Linear equations in Exotic Algebra
Utility, Results & Applications

October 3, 2001

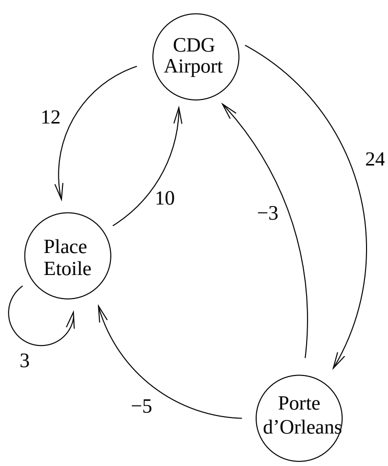
**Abstract**

Stress is made here to present the utility of exotic algebra, and methods built on it, in particular the spectral theory of $(\max, +)$ matrices , that helps in their applications. No mathematical background is specially needed. Further results, different areas of application, as well as extension of the framework are finally given.

# 1 $(\max, +)$ Algebra introduced by an application

## 1.1 Starting from a taxicab taking you from CDG to Paris.

We consider the day of a taxicab driver in Paris. He may go to the airport, or drive passenger within the city. We assume that the decision where to go from each place, is entirely left to driver's will among the different possibilities offered by clients.




- In "CGD airport", there is always passenger waiting to go anywhere to "Place de l'Etoile", or to "Porte d'Orleans", which is further.
- Near to "Place de l'Etoile" there is always passenger waiting to go to the airport, also some are waiting to be driven in the neighborhood.
- In "Porte d'Orleans" nobody want to use a cab and people always take buses. The taxicab can choose to go back to any other places, some negative gain is counted because of gas expense.

Figure 1: The Taxicab driver

We are wondering what can be the optimal gain of the taxi in $k$ steps, depending on its initial position $i$, we denote this maximal gain by $v_i(k)$.

One can deduce the value by recurrence :

$$v_i(0) = 0 \text{ and } v_i(k) = \max_{j=1,2,3} \{\texttt{Gain from } i \texttt{ to } j + v_j(k-1)\}$$

We are in particular interested by the asymptotic gain per drive, starting from initial position $i$ :

$$\chi_i = \lim_{k \to +\infty} \frac{v_i(k)}{k}$$

**A large class of system** found in decision process and computer science can be written with similar equation. Examples are given in the last section.

## 1.2 $(\max, +)$ algebra

When we look at operation used in the writing of the taxi driver problem, operators max and + are only used. When looking at the two operations max and + one can see that there is a distributive property,

$$\text{for } a, b, c \text{ three real numbers, } \max(a, b) + c = \max(a + c, b + c)$$

This holds in the case where $a, b, c$ can take value $-\infty$ as well.

It then makes sense to define the structure $\mathbb{R}_{\max} = (\mathbb{R} \cup \{-\infty\}, \max, +)$. This is NOT a ring, as the "addition part" given by $(\mathbb{R} \cup \{-\infty\}, \max)$ is far from being a group. In fact The operation "addition" is similar to a truncation. This feature differs totally from conventionnal algebra, creating a different behavior of linear system.

What can we say about properties of $\mathbb{R}_{\max}$ ?

- the "addition" operation max denoted by $\oplus$ is associative, commutative and there exists a neutral element ($-\infty$ that we now denote by $\varepsilon$).
- the "multiplication" operation + denoted by $\otimes$ is associative, with a unit (0 denoted by $e$),
- $\otimes$ distributes over $\oplus$ and $\varepsilon$ is absorbing.

This gives all the properties needed to be a *Semi Ring*. We also observed that the considered Semi Ring is here *commutative*.

Note also that $\mathbb{R}_{\max} \setminus \{\varepsilon\}$ provided with operation $\otimes$ is a group, making $(\mathbb{R}_{\max}, \oplus, \otimes)$ a *Semi field*.

Once the Semi-Ring $\mathbb{R}_{\max}$ is defined, we can consider the Semi-Ring $\mathbb{R}_{\max}^{n,n}$ made of the matrices $n \times n$ on $\mathbb{R}_{\max}$, provided with addition, multiplication by a scalar, and multiplication of matrices. Note that this Semi Ring is no more commutative. and that multiplication by non null scalar can be inverted.

