min-plus linear systems seen as a kind of perfect gas. It transforms probability calculus in what we call decision calculus. Then, dynamic programming equations, which are min-plus linear recurrences, may be seen as min-plus Kolmogorov equations for Markov chains. An ergodic theorem for Bellman chains, analogue of Markov chains, is given. The min-plus counterparts of aggregation, coherency, and reversibility of Markov chains are then studied. They provide new decomposition results to compute solutions of dynamic programming equations. Finally, some links between Wentzell-Freidlin asymptotics and min-plus algebra are described.

1. INTRODUCTION

Min-plus algebra, which is the set of real numbers endowed with the min and the plus operations, has been studied for a long time mainly in operations research. Within this mathematical structure, dynamic programming or Hamilton Jacobi equations become linear equations (for example see [37, 36]).

This algebra has been used to describe, linearly, systems in which synchronization is the main driving mechanism. Applications may be found in production systems, transportation and parallel computations [11]. For example, to achieve a task, in a production system, a machine and a part are needed. A task can start only at the supremum of the availability times of the machine and the part.

Min-plus algebra appears also in asymptotic computations. Indeed

$$\epsilon^n + \epsilon^m \simeq \epsilon^{\min(n,m)} ,$$

when ϵ is small. Large deviations to the law of large numbers [46, 24, 21], where such kind of asymptotics are used, suggests a duality between probability calculus and optimization theory. In some recent studies this duality has been formalized [43, 22, 23, 13, 6, 7, 1, 2]. Moreover, large deviations are related to statistical mechanics (for example [24]).

Key words and phrases. Max-Plus Algebra, Statistical Mechanics, Decision Theory, Large Deviation, Aggregation, Reversibility.

The current members of the Max-Plus working group are M. Akian, G. Cohen, S. Gaubert, M. Mc Gettrick, J.P. Quadrat and M. Viot. INRIA Domaine de Voluceau, Rocquencourt, 78153 Le Chesnay cedex (France).

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

In this paper we revisit some results on min-plus linear systems following the most elementary ideas used in statistical mechanics. We first recall the min-plus terminology (Section 2.1) and a Perron Frobenius like theorem (Section 2.2). Then we show that a min-plus system can be seen as a mechanical system and that the min-plus eigenvalue corresponds to the energy. The graph characterization of this eigenvalue is seen as a kind of Mariotte law, or, more precisely, as the adiabatic invariant of a mechanical system (Section 2.3). Then, a collection of independent min-plus systems with finitely many possible dynamics can be seen as a "perfect gas" (Section 3.1) composed of different kind of "molecules". Since the dynamic of the complete system is the tensor min-plus product of the individual subsystems, its eigenvalue is the sum of the individual eigenvalues. Then the Gibbs distribution can be introduced as the most likely distribution of the population of min-plus linear subsystems compatible with the observed eigenvalue of the complete system. In a standard way, the Cramér transform appears in the computation of the Gibbs distribution. The properties of the Cramér transform (Section 3.2) show clearly the duality existing between probability calculus and optimization.

The min-plus analogue of probability calculus, called decision theory, is recalled (Section 4.1). An ergodic theorem for the analogue of Markov chains, called Bellman chains, is given (Section 4.2). In Section 5, aggregation, coherency and reversibility of Bellman chains are introduced by analogy with Markov chains. When some of these properties are true, it is possible to decompose the computation of the eigenvector of the min-plus system when it is unique (that is, to decompose the computation of corresponding value function). This, perhaps new result, illustrates the interest of this duality.

Finally, in section 6, we briefly indicate how min-plus algebra naturally arises in large deviation asymptotics. For instance, we show how the Perron eigen-elements of matrices of the form $(\exp -(A_{ij}/\epsilon))$ converge in the large deviation sense, when ϵ decreases to 0, towards the corresponding min-plus eigen-elements. Such results, detailed in [5], suggest that deep connections exist between min-plus spectral theory and Wentzell-Freidlin perturbation theory.

2. MIN-PLUS LINEARITY GEODESICS AND THERMODYNAMICS

2.1. MIN-PLUS STRUCTURES AND PATHS OF MINIMAL WEIGHT IN A GRAPH

A *semiring* $\mathcal{K}$ is a set endowed with two operations denoted $\oplus$ and $\otimes$ where $\oplus$ is associative, commutative with zero element denoted ε, $\otimes$ is associative, admits a unit element denoted e, and distributes over $\oplus$; zero is absorbing ($\varepsilon \otimes a = a \otimes \varepsilon = \varepsilon$ for all $a \in \mathcal{K}$). This semiring is *commutative* when $\otimes$ is commutative. A module on a semiring is called a *semimodule*. A *dioid* $\mathcal{K}$ is a semiring which is idempotent ($a \oplus a = a,\ \forall a \in \mathcal{K}$). A [commutative, resp. idempotent] *semifield* is a [commutative, resp. idempotent] semiring whose nonzero elements are invertible.

The set $\mathbb{R} \cup \{+\infty\}$ endowed with the two operations $\oplus = \min$, $\otimes = +$, is denoted $\mathbb{R}_{\min}$. This structure is traditionally called min-plus algebra. It is an idempotent semifield with $\varepsilon = +\infty$ and $e = 0$. The structure $\mathbb{R}_{\min}$, completed with $-\infty$, with the convention $+\infty - \infty = +\infty$, is a dioid denoted $\overline{\mathbb{R}}_{\min}$.

We denote $\mathcal{M}_{np}(\mathcal{K})$ the semimodule of (n, p)-matrices with entries in the semiring $\mathcal{K}$. When $n = p$, $\mathcal{K} = \mathbb{R}_{\min}$, we write $\mathcal{M}_n$. It is a dioid and the matrix product

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

With a matrix C in $\mathcal{M}_n(\mathcal{K})$, we associate a *precedence graph* $\mathcal{G}(C)$ with nodes $\mathcal{N}(C) = \{1, 2, \cdots, n\}$, and arcs $\mathcal{A}(C) = \{(x, y) \mid C_{yx} \neq \varepsilon\}$. The *weight of the arc* (x, y) is C_{yx}.

A path p of length l, with origin x and end y, is an ordered set of nodes $p = (u_0 = x, \cdots, u_l = y)$. The paths of length 0 can be identified with the nodes. The *weight of path* p, denoted $w(p)$, is the $\otimes$-product of the weights of its arcs. A path with the same origin and end is called a *circuit*. The set of all paths of length l [resp. arbitrary length] with origin x and end y is denoted $\mathcal{P}^l_{xy}(C)$ [resp. $\mathcal{P}_{xy}(C)$]. The set of all paths [resp. circuits] is denoted $\mathcal{P}(C)$ [resp. $\mathcal{C}(C)$]. We have the following interpretation of the matrix product.

PROPOSITION 2.1. *For $C \in \mathcal{M}_n$ we have*

$$\inf_{p \in \mathcal{P}^l_{xy}(C)} w(p) = C^l_{yx} \ .$$

The matrix $C^* \stackrel{\text{def}}{=} \bigoplus_{i=0}^{\infty} C^i$ exists if we accept entries in $\overline{\mathbb{R}}_{\min}$. The entry C^*_{yx} is the infimum of the weights of the paths of arbitrary length connecting x to y.

