The (max,+) semiring. An introduction

Stéphane Gaubert
INRIA, Rocquencourt

March 11, 1996

[summary by Marianne Akian]

Abstract

Endowing real (or natural) numbers with max and + laws leads to an idempotent semiring which has been reinvented in many domains: graph optimization, language theory, statistical physics, quantum mechanics, discrete event systems, etc. The talk presents applications together with basic results of the so-called (max,+) algebra.

Introduction

We say that $(\mathbb{S}, \oplus, \otimes)$ is an idempotent semiring or dioid [19, 2] if $\oplus$ and $\otimes$ are associative laws on $\mathbb{S}$ with neutral elements $\mathbf{0}$ and $\mathbf{1}$ respectively, $\oplus$ is commutative and idempotent, that is $a \oplus a = a$, $\otimes$ is distributive with respect to the $\oplus$ law and $\mathbf{0}$ is absorbing with respect to the $\otimes$ law. By the idempotency property, $a \oplus b = \mathbf{0}$ implies $a = \mathbf{0}$. Then, the $\oplus$ law is not symmetrizable (and not simplifiable). However, idempotency leads to "simplifications" that partially compensate the non simplifiability. An idempotent semiring is said commutative when $\otimes$ is commutative and it is a semifield if the $\otimes$ law is invertible. Examples of commutative idempotent semifields are $\mathbb{R}_{\max} = (\mathbb{R} \cup \{-\infty\}, \max, +)$ with $\mathbf{0} = -\infty$ and $\mathbf{1} = 0$, $\mathbb{R}_{\min} = (\mathbb{R} \cup \{+\infty\}, \min, +)$, $(\mathbb{R}^+, \max, \times)$ which are isomorphic. They are called respectively $(\max, +)$, $(\min, +)$ and $(\max, \times)$ algebra and are used in operations research [7], graph theory [19], discrete event systems [2, 14, 13], dynamic programming, Hamilton-Jacobi-Bellman equations [28, 1, 8], exponential asymptotics [29, 23, 5, 4]. The subsemiring $\mathbb{N}_{\min} = (\mathbb{N} \cup \{+\infty\}, \min, +)$ of $\mathbb{R}_{\min}$, called tropical semiring, is used in language theory [21, 22, 33, 34, 25, 24]. Concerning theoretical results on $(\max, +)$ algebra, an historical reference is [7]. More recent accounts can be found in [2, 28], collections of survey papers will be presented in [20] and a general and complete bibliography can be found in [26].

1. Some applications

1.1. Shortest path problem. The traditional application of the $(\min, +)$ algebra concerns the shortest path problem in a graph [19]. Let G be a graph with nodes denoted $\{1, \ldots, n\}$ representing towns and arcs representing roads. Let A_{ij} denote the time to go from i to j (or the length of arc (i, j)) with $A_{ij} = +\infty$ when there is no arc. If $A = (A_{ij})$ is considered as a $(\min, +)$ matrix,

$$(A^k)_{ij} = \bigoplus_{i_1, \ldots, i_{k-1}} A_{ii_1} \otimes \cdots \otimes A_{i_{k-1}j} = \min_{i_1, \ldots, i_{k-1}} A_{ii_1} + \cdots + A_{i_{k-1}j}$$

represents the minimal time from i to j (or the minimal distance between i and j) in k steps. If $A^* = \oplus_{k=0}^{\infty} A^k$, then $(A^*)_{ij}$ represents the minimal time from i to j.

A similar problem arises in discrete deterministic optimal control. Let now A_{ij} represent the cost of i to j transition, b_i the final cost in state i at time N and let v_i^n denote the minimal cost of a trajectory starting in i at time $n \leq N$. The value function v^n satisfies the backward dynamic programming (or Hamilton-Jacobi-Bellman) equation

$$v_i^n = \min_j A_{ij} + v_j^{n+1}, \qquad v_i^N = b_i$$

that is $v^n = A \otimes v^{n+1}$ with $v^N = b$, which is the $(\min, +)$ analogue of the Kolmogorov or backward Fokker-Planck equation, (final or transition) costs replacing probabilities [1, 8]. More generally, dynamic programming equations with continuous time and state are solved using $(\min, +)$ algebra in [28, 23].

