Min-Max Functions

Jeremy Gunawardena[*]
Department of Computer Science,
Stanford University, Stanford, CA 94305, USA.

Tel : +1 415 723 4638
Fax : +1 415 725 7411
Elm : jeremy@cs.stanford.edu

March 29, 1994

Abstract

A variety of problems in operations research, control theory, computer science, etc, can be modelled as discrete event systems with maximum and minimum constraints. When these systems require only maximum constraints (or, dually, only minimum constraints) they can be studied by linear methods based on max-plus algebra. Systems with mixed constraints, however, are non-linear from this perspective and relatively little is known about their behaviour. The paper lays the foundations of the theory of discrete event systems with mixed constraints. We introduce min-max functions, $F : \mathbf{R}^n \to \mathbf{R}^n$, which are constructed using finitely many operations of min, max and +, and study them as dynamical systems. Among other results, we give a complete account of the periodic behaviour of functions of dimension 2; we introduce and characterize the concept of balance which generalizes irreducibility in the linear theory; and we give a formula for the cycle time (eigenvalue) of a min-max function which generalizes the maximum cycle mean formula.

Keywords: Cycle time, discrete event system, dynamical system, eigenvalue, fixed point, max-plus algebra, min-max function.

[*]Current address: Hewlett-Packard Labs, Filton Road, Stoke Gifford, Bristol BS12 6QZ; jhcg@hplb.hpl.hp.com

1 Introduction

A variety of problems arising in operations research, control theory, computer science, etc, can be modelled in terms of the evolution of a discrete event system. A variety of mathematical techniques have been developed to deal with the dynamics of such systems, [Ho89]. In this paper we study systems whose evolution can be specified by maximum and minimum constraints and we develop new methods for analysing them.

The study of discrete event systems with only maximum constraints (or, dually, only minimum constraints) has a long and complicated history. Researchers from different fields have approached the same underlying mathematics from widely differing viewpoints, usually in complete ignorance of each other's work. A detailed account of the historical development would tax the resources of a professional historian. We content ourselves with a sketch of the main strands of work and some pointers to the main sources in the literature.

From the computer science viewpoint, the basic problem can be formulated as follows. Consider a finite directed graph, G, whose edges are annotated with real numbers. The vertices of G represent events which can occur repeatedly while the numbers on the edges represent delays. (Negative delays are mathematically feasible although the specific application may rule them out.) Suppose that the vertices are $a^1, a^2, \cdots, a^n$. If there is an edge from a^2 to a^1 which is annotated with r—which we denote $a^2 \xrightarrow{r} a^1$—then each occurrence of a^1 is constrained to wait until r time units after the previous occurrence of a^2. The time of the s-th occurrence of a^1 is hence given by a maximum constraint over all the edges which enter a^1:

$$t(a^1_s) = \max\{t(a^k_{s-1}) + r \mid a^k \xrightarrow{r} a^1\}. \tag{1}$$

(This is a least fixed point semantics in an appropriate sense; see [Gun93b, Proposition 3.1] for more details.) To start the system off, it is necessary to assume a vector of initial time values, $\vec{v} = (v_1, \cdots, v_n)$, so that $t(a^i_1) = v_i$ for each vertex a^i. These are similar to the boundary conditions in a difference or differential equation. We shall use a subscript $t_{\vec{v}}$ to indicate the dependence of the time evolution on the initial conditions. Many authors start the system off at $(0, \cdots, 0)$ and do not concern themselves with behaviour under other initial conditions; in this case we shall simply write t. The following basic result about max-only systems has been proved repeatedly (see, among others, [Rei68, §3A], [RH80, Theorem 2], [Bur90, Theorem 2.8]):

Theorem 1.1 *If G is strongly connected then the limit*

$$\lim_{s\to\infty} \frac{t(a^i_s)}{s} \tag{2}$$

exists, is independent of the vertex a^i, and is equal to the maximum cycle mean of the graph.

A cycle mean is simply the sum of the annotations on a directed cycle, divided by the number of edges in the cycle. The maximum such number over all cycles in the graph (it is sufficient to consider elementary cycles in which no vertex is repeated) is called the maximum cycle mean. Karp has given an elegant formula for computing this, [Kar78], whose time complexity is $O(n.e)$ where n is the number of vertices and e the number of edges in the graph.

The limit (2) can be interpreted as the asymptotic average time to the next occurrence of the event a^i. After s occurrences, the average time to the next occurrence is given by:

$$\frac{t(a^i_2) - t(a^i_1) + \cdots + t(a^i_s) - t(a^i_{s-1})}{s-1} = \frac{t(a^i_s) - t(a^i_1)}{s-1},$$

which tends to (2) as $s \to \infty$. It is reasonable, therefore, to refer to the limit in question as the cycle time of the event a^i. The reciprocal of the cycle time is a measure of the rate of occurrence of a^i (number of occurrences per unit time).

This strand in the history of the subject is characterized by a concentration on asymptotic averages. Since the equations (1), which govern the evolution of the system, are entirely deterministic, we might well ask for more information about the exact time behaviour of the system. Does $t(a^i_s)/s$ jitter for ever around its asymptotic value, or does it eventually settle down into some periodic pattern? We might also like to know what happens if the system is started from somewhere other than $(0, \cdots, 0)$. Does this make a difference to the cycle time? Does it make a difference to whether or not the system reaches a periodic regime? The study of the exact behaviour of max-only systems forms a parallel, or, perhaps, slowly converging strand in the historical record. It begins with the following observation: the non-linear equation (1) can be regarded as a linear equation over a new "max-plus" algebra $\mathbf{R} \cup \{-\infty\}$ in which '+' is defined as maximum and "×" as addition. Let A denote the $n \times n$ matrix in which $A_{ij} = r$ if there is an edge from a^j to a^i annotated with r and $A_{ij} = -\infty$ if there is no edge from a^j to a^i. In max-plus algebra, the evolution of the system can be described by means of the vector equation:

$$(t_{\vec{v}}(a^1_{s+1}), \cdots, t_{\vec{v}}(a^n_{s+1}))^T = A^s \vec{v}^T \qquad (s \geq 0),$$

(where vectors are written as row vectors). In other words, the evolution is captured by a transformation: $A : (\mathbf{R} \cup \{-\infty\})^n \to (\mathbf{R} \cup \{-\infty\})^n$, which can be regarded as linear in max-plus algebra. This beautiful idea opens the way to an algebraic treatment of the dynamics of max-only systems. There is a close relationship between this matrix representation and the graph representation discussed earlier. The graph G is the so-called precedence graph of the matrix A and G is strongly connected if, and only if, A is irreducible, in the usual sense of matrix theory. In this case the cycle time (2) can be identified as an eigenvalue of A, which turns out to also be the spectral radius of A, [BCOQ92, Theorem 3.23].

As Cuninghame-Green points out, [CG79, §1.1], the discovery of max-plus algebra was made independently by several people: see the references cited in [CG79, §29] for more information. It would be appropriate, however, to cite Cuninghame-Green's early papers (see [CG62] and references therein) as the starting point for this strand of the subject. They contain a clear realisation, [CG62, §4], of the emergence of a new area of research: linear algebra over max-plus. Cuninghame-Green's work originated from problems in operations research.