## 1.3 Using these structures to write our problem

We can write explicitly the product of $(\max, +)$ matrices :

$$(AB)_{i,j} = \bigoplus_{l=1..n} A_{i,l} \otimes B_{l,j} = \max_{l=1..n} (A_{i,l} + B_{l,j})$$

Consider the matrix $A$ (elements of $\mathbb{R}_{\max}^{n,n}$) given by : "`is the gain for the taxicab in going from` $i$ `to` $j$." — i.e. "$A_{i,j}$ is the gain for the taxicab in going from $i$ to $j$."

**Convention** In the case where no path in the graph is leading from $i$ to $j$, the coefficients is set to $\varepsilon = -\infty$, this is a good convention as a path with gain $\varepsilon = -\infty$ will never be chosen by the system.

then considering the vector $V(k)$ as "$V_i(k)$ `is the optimal gain starting from` $i$ `in k steps`". We can write our previous evolution equation as :

$$V(0) = \begin{pmatrix} e \\ \vdots \\ e \end{pmatrix} \quad \text{and} \quad V(k+1) = AV(k)$$

$$\text{all that can be written } V(k) = A^k e \ \text{ for } k = 0, 1, 2, \ldots$$

**Some Interpretation in term of paths in the graph :**

- Power of the matrix $A$ can be referred to the graph :

  $A^k_{i,j}$ is the optimal gain going from $i$ to $j$ in a path with length $k$.

- Trace can be referred to loop in the graph :

  $Tr(A^k) = \bigoplus_i A^k_{i,i} = \max_i A^k_{i,i}$ is the optimal gain of a loop of length $k$.

**Notation Warning :** We are no more writing the $\otimes$ operation in the $(\max, +)$ algebra, as it is usual for conventionnal algebra to write product implicitly. It is then necessary for each writing to know whether we place ourselves in $(\max, +)$ or in normal algebra. A symbol max or + indicates conventionnal algebra, $\oplus$ or nothing at all indicates $(\max, +)$ algebra.
Also we will use for $\lambda$ a real number, and $V$ a vector, the notation

$$\lambda V = \lambda + V = \begin{pmatrix} \lambda + V_1 \\ \vdots \\ \lambda + V_n \end{pmatrix}$$

(Note the first expression is written in $(\max, +)$, the second and third are in normal algebra).

**Order and compatibility :** Note that in $\mathbb{R}_{\max}$, the multiplication is compatible with the natural order on real number.

$$\text{for any } c \text{ in } \mathbb{R}_{\max} : \quad a \leq b \Rightarrow ca \leq cb$$

Similarly in $\mathbb{R}_{\max}^{n,n}$, multiplication by any matrix preserves the partial order composant wise on vector.

$$\text{for any } (\max, +) \text{ matrix } A : \quad V \leq W \ (i.e.\ V_i \leq W_i \text{ for all } i) \Rightarrow AV \leq AW$$

# 2 Asymptotics of $(\max, +)$ matrices

In this section, we are presenting one method, based on spectral results, inspired by the Perron Frobenius theory of non negative matrices, to compute the asymptotic gain per step previously introduced.

$$\chi = \lim_{k \to +\infty} \frac{V(k)}{k} \text{ (written in } (\max, +) \text{ as } \lim_{k \to +\infty} (A^k e)^{\frac{1}{k}})$$

## 2.1 Spectral properties give asymptotic behavior

**FACT 1 :** A GOOD EIGENVECTOR GIVES THE ASYMPTOTIC GAIN PER STEP.

First we need to precise what do we mean by *good eigenvector*, that is a vector $V$ in $\mathbb{R}_{\max}$ that is verifying the two following properties :

- $V$ is eigenvector for the matrix : On this element, the multiplication by the matrix reduces to a scalar multiplication : $AV = \lambda V$.

  $$\text{or written in normal algebra : } \max_{j=1..n} (A_{i,j} + V_j) = \lambda + V_i \text{ for all } i$$

- $V$ does not have a null composant in $(\max, +)$. As the new zero is $\varepsilon = -\infty$, it means that all composants of $V$ are finite.

