# From max-plus algebra to nonexpansive mappings: a nonlinear theory for discrete event systems

Jeremy Gunawardena
BRIMS, Hewlett-Packard Labs, Filton Road, Stoke Gifford, Bristol BS34 8QZ, UK
Tel : +44 117 922 8216
Fax : +44 117 922 9190
Email : jhcg@hplb.hpl.hp.com

March 7, 1999

**Abstract**

Discrete event systems provide a useful abstraction for modelling a wide variety of systems: digital circuits, communication networks, manufacturing plants, etc. Their dynamics—stability, equilibrium states, cyclical behaviour, asymptotic average delays—are of vital importance to system designers. However, in marked contrast to continuous dynamical systems, there has been little systematic mathematical theory that designers can draw upon. In this paper we survey the development of such a theory, based on the dynamics of maps which are nonexpansive in the $\ell^\infty$ norm. This has its origins in linear algebra over the max-plus semiring but extends to a nonlinear theory that encompasses a variety of problems arising in other mathematical disciplines. We concentrate on the mathematical aspects and set out several open problems.

**Keywords:** cycle time, discrete event system, fixed point, max-plus semiring, nonexpansive map, topical function.

## 1 Introduction

In this paper we shall study functions $F : \mathbb{R}^n \to \mathbb{R}^n$ which are homogeneous

- $\forall x \in \mathbb{R}^n, \quad \forall h \in \mathbb{R}, \quad F(x_1 + h, \cdots, x_n + h) = (F_1(x) + h, \cdots, F_n(x) + h)$ H

and monotonic with respect to the usual product ordering on $\mathbb{R}^n$

- $\forall x, y \in \mathbb{R}^n, \quad x \leq y \;\Rightarrow\; F(x) \leq F(y)\,.$ M

As we shall see, such functions must necessarily be nonexpansive in the $\ell^\infty$ norm

- $\forall x, y \in \mathbb{R}^n, \quad \|F(x) - F(y)\| \leq \|x - y\|$ N

where $\|x\| = \max_{1 \leq i \leq n} |x_i|$. Functions of this kind have arisen recently in several contexts, [2, 30, 33, 37, 47], and we follow Gunawardena and Keane in calling them topical functions. We shall be concerned with their dynamics: with the behaviour of the functions under iteration and with such related questions as the existence of fixed points and the behaviour of trajectories $F^k(x)$ as $k \to \infty$.

Topical functions encompass maps and operators arising in a remarkable variety of mathematical disciplines: matrices over the max-plus semiring, nonnegative matrices of classical Perron-Frobenius theory (after suitable transformation), Leontieff substitution systems of mathematical economics, dynamic programming operators of games and of Markov decision processes, nonlinear operators arising in matrix scaling problems and demographic modelling, renormalisation operators associated to diffusions on fractals, etc (see §2).

The motivation for developing a theory of topical functions comes, in part, from the problem of modelling discrete event systems. These are best defined informally as systems comprising a finite set of events each of which can occur repeatedly. This is a convenient abstraction by which to study a variety of systems arising in real life: digital circuits, in which the events might be the voltage transitions on the wires in the circuit; communication networks, in which the events might be the arrival of packets at nodes in the network; manufacturing systems, in which the events might be the completion of a job at a machine. In designing such systems, engineers have to grapple with dynamical questions: the existence of equilibria or cyclical behaviour, whether or not the system is stable, how fast or slow the system is operating and what it might do "in the long term". However, in marked contrast to continuous dynamical systems, there has been little systematic mathematical theory that designers can draw upon.

To see why topical functions might be relevant to answering such questions, consider the following scenario. Choose some ordering of the events in the discrete event system, so that each event is associated to one of the numbers $\{1, \cdots, n\}$. Let $x_i$ denote the time of occurrence of event $i$, relative to some arbitrarily chosen origin of time. Suppose that the time evolution of the system is such that, for some function $F : \mathbb{R}^n \to \mathbb{R}^n$, the time of next occurrence of event $i$ is given by $F_i(x)$. In this case, the evolution of the system is captured by the dynamics of the function $F$, which conceals within itself the details of the system.

Given this scenario, what properties should be expected of $F$? First, since the origin of time is irrelevant, the times of occurrences of all events can all be changed by the same amount. This is exactly the property of homogeneity. Second, it is not unreasonable to ask that, if the times of occurrences of some events are delayed then this cannot cause any event to occur more quickly. This corresponds exactly to monotonicity with respect to the product ordering on $\mathbb{R}^n$. In this scenario, topical functions arise very naturally.

There are discrete event systems of practical importance which can be modelled even in this restricted manner. In particular, matrices over the max-plus semiring have provided a linear algebraic foundation for an important class of discrete event systems (see §2.1.1). Topical functions may be seen as a nonlinear generalisation of this. However, there are obvious limitations to the scenario above. It does not allow for nondeterminism—from any given state $x \in \mathbb{R}^n$, the system evolves to one and only one state, $F(x)$—nor for stochastic uncertainty in the system description. The theory of a single topical function can be broadened to accommodate both of these. Nondeterminism can be modelled in terms of the semigroup generated by a set of topical functions: $\{F_a \mid a \in A\}$. The state of the system can then change to any of the states $F_a(x)$. This is exactly the way in which nondeterminism is modelled in automata theory. Stochastic behaviour can be modelled by using random topical functions: the single topical function $F$ is replaced by a random variable from some suitable measure space into the space of topical functions.

Both of these directions lie outside the scope of the present paper and are discussed further in [29]. This paper concerns itself with the theory of a single topical function. This provides the foundation for all broader applications and already presents challenging unsolved problems.

A more serious difficulty with the scenario above is the assumption of monotonicity. This is not as compelling as that of homogeneity: it is easy to construct systems in which monotonicity is not satisfied. Nevertheless it is a convenient assumption which holds for many systems of practical interest. Glasserman and Yao have used a form of monotonicity as the foundation for their treatment of discrete event systems in [21]. In our context, it is crucially related to the property of nonexpansiveness, as shown by the following result of Crandall and Tartar.

**Proposition 1.1** ([12, Proposition 2]) *If $F : \mathbb{R}^n \to \mathbb{R}^n$ satisfies* $\mathsf{H}$ *then it satisfies* $\mathsf{M}$ *if, and only if, it satisfies* $\mathsf{N}$.

A proof is given in §2. Nonexpansiveness lies at the heart of the present paper. It implies that all trajectories are asymptotically equivalent (see (13)). We shall use this to define functionals which are independent of the starting conditions or the trajectory taken (Lemmas 3.1 and 4.1). We shall further show that these functionals encode much information about the dynamical behaviour of topical functions, in particular about the existence of fixed points, or equilibria, (Corollary 3.1 and Theorem 4.2).

The usual notion of a fixed point, $F(x) = x$, is inappropriate for discrete event systems. In the scenario above the events would have to occur infinitely fast! Definition 3.2 will allow for the possibility that there is some $h \in \mathbb{R}$ such that $F_i(x) = x_i + h$ for all $0 \leq i \leq n$. This is mathematically appropriate in the light of property $\mathsf{H}$ and is a reasonable model of equilibrium for a discrete event system: each event occurs at the same rate. The number $h$ amounts to an additive eigenvalue and will be recovered through the cycle time vector, one of the functionals mentioned above.

The results of §3 and §4 suggest a new perspective on the study of fixed points of nonexpansive functions. We recall that for a contraction, for which $\|F(x) - F(y)\| \leq \alpha\|x - y\|$ with $0 < \alpha < 1$, the Banach Contraction Principle tells us that there is a unique fixed point to which all trajectories converge at an exponential rate, [22, Chapter 2]. When the function is only nonexpansive, so that $\alpha = 1$, the existence of fixed points is a classical problem of functional analysis. Work on this has developed in two main directions. One, arising out of the work of Browder and others in the 1960s, seeks geometric conditions (usually convexity properties) on the ambient Banach space which imply that every nonexpansive function has a fixed point, [22]. The other, arising originally from attempts to extend the Brouwer fixed point theorem, has sought properties of the function (such as the vanishing of the Leray-Schauder degree) which imply the existence of a fixed point, [6].

The present paper shows that dynamical properties, in the form of averages over trajectories, also give information on the existence of fixed points (see Corollary 3.1). It remains unclear, at present, to what extent this is special to the theory of topical functions and the $\ell^\infty$ norm or is part of a broader approach to the classical fixed point problem for nonexpansive functions.