The max-plus algebra, $\mathbf{R} \cup \{-\infty\}$, is but one example of a dioid (sometimes, dioïd) or idempotent semiring. That is, a semiring in which the addition is idempotent: in $\mathbf{R} \cup \{-\infty\}$, $\max\{a, a\} = a$. Much of the linear theory can be carried out for a general dioid although additional axioms are necessary for certain aspects. The idempotency requirement introduces a semi-lattice structure on the ring and it is the interplay between order and algebra which gives the theory much of its character. Systematic treatments have been given by Cuninghame-Green, see [CG79] and the more recent survey article, [CG91]; Gondran and Minoux, see the survey article, [GM]; Cao, Kim and Roush, [CKR84], who deal with dioids in which the incline axiom, $ab \leq a$, holds; and Zimmermann, [Zim81].

The relevance of these ideas for discrete event systems was first recognised in performance problems arising in flexible manufacturing systems, [CDQV85]. The Max-Plus group, together with their co-workers and students, have gone on to solve many of the outstanding problems in the deterministic theory and to systematically investigate the stochastic aspects. This has culminated in their recent book, [BCOQ92]. The survey paper, [CMQV89], gives a snapshot of the deterministic theory.

Independently of this, and starting from problems in mathematical physics, economics and control theory, Maslov, Samborskiĭ and others have made deep contributions to "idempotent analysis", [MS92]. This work stems from the observation of Maslov, [Mas87], that certain non-linear differential equations that arise in mathematical physics and control theory, may be solved by superposition—usually only possible for linear equations—provided the superposition is done in an idempotent context. (Very much earlier, in 1967, Romanovskiĭ had already shown the existence of eigenvectors for endomorphisms of semimodules over idempotent semirings; see the references in [MS92].) As the name "idempotent analysis" suggests, and in contrast to most of the approaches discussed above, the Russian school have gone from linear algebra to functional analysis. (Cao, Kim and Roush, [CKR84, §4.7] discuss Hilbert spaces over an incline algebra.) It is in keeping with the traditions of the subject that the existence of the Russian school was unknown to others working on similar problems (including the author of the present paper) until very recently[1]. For instance, the relevant literature is not cited in [BCOQ92]. The collection of papers in [MS92] provides an excellent and accessible survey of the school's work.

The work described above gives a great deal of insight into the exact behaviour of max-only systems and provides a body of mathematical results applicable to a remarkably wide variety of practical problems. However, systems with mixed constraints, maxima and minima, are non-linear from this perspective and appear to be inaccessible to such methods. The final strand in our historical sketch begins with the work of Olsder on the existence of eigenvectors in certain mixed systems, [Ols91], [BCOQ92, §9.6]. These systems are of a restricted kind but can be of arbitrarily high dimension. Somewhat later, motivated by problems in timing analysis of digital circuits, [Gun93c], we gave the first proof of eventual periodicity in systems with mixed constraints, [Gun93b]. Olsder has recently studied several questions related to periodicity for the same class of systems which he considered earlier, [Ols93].

The present paper attempts to lay the foundations of the theory of discrete event systems with mixed constraints[2]. Instead of considering linear functions, $A : (\mathbf{R} \cup \{-\infty\})^n \to (\mathbf{R} \cup \{-\infty\})^n$ over max-plus algebra, we introduce min-max functions $F : \mathbf{R}^n \to \mathbf{R}^n$ in which each component $F_i : \mathbf{R}^n \to \mathbf{R}$ is built up in a specific way by using the operations of maximum, minimum and addition (see Definition 2.1 for precise details). Min-max functions provide a simple and convenient mathematical setting in which to describe systems with mixed constraints. Instead of taking an algebraic approach we study the behaviour of F as a dynamical system. That is, we study the eventual behaviour of the sequence

$$\vec{v},\ F(\vec{v}),\ F^2(\vec{v}),\ \cdots, \quad .$$

Our approach leads to a different perspective on the behaviour of systems to the linear case. As one might expect, generalization leads to a clarification of earlier results.

[1] We are deeply grateful to an anonymous reviewer for pointing it out to us.

[2] The present paper is an extended version of the technical report [Gun93a].

In the next section we give the basic definitions of conjunctive normal form, periodic point and cycle time for an arbitrary min-max function. These definitions provide the main themes for the rest of the paper. In §3 we prove the first of our main results, Theorem 3.3, which gives a necessary and sufficient condition for the existence of eventual periodicity in dimension 2. As pointed out in [BCOQ92, page 457], *"Such a property has not been shown ... though simulations do point in this direction."*. In §4 we introduce the balance condition for min-max functions in an attempt to identify a class of functions with good periodicity properties. We characterize balance in dimension 2 (Proposition 4.1) and show that for max-plus matrices, balance is a natural generalization of irreducibility (Theorem 4.1). Indeed, the main result on the spectrum of an irreducible max-plus matrix, [BCOQ92, Theorem 3.23], holds verbatim for balanced matrices (Proposition 4.2). This treatment sheds a new light on aspects of the linear theory. In §5 we study the cycle time of a min-max function and derive a formula for calculating it, Theorem 5.1. This generalizes the maximum cycle mean formula of [BCOQ92, Theorem 3.23]. Applications are discussed elsewhere, [Gun93c].

It is a pleasure to thank David Dill for introducing us to the work of Burns, [Bur90], which provided the initial motivation for this study, and for his continued interest in and encouragement of this work. Thanks are also due to Gerard Hoffman for pointing out the work of the Max-Plus group and to Nieke Tholen for ferreting out several misprints and confusions. We are also very grateful to various members of the Max-Plus group, and particularly to Geert-Jan Olsder, for kindly keeping us informed of their most recent work. The comments of three anonymous reviewers improved aspects of the presentation and clarified several obscurities. Any indiscretions that remain must be laid at the author's door. The work presented here was undertaken as part of project STETSON, a joint project between Hewlett-Packard Laboratories and Stanford University on asynchronous hardware design.

2 The basic definitions

We begin this section by introducing min-max expressions which are the components out of which min-max functions are built. After discussing some necessary technicalities we introduce min-max functions and give the definitions of periodic point and cycle time.

It will be convenient to use the infix operators $a \vee b$ and $a \wedge b$ to stand for maximum (least upper bound) and minimum (greatest lower bound) respectively: $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. It is easy to see that addition distributes over both maximum and minimum:

$$h + (a \vee b) = h + a \vee h + b, \qquad h + (a \wedge b) = h + a \wedge h + b. \tag{3}$$

In expressions like these we always assume that $+$ has higher binding than $\wedge$ or $\vee$.

Definition 2.1 *A min-max expression, f, is a term in the grammar:*

$$f := x_1, x_2, \cdots \mid f + a \mid f \wedge f \mid f \vee f$$

where $x_1, x_2, \cdots$ are variables and $a \in \mathbf{R}$ is referred to as a parameter.

The expressions $x_1 + 5 \wedge x_2 - 1$ and $(x_1 + 1 \vee x_1 + 2) \wedge x_1 + 3.14159$ are both min-max expressions. However, neither $(x_1 + x_2) \wedge x_3 + 2$ nor $x_1 \vee 2$ are legal terms in the min-max grammar. An expression which uses only $\vee$ and $+$ is a max-only expression; dually, a min-only expression uses only $\wedge$ and $+$.