PROPOSITION 2.2. *For all $C \in \mathcal{M}_n$ we have*

$$\inf_{p \in \mathcal{P}_{xy}(C)} w(p) = C^*_{yx} \ . \tag{2.1}$$

Moreover if

$$C_{xy} = C_{yx} > 0, \ C_{xx} \geq 0, \ \forall y \neq x \in \mathcal{N}(C) \ ,$$

C^*_{yx} is a distance.

Proof. Equation (2.1) follows from the interpretation of the matrix product. It is easy to check that $C^*_{xx} = 0$ and $C^*_{yx} \leq C^*_{yz} + C^*_{zx}$. □

A path achieving the optimum in (2.1) is a *geodesic* joining x to y in $\mathcal{G}(C)$.

2.2. EIGENVALUES AND TURNPIKE

An *eigenvalue* λ and an *eigenvector* X are solution of

$$\lambda X = CX, \ X \neq \varepsilon \ .$$

As soon as C is *irreducible*[1] (see [28, 29] for the general reducible case) there exists a unique eigenvalue. The eigenvalue has the following graph interpretation.

THEOREM 2.1. *For $C \in \mathcal{M}_n$, irreducible, one has that*

$$\lambda = \min_{c \in \mathcal{C}(C)} \frac{w(c)}{l(c)} \ . \tag{2.2}$$

Proof. See [11, Th.3.23]. □

[1] $\forall x, y \in \mathcal{N}(C), \ \mathcal{P}_{xy}(C) \neq \emptyset$.

Circuits achieving the optimum in (2.2) are called *critical circuits*. The subgraph which is the union of the nodes and arcs of the critical circuits is called *critical graph* and denoted $\mathcal{G}_c$. It may have several maximal strongly connected subgraphs (m.s.c.s.) $\overline{z}_1, \cdots, \overline{z}_g$ called *critical classes*.

There may exist several eigenvectors associated with one eigenvalue. Let us choose in each critical class $\overline{z_i}$ a node denoted z_i (called the representative) and denote $\mathcal{Z} \stackrel{\text{def}}{=} \{z_1, \cdots, z_g\}$. The eigen-semimodule of an irreducible matrix C is generated by the eigenvectors $\{X^z \stackrel{\text{def}}{=} [C_\lambda]^*_{.z},\ z \in \mathcal{Z}\}$ where $C_\lambda \stackrel{\text{def}}{=} \lambda^{-1}C$ (see [11, Th.3.2]). These eigenvectors satisfy $X^z_z = e$.

THEOREM 2.2. *The vectors X^z, for $z \in \mathcal{Z}$, form a minimal generating family of the right eigen-semimodule of the irreducible matrix C.*

Similarly the vectors $Y^z \stackrel{\text{def}}{=} [C_\lambda]^*_{z.}$, for $z \in \mathcal{Z}\}$, form a generating family of the left eigen-semimodule.

PROPOSITION 2.3. *If $C \in \mathcal{M}_n$ is such that all its eigenvalues λ are nonnegative, then $C^* = \bigoplus_{i=0}^{n-1} C^i$.*

Proof. Any path of length larger than n contains a circuit with a nonnegative weight therefore $C^n \geq \bigoplus_{i=0}^{n-1} C^i$. □

If the eigenvalue of an irreducible matrix C is negative, C^k goes to $-\infty$ when k goes to $+\infty$ and C^* is identically equal to $-\infty$. We have the following precise asymptotics.

THEOREM 2.3 (TURNPIKE). *For $C \in \mathcal{M}_n$ irreducible*

$$\exists k_0 \geq 0, \rho > 0 \ :\ \forall k \geq 0,\ q = \rho(k + k_0),\ \ C^q = \lambda^q \left(\bigoplus_{z \in \mathcal{Z}} X^z Y^z \right) , \qquad (2.3)$$

where $\mathcal{Z}$ denotes the set of the representatives of the critical classes, X^z and Y^z are respectively the generating families of the right and left eigen-semimodule of C^ρ.

Proof. This follows from Th.3.104, 3.109 and 3.112 of [11]. □

When the critical graph has only one critical class, (2.3) becomes, in standard notation,

$$C^k_{yx} = X^z_y + k\lambda + Y^z_x \ .$$

This result means that, for k large enough, the optimal path of length k joining x to y can be decomposed in three optimal paths. The first path connects x to an arbitrary node z of the critical graph. The second is a circuit in the critical graph starting and ending at z. The third connects z to y.

This asymptotic result on the min-plus linear recurrences $X_{k+1} = CX_k$ can be extended to the more general recurrences

$$X_k = \bigoplus_{i=0}^{m-1} C_i X_{k-i} \ .$$

Using the delay operator δ ($(\delta X)_k \stackrel{\text{def}}{=} X_{k-1}$), this recurrence can be written

$$X = C(\delta)X \ , \qquad (2.4)$$

with

$$C(\delta) = \bigoplus_{i=0}^{m-1} \delta^i C_i \ .$$

These recurrences are sometimes used to describe the dynamics of timed event graphs (a special class of timed Petri nets such that each place has only one arc upstream and one arc downstream, see [11, ch.2]). In this case, the vector X_k has the interpretation of the numbers of transition firings up to date k.

We can associate a precedence graph $\mathcal{G}(C(\delta))$ with the matrix $C(\delta)$. The weights of its arcs are now min-plus polynomials in δ. Let us suppose that they are monomials[2] in δ. Then, the weight of a path is also a min-plus monomial $w(p) = c \otimes \delta^t$ and we define $w_c(p) \stackrel{\text{def}}{=} c$ and $w_e(p) \stackrel{\text{def}}{=} t$. We still call eigenvalue λ and eigenvector X a pair satisfying

$$X = C(\lambda^{-1})X \ .$$

We have the following graph interpretation of the eigenvalue.

PROPOSITION 2.4. *An irreducible matrix $C(\delta)$, with monomial entries, whose circuits $\gamma \in \mathcal{G}(C(\delta))$ satisfy $w_e(\gamma) > 0$, admits the unique eigenvalue :*

$$\lambda = \inf_{\gamma \in \mathcal{C}(C(\delta))} \frac{w_c(\gamma)}{w_e(\gamma)} \ . \tag{2.5}$$

Proof. See [11, Th.3.28]. □

In the case of a strongly connected event graph, the eigenvalue is the number of firings by unit of time of an arbitrary transition. Equation (2.5) says that the "throughput" is equal to the infimum, among all the circuits, of the number of tokens in the circuit divided by the total amount of time that the tokens have to spend in the places of the circuit (see [11, Sect.3.2.5]).