The $\ell^\infty$ norm is well known to have singular properties. It is a polyhedral norm: the unit ball is a polyhedron and hence a combinatorial object. This discreteness is particularly appropriate to the study of discrete event systems. It also limits the cyclical behaviour that a system can have. A point $x \in \mathbb{R}^n$ is said to be a periodic point of $F$ with period $p$ if $F^p(x) = x$ and $F^k(x) \neq x$ for all $0 < k < p$. The following result is due independently to Sine and Nussbaum.

**Theorem 1.1** ([38, 45]) *There exists $M \in \mathbb{N}$, depending only on the dimension $n$, such that if $F : \mathbb{R}^n \to \mathbb{R}^n$ is any function satisfying* $\mathsf{N}$ *and $p$ is the period of any periodic point of $F$, then $p \leq M$.*

This should be contrasted with the case of the $\ell^2$ (Euclidean) norm. Even when $n = 2$ there are rotations (which, being isometries, are necessarily nonexpansive) having periodic points with arbitrarily large periods. Nussbaum has conjectured that $M$ can be taken to be $2^n$, which would be best possible. This remains an important open problem and a more detailed discussion of the literature surrounding it can be found in [36]. For topical functions, it is conjectured that the best upper bound on the period is the binomial coefficient ${}^nC_{[n/2]}$, [31].

We were originally led to study topical functions because they arose naturally from attempts to model certain discrete event systems, [26]. However, as pointed out above, it is not the case that all discrete event systems, or even most discrete event systems, can be modelled in this way.

Attempts to study discrete event systems are haunted by the bewildering complexity of real engineering practice. So many additional features must be incorporated to describe specific engineering situations that mathematical generality is all too often lost. While this may still have value for specific problems, it sacrifices the ultimate goal of identifying general theorems that engineers can use in their daily work. We take a different approach here. The mathematical understanding of the dynamics of discrete event systems is extremely incomplete. The strategic need to improve this situation outweighs, in the long term, the tactical gains to be had in modelling individual systems. In the work described here we limit our attention to some simple assumptions, which nevertheless occur in practice, and try to answer the kinds of questions that confront engineers.

The next section introduces the main examples that will be studied in this paper: the affine hierarchy, and, at the other extreme, nonnegative matrices. §3 introduces the cycle time vector and sketches the proof of the main result on the affine hierarchy, Theorem 3.2. §4 speculates on how this result can be extended to nonnegative matrices and general topical functions.

The results discussed here draw on the work of the author and several collaborators, in particular Stéphane Gaubert, Michael Keane and Colin Sparrow. It is a pleasure to thank them as well as the many others who have contributed to the development of this area. The support of the European Commission through the research network ALAPEDES (Algebraic Approach to Performance Evaluation of Discrete Event Systems) is also gratefully acknowledged.

## 2 Examples and applications

We begin with some notation and then give a short proof of Proposition 1.1. We then exhibit a series of examples, summarising along the way the role that some of them play in the applications to discrete event systems. These examples provide the raw material for the discussions in §3 and §4.

We use throughout this paper the following vector-scalar convention: if, in a binary operation or a relation, a vector and a scalar appear together, the corresponding operation is applied to, or the corresponding relation is taken to hold, on each component of the vector. For instance, if $x \in \mathbb{R}^n$ and $h \in \mathbb{R}$, then $x + h$ will denote the vector $(x_1 + h, \cdots, x_n + h)$. This allows us to restate the property of homogeneity more succinctly as follows:

$$F(x + h) = F(x) + h\,.$$

Similarly, $x \leq h$, means $x_i \leq h$ for all $1 \leq i \leq n$. The symbol $h$ always stands for a real number.

The standard partial order on $\mathbb{R}$ is denoted by $a \leq b$. We use infix operators for the lattice operations of least upper bound and greatest lower bound: $a \vee b = \mathrm{lub}(a, b)$ and $a \wedge b = \mathrm{glb}(a, b)$. The same notations are used for partially ordered sets derived from $\mathbb{R}$, such as the function space $X \to \mathbb{R}$, where the partial order is taken pointwise: $f \leq g$ if, and only if, $f(x) \leq g(x)$ for all $x \in X$. If $\mathbb{R}^n$ is identified with $\{1, \cdots, n\} \to \mathbb{R}$, then this corresponds to the usual product ordering on $\mathbb{R}^n$.

We assume that $+$ always has higher binding than either $\vee$ or $\wedge$. Hence, $x_1 + 2 \vee x_2 - 1 = (x_1 + 2) \vee (x_2 - 1)$.

It is helpful to pick out the following functions: $\mathsf{t}, \mathsf{b} : \mathbb{R}^n \to \mathbb{R}$ (top and bottom, respectively), where

$$\begin{array}{rcl} \mathsf{t}(x) & = & x_1 \vee \cdots \vee x_n \\ \mathsf{b}(x) & = & x_1 \wedge \cdots \wedge x_n\,. \end{array}$$

Note that $\|x\| = \mathsf{t}(x) \vee -\mathsf{b}(x)$ and that both $\mathsf{t}$ and $\mathsf{b}$ satisfy the properties $\mathsf{H}$ and $\mathsf{M}$. In particular, addition distributes over both $\vee$ and $\wedge$: $\mathsf{t}(x+h) = \mathsf{t}(x)+h$ and $\mathsf{b}(x+h) = \mathsf{b}(x)+h$.

We make extensive use of the vector-scalar convention in the following argument, which follows that given in [30, §1].

**Proof** (of Proposition 1.1): Let $F : \mathbb{R}^n \to \mathbb{R}^n$ satisfy $\mathsf{H}$. Assume first that $F$ also satisfies $\mathsf{M}$. Choose $x, y \in \mathbb{R}^n$. Since $y + \mathsf{b}(x-y) \leq x \leq y + \mathsf{t}(x-y)$ it follows by $\mathsf{M}$ and $\mathsf{H}$ that $F(y) + \mathsf{b}(x-y) \leq F(x) \leq F(y) + \mathsf{t}(x-y)$. Hence,

$$\mathsf{b}(x-y) \leq F(x) - F(y) \leq \mathsf{t}(x-y) \tag{1}$$

and so $\|F(x) - F(y)\| \leq \|x - y\|$ as required.

Now assume that $F$ satisfies $\mathsf{N}$. It is simplest to show first that (1) must hold. So choose $x, y \in \mathbb{R}^n$. Let $h = \mathsf{t}(y-x)$ and $z = x+h$. Note that $y \leq z$ so that $\|z - y\| = \mathsf{t}(z-y)$. By $\mathsf{H}$ and $\mathsf{N}$, $F(x) - F(y) = F(z) - F(y) - h \leq \mathsf{t}(z-y) - h = \mathsf{t}(x-y)$. Since $\mathsf{t}(x) = -\mathsf{b}(-x)$ and $x, y$ were chosen arbitrarily, it is easy to see that the other inequality, $\mathsf{b}(x-y) \leq F(x) - F(y)$, must also hold, thereby showing (1).

Now suppose that $x \leq y$. Then $0 \leq \mathsf{b}(y-x) \leq F(y) - F(x)$ and so $F(x) \leq F(y)$. This completes the proof.

□

As the proof shows, rather more is true that is stated in the result of Crandall and Tartar; see [30, Proposition 1.1] for a more detailed account.

We can now begin to explore the space of topical functions. Let $\mathsf{Top}(n,n)$ denote the set of topical functions in dimension $n$. It is easy to construct simple examples, such as coordinate substitutions.

**Definition 2.1** *A function $F : \mathbb{R}^n \to \mathbb{R}^n$ is said to be simple if each component $F_i$ has the form $F_i(x) = x_j$ for some $1 \leq j \leq n$. The set of simple functions in dimension $n$ is denoted $\mathsf{Sim}(n,n)$.*

Simple functions do not have to be permutations: the same $x_j$ may be used for different $x_i$. To put it another way, the matrix corresponding to this linear function has a 1 in each row but not necessarily in each column.

It is also easy to see that a number of operators preserve the properties $\mathsf{H}$ and $\mathsf{M}$.