**Proof :** Imagine we are provided with a good eigenvector $V$ for the matrix $A$, with eigenvalue $\lambda$, it is then possible to have the comparison, written in $(\max, +)$ :

$$\mu V \leq e \leq \nu V \text{ where } \mu \text{ and } \nu \text{ are two real numbers}$$

It then follows, as matrix multiplication preserves comparison on vector.

$$\lambda \mu V \leq Ae \leq \lambda \nu V, \text{ and after } k \text{ iterations : } \lambda^k \mu V \leq V(k) \leq \lambda^k \mu V$$

Writing the last comparison in normal algebra, dividing it by $k$, and looking at $k \to +\infty$, we then observe that the asymptotic per unit of time is equal to the eigenvalue, whatever is the initial position.

## 2.2 Star method, Spectral diameter and Spectral result

We need now to extract from the matrix an eigenvector that respect conditions given. One method to find an eigenvector is given by the iterated matrix of $A$. We will extract conditions for suitability of this method, and go back to this case by homothety.

### 2.2.1 The star of a matrix

We are considering the matrix $A^*$ given by $A^*_{i,j}$ is the optimal gain from $i$ to $j$ over a path leading from $i$ to $j$ in any number of steps $k = 0, 1, 2, \ldots$.

The matrix can be explicited in $(\max, +)$ as : $A^* = Id \oplus A \oplus A^k \oplus \ldots = \bigoplus_{k \geq 0} A^k$

In particular we have formally the equation :

$$Id \oplus AA^* = A^*$$

Such that a column of the iterated matrices $(A^*)_{.,j}$ gives an eigenvector for eigenvalue $e = 0$, provided that the two following conditions are verified :

(i) The matrices $A^*$ can be well defined, with coefficients in $\mathbb{R}_{\max}$, i.e. none of them is $+\infty$.

(ii) The operation $Id \oplus .$ does not have an impact, at least for one column of the matrix $AA^*$.

We will now provide conditions to be in this situation, we have :

- (i) holds iff all loop in the graph have non positive gain.
- (ii) holds iff one loop with positive length has a gain equal to $e$.

**Proof :** About the condition (i),

It is clear that if one loop has a positive gain, $A^*_{i,i}$ will be infinite if $i$ belongs to this loop (as one path can be made using repetitively an arbitrary number of time this loop). There is no hope to have (i) without this condition.

If no loop has a positive gain, as any path with length greater or equal than $n$ contains $n+1$ summits. One is here twice, and so a part of the path is a loop, its gain being non positive, we can cancel the loop and end with a path whose gain is greater or equal to the initial one. Then the series $\bigoplus_{k\geq 0} A^k$ defining $A^*$ does not evolve after index $n-1$. Its infinite sum is then always well defined and finite.

About the condition (ii)

Remember that $\oplus$ is a max operation, it then has no impact on the greater element : $a \leq b \iff a \oplus b = b$. Condition (ii) can be then interpreted as a comparison on the columns $(Id)_{.,j} \leq (AA^*)_{.,j}$.

This comparison is always true on coefficients of the column that do not correpond to $j$, as for $i \neq j$, $(Id)_{i,j} = \varepsilon$ which is the smaller element possible. The comparison then reduces to comparison of the "diagonal" coefficient : $e \leq (AA^*)_{j,j}$, than we just need to interpret.

$AA^* = A \oplus A^2 \oplus \ldots$ so that $(AA^*)_{i,j}$ is just the optimal gain for a path going from $i$ to $j$ in a number of step greater or equal to one (no zero-path allowed). We have then proved the announced fact.

### 2.2.2 The Spectral diameters

Now that we have manipulate $(\max, +)$ matrices and their coefficients, we will extract one quantity to describe the matrix, that will write very compactly conditions already considered.

let's define the spectral diameters $\rho(A)$ of a $(\max, +)$ matrix $A$ by :

$$\rho(A) = \max_{k\geq 1} \max_{i_1, i_2, \ldots, i_k} \frac{A_{i_1,i_2} + \ldots + A_{i_k-1,i_k} + A_{i_k,i_1}}{k} = \bigoplus_{k\geq 1} (tr(A^k))^{\frac{1}{k}}$$

**Interpretation in terms of path** spectral diameter is the maximum over all loop, with positive length, of the total gain produced by the loop and divided by its length.