A min-max expression of n variables, $f(x_1, \cdots, x_n)$, gives rise to a real-valued function, $f : \mathbf{R}^n \to \mathbf{R}$, for which we use the same notation. Note that an expression of n variables does not have to use each of the variables $x_1, \cdots, x_n$: we may consider $x_1 + 2$ as an expression of n variables for any $n \geq 1$. We refer to n as the dimension of f and this number depends on the context in which f is being used. In general, we are only concerned with f as a real-valued function, $f : \mathbf{R}^n \to \mathbf{R}$, and we shall freely use the associativity and commutativity of $\wedge$ and $\vee$, as well as (3), to simplify expressions. If we wish to emphasize the specific syntactic form of f (which is sometimes necessary) then we shall use $f \equiv g$ to indicate syntactic identity of expressions. The notation $f = g$ will always mean equality as real-valued functions: $f(\vec{x}) = g(\vec{x})$ for all $\vec{x} \in \mathbf{R}^n$. Hence, $(x_1 + 1 \vee x_1 + 2) \wedge x_1 + 3.14159 = x_1 + 2$. In like manner, the notation $f \leq g$ will mean that $f(\vec{x}) \leq g(\vec{x})$ for all $\vec{x} \in \mathbf{R}^n$. To avoid cluttering up the notation we shall write $f_1 \vee \cdots \vee f_n$ for $(f_1 \vee (f_2 \vee (\cdots (f_{n-1} \vee f_n))))$ and similarly for $\wedge$.

If f is a max-only expression of n variables, it is easy to see that it can be placed in the following form:

$$f = (a_1 + x_1 \vee \cdots \vee a_n + x_n), \tag{4}$$

where we now allow $a_i \in \mathbf{R} \cup \{-\infty\}$. A term of the form $-\infty + x_i$ merely indicates the absence of the variable x_i. Since each min-max expression must have at least one variable in it, there must exist $a_i \neq -\infty$. (When we write expressions such as (4) we shall sometimes leave out the terms with $a_i = -\infty$.) Suppose that g is another max-only expression which has also been placed in this form:

$$g = (b_1 + x_1 \vee \cdots \vee b_n + x_n),$$

Lemma 2.1 *If $f \leq g$ then $a_i \leq b_i$ for $1 \leq i \leq n$.*

Proof: If $a_i = -\infty$ then there is nothing to prove, so choose i such that $a_i \neq -\infty$. Fix arbitrary values for each variable other than x_i. By making x_i sufficiently large, we can always ensure that $f = a_i + x_i$. If $b_i = -\infty$ then, for all sufficiently large x_i, g will have a constant value. Since $f \leq g$, by hypothesis, this is clearly impossible. Hence, $b_i \neq -\infty$. Furthermore, for all sufficiently large x_i,

$$a_i + x_i = f \leq g = b_i + x_i.$$

Hence $a_i \leq b_i$.

QED

It follows directly from Lemma 2.1 that if $f = g$, then $a_i = b_i$ for $1 \leq i \leq n$. Hence $f \equiv g$. In other words, the form (4) is unique for max-only expressions. We shall refer to (4) as conjunctive form (or conjunctive normal form) for max-only expressions. The choice of the word conjunctive comes by analogy with normal forms in propositional calculus, [Gal86, Definition 3.4.7].

If f is a min-max expression, it can also be placed in conjunctive form:

$$f = f_1 \wedge \cdots \wedge f_m, \tag{5}$$

where each f_i is a max-only expression in conjunctive form. The form (5) is not unique. For instance, $1+x_1 = (1+x_1) \wedge (2+x_1)$. However, we can always remove redundant terms so that

$$i \neq j \implies f_i \not\leq f_j. \tag{6}$$

One way to do this is to note that the terms f_i form a partially ordered set under $\leq$. The minimal elements of this poset give a conjunctive form for f and also satisfy (6). Note that Lemma 2.1 provides a simple test for $f_i \leq f_j$.

Definition 2.2 *If f is a min-max expression of n variables then the expression (5) is a conjunctive normal form for f if each f_i is a max-only expression in conjunctive form and condition (6) holds.*

Theorem 2.1 *Conjunctive normal form is unique up to re-ordering of the f_i.*

Proof: The proof of this is relegated to the Appendix, in order not to delay the exposition at this point.

QED

There is a slight asymmetry between $\vee$ and $\wedge$ in (5). The number of terms, $a+x_i$, in each max-only expression, f_i, is fixed and is equal to the dimension of f. We need to allow $a = -\infty$ to indicate the absence of the corresponding term. However, the number of max-only expressions, m, is not fixed; if a max-only expression does not contribute to f, it is simply omitted. Hence we have no need for the constant $+\infty$. In the dual form, disjunctive form, which is similar to (5) but with $\vee$ and $\wedge$ interchanged, we would have $a \in \mathbf{R} \cap \{+\infty\}$ and we would no longer require $-\infty$.

Many definitions and results in this paper, such as Definition 2.2 and Theorem 2.1, have obvious duals in which the roles of maximum and minimum are interchanged and the directions of certain inequalities are reversed. Since $-(a \wedge b) = -a \vee -b$ dual results do not require separate proof. In the remainder of this paper we shall only state one version of these and will leave it to the reader to formulate the duals.

There is a simple algorithm for moving back and forth between conjunctive and disjunctive form which is useful in practice. We explain it here by working through an example. Consider the min-max expression of 2 variables,

$$f = (a + x_1 \vee b + x_2) \wedge c + x_1, \tag{7}$$

where $a, b, c \in \mathbf{R}$. This is effectively in conjunctive form but to be more precise we should write f as

$$(a + x_1 \vee b + x_2) \wedge (c + x_1 \vee -\infty + x_2).$$

To express f in disjunctive form we go back to the initial min-max expression (7) and rewrite each individual term $a_i + x_i$ in disjunctive form. This gives

$$((a + x_1 \wedge +\infty + x_2) \vee (+\infty + x_1 \wedge b + x_2)) \wedge (c + x_1 \wedge +\infty + x_2).$$

We now use the distributivity of $\wedge$ over $\vee$ to interchange the order of the two operations. This gives

$$((a \wedge c) + x_1 \wedge +\infty + x_2) \vee (c + x_1 \wedge b + x_2), \tag{8}$$

which is in disjunctive form. Be warned, however, that this algorithm does not take conjunctive normal form to disjunctive normal form. For example,

$$(1 + x_1 \vee 4 + x_2) \wedge (2 + x_1 \vee 3 + x_2)$$

is in conjunctive normal form. If we apply the algorithm above, we get the disjunctive form

$$(1 + x_1 \wedge +\infty + x_2) \vee (1 + x_1 \wedge 3 + x_2) \vee (2 + x_1 \wedge 4 + x_2) \vee (+\infty + x_1 \wedge 3 + x_2)$$

which is certainly not normal—the second min-only expression is redundant. This example should help clarify any confusion that may arise from the use of these forms.

Lemma 2.2 *If $f(x_1, \cdots, x_n)$ is a min-max expression and $h \in \mathbf{R}$ then*

$$f(x_1 + h, x_2 + h, \cdots, x_n + h) = f(x_1, x_2, \cdots, x_n) + h.$$

Proof: If $f \equiv x_j$ the result is immediate and if the result holds for f it certainly holds for $f + a$. By (3), we see that the result holds for $f_1 \wedge f_2$ or $f_1 \vee f_2$ if it holds for f_1 and f_2 separately. The result follows by structural induction.