2.3. MECHANICAL ANALOGY

Let us make an attempt to connect the objects discussed previously with quantities appearing, classically, in mechanics. Let us consider the one dimensional harmonic oscillator with Lagrangian $L(\dot{x}, x) = (\dot{x}^2 - x^2)/2$. Its Hamiltonian, defined by $H(p, x) = \sup_{\dot{x}}(p\dot{x} - L(\dot{x}, x))$, is $H(p, x) = (p^2 + x^2)/2$. We denote by $v(t, y)$ the extremum of the action

$$\mathcal{A}(t, x()) = \int_0^t L(\dot{x}(s), x(s))ds + \phi(x(0)) \ ,$$

among the continuous piecewise derivable trajectories satisfying $x(t) = y$, for a given initial cost ϕ. It is solution of the Hamilton-Jacobi Bellman (HJB) equation :

$$\frac{\partial v}{\partial t} + H\left(\frac{\partial v}{\partial x}, x\right) = 0, \quad v(0, x) = \phi(x) \ .$$

For t small enough, v is indeed the infimum of the action. Then we have

$$(R_t\phi)(z) \stackrel{\text{def}}{=} v(t, z) = \bigoplus_y r_t(z, y) \otimes \phi(y) \ ,$$

[2] In fact this assumption subsumes no loss of generality if we accept to change the realization of the dynamical system.

where

$$r_t(z,y) = \bigoplus_{x(),\, x(0)=y}^{x(t)=z} \mathcal{A}(t,x()) \ .$$

Therefore, R_t is a min-plus linear operator. We have $r_t(y,0) = s(t)y^2/2$, where s is solution of the Riccati equation

$$\dot{s} = -(1+s^2), \quad s(0) = +\infty \ .$$

Then, $s(t) = \text{cotan } t$ for $0 \leq t < \pi$. For $t \geq \pi$ and $y \neq 0$, $r_t(y,0) = -\infty$. The solution of the HJB equation gives an extremum of the action but not an infimum anymore. Nevertheless the effective trajectories follow the characteristic curves of the HJB equation. The dynamics describing the extremal trajectories are given by the Hamiltonian system

$$\dot{x} = \frac{\partial H(p,x)}{\partial p} = p \ ,$$
$$\dot{p} = -\frac{\partial H(p,x)}{\partial x} = -x \ .$$

The trajectories in the phase space (the space of pairs $(x,\ p)$, that is $\mathbb{R}^2$) are circles centered in 0 with radius equal to $\sqrt{2E}$. The extremal trajectories are $x(t) = \sqrt{2E}\sin(t+\alpha)$ and $p(t) = \sqrt{2E}\cos(t+\alpha)$, where E is the energy of the system. This energy E can be seen as the opposite of an eigenvalue of the HJB equation. Indeed, if we search for a solution of the form $v(t,x) = -Et + w_E(x)$ to the HJB equation, we have to solve

$$E = H\left(\frac{\partial w_E}{\partial x}, x\right) \ .$$

Two independent real eigenvectors exist, w_E and $-w_E$, with

$$w_E(x) = E\arccos\left(x/\sqrt{2E}\right) - \left(\sqrt{2E - x^2}\right)x/2 \ ,$$

which is defined only for $-\sqrt{2E} \leq x \leq \sqrt{2E}$.

The action computed along an extremal circuit of energy E, in the phase space, is 0. But $A(E) \stackrel{\text{def}}{=} \int_0^T p(t)dx(t)$, where the integral is computed along the extremal curve of energy E, and $T = 2\pi$ (the time to cover a circuit in the phase space), is equal to $2\pi E$ (the surface of the circle of radius $\sqrt{2E}$). The integrand $p(t)dx(t)/dt$ is twice the kinetic energy and the integral $A(E)$ has the unit of an action. Therefore, we have $E = A/T$ which is analogous to the graph interpretation of the eigenvalue of an irreducible min-plus matrix (the unit of A corresponds to the unit of the entries of the min-plus matrices). Fore more general situations we have $dA(E)/dE = T$ (see [9, Sect. 50]).

Consider a more general harmonic oscillator of Lagrangian

$$L(\dot{x},x) = (m(t)\dot{x}^2 - k(t)x^2)/2 \ ,$$

where $m(t)$ and $k(t)$ may vary with time, but, very slowly with respect to the speed of the oscillator motion (for fixed m and k). In the phase space, the trajectories look like ellipses varying slowly with the time. But, $A(E(t))$ stays constant in first approximation with respect to the coefficient measuring the slowness of the variation of m and k. It is called adiabatic invariant (see [9, ch.10, sect. E]). This adiabatic invariant can be seen as a Mariotte law for one particle. This is clearer on the example of a particle with mass m, speed v, in a one dimensional

box of length l with perfectly elastic walls. In this case, the motion in the phase space is a rectangle and the adiabatic invariant is $2mvl$ which is equal to twice the kinetic energy $1/2mv^2$ (which stays constant along the motion including the impacts) multiplied by $T = 2l/v$ (the time spent to cover the circuit in the phase space). Therefore, we have $2E = A/T = l(2mv/T)$ where $2mv/T$ has the unit of a force (corresponding to the pressure in the one dimensional case) exerted on the wall. We note that the pressure times the volume is equal to a constant times the kinetic energy of the particle, that is, its temperature.

In the case of event graphs, this adiabatic invariant appears when the transition timings change while the number of tokens stays constant. The Mariotte law is the graph interpretation of the eigenvalue during the variation. If the critical circuit stays constant, we have $N = \lambda T$ (with N the number of tokens of the critical circuit, T the time spent in the critical circuit and λ the throughput of the event graph). A thermodynamic theory may be developed based on this equality. For the time being, the interest of this kind of thermodynamic theory is not clear.

3. Statistical mechanics and duality between probability and optimization

If we think in terms of statistical mechanics, the previous section was concerned with one particle. In this section, 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 give the Gibbs distribution of the min-plus subsystems. This leads to introduce naturally the Cramér transform which plays an important role in the duality between probability calculus an optimization.

3.1. Min-plus perfect gas

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 3.1. *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*

$$(\odot_i A_i)(\odot_i X_i) = (\otimes_i \lambda_i)(\odot_i X_i) \ , \tag{3.1}$$

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

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 (3.1), it is given by:

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

that is

$$\sum_i p_i \lambda_i = U \stackrel{\text{def}}{=} E/N \ . \tag{3.2}$$

Then, in a standard way, the *Gibbs distribution* is defined as the one maximizing S among all the distributions satisfying the constraint (3.2).

THEOREM 3.1. *The Gibbs distribution is given by*

$$p_i(\theta) = \frac{e^{\theta\lambda_i}}{\sum_j e^{\theta\lambda_j}} \ , \tag{3.3}$$

where θ achieves the optimum in

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

where λ is a random variable taking the value λ_i with probability $1/k$.

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 (3.3).

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 $(\lambda_i)_{i=1,\cdots,k}$. □

3.2. CRAMÉR TRANSFORM

The *Cramér transform* $\mathcal{C}_r$ associates the convex function

$$c_\mu : U \mapsto \sup_\theta [\theta U - \log \mathbb{E}_\mu(e^{\theta\lambda})]$$

with the probability law μ of a random variable λ. It has appeared naturally (in the special case where μ is the uniform law) in computing the parameter θ of the Gibbs distribution. Let us recall its well known, important, properties.

We remark that the Cramér transform can be written $\mathcal{C}_r \stackrel{\text{def}}{=} \mathcal{F} \circ \log \circ \mathcal{L}$, where $\mathcal{L}$ is the Laplace transform and $\mathcal{F}$ the Fenchel transform defined by

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

Using the properties of the Laplace and Fenchel transforms, we have

$$\mathcal{C}_r(\mu * \nu) = \mathcal{C}_r(\mu) \star \mathcal{C}_r(\nu) \ ,$$

where $*$ denotes the convolution operator and $\star$ the inf-convolution operator defined by

$$[f \star g](y) = \inf_x [f(x) + g(y - x)] \ ,$$

for f and g two functions from $\mathbb{R}$ into $\mathbb{R}_{\min}$.