**Proposition 2.1** *Suppose that $F, G \in \mathsf{Top}(n,n)$. Choose $u \in \mathbb{R}^n$ and choose $\lambda, \mu \in \mathbb{R}$ such that $0 \leq \lambda, \mu$, $\lambda + \mu = 1$. Then $F + u$, $F \vee G$, $F \wedge G$, $\lambda F + \mu G$, $FG \in \mathsf{Top}(n,n)$.*

We now have a number of ways to construct classes of topical functions by starting with simple functions and closing under some sequence of operators from Proposition 2.1. It is helpful to have some notation for this, which we take from [19]. Let $\mathcal{A}$ denote the following set of operator symbols, $\mathcal{A} = \{\max, \min, +, \mathbb{E}\}$. If $S \subseteq \mathsf{Top}(n,n)$ then define the following constructions which, by Proposition 2.1, all yield further subsets of $\mathsf{Top}(n,n)$.

$$\max(S) \quad = \quad \{ \bigvee_{F \in A} F \mid A \subseteq S, \text{ A finite} \}$$

$$
\begin{aligned}
\min(S) &= \{\bigwedge_{F\in A} F \mid A \subseteq S,\ \text{A finite}\} \\
\mathbb{E}(S) &= \{\sum_{F\in A} \lambda_F F \mid A \subseteq S,\ \text{A finite},\ 0 \le \lambda_F,\ \textstyle\sum_{F\in A} \lambda_F = 1\} \\
+(S) &= \{F + u \mid F \in S,\ u \in \mathbb{R}^n\}
\end{aligned}
$$

If $\alpha_1, \cdots, \alpha_p$ is a sequence of operators symbols with $\alpha_i \in \mathcal{A}$, then the notation $(\alpha_1, \cdots, \alpha_p)$ will denote the subset $\alpha_1(\alpha_2(\cdots(\mathsf{Sim}(n,n))\cdots)) \subseteq \mathsf{Top}(n,n)$.

## 2.1 The affine hierarchy

We now examine the simpler classes of topical functions arising from Proposition 2.1, with specific examples mostly drawn from dimension 2. We then show how these can be viewed as elements of a hierarchy.

### 2.1.1 $(\max, +)$: matrices over the max-plus semiring

$$
\begin{aligned}
F_1(x_1,x_2) &= x_1 + 0.2 \vee x_2 - 1 \\
F_2(x_1,x_2) &= x_1 + 1.6
\end{aligned}
\tag{2}
$$

The standard properties of $\vee$ can be used to reduce all elements of $(\max, +)$ to the general form exemplified above. Consider now the following trick, whose origins go back at least as far as Cuninghame-Green, [13] (see also [14] for historical references). Adjoin the element $-\infty$ to $\mathbb{R}$ and redefine the operations of addition and multiplication to be maximum and addition, respectively. Note, as mentioned above, that addition distributes over maximum and that, furthermore, $-\infty$ acts as a zero for maximum. The resulting semiring is called the max-plus semiring and denoted $\mathbb{R}_{\max}$. It is easy to see that (2) can now be rewritten as a matrix equation over $\mathbb{R}_{\max}$: $F(x) = Ax$, where $x$ is now a column vector and $A$ is the max-plus matrix

$$
A = \begin{pmatrix} 0.2 & -1 \\ 1.6 & -\infty \end{pmatrix} . \tag{3}
$$

More precisely, there is a one-to-one correspondence between elements of $(\max, +)$ in dimension $n$ and $n \times n$ matrices over $\mathbb{R}_{\max}$ satisfying the following non-degeneracy condition:

$$
\forall 1 \le i \le n,\ \ \exists 1 \le j \le n,\ \ \text{such that}\ \ A_{ij} \ne -\infty\,. \tag{4}
$$

In the remainder of this paper we shall sometimes use max-plus notation (so that customary symbols or abbreviations will have their max-plus meanings) and sometimes ordinary notation (customary symbols have their customary meanings); the context will make clear which interpretation is intended.

A great deal is now understood about the spectral theory—the theory of eigenvectors and eigenvalues—of matrices over $\mathbb{R}_{\max}$. Part of the impetus for studying this has come from the realisation that eigenvalues of max-plus matrices give performance measures for discrete event systems. This is the most highly developed area of application and we outline some of this material here with references to the literature.

If $A$ is an $n \times n$ matrix over $\mathbb{R}_{\max}$, then $x \in (\mathbb{R}_{\max})^n$ is said to be an eigenvector of $A$ with eigenvalue $\lambda \in \mathbb{R}_{\max}$, if $Ax = \lambda x$. For purposes of illustration, assume that $A$ satisfies (4), so that $A : \mathbb{R}^n \to \mathbb{R}^n$, and restrict attention to real eigenvectors: eigenvectors lying in $\mathbb{R}^n$. Suppose that $x, y \in \mathbb{R}^n$ are two such real eigenvectors of $A$ with eigenvalues $\lambda$ and $\mu$ respectively. By matrix multiplication we see that $A^k x = \lambda^k x$ and $A^k y = \mu^k y$. By property $\mathsf{N}$, however, $\|A^k x - A^k y\| \leq \|x - y\|$. Rewriting this in ordinary notation, we see that $\|(k\lambda + x) - (k\mu + y)\| \leq \|x - y\|$ for any $k$. Dividing throughout by $k$ and letting $k \to \infty$, we see that $\lambda = \mu$. Any two real eigenvectors have the same eigenvalue.

Use of the nonexpansiveness property in this manner—to show that certain functionals are well defined and independent of initial conditions—will recur throughout this paper (see Lemma 3.1 and Lemma 4.1). Similar results hold for the other eigenvalues, corresponding to eigenvectors in $(\mathbb{R}_{\max})^n$, but different arguments are needed. The full spectral theory, applicable to arbitrary eigenvectors, is described by Wende *et al* in [48] and independently in Gaubert's thesis, [18].

What can be said about the eigenvalue corresponding to a real eigenvector? Unlike classical Perron-Frobenius theory, in which only bounds are known for the eigenvalues of a nonnegative matrix, [35, Chapter 2], there are formulae for the eigenvalues of a max-plus matrix. These emerge from the close relationship between matrices and graphs.

The precedence graph of $A$, denoted $\mathcal{G}(A)$, is the directed graph with labelled edges which has nodes $\{1, \cdots, n\}$ and an edge from $j$ to $i$ if, and only if, $A_{ij} \neq -\infty$. The label on this edge is then the real number $A_{ij}$. (The opposite convention for the direction of edges is sometimes used.) The existence of an edge from $j$ to $i$ is denoted $i \leftarrow j$. A path in this graph has the usual meaning of a chain of directed edges: a path from $i_m$ to $i_1$ is a sequence of nodes $i_1, \cdots, i_m$ such that $1 < m$ and $i_j \leftarrow i_{j+1}$ for $1 \leq j < m$. A circuit is a path which starts and ends at the same node: $i_1 = i_m$. A circuit is elementary if the nodes $i_1, \cdots, i_{m-1}$ are all distinct. The matrix $A$ is said to be irreducible if $\mathcal{G}(A)$ is strongly connected: there is a path connecting any pair of distinct nodes. Equivalently, $A$ is irreducible if there is no permutation of the rows and columns which brings it into upper triangular block form. The matrix of example (3) is irreducible.

The weight of a path $p$, $|p|_{\mathsf{w}}$, is the product in $\mathbb{R}_{\max}$ of the labels on the edges in the path, or, in ordinary notation:

$$|p|_{\mathsf{w}} = \sum_{j=1}^{m-1} A_{i_j i_{j+1}} .$$

Matrix multiplication has the following interpretation in terms of path weights: $A^s_{ij}$ is the maximum weight among all paths of length $s$ from $j$ to $i$. Hence problems of optimal path finding in graphs can be treated by methods of matrix algebra over $\mathbb{R}_{\max}$, [3, 9, 23].

The length of a path, $|p|_\ell$, is the number of edges in the path: $|p|_\ell = m - 1$. If $g$ is a circuit, its cycle mean, denoted $\mathsf{m}(g)$, is defined, in ordinary notation, by $\mathsf{m}(g) = |g|_{\mathsf{w}}/|g|_\ell$. Let $\mu(A) \in \mathbb{R}_{\max}$ denote the maximum cycle mean:

$$\mu(A) = \max\{\mathsf{m}(g) \mid g \text{ a circuit}\} . \tag{5}$$

This is well defined: by virtue of (4), $\mathcal{G}(A)$ has at least one circuit, and although it therefore has infinitely many, it is easy to see that only the elementary ones are needed to determine $\mu(A)$. For the matrix of example (3), $\mu(A) = 0.3$.

**Proposition 2.2** ([1, Theorem 3.23]) *Let $A$ be any $n \times n$ matrix over $\mathbb{R}_{\max}$. The eigenvalue of any real eigenvector is $\mu(A)$ and this is the largest eigenvalue of $A$. Furthermore, if $A$ is irreducible, it has a real eigenvector.*

$\mu(A)$ is sometimes called the spectral radius or Perron root of $A$, through the close analogy between Proposition 2.2 and the first Perron-Frobenius theorem for nonnegative matrices. Proposition 2.2 is one of the basic results of max-plus spectral theory and has been rediscovered so many times that it is hard to ascribe priority to any particular source. The stated reference is to one of the standard texts in the subject.