**Rewriting the condition (i) and (ii) :** Thanks to the effort made in last subsection, we can now easily rewrite precedent conditions.

- (i) holds iff $\rho(A) \leq e$.
- (ii) holds iff $\rho(A) \geq e$.

Conclusion is that we can provide any matrix $A$ with an eigenvector for eigenvalue $e$, provided that $\rho(A) = e$.

### 2.2.3 Extracting an eigen vector

For any matrix $A$, if $\rho(A)$ is non null (i.e. non equal to $\varepsilon$), we can write its associated normalized matrix :

$$\tilde{A} = (\rho(A))^{-1}A \text{ (or written in normal algebra : } \tilde{A}_{i,k} = A_{i,j} - \rho(A))$$

This matrix has a spectral diameters equal to $\rho(\tilde{A}) = (\rho(A))^{-1}\rho(A) = e$, as a very simple consequence of the definition of $\rho(A)$ is that we have $\rho(\lambda A) = \lambda\rho(A)$ for any real number $\lambda$. Then $\tilde{A}$ has an eigenvector $V$ for eigenvalue $e$, which is then an eigenvector for matrix $A$ with associated eigenvalue $\rho(A)$. This leads to the following ...

**FACT 2 :** A $(\max, +)$ MATRIX $A$, WITH $\rho(A) \neq -\infty$, ALWAYS ADMITS AN EIGENVECTOR FOR THE VALUE $\rho(A)$.

**remark** The case where $\rho(A) = \varepsilon$ does not create a real difficulty. In this case, there is no loop with non null gain (i.e. gain not equal to $-\infty$), meaning that the system always leads to a dead lock in finite time. All composant of $V(k)$ will be equal to $\varepsilon$ after a finite number of iterations.

### 2.2.4 Perron Frobenius irreductibility

We need now to make sure that the eigenvector is a *good eigenvector* and has no null composant. With similarity with the Perron Frobenius theory of irreductibility, one can see that the following conditions suffices :

**Irreductible matrices** are one where summits of the associated graph are strongly connected (i.e. A path links any summit of the graph to any other one).

**remark** One should keep here in mind the convention that everypath are supposed to have gain non equal to $\varepsilon$. A gain equal to $\varepsilon$ is precisely how we represent the absence of path.

$(\max, +)$ **Writing :** $A$ is irreductible iff for all $i$ and $j$, there exists $k \geq 0$ such that $A^k_{i,j} \neq \varepsilon$. It can be shown to be equivalent to $(Id \oplus A)^{n-1}$ has no null coefficient.

**FACT 3 :** AN EIGENVECTOR FOR AN IRREDUCTIBLE MATRIX HAS NO NULL COMPOSANT.

**Proof** Let $V$ be an eigenvector, a non null composant $V_j$ can then be found. From $AV = \lambda V$ we deduce for any $i$

$$\lambda V_i = \bigoplus_k A_{i,k} V_k \geq A_{i,j} V_j$$

such that for any $i$ leading to $j$ in a path of length 1, $V_i \neq \varepsilon$ as $A_{i,j}$ is non null.

We can iterate this methods, showing that any $i$ leading to $j$ will satisfy $V_i \neq \varepsilon$. Irreductibility allows us to conclude.

This conclude the proof that for an irreductible matrices $n \times n$ on $\mathbb{R}_{\max}$, the asymptotic increase of the gain does not depend on the initial position and is equal to $\chi = \rho(A)$, provided that this is not $-\infty$.

**In particular** for the example we have studied before, the asymptotic gain per unit of time would be 11.

—

# A few more step in the new linear structure.

Method presented here in this paper has been chosen to illustrate new method provided by this formalism. Conclusion on the simple case we have presented may not be very spectacular, we would like to stress a few points to the reader. First is that on many other similar system the answer may not be as easy to find intuitively.

## Study of $(\max, +)$ linear can be continued

Here is two results that follows from same type of study, not made here.