Let μ be the probability law of a random variable X with mean m and variance v. From the involution property of the Fenchel transform on l.s.c. (lower semi continuous) proper convex functions, we have $\mathcal{F}(c_\mu) = \log \circ \mathcal{L}(\mu)$, from which it is easy to deduce that

$$c_\mu(m) = \min_x c_\mu(x), \quad v = 1/c''_\mu(m) \ .$$

Moreover, if we denote

$$\mathcal{M}^p_{m,\sigma}(x) \stackrel{\text{def}}{=} \frac{1}{p}(|x - m|/\sigma)^p, \ \ p \geq 1 \ ,$$

a simple calculation shows that

$$\mathcal{M}^p_{m,\sigma} \star \mathcal{M}^p_{\bar{m},\bar{\sigma}} = \mathcal{M}^p_{m+\bar{m},\hat{\sigma}} \ ,$$

with

$$\hat{\sigma} = [\sigma^{p'} + \bar{\sigma}^{p'}]^{1/p'}, \quad 1/p + 1/p' = 1 \ .$$

These properties suggest the existence of a calculus similar to the probability calculus, in the min-plus context.

4. ERGODIC THEOREMS FOR BELLMAN CHAINS

From the previous remarks on the Cramér transform and the analogy between Markov transition probabilities matrices and min-plus transition cost matrices, it is clear that a duality exists between probability calculus and optimization. A min-plus probability theory has been formalized and developed in [13, 23, 22, 6, 2, 7, 30]. It uses the theory idempotent measures and integrals of Maslov [37] and is based on probabilities with values in min-plus algebra, called cost measures. We recall here basic definitions and results. Then, we give an ergodic theorem for finite state Bellman chains which are the min-plus analogue of Markov chains.

4.1. DECISION THEORY

DEFINITION 4.1. Let U be a topological space and $\mathcal{G}$ the set of its open sets. A finite min-plus idempotent measure on $(U, \mathcal{G})$ is an application $\mathbb{K}$ from $\mathcal{G}$ to $\mathbb{R}_{\min}$ such that

1. $\mathbb{K}(\emptyset) = \varepsilon$
2. $\mathbb{K}(\bigcup_n G_n) = \inf_n \mathbb{K}(G_n)$ for any $G_n \in \mathcal{G}$.

It is a min-plus probability or *cost measure* if in addition $\mathbb{K}(U) = e$.

Let c be a bounded function from U to $\mathbb{R}_{\min}$ (that is, lower bounded, since ε is the maximal element of $\mathbb{R}_{\min}$). Then, $\mathbb{K}(G) = \inf_{u \in G} c(u)$ is a min-plus idempotent measure. If $\mathbb{K}$ has this form, c is called a *density* of $\mathbb{K}$. Any cost measure $\mathbb{K}$ on $(U, \mathcal{G})$ admits a minimal extension $\mathbb{K}_*$ to the power set $\mathcal{P}(U)$ of U :

$$\mathbb{K}_*(A) = \sup_{G \supset A, G \in \mathcal{G}} \mathbb{K}(G).$$

If U is a separable metrizable space, $\mathbb{K}$ has necessarily a density. Its minimal density is equal to $c^*(x) = \mathbb{K}_*(\{x\})$ and is lower semicontinuous (l.s.c.) (see [1] or [33] for a weaker result, see also the related results on capacities in [35]).

In the sequel, χ_A denotes the min-plus characteristic function of the set A : $\chi_A(x) = e$ if $x \in A$ and $\chi_A(x) = \varepsilon$ otherwise. Given any cost measure $\mathbb{K}$ on $(U, \mathcal{G})$, the Maslov integral with respect to $\mathbb{K}$ is the unique $\mathbb{R}_{\min}$-linear form $\mathbb{V}$ on the set of lower bounded upper semicontinuous (u.s.c.) functions $f : U \to \mathbb{R}_{\min}$ such that $\mathbb{V}(f_n)$ decreases, and converges towards $\mathbb{V}(f)$ when f_n decreases and converges towards f and $\mathbb{V}(\chi_A) = \mathbb{K}(A)$ for $A \in \mathcal{U}$ (see [37, 1]). The integral $\mathbb{V}(f)$ is called the *value* of f : it is one analogue of the expectation. When confusion may occur, we denote it $\mathbb{V}_{\mathbb{K}}(f)$ or simply $\mathbb{K}(f)$. If the cost measure $\mathbb{K}$ has a density and c^* is its minimal density, $\mathbb{V}(f) = \inf_{u \in U}(f(u) + c^*(u))$. Therefore, the min-plus equivalent of the Dirac measure in point x is the cost measure with density χ_x.

Using this formalism, weak convergence and tightness of cost measures is defined as usual.

DEFINITION 4.2. We say that $\mathbb{K}_n$ weakly converges towards $\mathbb{K}$, ($\mathbb{K}_n \xrightarrow{w} \mathbb{K}$), if $\mathbb{K}_n(f) \to_n \mathbb{K}(f)$ for any bounded continuous[3] function $f : U \to \mathbb{R}_{\min}$.

DEFINITION 4.3. A set $\mathbb{K}$ of cost measures is tight iff

$$\sup_Q \inf_{\mathbb{K} \in \mathcal{K}} \mathbb{K}(Q^c) = \varepsilon = +\infty ,$$

where Q are compact sets.

Equivalent definitions of weak convergence, together with compactness results using tightness may be find in [35, 42, 41, 7]. These results are similar to that of Billingsley [14] on the weak convergence of probabilities. Weak convergence of cost measures is also related to the epiconvergence of their densities [7] (see [10] for definitions and results on epiconvergence).

Since the minimal extension of a cost measure is a cost measure on the set of all subsets of U, the minimal extension of its integral exists and is equal to the integral with respect to $\mathbb{K}_*$: it is defined, linear and continuous on all functions f. We

[3] endowed with the topology defined by the order relation (i.e. by $\lim_n x_n = x$ iff $\limsup_n x_n = \liminf_n x_n = x$).

denote it also by $\mathbb{K}$ or $\mathbb{V}$. We will only consider minimal extensions and densities, and omit the star.

These results allow us to define all the notions of probability theory; sometimes with a change of name. The analogue of conditional probability is called *conditional cost excess* : $\mathbb{K}(A|B) = \mathbb{K}(A \cap B) - \mathbb{K}(B)$, for any sets A, B such that $\mathbb{K}(B) \neq \varepsilon$. A *decision variable* (d.v.) with values in a topological space E is any application X from U to E. Its cost measure $\mathbb{K}_X$ is the minimal extension of its restriction to the topology of E defined by $\mathbb{K}_X(V) = \mathbb{K}(X^{-1}(V))$. Its cost density is the minimal density c_X of $\mathbb{K}_X$ (when it exists). It is the l.s.c. envelope of the function $\tilde{c}_X(x) = \inf\{c(u), u \in U \text{ and } X(u) = x\}$. Independence of d.v. is defined using open sets, *conditional cost excess* of a d.v. with respect to another is defined using minimal densities by $c_{X|Y}(x, y) = c_{X,Y}(x, y) - c_Y(y)$; clearly, when X and Y take a finite number of values, $c_{X|Y}(x, y) = \mathbb{K}(X = x|Y = y)$. The conditional value may be defined using the conditional cost. Weak convergence of decision variables corresponds to that of their cost measures.

A negligible set is such that its cost is equal to ε, that is to $+\infty$. Then, a sequence of decision variables X_n *converges almost surely* towards X iff $X_n(u) \to X(u)$ for all u with finite cost $c(u) < +\infty$. Contrary to classical probability theory, this convergence is implied by the *convergence in cost* (the analogue of the convergence in probability), which implies (resp. is equivalent to) the weak convergence when the limit is tight (resp. a constant) [22, 1].