Max-plus matrices can be used to describe the time evolution of discrete event systems in which the timing constraints are all maximum ones. These are the timed versions of systems in which the causal relationships between events are represented by a partially ordered set: the AND causality between events being transformed naturally into a maximum timing constraint. A well-known model with this property is that of event graphs (sometimes called marked graphs). These are Petri nets in which each place is the input to at most one transition and the output to at most one transition. This leads to the observation that *"timed event graphs are max-plus linear systems"*, [1, Theorem 2.58]: the evolution of a timed event graph can be described by a linear equation over $\mathbb{R}_{\max}$ of the form

$$x(k) = A_0 x(k) + A_1 x(k-1) + \cdots + A_s x(k-s)\,.$$

The vector $x(k) \in \mathbb{R}^n$ describes the times at which the $n$ transitions in the event graph fire for the $k$-th time, as in the scenario described in the Introduction. This approach and its consequences are described further in [1, 11, 24].

Because of the widespread importance of event graphs and related models, many special cases and *ad hoc* results about their timing behaviour have appeared in the literature, [8, 16, 40, 41, 42]. More recently, max-plus matrix methods have been used more directly: by Ferrari and Montanari in developing cost calculi for communicating processes, [17], and by Hulgaard *et al* in studying the time separation of events problem, [32]. A recurring theme in some of this work is the construction of performance measures which turn out to be nothing other than the spectral radius, $\mu(A)$, of the underlying max-plus matrix. We shall recover this performance measure for general topical functions in the guise of the cycle time vector of §3.

The max-plus semiring is an example of a dioid, or idempotent semiring: a semiring in which addition satisfies the idempotent law, $a + a = a$. Another example well known to computer scientists is the dioid of formal languages over an alphabet $A$, in which addition corresponds to union of languages and multiplication to concatenation. The subject of idempotency, which encompasses both such examples, is discussed further in [25].

### 2.1.2 $(\min, +)$: matrices over the min-plus semiring

This is dual to the case of $(\max, +)$. The min-plus semiring, $\mathbb{R}_{\min}$, is the set $\mathbb{R} \cup \{+\infty\}$ with addition and multiplication defined as minimum and addition, respectively. The map $x \to -x$ establishes the duality between $\mathbb{R}_{\max}$ and $\mathbb{R}_{\min}$ and is also an isomorphism of idempotent semirings.

### 2.1.3 $(\min, \max, +)$: min-max functions

$$
\begin{array}{rcl}
F_1(x_1, x_2) & = & (x_1 - 2 \vee x_2 + 1) \wedge (x_1 - 3 \vee x_2 + 3) \wedge (x_1 - 4 \vee x_2 + 2) \\
F_2(x_1, x_2) & = & (x_1 - 2 \wedge x_2 + 1) \vee (x_1 \wedge x_2 - 2)
\end{array}
\tag{6}
$$

The properties of $\vee$ and $\wedge$ can be used to reduce any element of $(\min, \max, +)$ to the general form above, where there is no finite bound on the number of terms which may appear. Topical functions of this kind are called min-max functions. Special cases of them were considered by Olsder in [39] while min-max functions themselves were introduced in [28].

Min-max functions appear in analysing the timing behaviour of digital circuits. For instance, Sakallah, Mudge and Olukotun (SMO) developed a model for analysing circuits containing storage latches controlled by a central clock, [44]. In such circuits, the incoming signal at a latch must arrive and stabilise within a certain setup time in order for the signal to be correctly stored when the latch opens. The opening and closing of the latches is controlled through a clocking schedule, which allows overlapping of latch operation so as to optimise performance. Once a clocking schedule has been chosen, it must be verified to meet the setup constraints.

The SMO model can be formulated as follows, [46]. Assume that the latches are numbered from 1 to $n$ and that $j \to i$ is the "fans out to" relation on latches. That is, $j \to i$, if, and only if, there is a path of combinational logic from the output of $j$ to the input of $i$. Define the min-max functions $D, d : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ by the following equations:

$$
\begin{array}{rcll}
D_i(x) & = & (\vee_{j \to i}\{x_j + \Lambda_{j,i}\}) \wedge (x_{n+1} + B_i) & \text{for } 1 \leq i \leq n \\
D_{n+1}(x) & = & x_{n+1} & \\
d_i(x) & = & (\wedge_{j \to i}\{x_j + \lambda_{j,i}\}) \wedge (x_{n+1} + B_i) & \text{for } 1 \leq i \leq n \\
d_{n+1}(x) & = & x_{n+1} &
\end{array}
$$

where $\Lambda_{i,j}$, $\lambda_{i,j}$ and $B_i$ are constants defined by the clocking schedule and the minimum and maximum delays through the combinational logic, [46, Figure 2]. $x_{n+1}$ is a dummy variable whose only purpose is to make it clear that $D$ and $d$ are genuine min-max functions. While $d$ is min only, $D$ is min-max.

It can be shown that if $D(x) = x$ then $x_i - x_{n+1}$ is the latest signal departure time from latch $i$, [46, Figure 2]. Similarly, if $d(x) = x$ then $x_i - x_{n+1}$ is the earliest signal departure time from latch $i$. If these fixed points can be found and the arrival times can be shown to satisfy the setup constraints, then the clock schedule is verified. Hence the problem of clock schedule verification can be reduced to that of finding a fixed point for a min-max function. See [26] for further discussion.

Another example of the use of min-max functions is provided by the work of Hulgaard, Burns, Amon and Borriello, [32]. If $t_a^k$ denotes the time of $k$-th occurrence of event $a$ in a discrete event system, then the time separation of events problem asks for bounds on the separation, $t_a^k - t_b^l$, between the $k$-th occurrence of $a$ and the $l$-th occurrence of $b$. The data for this kind of problem consists not of individual delays but instead of delay bounds, $[u, v]$, where $0 \leq u \leq v < +\infty$. This is commonplace in hardware systems, as in the clock schedule verification problem above, where manufacturers provide minimum and maximum propagation delays in component specifications. The presence of both minimum and maximum constraints leads naturally to a min-max formalism and Hulgaard *et al* make use of an algebra of min-max functions to calculate exact bounds for the time separation problem.

### 2.1.4 $(\mathbb{E})$: row stochastic matrices

Suppose that $\{a, b, c, d\} \subseteq [0, 1]$ and that $a + b = 1$ and $c + d = 1$. It is easy to verify that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = a\left[c\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} + d\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right] + b\left[c\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + d\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}\right]$$

A glance at this should convince the reader of the following elementary observation, which is amusingly similar to Birkhoff's famous result on the structure of doubly stochastic matrices, [35, §5.3].

**Lemma 2.1** *A matrix is row stochastic if, and only if, it is a finite convex combination of substitution matrices.*

It follows that $(\mathbb{E})$ in dimension $n$ is exactly the set of $n \times n$ row stochastic matrices.

### 2.1.5 $(\min, +, \mathbb{E})$: Bellman operators of Markov decision processes

Let $U$ be a finite index set and, for each $u \in U$, let $c^u \in \mathbb{R}^n$ and $A^u$ be a $n \times n$ row stochastic matrix. In the light of Lemma 2.1, the function

$$F(x) = \bigwedge_{u \in U} c^u + A^u x \tag{7}$$

gives the general form of an element in $(\min, +, \mathbb{E})$. Functions of this kind are well known in optimal control as the dynamic programming (Bellman) operators associated to Markov decision processes. $U$ indexes the set of possible decisions. The choice of decision influences the transition probabilities, $A^u$, of moving between the $n$ states of the process. Each decision comes with an associated state-dependent cost, $c^u \in \mathbb{R}^n$. $F_i(x)$ is the minimal expected cost of a single process step starting from state $i$, given that $x_j$ is the cost of ending in state $j$. By Bellman's optimality principle, $F^k(x)$ gives the corresponding minimal expected cost associated to $k$ process steps. A fixed point of $F$ corresponds to an optimal cost vector, whose existence and calculation are important problems in decision theory, [5].

This completes our survey of the simpler examples arising from $\mathsf{Sim}(n, n)$ through repeated use of Proposition 2.1. On the face of it, there are many more complex functions arising from longer sequences of operator symbols. It is helpful to have a linear hierarchy in which all these functions eventually appear. To this end, define the sequence $\mathcal{A}^0 \subseteq \mathcal{A}^1 \subseteq \cdots \subseteq \mathsf{Top}(n, n)$ as follows.

$$\begin{aligned} \mathcal{A}^0 &= \mathsf{Sim}(n, n) \\ \mathcal{A}^i &= \bigcup_{\alpha \in \mathcal{A}} \alpha(\mathcal{A}^{i-1}) \text{ for } 1 \leq i \\ \mathcal{A}^* &= \bigcup_i \mathcal{A}^i \end{aligned}$$

The following result shows that this hierarchy collapses sooner than one might think and that many of the functions in $\mathcal{A}^*$ have already arisen in the examples above. The proof is straightforward but tedious and is left to the reader and [19, §2.2].