1.2. Synchronization problems. Let us consider a manufacturing system where 2 types of parts are assembled, taking a fixed duration τ. Let $u_i(t)$ denote the number of parts of type $i = 1, 2$ arrived at time t and $y(t)$ the number of parts assembled. Then

$$y(t) = \min(u_1(t-\tau), u_2(t-\tau)) = u_1(t-\tau) \oplus u_2(t-\tau)$$

in the $(\min, +)$ algebra. If now $u_i(n)$ (resp. $y(n)$) denotes the date of the n-th arriving of part i (resp. of the n-th assemblage of parts), we obtain

$$y(n) = \tau + \max(u_1(n) + u_2(n)) = \tau \otimes (u_1(n) \oplus u_2(n))$$

in $(\max, +)$ algebra. More generally, any problem that can be modelled by a timed event graph (a subclass of timed Petri nets modelled synchronization features) can also be represented by a $(\min, +)$-linear dynamical system (for counter variables)

$$\begin{cases} x(t) = A \otimes x(t-1) \oplus B \otimes u(t), \\ y(t) = C \otimes x(t) \end{cases}$$

or by a $(\max, +)$-linear dynamical system (for dater variables $y(n)$, $x(n)$ and $u(n)$). A linear system theory in $(\min, +)$ and $(\max, +)$ algebras analogous to the classical linear control theory is developed in [2].

1.3. Exponential asymptotics. Let us consider a one-dimensional system of n atoms with energy $H_n(q_1, \ldots, q_n) = V(q_1) + \sum_{k=2}^n K(q_{k-1}, q_k)$, where q_n is the position (state) of the n-th atom with $q_1 < \cdots < q_n$ and $K(q, q') = V(q') + W(q' - q)$ is the sum of the potential V in position q and the potential energy W linking nearest neighbours. The Gibbs distribution of this system has density $\exp(-\beta H_n(q_1, \ldots, q_n))/Z_n$, where β is the inverse of the temperature and $Z_n = \sum_{q_1, \ldots, q_n} \exp(-\beta H_n(q_1, \ldots, q_n))$ is the partition function. Let T be the transfer matrix

$$T_{qq'} = \exp\left(-\beta K(q, q')\right),$$

Q be the row vector with entries $Q_q = \exp(-\beta V(q))$ and e the vector with entries 1. Then $Z_n = QT^{n-1}e$ and the probability for the first atom to be in position q is $P(q) = Q_q (T^{n-1}e)_q / Z_n$. For good matrices T, $P_n(q)$ tends to $P(q) = Q_q R_q$ when n goes to infinity, where R is a right eigenvector of the transfer matrix such that $Q \cdot R = 1$. Similarly, the probability of the n-th atom tends to L_q, where L is a left eigenvector of the transfer matrix such that $L \cdot e = 1$. Moreover, for any transfer matrix, $\log Z_n / n$ tends to $\log \rho$, where ρ is the Perron root of T. The free energy by atom is then $\lambda = \log \rho / \beta$.

If now the temperature is zero ($\beta = +\infty$), either the previous results have to be obtained passing to the limit in β using the property that the $(\min, +)$ algebra is the limit of the $(\mathbb{R}^+, +, \times)$ semifield:

$$\lim_{\beta \to +\infty} \frac{-1}{\beta} \log(e^{-\beta a} + e^{-\beta b}) = \min(a, b), \qquad \frac{-1}{\beta} \log(e^{-\beta a} \cdot e^{-\beta b}) = a + b;$$

or a similar reasoning has to be done directly in the $(\min, +)$ algebra. In this last case, the transfer matrix method is replaced by the effective potential method [5, 4]. Let us consider the $(\min, +)$-matrix K in place of T. The effective potential of the extremal atom of a semi-infinite chain of atoms extending to the right (resp. left) is equal to $F(q) = V(q) + R_q$ (resp. $F(q) = L_q$), where R and L are right and left $(\min, +)$-eigenvectors of K such that $\min_q V(q) + R_q = \min_q L_q = 0$. The energy by atom for a minimum-energy configuration is then the $(\min, +)$-eigenvalue λ of K: $K \otimes R = \lambda \otimes R = \lambda + R$, $L \otimes K = \lambda \otimes L = \lambda + L$. Exponential asymptotics also occur in large deviations and asymptotics of Schrödinger equations (WKB method) [29, 23].