In addition to classical notions of probability, we define the *optimum* $\mathbb{O}(X)$ of a d.v. X : $\mathbb{O}(X) = \{x \in E,\ c_X(x) = 0\}$. It is another (different from the value $\mathbb{V}$) analogue of the expectation. Indeed, for a d.v. X which is the image $\mathcal{C}_{\mathrm{r}}(X')$ by the Cramer transform of a random variable X' (in the sense that the cost density of X is the image of the law of X'), the optimum of X is equal to the expectation of X' (see Section 3.2). If f is continuous and X is tight (that is if $\mathbb{K}_X$ is tight), $\mathbb{O}(f(X)) = f(\mathbb{O}(X))$ ($\mathbb{O}(X)$ is compact). Since the optimum of a d.v. only depends on its cost measure, we can define the *conditional optimum* $\mathbb{O}(X|Y) : y \mapsto \{x \in E,\ c_{X|Y}(x, y) = 0\}$.

4.2. Ergodic theorems for Bellman chains

The analogue of a Markov chain is called a *Bellman chain*. Let X_n be a Bellman chain with values in a finite state space E, initial cost density ψ and conditional cost excess $\mathbb{K}(X_{n+1} = y|X_n = x) = C_{yx}$. Since $E^{\mathbb{N}}$ endowed with the product topology is a separable and metrizable topological space, we see that the decision variable $X = (X_0, X_1, \dots) \in E^{\mathbb{N}}$ has a cost density $c_X(x) = \sum_{n=0}^{\infty} C_{x_{n+1},x_n} + \psi(x_0)$, where $x = (x_0, x_1 \dots)$ (the sum may be equal to $+\infty$ which is the zero of $\mathbb{R}_{\min}$). The *initial cost* of a chain starting at $x \in E$ is $\psi = \chi_x$.

We study here the ergodic mean of a function of a Bellman chain X_n, using the spectral min-plus theory recalled in Section 2.2. Proofs and generalization will be given in [3]. Results about return time to a state will be given in [45].

For a circuit $c = (x_0, \dots, x_l = x_0) \in \mathcal{C}(C)$ and a function $f : E \to F$ with values in a finite dimensional normed vector space F, we denote

$$c(f) = \frac{f(x_1) + \cdots + f(x_l)}{l} .$$

For a subgraph $\mathcal{G}$ of $\mathcal{G}(C)$, we denote

$$\mathcal{G}(f) = \mathrm{conv}\{c(f), c \in \mathcal{C}(\mathcal{G})\} ,$$

where $\mathrm{conv}(A)$ is the convex hull of $A \subset F$.

THEOREM 4.1. *Let X_n be a Bellman chain with values in a finite state space E, starting at $x \in E$ and with conditional cost C. If C irreducible, with a unique critical class $\mathcal{G}_c$, then*

$$\frac{f(X_1) + \cdots + f(X_n)}{n} \stackrel{\mathrm{w}}{\to} Z, \text{ when } n \to +\infty ,$$

where Z is a d.v. with cost density $\chi_{\mathcal{G}_c(f)}$ (this is the uniform cost on $\mathcal{G}_c(f)$), independently of x.

In order to compare Theorem 4.1 with the ergodic theorem for Markov chains, we need to relate the limit $\mathcal{G}_c(f)$ with some expectation of f with respect to the invariant cost measure of the Bellman chain. The unique invariant cost density γ, which satisfies $C\gamma = \gamma$, has the nodes of $\mathcal{G}_c$ as optimum. Indeed, $\gamma_x = C^*_{zx}$ for any $z \in \mathcal{G}_c$ and $\gamma_x > 0$ when $x \notin \mathcal{G}_c$. If Y is a decision variable with $c_Y = \gamma$ then $\mathbb{O}(f(Y)) = f(\mathcal{G}_c)$ and $\mathcal{G}_c(f) \subset \mathbb{O}(f(Y))$.

COROLLARY 4.1. *Let γ be the unique invariant cost density of the Bellman chain of Theorem 4.1. If Y, a d.v. of density γ, is such that $\mathbb{O}(f(Y))$ is reduced to one point then*

$$\frac{f(X_1) + \cdots + f(X_n)}{n} \to \mathbb{O}(f(Y)), \text{ when } n \to +\infty ,$$

where the convergence holds weakly, in cost and almost surely.

A sequence X_n of independent d.v. with same cost measure ψ is the particular case of Bellman chain when $C_{yx} = \psi_y$. The invariant cost measure is ψ, $\mathbb{O}(f(Y)) = \mathbb{O}(f(X_1))$ and $\mathcal{G}_c(f) = \mathrm{conv}(\mathbb{O}(f(X_1)))$. This leads to the following law of large numbers which generalizes the results of [43, 7, 22], where the optimum was supposed to be unique.

COROLLARY 4.2. *Let X_n be independent d.v. taking a finite number of values in F, and let Y be a d.v. with uniform cost on* $\mathrm{conv}(\mathbb{O}(X_1))$, *then*

$$\frac{X_1 + \cdots + X_n}{n} \stackrel{\mathrm{w}}{\to} Y, \text{ when } n \to +\infty .$$

Another case where the limit is "unique" is the following.

COROLLARY 4.3. *If the critical graph of the Bellman chain X_n of Theorem 4.1 is reduced to one circuit c, we have*

$$\frac{f(X_1) + \cdots + f(X_n)}{n} \to c(f), \text{ when } n \to +\infty ,$$

where the convergence holds weakly, in cost and almost surely.

In this case, the set $\mathcal{G}_c(f)$ is reduced to one point $c(f)$. Any optimal trajectory of the Bellman chain starting at x, is deterministic after some finite time, and the ergodic theorem is reduced to the classical ergodic theorem for the "deterministic" application $x_i \mapsto x_{i+1}$ in the critical class $c = (x_0, \ldots, x_l)$. The (classical) invariant measure of this application is here the uniform measure on c.

However, when the critical graph has more than one circuit, an optimal trajectory has to choose between several directions at each intersection of circuits. If we assign a probability law to choose, at random, between these directions, the trajectory becomes a Markov chain, and the ergodic theorem says that the limit is the mean of f with respect to the invariant measure. Theorem 4.1 says that this mean

is always an element of $\mathcal{G}_c(f)$, but depends on the probability law assigned to the directions.

If the Bellman chain is irreducible, but with at least two critical classes, the invariant cost density is not unique, and the limit of the mean of f on the chain depends on the initial point. The corresponding more difficult results will be given in a forthcoming paper [3].

5. Lumpability Coherency and Reversibility of Bellman Chains

Statistical mechanics is useful to study very large systems. For moderate size systems the only methods are aggregation or separation of variables. We study here aggregation and separation of variables in the context of Bellman chains.

5.1. Residuation, Linear Projection, Aggregation and Coherency

The only invertible min-plus matrices are the diagonal matrices multiplied by permutation matrices. Fortunately, we can use the monotonicity properties of min-plus linear operators to define a minimal supersolution of a linear min-plus system.

For $A \in \mathcal{M}_{np}$, $B \in \mathcal{M}_{nq}$ with entries in $\overline{\mathbb{R}}_{\min}$, we define

$$X = A\backslash B \stackrel{\text{def}}{=} \min\{X \in \mathcal{M}_{pq} \mid AX \geq B\} ,$$

which does exist. We have $X_{lk} = \max_j (B_{jk} - A_{jl})$.