**Proposition 2.3** ([19, Proposition 12]) $\mathcal{A}^* = \mathcal{A}^4 = (\min, \max, +, \mathbb{E})$.

Functions in $\mathcal{A}^*$ consist of a finite number of affine pieces. Hence the phrase "affine hierarchy" to describe $\mathcal{A}^*$. This property plays an important role in the proof of Theorem 3.2.

We now turn to a quite different class of topical functions, which may be thought of as lying at the opposite extreme to the affine hierarchy within $\mathsf{Top}(n, n)$.

### 2.2 Functions on the positive cone

Let $\mathbb{R}^+$ denote the positive reals. Let $\exp : \mathbb{R}^n \to (\mathbb{R}^+)^n$ and $\log : (\mathbb{R}^+)^n \to \mathbb{R}^n$ be defined componentwise: $\exp(x)_i = \exp(x_i)$ and $\log(x)_i = \log(x_i)$. These establish a bijective correspondence between $\mathbb{R}^n$ and $(\mathbb{R}^+)^n$. This correspondence is an isometry from $\mathbb{R}^n$ with the $\ell^\infty$ norm to $(\mathbb{R}^+)^n$ equipped with Thompson's metric, [37, Proposition 1.6]. Let $A : (\mathbb{R}^+)^n \to (\mathbb{R}^+)^n$ be any function on the positive cone and let $\mathcal{E}(A) : \mathbb{R}^n \to \mathbb{R}^n$ denote the function $\log(A(\exp))$. The functional $\mathcal{E}$ transports functions on the positive cone bijectively to functions on $\mathbb{R}^n$. It is easy to see that $\mathcal{E}(AB) = \mathcal{E}(A)\mathcal{E}(B)$. In particular, $\mathcal{E}(A^k) = \mathcal{E}(A)^k$ so that the dynamic behaviour of $A$ and $\mathcal{E}(A)$ are equivalent and interchangeable and the dynamics may be studied either on $\mathbb{R}^n$ or on $(\mathbb{R}^+)^n$.

If $x, y \in (\mathbb{R}^+)^n$, then $x \leq y$ will denote the product ordering on $(\mathbb{R}^+)^n$. (This coincides with the cone ordering, [37, Chapter 1].) The properties of monotonicity and homogeneity have obvious counterparts on $(\mathbb{R}^+)^n$. Let $A : (\mathbb{R}^+)^n \to (\mathbb{R}^+)^n$. It follows easily from the well-known properties of exp and log that $\mathcal{E}(A)$ satisfies $\mathsf{H}$ if, and only if, $A$ satisfies

$$\forall x \in (\mathbb{R}^+)^n,\ \forall \lambda \in \mathbb{R}^+, \quad A(\lambda x) = \lambda A(x)$$

and that $\mathcal{E}(A)$ satisfies $\mathsf{M}$ if, and only if, $A$ satisfies

$$\forall x, y \in (\mathbb{R}^+)^n, \quad x \leq y \ \Rightarrow \ A(x) \leq A(y)\,.$$

Nonexpansiveness in the $\ell^\infty$ norm in $\mathbb{R}^n$ corresponds to nonexpansiveness with respect to Thompson's metric in $(\mathbb{R}^+)^n$ but this will not be needed here.

This correspondence between functions on $\mathbb{R}^n$ and functions on $(\mathbb{R}^+)^n$ immediately suggests a fresh source of topical functions.

#### 2.2.1 Nonnegative matrices

Suppose that $A : (\mathbb{R}^+)^n \to (\mathbb{R}^+)^n$ is represented by a nonnegative matrix with respect to the standard basis. A given nonnegative matrix will define a function on the positive cone if, and only if, it satisfies a similar non-degeneracy condition to (4):

$$\forall i,\ \exists j, \text{ such that } A_{ij} \neq 0\,. \tag{8}$$

It follows from the discussion above that $\mathcal{E}(A)$ is a topical function. For instance, the matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 0 \\ 5 & 0 & 6 \end{pmatrix}$$

becomes the topical function

$$\begin{aligned} \mathcal{E}(A)_1 &= \log(\exp(x_1) + 2\exp(x_2) + 3\exp(x_3)) \\ \mathcal{E}(A)_2 &= \log(4) + x_2 \\ \mathcal{E}(A)_3 &= \log(5\exp(x_1) + 6\exp(x_3)) \,. \end{aligned}$$

Although $A$ is affine on $(\mathbb{R}^+)^n$, $\mathcal{E}(A)$ is very far from affine on $\mathbb{R}^n$. Lemma 2.1 can now be used to create new families of topical functions and this process once again yields examples that have been studied extensively in other fields. The Leontieff substitution systems of mathematical economics, for instance, arise as maxima of sets of nonnegative matrices, [7].

This completes our discussion of examples of topical functions. There are many others (see, for instance, Theorem 3.1 below) and a more detailed map of the geography of topical space can be found in [31].

## 3 Cycle times and fixed points

The examples and applications discussed in the previous section reveal two related general questions. First, if a topical function represents the time evolution of a discrete event system then how can the performance of the system be measured? Second, under what circumstances does a topical function have a fixed point? The main result of this section is Theorem 3.2 which gives a detailed answer to these questions for functions in $\mathcal{A}^*$.

Suppose that $F : \mathbb{R}^n \to \mathbb{R}^n$ is a topical function which models a discrete event system under the scenario described in the Introduction. The delay between event occurrences can be measured by $F(x) - x$. If this is averaged over several occurrences, it yields

$$\frac{(F^k(x) - F^{k-1}(x)) + \cdots + (F(x) - x)}{k} = \frac{F^k(x) - x}{k}$$

which reduces asymptotically to $\lim_{k\to\infty} F^k(x)/k$, provided the limit exists. Assume for the moment that it does exist and that its value is $\chi \in \mathbb{R}^n$. This measures the asymptotic average delay between event occurrences assuming the system is started at the times $x \in \mathbb{R}^n$. What happens at other starting times? Choose $y \in \mathbb{R}^n$ and $\epsilon > 0$. We may choose $k$ so large that $\|y - x\|/k \leq \epsilon/2$ and $\|F^k(x)/k - \chi\| \leq \epsilon/2$. Then, by nonexpansiveness,

$$\|F^k(y)/k - \chi\| \leq \|(F^k(y) - F^k(x))/k\| + \|(F^k(x)/k - \chi)\| \leq \|y - x\|/k + \epsilon/2 \leq \epsilon$$

for all sufficiently large $k$. It follows that $\lim_{k\to\infty} F^k(y)/k = \chi$. We have proved

**Lemma 3.1** *Let $F \in \mathsf{Top}(n,n)$. If $\lim_{k\to\infty} F^k(x)/k$ exists for some $x \in \mathbb{R}^n$, then it exists for every $x \in \mathbb{R}^n$ and has the same value.*

**Definition 3.1** *The cycle time vector of a topical function, denoted* $\chi(F)$*, is defined to be* $\lim_{k\to\infty} F^k(x)/k \in \mathbb{R}^n$*, when this limit exists, and is undefined otherwise.*

The cycle time vector does not always exist and $\chi$ defines only a partial functional from $\mathsf{Top}(n,n)$ to $\mathbb{R}^n$. However, all the examples discussed in §2 can be shown to have cycle time vectors (see below) and examples for which it does not exist have to be constructed carefully.

**Theorem 3.1** ([30, Theorem 3.1]) *Let* $a_1, a_2, \cdots \in [0,1]$ *be any sequence of numbers drawn from the unit interval. There exists* $F \in \mathsf{Top}(3,3)$ *such that*

$$F^k(0,0,0)_2 = a_1 + \cdots + a_k \,.$$

It follows, by suitably choosing the sequence $\{a_i\}$, that there are topical functions $F$ for which $\chi(F)$ does not exist. It remains an important open problem to characterise those topical functions for which $\chi(F)$ does exist. There is not even a reasonable conjecture regarding this, although there is unpublished evidence relating the existence of the cycle time to the possibility of extending $\mathcal{E}^{-1}(F) : (\mathbb{R}^+)^n \to (\mathbb{R}^+)^n$ continuously to the boundary of the positive cone.

The cycle time vector is a performance measure for discrete event systems under the scenario described in the Introduction. It is also the appropriate generalisation of the eigenvalue to a nonlinear context. To see this, it is helpful to first introduce fixed points. As mentioned in the Introduction, it is appropriate to broaden the usual definition.