- Study can be refined to deal with reductibility of the matrix.
- It is possible to extract from the eigenvector the optimal strategy for the type of decision process described in the example.
- Introducing new conditions on the graph, or equivalently on the $(\max, +)$ matrix associated with the system, it is possible to show a finite time convergence to a periodic behavior of the system.

## A large class of system fits into the $(\max, +)$ framework

- As we have seen, Markov decision process are $(\max, +)$ linear, for optimization of gain as well as for minimisation of cost (with a few changes in structures' definition, all formal methods can be used in the algebra $\mathbb{R}_{\min} = (\mathbb{R} \cup \{+\infty\}, \min, +)$).
- More generally, the behavior of any timed event graph can be written in this new algebra :

A timed event graphs is a particular Petri nets, where there is at most one transition leading to any particular place, and any place leads to at most one transition. Each place can contain a certain number of token. The computation of the graph is as any Petri nets : A transition $i$ is "executed" as soon as there is one at least one token in every preceding place, and it first takes one token from of these places. Then for each place $p$ whose preceding transition is $i$, one token is created after a delay $D_p$.

Introducing $x_j(k)$ the date where start the k_execution of transition $j$, we have :

$$x_j(k) = \max_i(x_t(k - l_{p(i,j)}) + D_{p(i,j)})$$

where $p(i,j)$ is the place with preceding transition $i$, and that leads to $j$. The case where there can be two places for $p(i,j)$ is not more general for our purpose, if there is no places for $p(i,j)$, we do not consider $i$ in the max, or equivalently we choose the convention $D_{p(i,j)} = -\infty$.

Most of the times, one can then deduce a relation with the following shape :

$$X(k) = DX(k-1)$$

where $D$ is made of delay in the places and $X(1), X(2), \ldots$ from execution dates of transition. "Asymptotic gain per step" in MDP can be reinterpreted here as the inverse of an asymptotic average speed of transition.

In particular reliable communication protocol as found in packet transmission network have been written in this framework.

- Finally undistinguishable ressources arrangment for task making, as written in heaps of pieces is $(\max, +)$ linear. The most famous example of this type of system is the Tetris game.

## Fluctuation in $(\max, +)$ linear systems applied to reliability

Another extension of the analytical method is the introduction of random into the system. Here we briefly describe the case of reliable communication.

Taking the delay for places in the timed event graph to be random value, we then capture in the model fluctuations of queuing delays in the network due to congestion and other connections sharing the router. We assume independence - delays are independent, with same distribution, over the time - we are reduced to a product of sequence of $(\max, +)$ i.i.d. random matrices, case where $(\max, +)$ random matrices limit theorem [SynchLinear] can be used. We are then allowed to conclude that $\lim ||A_n A_{n-1} \ldots A_1||/n$ exists and is a deterministic value (where the norm is given by the highest coefficients in the matrix), rspresenting the inverse of an asymptotic troughput.

One important fact to observe is that the "corresponding deterministic value" - which is the troughput given by a model where every delay is supposed not to fluctuate and to be equal to its mean - is always optimistic. This is a clear consequence from the inegality true for each random values $X$ and $Y$,

$$\mathbb{E}[\max(X, Y)] \geq \max(\mathbb{E}[X], \mathbb{E}[Y])$$

which is most of the time a strict inegality (e.g. when $X$ and $Y$ are independent and none of them is a constant, or if they have the same law, and are not equal).

The deterministic value is a reference, always optimistic, that eludes a large part of the problem. We treat a simple example. Imagine a matrix $A$ whose coefficients are all independent exponential random value with parameter 1, and the vector V whose coefficients are all $e = 0$ : The "correspondent deterministic matrix" for $A$, is made of coefficients all equal to 1, and applying it to the vector $V$ gives you a vector made of coefficients 1. However the coefficients of the random vector $AV$ are the maximum of $n$ independent exponential random variables with same parameter 1, such that they depends on the size of the matrix $A$. In particular the expectation of each coefficients grows asymptotically with $n$ as a logarithm.

Similary the performance of a reliable multicast communication, where packets are sent from a source to $N$ receivers in a tree, will show a dependence in the size of the group - as well as the tree topology - owing to fluctuation of the delay, and this will be true even if all receivers suffer statistically the same delay.