QED

This would not be true, of course, for either $(x_1 + x_2) \wedge x_3 + 2$ or $x_1 \vee 2$. This illustrates a significant aspect of the chosen syntax for min-max expressions. Note that if $f(x_1, \cdots, x_n)$ is a min-max expression of dimension n and $g_1, \cdots, g_n$ are min-max expressions of dimension m then $f(g_1, \cdots, g_n)$ is a min-max expression of dimension m.

Definition 2.3 *A min-max function of dimension n is any function, $F : \mathbf{R}^n \to \mathbf{R}^n$, each of whose components, $F_i : \mathbf{R}^n \to \mathbf{R}$, is a min-max expression of n variables $x_1, \cdots, x_n$.*

If F and G are min-max functions of dimension n, it is easy to see that F composed with G is also a min-max function of dimension n. Hence F^k is a min-max function, for any $k \geq 0$.

Let $\vec{c}(h) = (h, \cdots, h)$ denote the constant vector each of whose components is h; the dimension of the vector should be clear from the context. The following observation is a trivial consequence of Lemma 2.2.

Lemma 2.3 *If F is a min-max function of dimension n, then*

$$F(\vec{x} + \vec{c}(h)) = F(\vec{x}) + \vec{c}(h). \tag{9}$$

The main concern of the theory of min-max functions is with the behaviour of a function as a dynamical system: with the asymptotic properties of the sequence

$$\vec{x},\ F(\vec{x}),\ F^2(\vec{x}),\ \cdots$$

for different starting points $\vec{x} \in \mathbf{R}^n$. However, the classical definitions of fixed point, periodic point, etc, [Dev89, Definition 3.2], are inappropriate in this context. It is unusual for a min-max function to have a fixed point, $F(\vec{x}) = \vec{x}$. Instead, it is more commonplace to find a point $\vec{x}$ for which there is some real number h such that $F(\vec{x}) = \vec{x} + \vec{c}(h)$. This observation motivates the following definitions. Let F be a min-max function of dimension n.

Definition 2.4 *$\vec{x}$ is a periodic point of F, of period $k \geq 1$, if, for some $h \in \mathbf{R}$, $F^k(\vec{x}) = \vec{x} + \vec{c}(h)$. The least k with this property is the prime period of $\vec{x}$. If $k = 1$ then $\vec{x}$ is a fixed point of F.*

Definition 2.5 *F is eventually periodic (EP) at $\vec{x}$, with period (respectively, prime period) k, if, for some $l \geq 0$, $F^l(\vec{x})$ is a periodic point of F of period (respectively, prime period) k.*

Suppose that F is eventually periodic at $\vec{x}$, so that $F^{l+k}(\vec{x}) = F^l(\vec{x}) + \vec{c}(h)$ for some $l \geq 0$. For $s \geq l$, we can use the Euclidean algorithm to write $s - l = i_s k + e_s$, where $0 \leq e_s < k$. By repeated application of (9), we see that $F^s(\vec{x}) = F^{l+e_s}(\vec{x}) + i_s\vec{c}(h)$ for all sufficiently large s. Since $i_s/s \to 1/k$ as $s \to \infty$, it follows that

$$\lim_{s\to\infty} \frac{F^s(\vec{x})}{s} = \vec{c}(\frac{h}{k}).$$

In view of the discussion in the Introduction, the following definition is quite natural.

Definition 2.6 *If F is eventually periodic at $\vec{x}$, as above, then the cycle time of F at $\vec{x}$, denoted by $\chi_F(\vec{x})$, is given by h/k.*

It is important to note that the cycle time exists at a point only if the function is eventually periodic at that point. The problem of how $\chi_F(\vec{x})$ varies with the choice of $\vec{x}$ will be addressed in §5. For the moment, let us note that if F is 1-dimensional then, by (5) or by (9), $F(x) = x + F(0)$ for all x. Hence, every point is a fixed point of F and $\chi_F(x) = F(0)$. The cycle time is independent of the choice of x.

With these preliminaries out of the way, we can embark on the proofs of the main results.

3 Periodicity in dimension 2

If F is a min-max function of dimension n, (9) indicates that the effective dimension of F may be reduced by 1. There are many ways in which this reduction can be accomplished; we shall concentrate on only one of them here. Let $(-)^* : \mathbf{R}^n \to \mathbf{R}^{n-1}$ denote the function which takes $(x_1, \cdots, x_n)$ to $(x_1 - x_n, \cdots, x_{n-1} - x_n)$. Define the auxiliary function $H : \mathbf{R}^{n-1} \to \mathbf{R}^{n-1}$ by $H(\vec{x}) = F(\vec{x}, 0)^*$ so that

$$H_i(x_1, \cdots, x_{n-1}) = F_i(x_1, \cdots, x_{n-1}, 0) - F_n(x_1, \cdots, x_{n-1}, 0), \quad (1 \leq i \leq n-1).$$

H can be thought of as the restriction and projection of F on the hyperplane $x_n = 0$. Because of (9), the behaviour of F can be reconstructed from H up to a scaling factor.

Lemma 3.1 *For all $s \geq 0$ and $\vec{x} \in \mathbf{R}^n$, there exists $h \in \mathbf{R}$, such that $F^s(\vec{x}) = (H^s(\vec{x}^*), 0) + \vec{c}(h)$.*

Proof: Note first that for any $\vec{y} \in \mathbf{R}^{n-1}$, $F(\vec{y}, 0) = (H(\vec{y}), 0) + \vec{c}(h)$ where $h = F_n(\vec{y}, 0)$. Now proceed by induction. For $s = 0$, $\vec{x} = (\vec{x}^*, 0) + \vec{c}(x_n)$. Assume the formula for $s \geq 0$ and apply F to it.

$$\begin{array}{rcll} F^{s+1}(\vec{x}) & = & F(H^s(\vec{x}^*), 0) + \vec{c}(h) & \text{by (9)} \\ & = & (H^{s+1}(\vec{x}^*), 0) + \vec{c}(h') & \text{by the first remark.} \end{array}$$

QED

It follows from this that $\vec{x}$ is a periodic point of F of period k if, and only if, $H^k(\vec{x}^*) = \vec{x}^*$, so that $\vec{x}^*$ is a periodic point of H in the usual sense of dynamical systems theory, [Dev89, Definition 3.2]. Furthermore, $F^l(\vec{x})$ is a periodic point of F if, and only if, $H^l(\vec{x}^*)$ is a periodic point of H. The dynamical behaviours of F and H are essentially the same and can be read off from each other.

The auxiliary function takes on a particularly simple form in dimension 2: $H(x) = F_1(x, 0) - F_2(x, 0)$ (using x in place of x_1). What can we say about such functions?

Definition 3.1 *A function* $\mathbf{R} \rightarrow \mathbf{R}$ *is said to be piecewise nice if it is everywhere continuous and its graph is composed of finitely many straight lines of slope +1, 0, or -1.*

If p is piecewise nice, we can always find a finite set of intervals,

$$(-\infty, u_1], \ [u_1, u_2], \ \cdots, \ [u_{n-1}, u_n], \ [u_n, +\infty), \tag{10}$$

such that, on any one of the intervals, either $p(x) = x + a$ or $p(x) = -x + a$ or $p(x) = a$, for some $a \in \mathbf{R}$. We do not necessarily assume that the derivative of p is discontinuous at the points u_i, although we can always choose the intervals to satisfy that condition.