**Definition 3.2** *Let* $F \in \mathsf{Top}(n,n)$*. The vector* $x \in \mathbb{R}^n$ *is said to be a fixed point of* $F$ *if there exists* $h \in \mathbb{R}$ *such that* $F(x) = x + h$*.*

If $F$ has a fixed point $x \in \mathbb{R}^n$, then homogeneity implies that $F^k(x) = x + kh$. It follows that $\chi(F) = h$. Hence, any function that has a fixed point must also have a cycle time vector. This immediately gives examples where the cycle time vector exists and shows, moreover, that it is a natural generalisation of the eigenvalue for both max-plus matrices and nonnegative matrices. For the former, if $A$ is a $n \times n$ matrix over $\mathbb{R}_{\max}$ satisfying the nondegeneracy condition (4), then a fixed point corresponds precisely to a real eigenvector of $A$: $x \in \mathbb{R}^n$ such that, in max-plus notation, $Ax = \lambda x$. It follows that $\chi(A) = \lambda$ and we know from Proposition 2.2 that $\lambda = \mu(A)$, the maximum cycle mean of $A$ defined in (5). If $A$ is a nonnegative matrix satisfying the nondegeneracy condition (8), then a fixed point of $\mathcal{E}(A)$ corresponds precisely to a positive eigenvector of $A$: $x \in (\mathbb{R}^+)^n$ such that, in the usual notation, $Ax = \lambda x$. It follows that $\log(x)$ is a fixed point of $\mathcal{E}(A)$ and that $\chi(\mathcal{E}(A)) = \log(\lambda)$. In both cases, it can be shown that the cycle time vector always exists, even when the function does not have a fixed point, [27, Proposition 2.1], [31].

The existence of a fixed point implies a stronger constraint on the cycle time because of its vectorial nature: each component must have the same value. In the context of discrete event systems, an equilibrium can only exist if each event occurs, asymptotically on average, at the same rate. It is interesting to ask to what extent this is also a sufficient condition for the existence of a fixed point. Is it the case that $\chi(F) = h$ if, and only if, there exists $x \in \mathbb{R}^n$ such that $F(x) = x + h$? For an important class of topical functions, an even stronger result is true.

**Definition 3.3** *Let $F \in \mathsf{Top}(n, n)$. The vector $x \in \mathbb{R}^n$ is said to be a generalised fixed point of $F$ if there exists $v \in \mathbb{R}^n$ such that $F^k(x) = x + kv$ for all $k$.*

It is clear that if $F$ has a generalised fixed point, then $\chi(F) = v$. However, because $v$ is no longer a scalar, the iterative behaviour of $F$ at $x$ cannot be deduced from the homogeneity property. The definition of a generalised fixed point requires, in principle, an infinite amount of information about the trajectory starting from $x$.

**Theorem 3.2** ([19, Theorem 15]) *Any function in $\mathcal{A}^*$ has a generalised fixed point.*

This result has a number of useful and interesting consequences.

**Corollary 3.1** *If $F \in \mathcal{A}^*$ then $\chi(F)$ exists. Moreover, $F$ has a fixed point if, and only if, $\chi(F) = h$ for some $h \in \mathbb{R}$.*

The problem of calculating $\chi(F)$ is not solved in general. However, for the class of min-max functions, Theorem 3.2 implies the positive solution of the Duality Conjecture, [20, Corollary 2], which yields a systematic method of calculation, albeit one of exponential complexity. The problem of calculating fixed points is also unsolved in general but for min-max functions an algorithm that works well in practice is described in [10, §2.3].

The proof of Theorem 3.2 starts from the following observation, inspired by the discounting arguments used in stochastic optimal control. Let $F \in \mathsf{Top}(n, n)$. Choose $0 < \alpha < 1$ and let $F_\alpha(x) = F(\alpha x)$. Then, by nonexpansiveness,

$$\|F_\alpha(x) - F_\alpha(y)\| \leq \alpha \|x - y\|\,,$$

which shows that $F_\alpha$ is a contraction. By the Banach Contraction Principle, [22, Theorem 2.1], $F_\alpha$ has a unique fixed point, $x_\alpha \in \mathbb{R}^n$, where $F_\alpha(x_\alpha) = x_\alpha$. In other words,

$$F(\alpha x_\alpha) = x_\alpha\,. \tag{9}$$

This procedure defines a function, $\alpha \to x_\alpha : (0, 1) \to \mathbb{R}^n$. The asymptotics of $x_\alpha$ as $\alpha \to 1$ from below (denoted $\alpha \to 1^-$) reveals a great deal about the cycle time vector. For instance, suppose that there are vectors $u_{-1}, u_0 \in \mathbb{R}^n$ such that, as $\alpha \to 1^-$,

$$x_\alpha = u_{-1}(1 - \alpha)^{-1} + u_0 + o(1)\,, \tag{10}$$

corresponding to a truncated Laurent series expansion about 1. The notation is intended to mean that the remainder term $x_\alpha - u_{-1}(1 - \alpha)^{-1} - u_0$ defines a function $g : (0, 1) \to \mathbb{R}^n$ such that $g(\alpha) \to 0$ as $\alpha \to 1^-$, [15]. Using (9) and nonexpansiveness, it is not difficult to show that

$$F(\alpha(1 - \alpha)^{-1} u_{-1} + u_0) = u_{-1}(1 - \alpha)^{-1} + u_0 + o(1)\,.$$

Suppose now that we take $\alpha = 1 - 1/k$ with $k \in \mathbb{N}$. As $k \to \infty$, $\alpha \to 1^{-1}$. Then

$$F(u_0 + (k-1)u_{-1}) = u_0 + ku_{-1} + o(1) \tag{11}$$

from which it can be deduced by induction that $F^k(u_0) = u_0 + ku_{-1} + o(1)$. It follows that $\chi(F) = u_{-1}$, so that the residue of $x_\alpha$—the coefficient of $(1-\alpha)^{-1}$—is the cycle time vector of $F$.

More can be said for functions in $\mathcal{A}^*$ because of their affine structure. If $F \in \mathcal{A}^*$ then any straight line, if extended far enough in $\mathbb{R}^n$, must be mapped by $F$ onto a straight line. This implies that the remainder term in (11) must vanish:

$$F(u_0 + (k-1)u_{-1}) = u_0 + ku_{-1}\,,$$

for all sufficiently large $k$. It follows that there is some $K \in \mathbb{N}$ such that $u_0 + Ku_{-1}$ is a generalised fixed point of $F$.

The proof of Theorem 3.2 now reduces to showing that $x_\alpha$ has a truncated Laurent series expansion of the form shown in (10). For functions in the class $(+, \mathbb{E})$, which are of the form $c + A$ where $A$ is a row stochastic matrix, this follows from standard results in matrix theory. For functions formed by taking minima of maxima of these, the result follows by using the fact that Laurent series like (10) are componentwise totally ordered under the lexicographic ordering on the pair $(u_{-1}, u_0)$. The full details are given in [19].

It is too much to expect that the fixed point theorem in Corollary 3.1 holds in general. Consider, for example, the nonnegative matrix

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \tag{12}$$

An easy calculation shows that $A^k(1,1)^{\mathrm{t}} = (k+1,1)^{\mathrm{t}}$. Hence, $\chi(\mathcal{E}(A)) = 0$. However, $A$ does not have a positive eigenvector. To understand this phenomenon it is necessary to go beyond the cycle time vector. In the next section we discuss some ideas in this direction.

# 4 Lower order asymptotics

If $F \in \mathsf{Top}(n,n)$, the cycle time vector of $F$ can be defined as the vector $\chi \in \mathbb{R}^n$ such that $F^k(x) = k\chi + o(k)$ as $k \to \infty$. As above, this notation is intended to mean that $(F^k(x) - k\chi)/k \to 0$ as $k \to \infty$, [15]. The argument of Lemma 3.1 was essential here to show that $\chi$ was well defined. The implication is that all trajectories are asymptotically linear, at least up to order $o(k)$. But what can be said about the growth of $F^k$ at rates slower than $k$?

Consider functions $u : \mathbb{N} \to \mathbb{R}$ such that $u(k) \to \infty$ as $k \to \infty$. We call these rate functions. The inclusion function will be denoted by 1: $1(k) = k$. If $F \in \mathsf{Top}(n,n)$ and $x, y \in \mathbb{R}^n$, the nonexpansiveness property implies that, for any rate function, $u$,

$$F^k(x) - F^k(y) = o(u(k)) \text{ as } k \to \infty\,. \tag{13}$$

In other words, all trajectories are asymptotically equivalent, at all rates. If $u$ and $v$ are rate functions, the notation $u \succ v$ will indicate that $v(k)/u(k) \to 0$ as $k \to \infty$. This relation is clearly transitive: if $u \succ v$ and $v \succ w$ then $u \succ w$.