1.4. Language theory. A finite automaton with cost or distance is an automaton with multiplicity over the tropical semiring $\mathbb{N}_{\min}$. For any rational language L over the finite alphabet Σ, a finite automaton with cost A can be constructed, recognizing $L^* = \cup_{n=0}^{\infty} L^n$ (where product means concatenation) and counting for each word $w \in L^*$ the least n such that $w \in L^n$. This has been used by Simon and Hashiguchi [21, 22, 33] to solve positively a long standing problem of J. A. Brzozowski, the decidability for a rational language of the finite power property (FPP) (a language L has the FPP iff there exists N such that $L^* = \cup_{n=0}^{N} L^n$). Indeed, the automaton A has only one initial state and one terminal state and since the language L has the FPP iff A is limited (that is costs of recognized words are bounded), the FPP is equivalent to the finite section property of a finitely generated subsemigroup of matrices of $\mathbb{N}_{\min}^{n \times n}$. Following this first application, other decidability properties for finitely generated subsemigroups of matrices over the tropical semiring and/or automata with cost have been studied [21, 22, 33, 34, 25, 24].

Similarly to cost automata, $(\max, +)$ automata can be also constructed. They allow to represent heaps of pieces and parallel (multitask, multiresource) discrete event systems [17, 16, 27].

2. $(\max, +)$ linear algebra

2.1. Solutions of linear equations and subsemimodules. Since the $\oplus$ law is not symmetrizable in a dioid, general linear equations are of the form $A \otimes x \oplus b = C \otimes x \oplus d$. Important particular cases are $A \otimes x = b$ and $x = A \otimes x \oplus b$. The following result is classical [7] and shows that the first particular equation is not easy to solve.

THEOREM 1. *$A \in \mathbb{R}_{\max}^{n \times n}$ is invertible iff $A = DS$, where D and S are diagonal and permutation matrices.*

THEOREM 2 ([30, 36]). *Any finitely generated subsemimodule of $\mathbb{R}_{\max}^n$ has a base (minimal generating family) which is unique up to invertible linear operations.*

THEOREM 3 ([3, 14]). *For any matrices $A, B \in \mathbb{R}_{\max}^{m \times n}$, the set of solutions of $A \otimes x = B \otimes x$ is a finitely generated semimodule.*

Let us solve $x = A \otimes x \oplus b$. To any dioid is associated a partial order: $a \preceq b \Leftrightarrow a \oplus b = b$. In $\mathbb{R}_{\max}$ it is the classical order $\leq$, in $\mathbb{R}_{\min}$ it is the opposite order $\geq$. The dioid $(\mathbb{S}, \oplus, \otimes)$ is complete if any set (even empty) has a least upper bound and if $\otimes$ is distributive with respect to infinite sums. $\mathbb{R}_{\max}$ is not complete but it may be completed in the complete dioid $\overline{\mathbb{R}}_{\max} = (\mathbb{R} \cup \{+\infty, -\infty\}, \max, +)$ with the convention $+\infty + -\infty = -\infty$ ($\mathbf{0}$ is absorbing).