Lemma 3.2 *If* p, q *are piecewise nice functions and* $a \in \mathbf{R}$ *then* $\pm x$, a, $p + a$, $p \wedge q$, $p \vee q$, $-p$ *and* $p(q)$ *(*p *composed with* q*) are all piecewise nice.*

Proof: Exercise for the reader.

QED

Corollary 3.1 *If* F *is a min-max function of dimension 2, then the auxiliary function* H *is piecewise nice.*

Proof: By structural induction using Lemma 3.2, it is easy to see that $F_1(x, 0)$ and $F_2(x, 0)$ are piecewise nice functions whose graphs have no segments of slope -1. Hence, H is also piecewise nice.

QED

There is a converse to this which we shall state without proof. It is not needed for the development which follows and its proof is quite tedious. However, it does make clear that the class of piecewise nice functions exhibits exactly the behaviours that we are interested in.

Theorem 3.1 *Every piecewise nice function is the auxiliary function of some min-max function of dimension 2.*

The next result is one of the key contributions of this paper. Since the proof is a little involved, we add a few preparatory comments so that the reader does not lose sight of the wood amidst all the trees. Consider the piecewise nice function, $p_1(x) = -x$. Iterating p_1 at any point

$x \in \mathbf{R}$, immediately results in either a 2-period, $(p_1)^2(x) = p_1(x)$, or a fixed point, $p_1(x) = x$, in the usual sense. Now consider the function $p_2(x) = -\lambda x$, where $0 < \lambda < 1$. This function behaves differently under iteration. It is easy to see that $\lim_{n\to\infty}(p_2)^n(x) = 0$ for any $x \in \mathbf{R}$ but p_2 takes infinitely long to reach its fixed point at 0. We shall show that when a piecewise nice function, p, has a fixed point, its behaviour under iteration is qualitatively similar to p_1 and not to p_2. (By qualitative we mean simply that periodic behaviour is eventual and not immediate.) We shall establish this by exhibiting a constant $\delta > 0$ such that, if p is not eventually periodic at x then, by iterating p sufficiently far, we can always continue, eventually, to move in the same direction by at least δ. This quickly yields a contradiction when p has a fixed point. The constant emerges from the finite structure of the graph of p. The fine print in the proof is a series of trivial observations which establish that δ exists and has the right properties. We hope these remarks will aid the reader in navigating through the proof. It may help to have a piece of squared graph paper to hand while working through the details.

Theorem 3.2 *Let p be a piecewise nice function. If p has a fixed point, $v = p(v)$, then, for any $x \in \mathbf{R}$, there exists $k \geq 0$ such that, $p^k(x) = p^{k+2}(x)$. Conversely, if, for some $v \in \mathbf{R}$, the set $\{p^i(v) \mid i \geq 0\}$ is finite, then p has a fixed point.*

Proof: We first establish some conventions which will be used during the proof. We shall work in the (x, y) plane where $y = p(x)$ represents the graph of the piecewise nice function p. It will be convenient to use the phrase "the point x" as a shorthand for "the point $(x, p(x))$ on the graph of p". We shall often refer to "horizontal distance" meaning by that, distance measured along the x axis. The horizontal distance between the point (c, d) and the line $y = x + a$ is $|c + a - d|$. The horizontal distance between parallel lines, say $y = -x + a$ and $y = -x + b$, is the horizontal distance between any point on one line and the other line; in this case it is given by $|a - b|$. When we speak of "left" or "right" we mean with respect to the usual orientation of the x and y axes on the page: $+x$ going to the right and $+y$ going upwards. If a is a point on the graph of p, the piecewise nice property implies that the graph must lie within or on the boundaries marked out by lines of slope +1 and -1 through a. If it is known that, say, the graph to the left of a has no segment of slope +1, then the graph is further restricted to lie in the area on or above the line $y = p(a)$. These observation will be called the "cone restrictions". We shall refer to the line $y = x$ as the main diagonal. There is a simple and helpful geometric rule for finding the point $p(a)$ (ie: $(p(a), p^2(a))$) from the point a (ie: $(a, p(a))$). Move horizontally from a until the main diagonal is reached and then move vertically until the graph of p is reached. The point of intersection is the point $p(a)$. We shall call this trick "the reflection principle".

We are now ready to embark on the first part of the proof. Divide the graph of p into three regions, the left region, where $p(x) > x$, the centre, where $p(x) = x$, and the right, where $p(x) < x$. The hypothesis of the theorem guarantees that the centre region is non-empty. If both other regions are empty then every point is a fixed point and the result holds. Hence, we may take it that either the left or the right region is non-empty. Without loss of generality assume that the left region is non-empty.

The graph of p consists of line segments of slope +1, 0 and -1. Let δ_1 denote the minimum horizontal distance between a line segment of slope +1, which is not coincident with the main diagonal, and the main diagonal itself. If no such line segments exist, then $\delta_1 = 0$. Let δ_2 denote the minimum horizontal distance between any two line segments of slope -1. If none or only one such exists then $\delta_2 = 0$.

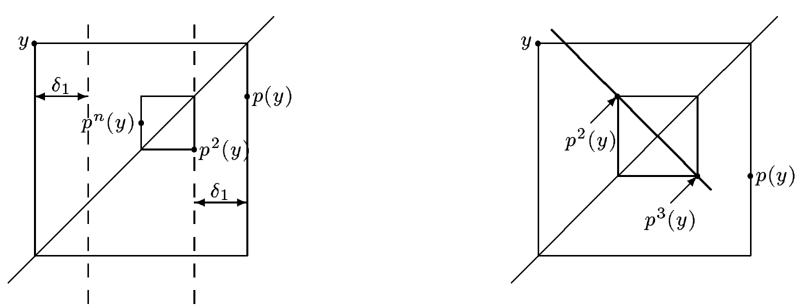
Now suppose that the conclusion of the theorem does not hold. That is, assume that, for some $y \in \mathbf{R}$, it never happens that $p^k(y) = p^{k+2}(y)$ We may assume that y is in the left region. Then $p(y)$ can either lie in the left or the right region. (It clearly cannot lie in the center, for then the conclusion of the theorem would hold.)

Assume first that $p(y)$ lies in the left region. We claim that, in this case, $\delta_1 > 0$ and $p(y)$ moves to the right by at least δ_1. For suppose this is not the case. If $\delta_1 = 0$ then it may be that there are no line segments of slope $+1$. But then the cone restrictions imply that $p(y)$ must be either in the centre or in the right region. It could also happen that there is only 1 line segment of slope $+1$ and it coincides with the main diagonal. In this case, the cone restrictions show once again that $p(y)$ is either in the center region or in the right region. By assumption, we can discount either of these possibilities. So we may take it that $\delta_1 > 0$. Now suppose that $p(y) - y < \delta_1$. Then it must be the case that there are no line segments of slope $+1$ strictly between the point y and the main diagonal, for otherwise, y would be horizontally distant by at least δ_1 from the main diagonal. But then, by the cone restrictions again, $p(y)$ must be either in the centre region or in the right region, which we can discount as before. We conclude that, if y and $p(y)$ are both in the left region, then

$$p(y) - y \geq \delta_1 > 0.$$

If the right region is empty then this inequality is all that we shall need to finish the argument. So assume now that the right region is also non-empty. Note that we may apply a similar argument to the one given above if y starts off in the right region. If y and $p(y)$ are both in the right region, then $p(y)$ has moved to the left by at least δ_1.