For $A \in \mathcal{M}_{pn}$, $B \in \mathcal{M}_{qn}$, we define also

$$X = B/A \stackrel{\text{def}}{=} \min\{X \in \mathcal{M}_{qp} \mid XA \geq B\} .$$

We have $X_{kl} = \max_j (B_{kj} - A_{lj})$.

Given $B : \overline{\mathbb{R}}_{\min}^p \to \overline{\mathbb{R}}_{\min}^n$ and $C : \overline{\mathbb{R}}_{\min}^n \to \overline{\mathbb{R}}_{\min}^q$, we denote $\operatorname{im} B$ the image of B and define $(\ker C)_x = C^{-1}C(x)$ which yields a fibration of $\overline{\mathbb{R}}_{\min}^n$.

The *projection* of $x \in \overline{\mathbb{R}}_{\min}^n$ on $\operatorname{im} B$ parallel to $\ker C$, denoted Px, is defined by $\operatorname{im} B \cap (\ker C)_x$ when this set is nonempty and contains a unique element. In this case we say that $\operatorname{im} B$ and $\ker C$ are *transverse*.

A necessary and sufficient condition of transversality [17] is that

$$CB((CB)\backslash C) = C, \text{ and } B = (B/(CB))CB .$$

Then we have

$$P = B((CB)\backslash C) = (B/(CB))C .$$

Given B [resp. C], there does not always exist C [resp.B] such that $\operatorname{im} B$ and $\ker C$ are transverse. There exists C [resp. B] iff B [resp. C] is *regular*, that is, if there exists a *generalized inverse* X to B, which satisfies, by definition $BXB = B$ [resp. $CXC = C$] (see [16, 17]).

We say that A is *aggregable* by C if there exists A_C such that $CA = A_C C$. In this case, the dynamic system $X_{k+1} = AX_k$ admits aggregate variables $Y_k = CX_k$ satisfying a reduced order dynamic $Y_{k+1} = A_C Y_k$.

THEOREM 5.1. *The matrix A is aggregable by the matrix C iff $CA = (AC/C)C$. Moreover if C is regular, A is aggregable with C iff there exists B satisfying $PA = PAP$, where P is the projector on* $\operatorname{im} B$ *parallel to* $\ker C$. *Then, we have $A_C = CA(B/(CB))$.*

Proof. The first part follows directly from the definition of the residuation. Let us prove the second part.

Since C is regular, we know [17] that there exist B and P such that $CP = C$ and $PB = B$.

The sufficiency condition is obtained by left multiplying $PA = PAP$ by C. We obtain

$$CPA = CPAP = CAP = CA(B/(CB))C \ .$$

The necessary condition is obtained by left multiplying $CA = A_C C$ by (B/CB). We obtain

$$PA = (B/(CB))CA = (B/(CB))A_C C = (B/(CB))A_C CP$$

$$= (B/(CB))CAP = PAP \ .$$

□

We say that B is *coherent* with A if there exists A_B such that $AB = BA_B$. Then, if $X_0 \in \operatorname{im} B$, the dynamical system $X_{k+1} = AX_k$ admits coherent variables U_k such that $X_k = BU_k$. The coherent variables follow a reduced order dynamic $U_{k+1} = A_B U_k$.

THEOREM 5.2. *The matrix A is coherent with the matrix B iff $AB = B(B\backslash AB)$. Moreover, if B is regular, the matrix A is coherent with B iff there exists C satisfying $AP = PAP$, where P is a projector on* $\operatorname{im} B$ *parallel to* $\ker C$*. Then, we have $A_B = ((CB)\backslash C)AB$.*

Proof. The proof is dual to the proof of the previous theorem. □

All the results about lumpability and reversibility given in [20] can be extended to the min-plus context because they are purely combinatorial results. We recall them here because they give new results about aggregation and decomposition of dynamic programming equations.

5.2. REVERSIBILITY

Let us consider an irreducible matrix $A \in \mathcal{M}_n$ with a unique critical class. The eigen-semimodule associated to its unique eigenvalue is generated by only one eigenvector v. Denoting $V = \operatorname{diag} v$, we have $v = VE$, where E is the n-vector with entries e. We have $AVE = VE$, therefore $V^{-1}AVE = E$ which means that $\hat{A} \stackrel{\text{def}}{=} VA'V^{-1}$ has E' as left eigenvector. It is the transition matrix of a Bellman chain.

The matrix A is said *reversible* when $A = \hat{A}$. A reversible matrices is normalized. It is quite easy to compute the right eigenvector of a reversible matrix.

THEOREM 5.3. *For a reversible matrix A (with right eigenvector v) and a path p from i to j we have*

$$\frac{v_j}{v_i} = \bigotimes_{(k,k')\in p} \frac{A_{k'k}}{A_{kk'}} \ .$$

Proof. The proof is immediate from the equality $AV = VA'$ satisfied by A. □

5.3. LUMPABILITY AND COHERENCY

Let us consider a dynamic system with transition matrix A. Aggregation of A by the characteristic function C of a partition of the states is called lumpability.

More precisely, if we denote the state space by $\mathcal{E} = \{1, \cdots, n\}$ and if we consider a partition $\mathcal{U}$ of the states, the characteristic function U of the partition $\mathcal{U}$ is defined by

$$U_{iJ} = \begin{cases} e & \text{if} \quad i \in J, \\ \varepsilon & \text{if} \quad i \notin J, \end{cases} \quad \forall i \in \mathcal{E},\, J \in \mathcal{U} \ .$$

The matrix A is said *lumpable* if it is aggregable by the matrix $C = U'$.

Let us consider a weight diagonal matrix $W = \operatorname{diag}(w_1, \cdots, w_n)$, where w is normalized (that is $E'w = e$), with nonzero entries. Then, the matrix $S = U'WU$ is diagonal. Taking $B = WUS^{-1}$, we have $CB = e$, then B and C are transverse and $P = BC$. We have the following easy result.

THEOREM 5.4. *A is lumpable iff for all $J, K \in \mathcal{U}$,*

$$\bar{A}_{KJ} \stackrel{\text{def}}{=} \bigoplus_{k \in K} A_{kj} \ , \tag{5.1}$$

is independent of the choice of $j \in J$. Then, $A_C = \bar{A} = CAB$ for all admissible weights w.

In the following section we will only consider coherency with the matrix $B = WUS^{-1}$. It is important to see that this matrix is a conditional cost with respect to the partition $\mathcal{U}$. Indeed, we have

$$B_{jJ} = w^{\mathcal{U}}_{jJ} \stackrel{\text{def}}{=} \frac{w_j}{\bigoplus_{k \in J} w_k}, \text{ if } j \in J, \ \varepsilon \text{ otherwise} \ .$$

THEOREM 5.5. *If A is normalized and coherent (with weight w and partition $\mathcal{U}$), then there exists a right eigenvector q of A satisfying $q^{\mathcal{U}} = w^{\mathcal{U}}$.*

Proof. If A is B coherent, we have $AB = BA_B$. Denoting $\overline{q}$ any eigenvector of A_B, we see that $q = B\overline{q}$ is an eigenvector of A. The result follows from $q^{\mathcal{U}} = B = w^{\mathcal{U}}$. □

Denoting $\hat{A} = WA'W^{-1}$, it is clear that:

- if A is C-aggregable then $\hat{A}$ is B-coherent,
- if A is B-coherent then $\hat{A}$ is C-aggregable,
- if $AP = \hat{A}P$ then $\bar{A} = \bar{\hat{A}}$ and aggregability and coherency imply each other.