THEOREM 4. *In a complete dioid $\mathbb{S}$, the least solution of $x = a \otimes x \oplus b$ is $a^* \otimes b$, where $a^* = \oplus_{n\in\mathbb{N}} a^n = \sup_{n\in\mathbb{N}} a^n$. Similarly, the least solution of $x = A \otimes x \oplus b$ in $\mathbb{S}^n$ is $x = A^* b$. It can be computed by Gauss algorithm.*

In order to solve the general equation $A \otimes x \oplus b = C \otimes x \oplus d$, a symmetrization of $\mathbb{R}_{\max}$ seems necessary. Although no idempotent field or ring containing $\mathbb{R}_{\max}$ exists, a symmetrized idempotent semiring $\mathbb{S}_{\max}$ has been constructed. It contains positive numbers $x \in \mathbb{R}_{\max}$, negative numbers $\ominus x$, but also doted numbers $\dot{x} = x \ominus x$ which are not invertible. Symmetrizing linear equations in $\mathbb{R}_{\max}$, we obtain balance equations in $\mathbb{S}_{\max}$, where x balances y iff $x \ominus y$ is doted. In $\mathbb{S}_{\max}$, determinants can be calculated and linear balance equations can be solved using Cramer formula or Gauss-Seidel and Jacobi algorithms [2, 14, 31].

2.2. Subsolutions of linear equations: residuation.

DEFINITION 1. Let $f : (E, \leq) \to (F, \leq)$ be a nondecreasing application between lattices. f is residuable iff $\{x \in E, f(x) \leq b\}$ has a maximal element for any $b \in F$.

THEOREM 5. *If $f : \mathbb{S} \to \mathbb{S}'$ is an application between complete dioids such that $f(\mathbf{0}) = \mathbf{0}$ and $f(\sup_{x\in X} x) = \sup_{x\in X} f(x)$ for any subset X of $\mathbb{S}$, then f is residuable.*

As a corollary, any multiplication operation (by a scalar or a matrix) is residuable. Let us denote by $a\backslash b = \max\{x, a \otimes x \preceq b\}$ and $b/a = \max\{x, x \otimes a \preceq b\}$ the residuations of multiplications by the scalar a in any complete dioid. The residuation of the multiplication by a matrix in $\mathbb{R}_{\max}$, $A\backslash b = \max\{x \in \mathbb{R}^n_{\max}, A \otimes x \preceq b\}$ gives the vector with entries $(A\backslash b)_i = \inf_j A_{ji}\backslash b_j = \min_j -A_{ji} + b_j$, that is the $\mathbb{R}_{\min}$ product of the matrix $-A^T$ by b. Applications to system theory can be found in [2]. While linear operators represent the earliest behaviour of a system, the latest behaviour can be represented by a dynamical equation involving residuation.

2.3. Spectral theory. The most useful result of $(\max, +)$ linear algebra is perhaps the following analogue of Perron-Frobenius theorem.

THEOREM 6 ([7, 35, 32, 18, 6, 10]). *Any irreducible matrix $A \in \mathbb{R}^{n\times n}_{\max}$ has a unique eigenvalue $\rho(A)$ and*

$$\rho(A) = \oplus_{k=1}^{n} \operatorname{tr}(A^k)^{\frac{1}{k}} = \max_{k=1,\dots,n} \max_{i_1,\dots,i_k} \frac{A_{i_1 i_2} + \cdots + A_{i_k i_1}}{k}$$

If A is reducible, the previous formula gives the maximal eigenvalue.

The $(\min, +)$ eigenvalue is then the minimal mean cost (ergodic cost) of a control problem or the asymptotic production rate of a manufacturing system, etc. As in the statistical physics application of section 1.3, it can be obtained as the limit of the Perron root of a matrix.