Now assume that $p(y)$ lies in the right region. By the reflection principle, $p(y)$ must lie on the right hand boundary of the square whose top left corner is y and whose diagonal coincides with the main diagonal. This is illustrated in the left hand picture of (11). If $p^2(y)$ also lies in the right region, then, if $p^n(y)$ ever returns to the left region for some $n > 2$, it must have moved to the right from y by at least $\delta_1 > 0$. This is because, by the argument in the previous paragraph, each $p^j(y)$, for $2 \leq j < n$, is in the right region and so keeps moving to the left by at least δ_1. Hence, when $p^n(y)$ returns to the left region, it cannot be closer than δ_1 to the right of y. (In fact, it cannot be closer than $(n-2).\delta_1$ to the right but δ_1 is adequate for our purposes.) This is also illustrated in the left hand picture of (11).



(11)

It may, however, be the case that $p^2(y)$ returns immediately to the left region. We can assume that $p^2(y) \neq y$, since this would satisfy the conclusion of the theorem. Suppose first that there is at most one line segment of slope -1 between y and $p(y)$. The cone restrictions make it clear that then there must be at least one such line segment and that it must originate at or above y and must terminate at or below $p(y)$. The reflection principle then implies that $p^2(y) = p^4(y)$. The right hand picture of (11) illustrates this situation. By hypothesis, we may rule out this possibility. Hence, there are at least two lines of slope -1 between y and $p(y)$. Hence, $\delta_2 > 0$ and moreover, $p^2(y)$ has moved to the right with respect to y by at least δ_2. It follows that, if $p^n(y)$ ever returns to the left region, then either

$$p^n(y) - p(y) \geq \delta_1 > 0, \quad \text{or,} \quad p^n(y) - p(y) \geq \delta_2 > 0.$$

A similar argument applies if y is in the right region and $p(y)$ moves into the left region.

We can now summarize what we have learned. If $\delta_1 = 0$, the points $y, p(y), p^2(y), \cdots$ alternate between the left and right regions. Furthermore, $\delta_2 \neq 0$, for otherwise the points would quickly fall into a double period. Hence, we can certainly pick an infinite subsequence a_i such that all the a_i lie in the left region and

$$a_{i+1} - a_i \geq \delta_2 > 0.$$

If $\delta_1 > 0$ then some of the points may lie in the left region and some on the right. We can, however, still pick an infinite subsequence, a_i, lying entirely in one of the regions. Suppose, without loss of generality, that this is the left region. If δ_3 is defined by

$$\delta_3 = \begin{cases} \delta_1 & \text{if } \delta_2 = 0 \\ \min(\delta_1, \delta_2) & \text{otherwise} \end{cases}$$

then we can assert that

$$a_{i+1} - a_i \geq \delta_3 > 0.$$

So far we have not utilized the hypothesis that p has a fixed point. Let v be some fixed point. The cone conditions imply that any point to the left of v is either in the centre or the left region while any point to the right of v lies in the centre or in the right region. But, for sufficiently large i, we can clearly find a point of our subsequence which lies to the right of v and must hence be either in the centre or in the right region. But the subsequence was chosen to lie entirely in the left region! This contradiction establishes the first part of the theorem.

For the second part of the theorem, we can take it that none of the $p^i(v)$ are in the centre region. If all the $p^i(v)$ are in one region, either the left or the right, then the analysis above shows that $\delta_1 = 0$. But then, as argued above, $p(v)$ must be either in the centre or in the right region. This contradiction shows that there must be points in both regions. By continuity, there must be some $v \in \mathbf{R}$ where $p(v) = v$.

QED

We hope the details of the above proof do not hide its essential simplicity. It is much easier to convince oneself that the result is true than to write down a convincing proof. We can now easily deduce our main theorem on periodic behaviour in min-max functions.

Theorem 3.3 *Let F be a min-max function of dimension 2. The following statements then hold.*

1. *F is EP everywhere $\Leftrightarrow$ F is EP somewhere $\Leftrightarrow$ F has a fixed point.*
2. *Wherever F is EP, its prime period is at most 2.*

Proof: It is sufficient, by what was said above after Lemma 3.1, to establish the result for H in place of F. If H is EP everywhere, it is certainly EP somewhere. If H is EP somewhere then the converse conditions of Theorem 3.2 are satisfied. Hence, H has a fixed point. Conversely, suppose that H has a fixed point. By Theorem 3.2, H is EP everywhere. This proves the first part.

For the second part, if H is EP then, by the first part, H has a fixed point. By Theorem 3.2, H has prime period at most 2.

QED

Theorem 3.3 is the first demonstration of eventual periodicity for systems with mixed constraints. As was pointed out in [BCOQ92, page 457], *"Such a property has not been shown ... though simulations do point in this direction."*. It is interesting to note the all-or-nothing quality of periodic behaviour implied by Theorem 3.3: either the function is eventually periodic everywhere or it is nowhere eventually periodic. Note also that if F has a periodic point then it must necessarily have a fixed point. There is evidence to suggest that a similar result holds in higher dimensions, [BO93, Gun94a].

The second part of Theorem 3.3 suggests a connection between the maximum prime period and the dimension of F. This cannot be so simple in general. For instance, we can always construct a max-only function of dimension n which implements any permutation of the n variables. The prime period of this function, at most points in $\mathbf{R}^n$, is equal to the order of the permutation in the symmetric group on n letters. Of course there are many permutations whose order is greater than n: the permutation with cycle shape (123)(45) has order 6. Hence dimension 2 is a special case. We should also note that, in view of Sarkovskii's celebrated result on the dynamics of a continuous function, [Dev89, Theorems 10.1 and 10.2], it should come as no surprise that the prime period in dimension 2 is at most 2.

When does F have a fixed point? For high dimensional functions this is a difficult open problem but the methods of this section enable us to give a complete answer in dimension 2. First we need some extra terminology.

- If p is a piecewise nice function then the rightmost (respectively, leftmost) segment of p is that part of the graph of p which corresponds to the interval $[u_n, +\infty)$ (respectively, $(-\infty, u_1])$ in (10).

If $S \in \mathbf{R} \cup \{-\infty\}$ is a finite subset then, as usual, $\bigwedge_{a \in S} a$ denotes the minimum element of S. Note that if $-\infty \in S$ then $\bigwedge_{a \in S} a = -\infty$. If $S = \emptyset$ then, by convention, $\bigwedge_{a \in S} a = +\infty$; the intuition being that as S gets smaller, $\bigwedge_{a \in S} a$ gets larger. For similar reasons, $\bigvee_{a \in S} = -\infty$ if $S = \emptyset$.