When A is simultaneously aggregable and coherent, it is possible to decompose the computation of an eigenvector.

THEOREM 5.6. *For A lumpable and coherent with respect to the partition $\mathcal{U}$ and the weight w, there exists an eigenvector q satisfying*

$$q_j = \bar{q}_J q^J_j, \quad \forall j \in J, \ \forall J \in \mathcal{U} \ ,$$

$$\bar{q} = \bar{A}\bar{q}, \quad A^{JJ} q^J_+ = \bar{A}_{JJ} q^J_+ \ ,$$

where q^J_+ is the nonzero part of q^J, $\bar{A}_{JJ}$ is defined by (5.1) and A^{JJ} is the Jth diagonal block of A (having the size the number of elements of the set J).

Proof. We have only to prove that $A^{JJ}q_+^J = \bar{A}_{JJ}q_+^J$. The other facts amount to rephrasing Theorem 5.5. From the structure of B, $AB = BA_B$ and $q^{\mathcal{U}} = w^{\mathcal{U}}$ we see that $A^{JJ}q_+^J = (A_B)_{JJ}q_+^J$. Thanks to the lumpability assumption we know that the aggregate matrix is given by (5.1). □

5.4. PARTIAL REVERSIBILITY

It is possible to compute in a decomposed way the eigenvector (supposed to be unique) of a matrix A under another assumption. We say that the matrix A is *partially reversible* if it satisfies $AQU = QA'U$, with $Q = \operatorname{diag} q$, for $q = Aq$.

PROPOSITION 5.1. *The following statements are equivalent:*

1. *A is partially reversible;*
2. $\bigoplus_{j\in J} A_{kj}q_j = \bigoplus_{j\in J} A_{jk}q_k,\ \forall J, k$;
3. *$AP = \hat{A}P$, with $P = B(q)C$.*

Under partial reversibility, we can decompose the computation of q but the local problems that we have to solve are different from those of the previous section.

COROLLARY 5.1. *The right eigenvector q of a partially reversible matrix A satisfies $q_j = \bar{q}_J q_j^J$ with $\bar{A}\bar{q} = \bar{q}$ and $A^{JJ}q_+^J = D^J q_+^J$, where $D^J = \operatorname{diag}(\oplus_{j\in J}A_{jk},\ k \in J)$.*

Proof. This result is a rephrasing of statement 2 of the previous proposition. □

The following result gives the relation existing between aggregability coherency and reversibility.

THEOREM 5.7. *Under partial reversibility of a matrix A, aggregability and coherency imply each other and the aggregate matrix $\bar{A}$ is reversible.*

Under aggregability and coherency of the matrix A and the reversibility of the aggregate $\bar{A}$, we have

$$AP = PA = PAP = P\hat{A}P = \hat{A}P = P\hat{A}\ .$$

6. EPILOGUE: MIN-PLUS ALGEBRA AND LARGE DEVIATION ASYMPTOTICS

In the above sections, we tried to set up connections between statistical physics and discrete event systems, and between probability calculus and decision calculus. These connections were mostly motivated by *analogies*. As a conclusion, we would like to point out different connections, motivated by *asymptotics*. Such connections are presented in detail in [5].

As pointed out in the introduction, min-plus algebra can be seen as an asymptotic deformation of usual algebra. Formally, consider the semifield $\mathbb{R}_\epsilon = (\mathbb{R}\cup\{+\infty\}, \oplus_\epsilon, +)$, with

$$a \oplus_\epsilon b = -\epsilon \log(e^{-a/\epsilon} + e^{-b\epsilon})\ .$$

This semifield was already introduced by Maslov in [37] (see [34] for a more recent presentation) and by Pap in [39]. Clearly, for any non-zero value of ϵ, $\mathbb{R}_\epsilon$ is isomorphic to the ordinary semifield of non-negative integers, $(\mathbb{R}^+, +, \times)$. However,

$$\lim_{\epsilon\to 0^+} a \oplus_\epsilon b = \min(a, b)\ .$$

Hence, the min-plus semifield $\mathbb{R}_{\min}$ can be seen as the limit of the semifield $\mathbb{R}_\epsilon$. This elementary observation supports the following general "principle", that the objects which appear in large deviation type asymptotics have min-plus algebraic properties. As an illustration of this principle, we will show how min-plus spectral theory can help us to determine the asymptotics of non-negative spectral elements of non-negative matrices. A different illustration, which relates the large deviation principle with weak convergence of $\oplus_\epsilon$-additive measures in the family of semifields $\{\mathbb{R}_\epsilon\}$, is given in [4].

With a matrix $A \in (\mathbb{R} \cup \{+\infty\})^{n\times n}$, that is *irreducible* in the min-plus sense (see section 2.2), we associate the non-negative matrix

$$A_\epsilon = (e^{-\frac{A_{ij}}{\epsilon}}) \ ,$$

and look for the normalized Perron eigenvector U_ϵ and Perron eigenvalue λ_ϵ, which are defined by

$$A_\epsilon U_\epsilon = \lambda_\epsilon U_\epsilon \ , \qquad \forall i \in \{1, \dots, n\}, \ (U_\epsilon)_i \geq 0, \qquad \sum_{1\leq j\leq n} (U_\epsilon)_j = 1 \ .$$

In the sequel, we will use the notation $\log x \stackrel{\text{def}}{=} (\log x_1, \dots, \log x_n)$ for vectors. Consider

$$U = \lim_{\epsilon\to 0^+} -\epsilon \log U_\epsilon, \qquad \lambda = \lim_{\epsilon\to 0^+} -\epsilon \log \lambda_\epsilon \ . \tag{6.1}$$

Such limits arise for instance in the study of low-temperature asymptotics of one-dimensional Ising type models with transfer matrix A_ϵ. In this context, λ can be interpreted as the free energy per site at zero temperature, and U is an "effective potential". See e.g. [12], and [15] for a direct max-plus algebraic approach.

The existence of the limits in (6.1) can be derived from the existence of a (generalized) Puiseux expansion for λ_ϵ. Clearly, U and λ are solution of the min-plus spectral problem:

$$A \otimes U = \lambda \otimes U \ . \tag{6.2}$$

Therefore, we obtain as a consequence of the min-plus spectral theorem (Theorem 2.2) the following result.

THEOREM 6.1. *Let $A \in (\mathbb{R} \cup \{+\infty\})^{n\times n}$ be an irreducible*[4] *matrix. Then, $\lambda = \lim_{\epsilon\to 0^+} -\epsilon \log \lambda_\epsilon$ is the unique min-plus eigenvalue of A.*

The argument is detailed in [28, Ch.IV],[5].