Impact of fluctuations on (max,+) systems are a in a unique sense, and shows phenomena that cannot be seen by a deterministic model.

## A starting point to new linearity

Two examples of frameworks extended axioms of the (max,+) algebra :

**Dioïd** From the axioms of Semi Ring, and the idemptotency ($a \oplus a = a$) (structures known as *Dioïd* ). One can define a partial order given by $a \leq b$ if $a \oplus b = b$. Iterated and ordering reasonment can be made in this structure, inspired by the case of $(\max, +)$ linear system. Asymptotic in quantum physics and some topic of language theory have been shown to factorize into this type of structures.

**Monotone Homogeneous transformation** This starts with a new interpretation of $(\max, +)$ linearity :

- $f(x \oplus y) = f(x) \oplus f(y)$ : in particular

  $$\text{if } x \leq y,\ x \oplus y = y, \text{ such that } f(y) = f(x) \oplus f(y) \Rightarrow f(x) \leq f(y)$$

  Thus compatibility with $\oplus$ can be interpreted as a **monotony**.

- $f(\lambda x) = \lambda f(x)$ can be written on vector as

  $$f \begin{pmatrix} \lambda + x_1 \\ \vdots \\ \lambda + x_n \end{pmatrix} = \lambda + f(x)$$

  meaning that a translation on all the composant by $\lambda$ is transparent to the transformation $f$. This can be interpreted as an **homogeneity**.

This can be chosen as two axioms taken on a transformation, starting point of a new study. Actually, inspired by the case of $(\max, +)$ linear system that in particular respect these two conditions, it has been possible to show analytically extension of result from spectral theory, using a Collatz Wielandt type criteria, and notion of subeigenvalue, supereigenvalue.

Monotone Homogeneous transformations allows to consider a larger class of system, as $(\min, \max)$ function found in game theory.

### Bibliographical notes

For a more complete presentation of this domain, see the article [MethAppl] by Stephane Gaubert, available with other introductive text on $(\max, +)$ at the URL : `http://amadeus.inria.fr/gaubert/` Also a $(\max +)$ toolbox for Scilab, written by INRIA France, can be found from this webpage that contains lot of references and article.

About $(\max, +)$ applied to discrete event systems both deterministic and stochastic, [SynchLinear] can be consulted, with both technical part as well introductive chapters.

About idemptotency analysis, and study of Hamilton-Jacobi equation, and quasi clasic asymptotic of quantum physic : [IdemAnalysis], chapter 13 in [OperMeth].

About $(\max, +)$ and network engineering : A $(\max, +)$ writing of the TCP protocol has been written, and a study of it can be found in [TCPMax+]. A $(\max, +)$ study of reliable multicast congestion control may be found in [MultiCCMax+]. The network calculus group of epfl uses $(\min, +)$ algebra to give guarantee on network performance, [NetCal], [NetCalGuar], the URL: `http://ica1www.epfl.ch/PS_files/NetCal.htm` provides information on their work, as well as articles.

## References


[MethAppl] Stephane Gaubert, Methods and Applications of $(\max, +)$ Linear Algebra, STACS'97, LNCS 1200, Springer.

[SynchLinear] F. Baccelli, G. Cohen, G.J. Oldster and J.P. Quadrat, Synchronization and Linearity, Wyley, 1992.

[IdemAnalysis] V. Maslov and S. Samborskii, editors. Idemptotent analysis, volume 13 of Adv. in Sov. Math. AMS,RI 1992.

[OperMeth] V.P. Maslov, Operational Methods. Mir Moscow, 1973.

[TCPMax+] TCP is max-plus linear and what it tells us on its throughput, F. Baccelli and D. Hong, ACM Sigcomm 2000, Stockholm.

[MultiCCMax+] Impact of Network delay variation on multicast session with TCP like congestion control, A. Chaintreau, F. Baccelli and C. Diot, IEEE Infocom 2001, Anchorage.

[NetCal] J.Y. Le Boudec and P. Thiran, Network Calculus, Springer (2050).

[NetCalGuar] Application of Network Calculus To Guaranteed Service Networks, JY Le Boudec, IEEE Trans on Information theory, (44) 3, May 1998 (introductory material).