THEOREM 7 ([12, 11]). *Let A be any $n \times n$ matrix with entries in $\mathbb{R}^+$. If $\rho_{PF}(A)$ is the Perron-Frobenius root of A and $\rho_{(\max,\times)}(A) = \exp(\rho((\log A_{ij})))$ its $(\max, \times)$-eigenvalue, we have*

$$\rho_{(\max,\times)}(A) \leq \rho_{PF}(A) \leq n\rho_{(\max,\times)}(A).$$

COROLLARY 1. *Let $A^{\circ r} = (A^r_{ij})$ and $e^{\circ\beta A} = (\exp(\beta A_{ij})$ denote the r-th power of A and the exponential of βA for the Hadamard product. For any matrix with positive entries*

$$\rho_{(\max,\times)}(A) = \lim_{r\to+\infty} (\rho_{PF}(A^{\circ r}))^{\frac{1}{r}}$$

and for any matrix with entries in $\mathbb{R}_{\max}$

$$\rho(A) = \lim_{\beta \to +\infty} \frac{1}{\beta} \log \rho_{PF}(e^{\circ \beta A}).$$

THEOREM 8 ([6, 9]). *For any irreducible matrix* $A \in \mathbb{R}_{\max}^{n \times n}$*, there exists* c *and* $N \geq 1$ *such that* $A^{n+c} = \rho(A)^c A^n$ *for* $n \geq N$.

In the context of timed event graphs, this means that the system reaches after a finite transient behaviour (of length N) a periodic regime of period c in which the production rate is equal to the eigenvalue.

These periodicity results can also be dealt with using rational generating series over the $(\max, +)$ semiring [2, 15].

Bibliography

[1] Akian (M.), Quadrat (J. P.), and Viot (M.). – Duality between probability and optimization. In Gunawardena (J.) (editor), *Idempotency. Publication of the Isaac Newton Institute.* – Cambridge University Press, 1996. To appear.

[2] Baccelli (F.), Cohen (G.), Olsder (G. J.), and Quadrat (J. P.). – *Synchronization and Linearity.* – Wiley, 1992.

[3] Butkovic (P.) and Hegedűs (G.). – The elimination method for finding all solutions of the system of linear equations over an extremal algebra. *Ekonomicko-matematicky Obzor*, vol. 20, 1984.

[4] Chou (W.) and Griffiths (R. B.). – Effective potentials, a new approach and new results for one-dimensional systems with competing lenght scales. *Physical Review Letters*, vol. 56, 1986, pp. 1929–1931.

[5] Chou (W.) and Griffiths (R. B.). – Ground states of one dimensional systems using effective potentials. *Physical Review B*, vol. 34, 1986, pp. 6219–6234.

[6] Cohen (G.), Dubois (D.), Quadrat (J. P.), and Viot (M.). – *Analyse du comportement périodique des systèmes de production par la théorie des dioïdes.* – Rapport de recherche n° 191, Institut National de Recherche en Informatique et en Automatique, Le Chesnay, France, 1983.

[7] Cuninghame-Green (R. A.). – *Minimax Algebra.* – Springer Verlag, 1979, *Lecture notes in Economics and Mathematical Systems.*

[8] Del Moral (P.). – *Résolution particulaire des problèmes d'estimation et d'optimisation non-linéaires.* – Thèse, Université Paul Sabatier, Toulouse, 1994.

[9] Dudnikov (P.) and Samborskiĭ(S.). – Endomorphisms of finitely generated free semimodules. In Maslov (V.) and Samborskiĭ(S.) (editors), *Idempotent analysis.* – Americal Mathematical Society, Rhode Island, 1992.

[10] Dudnikov (P. I.) and Samborskiĭ(S. N.). – Spectra of endomorphisms of semimodules over semirings with an idempotent operation. *Soviet Mathematics Doklady*, vol. 40, n° 2, 1990, pp. 363–366.

[11] Elsner (L.), Johnson (C. R.), and Dias da Silva (J.). – The Perron root of a weighted geometric mean of nonnegative matrices. *Linear Multilinear Algebra*, vol. 24, 1988, pp. 1–13.

[12] Friedland (S.). – Limit eigenvalues of nonnegatives matrices. *Linear Algebra and Applications*, vol. 74, 1986, pp. 173–178.

[13] Gaubert (S.). – *Introduction aux systèmes dynamiques à événements discrets.* – Polycopié de cours donné à l'ENSTA, 1992.

[14] Gaubert (S.). – *Théorie des systèmes linéaires dans les dioïdes.* – Thèse, École des Mines de Paris, July 1992.