Now suppose that F is a min-max function of dimension 2 and each component of F is in conjunctive form:

$$\begin{array}{rcl} F_1(x_1, x_2) & = & (a_{11} + x_1 \vee a_{12} + x_2) \wedge \cdots \wedge (a_{n1} + x_1 \vee a_{n2} + x_2) \\ F_2(x_1, x_2) & = & (b_{11} + x_1 \vee b_{12} + x_2) \wedge \cdots \wedge (b_{m1} + x_1 \vee b_{m2} + x_2). \end{array}$$

Proposition 3.1 *F has a fixed point if, and only if, the following two conditions hold:*

$$\bigwedge_{1 \leq i \leq n} a_{i1} \leq \bigwedge_{b_{i1} = -\infty} b_{i2},$$

$$\bigwedge_{1\leq i\leq m} b_{i2} \;\leq\; \bigwedge_{a_{i2}=-\infty} a_{i1}.$$

Proof: As before, it is sufficient to deal with H instead of F. Consider the rightmost segment of H. It follows from the cone restrictions used in the proof of Theorem 3.2 that one way for H to fail to have a fixed point is if this segment looks like the graph of $y = x + a$ where $a > 0$. The only way for this to happen is if the corresponding rightmost segments of $F_1(x, 0)$ and $F_2(x, 0)$ look like $y = x + u$ and $y = v$, respectively, and $u > v$. By considering

$$\begin{aligned} F_1(x,0) &= (a_{11} + x \vee a_{12}) \wedge \cdots \wedge (a_{n1} + x \vee a_{n2}), \\ F_2(x,0) &= (b_{11} + x \vee b_{12}) \wedge \cdots \wedge (b_{m1} + x \vee b_{m2}) \end{aligned}$$

and letting x become large, it is easy to see that this implies $u = \bigwedge_{1\leq i\leq n} a_{i1} \neq -\infty$, $v = \bigwedge_{b_{i1}=-\infty} b_{i2} \neq +\infty$ and

$$\bigwedge_{1\leq i\leq n} a_{i1} > \bigwedge_{b_{i1}=-\infty} b_{i2}. \tag{12}$$

Conversely, if this inequality holds then both sides of it must be finite. For if not then the left side could only be $-\infty$ or the right side $+\infty$. In neither case can (12) hold. Hence the rightmost segment of H will have the required form. It follows that (12) is a necessary and sufficient condition for this form of fixed point failure

The only other way for H to fail to have a fixed point is if its leftmost segment looks like $x + a$ where $a < 0$. A similar analysis shows that this can happen if, and only if,

$$\bigwedge_{1\leq i\leq m} b_{i2} > \bigwedge_{a_{i2}=-\infty} a_{i1}.$$

Hence H, and therefore also F, has a fixed point if, and only if, the two conditions in the statement of the proposition hold.

QED

As an example of this, consider the following min-max function, which is of the form considered by Olsder in [Ols91, Ols93]:

$$\begin{aligned} F_1(x_1,x_2) &= a + x_1 \vee b + x_2 \\ F_2(x_1,x_2) &= c + x_1 \wedge d + x_2. \end{aligned} \tag{13}$$

Here we assume that $a, b, c, d \in \mathbf{R}$ which is the only interesting case. To apply Proposition 3.1 we need to convert F_2 to conjunctive form using the algorithm discussed in §2:

$$F_2(x_1,x_2) = (c + x_1 \vee -\infty + x_2) \wedge (-\infty + x_1 \vee d + x_2).$$

The first condition of proposition 3.1 reduces to $a \leq d$ while the second condition becomes $-\infty \leq +\infty$ which we may ignore. Hence the min-max function (13) has a fixed point if, and only if, $a \leq d$. This is Theorem 2.1 of [Ols91] restricted to functions of dimension 2. We shall return to examples like this in §5.

4 The balance condition

In example (13) it is clear that if we change the values of the parameters a, b, c and d, then the fixed point behaviour of F will sometimes change. However, the methods of the previous section also reveal the existence of a class of 2 dimensional functions whose fixed point behaviour is invariant under change of parameters (see Proposition 4.1 below). This is an attractive property and in this section we shall explore it further.

First we need to clarify what we mean by "change the values of the parameters". The parameters are part of the definition of F; if they change, then so does F. To contemplate changing the values of the parameters while keeping to the same function is an abuse of terminology. However, it is a very convenient abuse which should not mislead the reader. It could be formalized and, in fact, we shall do so for max-plus matrices in Definition 4.3 below. But the price in syntax is high for a general min-max function and the reader would surely not thank us for it. We should note, however, that parameters are always real numbers; although we may sometimes use $-\infty$ and $+\infty$ when writing expressions in conjunctive or disjunctive form, we never allow an infinite value to be changed to a finite value or vice versa. That would "really" change the function!

Definition 4.1 *Let F be a min-max function of arbitrary dimension. F is balanced if F has a fixed point for all values of its parameters.*

We shall show in this section that we can characterize the balance property in two cases of interest: functions of dimension 2 and max-only functions. We begin with the former.

Definition 4.2 *The characteristic of a piecewise nice function p, denoted $char(p)$, is a pair, $[u, v]$, where $u, v \in \{+1, 0, -1\}$ are the slope of the rightmost segment of p and the slope of the leftmost segment of p, respectively. If F is min-max function of dimension 2, its characteristic is that of its auxiliary function.*

We shall sometimes write $char(p) = [char_+(p), char_-(p)]$. Note that the characteristic of $p(x) = x$ is $[+1, +1]$ while that of $p(x) = -x$ is $[-1, -1]$. It is easy to compute the characteristic; the following Lemmas give the details. The values which appear in the characteristic are treated as numbers to which the usual operations, including $\wedge$ and $\vee$, can be applied.

Lemma 4.1 *With the same assumptions as in Lemma 3.2, let $char(p) = [p_+, p_-]$, $char(q) = [q_+, q_-]$. Then,*

$$\begin{aligned}
char(\pm x) &= [\pm 1, \pm 1] \\
char(a) &= [0, 0] \\
char(p + a) &= char(p) \\
char(p \vee q) &= [p_+ \vee q_+, p_- \wedge q_-] \\
char(p \wedge q) &= [p_+ \wedge q_+, p_- \vee q_-] \\
char(-p) &= [-p_+, -p_-] \\
char(p(-x)) &= [-p_-, -p_+]
\end{aligned}$$

Proof: Another exercise for the reader.

QED

Lemma 4.2 *If F is a min-max function of dimension 2 then $char_{\pm}(F) = char_{\pm}(F_1(x,0)) - char_{\pm}(F_2(x,0))$. Furthermore, $char(F)$ is independent of the values of the parameters in F.*

Proof: The first part is obvious. The second part follows from the third formula in Lemma 4.1.

QED

The significance of the characteristic is revealed in the following proposition.

Proposition 4.1 *Let F be a min-max function of dimension 2. F is balanced if, and only if, $char_{+}(F) < +1$ and $char_{-}(F) < +1$.*

Proof: Let H be the auxiliary function of F and suppose that the condition holds. Since $char_{+}(H) < +1$, the main diagonal, which has slope $+1$, must be above the graph of H for sufficiently large positive values of x. In other words, there must be a point for which $H(x) - x < 0$. Similarly, since $char_{-}(H) < +1$, there must be a point for which $H(x) - x > 0$. Since H is continuous, it must have a fixed point. Hence so does F. Since the characteristic is independent of the values of the parameters, F is balanced.

Now suppose that $char_{+}(F) = +1$. For large positive values of x, the graph of H must coincide with the line $x + u$ for some $u \in \mathbf{R}$. If $u > 0$, it is easy to see, as in Proposition 3.1, that H does not have a fixed point. So suppose that $u < 0$. Let $v = -2u$, which is positive. Express F_1 in conjunctive form, as in (5), and add v to all the parameters which appear in this expression for F_1. Keep F_2 the same as it was. It is easy to see, using (3), that the new function is $(F_1(x_1, x_2) + v, F_2(x_1, x_2))$. Hence, although the characteristic of F has not changed, its new auxiliary function behaves like $x - u$ for large x. As pointed out above, this new H cannot have a fixed point. It follows that, for some choice of parameters, F does not have a fixed point and hence is not balanced. A similar argument works if $char_{-}(F) = +1$.