If A has a unique *critical class* (see section 2.2), it has a unique eigenvector up to a proportionality factor, thus, the limit $U = \lim_{\epsilon\to 0^+} -\epsilon \log U_\epsilon$ is well characterized by (6.2), together with $\min_{1\leq i\leq n} U_i = 0$, which follows from the normalization condition on U_ϵ. However (and this is probably the main difference between min-plus spectral theory and Perron-Frobenius theory), an irreducible matrix A has in general several non-proportional eigenvectors (see Theorem 2.2). This raises the problem of deciding which particular min-plus eigenvector is selected by the limit procedure (6.1). The solution to this problem, given in [5], relies on an aggregation procedure which is analogous to the cycle aggregation in Wentzell-Freidlin theory (although specific technical difficulties arise due to the absence of "river networks" explicit formulæ for the principal cofactors of the non-negative but non-Markov matrix A_ϵ). At each aggregation level, a new spectral problem in min-plus algebra

[4] In the min-plus sense.

has to be solved, and the eigenvector U is finally obtained as a product of matrices, whose columns are formed by solutions of these min-plus spectral problems.

THANKS. We thank a referee for the remarks he made about the paper. Clearly, this paper is only a first attempt to connect min-plus theory with the most elementary notions of statistical mechanics. Deeper results and improvements of the point of view are still to be done.

REFERENCES

[1] M. Akian : "Densities of idempotent measures and large deviations", to appear AMS, and INRIA Report N. 2534 (1995).
[2] M. Akian : "Theory of cost measures: convergence of decision variables", INRIA Report N. 2611 (1996).
[3] M. Akian : "Ergodic theorems for Bellman chains", in preparation.
[4] M. Akian : "From probabilities to cost measures via large deviations," INRIA Report, to appear.
[5] M. Akian, R.B. Bapat and S. Gaubert : "Asymptotics of the Perron root and eigenvector using max-algebra". In preparation.
[6] M. Akian, J.P. Quadrat and M. Viot : "Bellman Processes" L.N. in control and Inf. Sciences N. 199, Ed. G. Cohen, J.P. Quadrat, Springer Verlag (1994).
[7] M. Akian, J.P. Quadrat and M. Viot : "Duality between probability and optimization" In J. Gunawardena (Editor), "Idempotency", Cambridge University Press (1997) to appear.
[8] M. Aoki: "Control of Large Scale Dynamic Systems by Aggregation," IEEE Trans. on Automatic Control, Vol. AC-13, p. 246–253 (1968).
[9] V. Arnold : "Méthodes Mathématiques de la Mécanique Classique", Éditions MIR, Moscou, (1976).
[10] H. Attouch : "Variational convergence for functions and operators", Pitman, (1984).
[11] F. Baccelli, G. Cohen, G.J. Olsder, and J.P. Quadrat : "Synchronization and Linearity", Wiley (1992).
[12] R.J. Baxter : "Exactly solved models in statistical mechanics", Academic Press, New York (1982).
[13] F. Bellalouna : "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).
[14] P. Billingsley : "Convergence of probability measures", J. Wiley & Sons, New York, (1968).
[15] W. Chou and R.B. Griffiths : "Ground states of one dimensional systems using effective potentials", Phys. Rev. B, 34, 6219–34 (1986).
[16] G. Cohen, S. Gaubert and J.P. Quadrat : "Kernels, Images and Projections in Dioids" Proceedings of WODES'96, Edinburgh (1996).
[17] G. Cohen, S. Gaubert and J.P. Quadrat : "Linear Projectors in the max-plus Algebra" IEEE Mediterranean Conference on Control, Chypre (August 1997).
[18] P. J. Courtois : "Decomposability Queuing and Computer System Applications", Academic Press, New York (1977).
[19] R. Cuninghame-Green : "Minimax Algebra", L.N. on Economics and Math. Systems N. 166, Springer Verlag (1979).
[20] F. Delebecque, P. V. Kokotovic and J. P. Quadrat : "Aggregation and Coherency in Networks and Markov Chains", Int. Jour. of Control, Vol. 35 (1984).
[21] A. Dembo and O. Zaitouni : "Large deviations techniques and applications", Jones and Barlett, Boston, MA, (1993).
[22] P. Del Moral : "Résolution particulaire des problèmes d'estimation et d'optimisation nonlinéaires", Thesis dissertation, Toulouse, France (1994).
[23] P. Del Moral, T. Thuillet, G. Rigal and G. Salut : "Optimal versus random processes : the nonlinear case", rapport de recherche LAAS, (1990)
[24] R.S. Ellis : "Entropy, Large Deviations and Statistical Mechanics", Springer Verlag (1985).
[25] G. Fayolle and J.M. Lasgouttes : "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).

[26] M.I. Freidlin and A.D. Wentzell : “Random Perturbations of Dynamical Systems”, Springer (1984).
[27] V. G. Gaitsgori and A. A. Pervozvanskii : “Aggregation of States in a Markov Chain with Weak Interactions,” Kybernetika, p.91–98, May-June (1975).
[28] S. Gaubert : “Théorie des systèmes linéaires dans les dioides”, Thesis dissertation, École des Mines de Paris, (1992).
[29] S. Gaubert and M. Plus : “Methods and applications of (max,+) linear algebra”, in R. Reischuk and M. Morvan, editors, *STACS'97*, number 1200 in LNCS, Lübeck, Springer, (1997).
[30] A. Hamm : “Uncertain dynamical systems defined by pseudomeasures”, Int. J. Math. Phys., vol.38, p.3081-3109, (1997).
[31] F. P. Kelly : “Reversibility and Stochastic Networks”, J. Wiley (1979).
[32] J. G. Kemeny and J. L. Snell : “Finite Markov Chains”, Van Nostrand, Princeton, N.J. (1967).
[33] V. N. Kolokoltsov and V. P. Maslov : “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, N.1, p.55–59 (1988).
[34] V. N. Kolokoltsov and V. P. Maslov : “Idempotent Analysis and Its Applications”, Kluwer (1997).
[35] G.L. O'Brien and W. Vervaat : “Capacities, large deviations and loglog laws” In S. Cambanis, G. Samorodnitsky and M.S. Taqqu (Editors) “Stable processes and related topics”, vol. 25 of Progress in Probability, p.43–83, Birkhauser, (1991).
[36] V.P. Maslov and S.N. Samborskii : “Idempotent Analysis”, AMS (1992).
[37] V.P. Maslov : “Méthodes Opératorielles”, Éditions MIR, Moscou (1987).
[38] L. Landau and E. Lifchitz : “Physique Statistique”, Editions MIR, Moscou (1967).
[39] E. Pap : “Null-Additive Set Functions”, Mathematics and Its applications 337, Kluwer academic Publishers, Dordrecht (1995).
[40] K. Petersen : “Ergodic Theory”, Cambridge University Press (1983).
[41] A. Pulhaskii : “Large deviations of semimartingales via convergence of the predictable characteristics”, Stochastic and stochastic reports, N. 49, p.27–85 (1994).
[42] A. Pulhaskii : “Large deviation analysis of the single server queue”, Queuing Systems, No.21, p.5–66, (1995).
[43] J.P. Quadrat : “Théorèmes asymptotiques en programmation dynamique”, Note CRAS 311, p.745-748, Paris (1990) .
[44] H. A. Simon and A. Ando : “Aggregation of Variables in Dynamic Systems,” Econometrica, Vol. 729, p.111–138, (1963).
[45] M. Viot : “Processus de renouvellement min-plus”. In preparation.
[46] S.R.S., Varadhan : “Large deviations and applications”, CBMS-NSF Regional Conference Series in Applied Mathematics N. 46, SIAM Philadelphia, Penn. (1984).
[47] P. Whittle : “Risk sensitive optimal control” John Wiley and Sons, New York (1990).