[15] Gaubert (S.). – Rational series over dioids and discrete event systems. In *Proceedings of the 11th Conference on Analysis and Optimization of Systems: Discrete Event Systems. Lecture notes in Control and Information Sciences.* – Sophia Antipolis, June 1994.

[16] Gaubert (S.). – Performance evaluation of (max,+) automata. *IEEE Transactions on Automatic Control*, vol. 40, n° 12, December 1995.

[17] Gaubert (S.) and Mairesse (J.). – Task resource systems and (max,+) automata. In Gunawardena (J.) (editor), *Idempotency*. – Cambridge University Press, March 1995. To appear in 1996.

[18] Gondran (M.) and Minoux (M.). – Valeurs propres et vecteurs propres en théorie des graphes. In *Problèmes combinatoires et théorie des graphes, Colloques internationaux CNRS*. – Orsay, 1976.

[19] Gondran (M.) and Minoux (M.). – *Graphes et algorithmes*. – Eyrolles, Paris, 1979. English translation *Graphs and Algorithms*, Wiley, 1984.

[20] Gunawardena (J.) (editor). – *Idempotency*. – Cambridge University Press, 1996, *Publications of the Newton Institute*. To appear.

[21] Hashiguchi (K.). – Limitedness theorem on finite automata with distance functions. *Journal of Computer and System Sciences*, vol. 24, n° 2, 1982, pp. 233–244.

[22] Hashiguchi (K.). – Improved limitedness theorems on finite automata with distance functions. *Theoretical Computer Science*, vol. 72, 1990, pp. 27–38.

[23] Kolokoltsov (V.) and Maslov (V.). – *Idempotent analysis and applications*. – Kluwer Academic Publisher, 1996. To appear.

[24] Krob (D.). – The equality problem for rational series with multiplicities in the tropical semiring is undecidable. *International Journal of Algebra and Computation*, vol. 3, 1993.

[25] Leung (H.). – Limitedness theorem on finite automata with distance function: an algebraic proof. *Theoretical Computer Science*, vol. 81, 1991, pp. 137–145.

[26] Litvinov (G.) and Maslov (V.). – *Correspondence principle for idempotent calculus and some computer applications*. – Technical Report n° IHES/M/95/33, IHES, Bures-sur-Yvette, France, April 1995.

[27] Mairesse (J.). – *Stabilité des systèmes à événements discrets stochastiques. Approche algébrique*. – Thèse, École Polytechnique, June 1995.

[28] Maslov (V.) and Samborskiĭ(S.) (editors). – *Idempotent analysis*. – American Mathematical Society, Rhode Island, 1992, *Advances in Soviet Mathematics*, vol. 13.

[29] Maslov (V. P.). – *Méthodes Opératorielles*. – Mir, Moscou, 1973. French translation 1987.

[30] Moller (P.). – *Théorie algébrique des Systèmes à Événements Discrets*. – Thèse, École des Mines de Paris, 1988.

[31] Plus (M.). – Linear systems in (max, +)-algebra. In *Proceedings of the 29th Conference on Decision and Control*. – Honolulu, December 1990.

[32] Romanovskiĭ(I. V.). – Optimization and stationary control of discrete deterministic process in dynamic programming. *Kibernetika*, vol. 2, 1967, pp. 66–78. – English translation in Cybernetics 3 (1967).

[33] Simon (I.). – Limited subsets of the free monoid. In *Proceedings of the 19th Annual Symposium on Foundations of Computer Science*. pp. 143–150. – IEEE, 1978.

[34] Simon (I.). – The nondeterministic complexity of a finite automaton. In Lothaire (M.) (editor), *Mots*. – Hermes, 1990.

[35] Vorobyev (N. N.). – Extremal algebra of positive matrices. *Elektronische Informationsverarbeitung und Kybernetik*, vol. 3, 1967. – In russian.

[36] Wagneur (E.). – Moduloids and pseudomodules. 1. dimension theory. *Discrete Mathematics*, vol. 98, 1991, pp. 57–73.