**Definition 4.1** *A sequence of rate functions $(u_1, u_2, \cdots, u_m)$ is said to be an asymptotic scale if $u_1 \succ u_2 \succ \cdots \succ u_m$.*

Let $F \in \mathsf{Top}(n,n)$ and choose an asymptotic scale $(1, u_1, u_2, \cdots, u_m)$. Suppose that for points $x, y \in \mathbb{R}^n$ there are vectors $\tau_0, \tau^1, \cdots, \tau^m \in \mathbb{R}^n$ and $\theta^0, \theta^1, \cdots, \theta^m \in \mathbb{R}^n$ such that, as $k \to \infty$,

$$\begin{array}{rcl} F^k(x) & = & k\tau^0 + u_1(k)\tau^1 + \cdots + u_m(k)\tau^m + o(u_m(k)) \\ F^k(y) & = & k\theta^0 + u_1(k)\theta^1 + \cdots + u_m(k)\theta^m + o(u_m(k))\,. \end{array} \tag{14}$$

**Lemma 4.1** *Under the circumstances above, $\tau^i = \theta^i$ for $0 \leq i \leq m$ and $\tau^0 = \chi(F) = \theta^0$.*

**Proof:** Let $u_0 = 1$. It follows from the transitivity of $\succ$ that the equations in (14) continue to hold with $i$ in place of $m$, for any $0 \leq i \leq m$. In particular, for $i = 0$, $F^k(x) = k\tau^0 + o(k)$ and $F^k(y) = k\theta^0 + o(k)$. It follows immediately from Lemma 3.1 that $\tau^0 = \chi(F) = \theta^0$. Furthermore, for any $0 \leq i \leq m$,

$$F^k(x) - F^k(y) = k(\tau^0 - \theta^0) + u_1(k)(\tau^1 - \theta^1) + \cdots + u_i(k)(\tau^i - \theta^i) + o(u_i(k))$$

Assume as an inductive hypothesis that $\tau^i = \theta^i$ for $0 \leq i < r$, where $0 < r \leq m$. Using (13), we see that $u_r(k)(\tau^r - \theta^r) = o(u_r(k))$. Hence, $\tau^r = \theta^r$ and the result follows by induction.

$\square$

We shall say that $F$ has cycle times with respect to the asymptotic scale $(u_1, \cdots, u_m)$ if an equation of the form in (14) holds. In this case the notation $\chi(F, u_i)$ will denote the corresponding vector $\tau^i$, which we have just shown to be uniquely determined by $F$ and the chosen asymptotic scale. The cycle time vector of Definition 3.1 corresponds to $\chi(F, 1)$. This notation is potentially misleading since the value of $\chi(F, u_i)$ depends on the chosen scale as well as on the specific rate function $u_i$. The scale should be clear from the context and we prefer to keep the notation lightweight.

The point of this additional complexity is that if $F$ has a fixed point, so that $F(x) = x + h$, then $F^k(x) = kh + o(u(k))$ for any rate function $u$. Hence, by Lemma 4.1, $F$ has cycle times for any asymptotic scale, $(1, u_1, \cdots, u_m)$, and, furthermore, $\chi(F, 1) = h$ and $\chi(F, u_i) = 0$ for all $1 \leq i \leq m$. This is a much stronger necessary condition than just requiring $\chi(F, 1) = h$. It suggests the form that a generalised fixed point theorem could take.

**Conjecture 4.1** *Let $C \subseteq \mathsf{Top}(n,n)$ be a subset of topical functions, such as the subset of nonnegative matrices. Does there exist an asymptotic scale $(1, u_1, \cdots, u_m)$ such that $F \in C$ has a fixed point if, and only if, $F$ has cycle times with respect to this scale, $\chi(F, 1) = h$ for some $h \in \mathbb{R}$ and $\chi(F, u_i) = 0$ for $1 \leq i \leq m$?*

We have allowed for the possibility that there is no single asymptotic scale that works for all topical functions. There is no *a priori* reason why scales should not be infinite or even continuous, as in the scale $(k^{1/a})$ where $1 \leq a \in \mathbb{R}^+$, but we have not allowed for this. The paragraph above is less a conjecture than a suggestion as to the form that an answer might take.

What evidence is there for Conjecture 4.1? Theorem 3.2 can be reinterpreted as saying that the scale (1) works for the class $\mathcal{A}^*$. The functions with affine structure require only the linear growth rate. More interestingly, two old results in the literature on nonnegative matrices can also be reinterpreted in this language and show that the conjecture does explain the fixed point behaviour of examples like (12).

First, Rothblum and Whittle investigated growth rates for Markov decision processes in [43]. We claim that their results can be reinterpreted as follows.

**Theorem 4.1** *If $A$ is a $n \times n$ nonnegative matrix satisfying (8) then $\mathcal{E}(A)$ has cycle times with respect to the asymptotic scale* $(1, \log)$.

In other words, as $k \to \infty$,

$$\mathcal{E}(A^k)(x) = k\chi(\mathcal{E}(A)) + \log(k)\chi(\mathcal{E}(A), \log) + o(\log(k))\,.$$

$\chi(\mathcal{E}(A))$ and $\chi(\mathcal{E}(A), \log)$ correspond to the geometric and algebraic growth rates, respectively, of [43].

Second, an old result, going back to Gantmacher in the 1950s, gives a necessary and sufficient condition for a nonnegative matrix to have a positive eigenvector, [4, Theorem 3.10]. We claim that this can be reinterpreted as follows.

**Theorem 4.2** *If $A$ is a nonnegative matrix satisfying (8) then $A$ has a positive eigenvector corresponding to its spectral radius $r$ if, and only if, $\chi(\mathcal{E}(A)) = r$ and $\chi(\mathcal{E}(A), \log) = 0$.*

Proofs of these assertions will appear in due course. It may be helpful to work out the $2 \times 2$ case, where the calculations can be easily done. Let $\{a, b, c\} \subseteq \mathbb{R}^+$. Consider the nonnegative matrix

$$B = \begin{pmatrix} a & b \\ 0 & c \end{pmatrix}.$$

(The other cases in dimension 2 are either trivial or uninteresting.) We can work out the eigenvector problem in 3 different ways. A simple calculation with the eigenvalue equation $B(x, y)^{\mathsf{t}} = \lambda(x, y)^{\mathsf{t}}$ shows that $B$ has a positive eigenvector if, and only if, $c > a$.

In the language of [4, Theorem 3.10], $B$ has two strongly connected components (classes) corresponding to the vertices 1 and 2 and $\{2\}$ is the unique final class. The spectral radius of $B$ is $c \vee a$. A class is basic if the spectral radius of the corresponding irreducible submatrix equals the spectral radius of $B$. [4, Theorem 3.10] states that $B$ has a positive eigenvector if, and only if, the basic classes coincide with the final classes. There are hence 3 possibilities

| Condition | Final classes | Basic classes | |
|---|---|---|---|
| $c > a$ | $\{2\}$ | $\{2\}$ | (15) |
| $c = a$ | $\{2\}$ | $\{1\}, \{2\}$ | |
| $c < a$ | $\{2\}$ | $\{1\}$ | |

of which only the first gives a positive eigenvector.

The asymptotics of the trajectories of $B$ can be calculated as follows. Write

$$B^k = \begin{pmatrix} a^k & r_k \\ 0 & c^k \end{pmatrix},$$

where $r_k = a^{k-1}b + r_{k-1}c$. It follows that $r_k/a^k = b/a + (r_{k-1}/a^{k-1})(c/a)$ and hence

$$r_k = \begin{cases} a^k \left(\frac{b}{a}\right)\left(\frac{1-(c/a)^k}{1-(c/a)}\right) & \text{if } a \neq c \\ a^k \left(\frac{b}{a}\right) k & \text{if } a = c \end{cases}.$$

It is now easy to check that $\mathcal{E}(B)$ has cycle times with respect to the asymptotic scale $(1, \log)$ and that their values are as follows.

| Condition | $\chi(\mathcal{E}(B))$ | $\chi(\mathcal{E}(B), \log)$ | |
|---|---|---|---|
| $c > a$ | $(\log(c), \log(c))^{\mathrm{t}}$ | $(0,0)^{\mathrm{t}}$ | |
| $c = a$ | $(\log(c), \log(c))^{\mathrm{t}}$ | $(1,0)^{\mathrm{t}}$ | (16) |
| $c < a$ | $(\log(a), \log(c))^{\mathrm{t}}$ | $(0,0)^{\mathrm{t}}$ | |

According to Theorem 4.2, $\mathcal{E}(B)$ has a fixed point if, and only if, $c > a$. Comparing (16) with (15) gives some insight into the workings of Theorems 4.1 and 4.2.

We can answer positively the question raised in Conjecture 4.1, at least for the subset of nonnegative matrices. This gives some encouragement that the ideas of asymptotic scales and cycle times are the right ones for formulating a general fixed point theorem for topical functions.

## 5 Summary

Engineers who build discrete event systems have to confront dynamical problems as a matter of course. For the most part, they have had little mathematical support to do this, despite the considerable understanding of dynamical systems arising from classical mechanics and the study of chaos. As we have seen in this paper, discrete event systems give rise to very different issues. They can lead naturally to dynamics which are nonexpansive in the $\ell^\infty$ norm. Nonexpansiveness constrains the dynamical behaviour and forces all trajectories to have the same asymptotics. It is this fundamental observation that allows performance measures like the cycle time vector to be defined.

What is interesting and unexpected is the close relationship that has emerged between cycle times and fixed points. This would not have been expected from the linear examples and there is little hint of it in Perron-Frobenius theory. The results surveyed in this paper show that the existence of fixed points can be deduced from the asymptotics of trajectories; a new dynamical insight into an old classical problem.

## References


[1] F. Baccelli, G. Cohen, G. J. Olsder, and J.-P. Quadrat. *Synchronization and Linearity.* Wiley Series in Probability and Mathematical Statistics. John Wiley, 1992.

[2] F. Baccelli and J. Mairesse. Ergodic theorems for stochastic operators and discrete event systems. Appears in [29].

[3] W. Backes. *The structure of longest paths in periodic graphs*. PhD thesis, Universität des Saarlandes, 1994.

[4] A. Berman and R. J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. Classics in Applied Mathematics. SIAM, 1994.

[5] D. P. Bertsekas and S. E. Shreve. *Stochastic Optimal Control: the Discrete Time Case*. Athena Scientific, 1996.

[6] R. F. Brown. *A Topological Introduction to Nonlinear Analysis*. Birkhäuser, 1993.

[7] E. Burmeister and R. Dobell. *Mathematical Theories of Economic Growth*. Macmillan, 1970.

[8] S. M. Burns. *Performance Analysis and Optimization of Asynchronous Circuits*. PhD thesis, California Institute of Technology, 1990.

[9] B. A. Carré. An algebra for network routing problems. *Journal of the Institute of Mathematics and its Applications*, 7:273–294, 1971.

[10] J. Cochet-Terrasson, S. Gaubert, and J. Gunawardena. A constructive fixed point theorem for min-max functions. To appear in *Dynamics and Stability of Systems*.

[11] G. Cohen, D. Dubois, J.-P. Quadrat, and M. Viot. A linear-system-theoretic view of discrete-event processes and its use for performance evaluation in manufacturing. *IEEE Transactions on Automatic Control*, 30(3):210–220, 1985.

[12] M. G. Crandall and L. Tartar. Some relations between nonexpansive and order preserving maps. *Proceedings of the AMS*, 78(3):385–390, 1980.

[13] R. A. Cuninghame-Green. Describing industrial processes with interference and approximating their steady-state behaviour. *Operational Research Quarterly*, 13(1):95–100, 1962.

[14] R. A. Cuninghame-Green. *Minimax Algebra*, volume 166 of *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, 1979.

[15] N. G. de Bruijn. *Asymptotic Methods in Analysis*. Dover, 1981.

[16] S. Even and S. Rajsbaum. Unison, canon and sluggish clocks in networks controlled by a synchronizer. *Mathematical Systems Theory*, 28:421–435, 1995.

[17] G. L. Ferrari and U. Montanari. Cost calculi for communicating processes. To appear.

[18] S. Gaubert. *Théorie des Systèmes Linéaires dans les Dioides*. PhD thesis, L'École Nationale Supérieure des Mines de Paris, 1992.

[19] S. Gaubert and J. Gunawardena. A non-linear hierarchy for discrete event dynamical systems. Proceedings of WODES'98, IEE, Cagliari, Italy, August 1998.

[20] S. Gaubert and J. Gunawardena. The duality theorem for min-max functions. *Comptes Rendus Acad. Sci.*, 326:43–48, 1998.

[21] P. Glasserman and D. D. Yao. *Monotone Structure in Discrete Event Systems.* Wiley Series in Probability and Mathematical Statistics. John Wiley, 1994.

[22] K. Goebel and W. A. Kirk. *Topics in Metric Fixed Point Theory*, volume 28 of *Cambridge Studies in Advanced Mathematics.* Cambride University Press, 1990.

[23] M. Gondran and M. Minoux. *Graphs and Algorithms.* Wiley-Interscience Series in Discrete Mathematics. John Wiley, 1984.

[24] Max-Plus Working Group. Max-plus algebra and applications to system theory and optimal control. In *Proceedings of the International Congress of Mathematicians, Zürich, 1994.* Birkhäuser, 1995.

[25] J. Gunawardena. An introduction to idempotency. Appears in [29].

[26] J. Gunawardena. Timing analysis of digital circuits and the theory of min-max functions. In *Digest of Technical Papers of the ACM International Workshop on Timing Issues in the Specification and Synthesis of Digital Systems.* ACM, 1993.

[27] J. Gunawardena. Cycle times and fixed points of min-max functions. In G. Cohen and J.-P. Quadrat, editors, *11th International Conference on Analysis and Optimization of Systems*, pages 266–272. Springer LNCIS 199, 1994.

[28] J. Gunawardena. Min-max functions. *Discrete Event Dynamic Systems*, 4:377–406, 1994.

[29] J. Gunawardena, editor. *Idempotency.* Publications of the Isaac Newton Institute. Cambridge University Press, 1998.

[30] J. Gunawardena and M. Keane. On the existence of cycle times for some nonexpansive maps. Technical Report HPL-BRIMS-95-003, Hewlett-Packard Labs, 1995.

[31] J. Gunawardena, M. Keane, and C. Sparrow. In preparation, 1998.

[32] H. Hulgaard, S. M. Burns, T. Amon, and G. Borriello. An algorithmm for exact bounds on the time separation of events in concurrent systems. *IEEE Transactions on Computers*, 44:1306–1317, 1995.

[33] V. N. Kolokoltsov. On linear, additive and homogeneous operators in idempotent analysis. Appears in [34].

[34] V. P. Maslov and S. N. Samborskiĭ, editors. *Idempotent Analysis*, volume 13 of *Advances in Soviet Mathematics.* American Mathematical Society, 1992.

[35] H. Minc. *Nonnegative Matrices.* Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley, 1988.

[36] R. D. Nussbaum. Periodic points of nonexpansive maps. Appears in [29].

[37] R. D. Nussbaum. Hilbert's projective metric and iterated nonlinear maps. *Memoirs of the AMS*, 75(391), 1990.

[38] R. D. Nussbaum. Omega limit sets of nonexpansive maps: finiteness and cardinality estimates. *Differential and Integral Equations*, 3:523–540, 1990.

[39] G. J. Olsder. Eigenvalues of dynamic max-min systems. *Discrete Event Dynamic Systems*, 1:177–207, 1991.

[40] S. Rajsbaum and M. Sidi. On the performance of synchronized programs in distributed networks with random processing times and transmission delays. *IEEE Transactions on Parallel and Distributed Systems*, 5:939–950, 1994.

[41] C. V. Ramamoorthy and G. S. Ho. Performance evaluation of asynchronous concurrent systems using Petri nets. *IEEE Transactions on Software Engineering*, SE-6(5):440–449, 1980.

[42] R. Reiter. Scheduling parallel computations. *Journal of the ACM*, 15(4):590–599, 1968.

[43] U. G. Rothblum and P. Whittle. Growth optimality for branching Markov decision chains. *Mathematics of Operations Research*, 7:582–601, 1982.

[44] K. A. Sakallah, T. N. Mudge, and O. A. Olukotun. Analysis and design of latch-controlled synchronous digital circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits*, 11(3):322–333, 1992.

[45] R. Sine. A nonlinear Perron-Frobenius theorem. *Proceedings of the AMS*, 109:331–336, 1990.

[46] T. Szymanski and N. Shenoy. Verifying clock schedules. In *Digest of Technical Papers of the IEEE International Conference on Computer-Aided Design of Integrated Circuits*, pages 124–131. IEEE Computer Society, 1992.

[47] J. M. Vincent. Some ergodic results on stochastic iterative DEDS. To appear in Journal of Discrete Event Dynamics Systems.

[48] C. Wende, Q. Xiangdong, and D. Shuhui. The eigen-problem and period analysis of the discrete event system. *Systems Science and Mathematical Sciences*, 3:243–260, 1990.