QED

We could also obtain a characterization of balance using Proposition 3.1. However, $char(F)$ contains the essential information on the balance property and is much more convenient to use in practice. For one thing, it does not require F to be expressed in conjunctive form. We recommend, when confronted with a 2 dimensional function whose fixed point behaviour is in question, to first calculate $char(F)$ and then fall back on Proposition 3.1 if F turns out to be unbalanced.

The geometric methods used above enable one to deduce much useful information about a 2 dimensional function. However, these methods do not extend to higher dimensions: there is no obvious analogue of piecewise nice functions in dimensions greater than 2. More precisely, there is no similar class of functions for which Theorem 3.1 is known to hold. In the remainder of this paper we shall be concerned with functions of any dimension and the reader will observe that the methods used are quite different and rely more heavily on the established techniques of max-plus algebra.

We now turn our attention to max-only functions. In the rest of this section we shall determine the conditions under which they are balanced. We begin with some notation, then state some lemmas about eigenvalues and eigenvectors and finally prove the main result. We conclude with a comparison between balance and irreducibility.

In what follows, we shall use both standard algebra and max-plus algebra. We shall adhere to the following rules in order not to confuse the reader. The operator $+$ will always have its standard meaning and $\vee$ and $\wedge$ will always mean maximum and minimum respectively. The following operations with vectors and matrices will be interpreted in terms of max-plus algebra: $A\vec{v}^T$, $\vec{x}\vec{v}^T$, $h\vec{v}$ and hA.

We shall need various notations and terminology from max-plus algebra. Most of the definitions given below can be found in [BCOQ92, §2.3], which should be consulted for more details.

- The precedence graph of A, denoted $\mathcal{G}(A)$, is the annotated directed graph on the vertices $\{1, \cdots, n\}$, where there is an edge from j to i if, and only if, $A_{ij} \neq -\infty$. This will be denoted by $j \to i$. The annotation on the edge $j \to i$ is then A_{ij}.
- Let $(\to)^*$ and $(\leftarrow)^*$ denote the reflexive, transitive closures of the edge relations $\to$ and $\leftarrow$, respectively. The equivalence relation on the vertices of $\mathcal{G}(A)$ given by $(\to)^* \cap (\leftarrow)^*$ will be denoted $\mathcal{R}$.
- The equivalence classes of vertices under $\mathcal{R}$ are the vertex sets of the maximal strongly connected subgraphs (MSCSs) of $\mathcal{G}(A)$.
- An MSCS of $\mathcal{G}(A)$ is said to be non-trivial if it contains a circuit. The trivial MSCSs are those with only a single vertex and no self-loop.
- The relation $\to$ on $\mathcal{G}(A)$ passes to the equivalence classes under $\mathcal{R}$, where it is clearly acyclic. Let $\rho(A)$ denote the corresponding partial order on the set of MSCSs.
- A path in $\mathcal{G}(A)$ is a sequence of edges, $i_1 \to i_2 \to \cdots \to i_m$, from i_1 to i_m. A circuit is a path in which $i_1 = i_m$.
- The weight of a path $p = i_1 \to i_2 \to \cdots \to i_m$ is the sum of the annotations on the path, $|p|_{\mathsf{w}} = A_{i_2 i_1} + \cdots + A_{i_m i_{m-1}}$. The length of a path is the number of edges in the path, $|p|_\ell = m - 1$.
- The cycle mean of a circuit, g, is $\omega(g) = |g|_{\mathsf{w}}/|g|_\ell$.

For matrices, it is easy to formalise the idea of changing the values of parameters and we shall borrow an idea from qualitative matrix theory, [Joh88], to do so.

Definition 4.3 *Let A and B be max-plus matrices. A and B have the same pattern if $A_{ij} = -\infty$ if, and only if, $B_{ij} = -\infty$.*

If A and B have the same pattern then $\mathcal{G}(A)$ and $\mathcal{G}(B)$ are isomorphic as directed graphs (forgetting the annotations). It follows that $\rho(A)$ and $\rho(B)$ are isomorphic as partially ordered sets.

If F is a max-only function, then, as we saw in §2, each F_i can be written in conjunctive normal form:

$$F_i(\vec{x}) = (A_{i1} + x_1 \vee \cdots \vee A_{in} + x_n) \quad 1 \leq i \leq n,$$

where $A_{ij} \in \mathbf{R} \cup \{-\infty\}$. Let A denote the $n \times n$ matrix in max-plus algebra whose i, j-th component is A_{ij}. Then $F(\vec{x})^T = A\vec{x}^T$. By the remarks following Lemma 2.1, A is uniquely associated to F. Not all matrices correspond to max-only functions in this way: they must have the property that $A_{i*} \neq \vec{c}(-\infty)$ for each row i.

The matrix A operates on vectors in the space $(\mathbf{R} \cup \{-\infty\})^n$. This results in an important distinction between eigenvectors, for which $A\vec{x}^T = h\vec{x}^T$, and fixed points of F, for which $F(\vec{x}) = \vec{x} + \vec{c}(h)$. Some (but not all) of the components of an eigenvector can be $-\infty$ but the components of a fixed point must all be real.

- A vector $\vec{v} \in (\mathbf{R} \cup \{-\infty\})^n$ is real if $\vec{v} \in \mathbf{R}^n$.

It follows that there is a one-to-one correspondence between fixed points of F and real eigenvectors of A. We shall need to construct eigenvectors of A and we shall repeatedly use the following technique. Let k be a vertex in $\mathcal{G}(A)$. Let $A^+_{ik} \in \mathbf{R} \cup \{-\infty, +\infty\}$ denote the maximum weight among all paths from k to i. If there are no paths from k to i, then $A^+_{ik} = -\infty$. If there is no maximum, then $A^+_{ik} = +\infty$. Alternatively, we can simply define $A^+_{ik} = \max\{A_{ik}, A^2_{ik}, A^3_{ik}, \cdots\}$, (compare [BCOQ92, (1.18)]). We shall say that the column vector A^+_{*k} "exists" if none of its entries are $+\infty$. In this case $(A^+_{*k})^T$ is a genuine vector in max-plus algebra. The next result is contained in Theorem 3.101 of [BCOQ92] and the proof is hence omitted.

Lemma 4.3 *Let A be an $n \times n$ matrix in max-plus algebra. If A^+_{*k} exists and $A^+_{kk} = 0$, then $(A^+_{*k})^T$ is an eigenvector of A with eigenvalue* 0.

It is worth noting that if $A^+_{kk} = 0$, then the vertex k must lie on a circuit in $\mathcal{G}(A)$ of weight 0. Hence, when constructing eigenvectors by the technique of Lemma 4.3 it is only necessary to consider those vertices which lie on circuits g in $\mathcal{G}(A)$ with $\omega(g) = 0$.

We now need some observations about the eigenvalues of a max-plus matrix. In the general case, when $\mathcal{G}(A)$ is not assumed to be strongly connected, the subject has been carefully studied by Stéphane Gaubert in his thesis, [Gau92]. Since these results are not yet generally available we have given an elementary proof of the aspects which are needed here, with references to the more complete results in [Gau92].

- ([Gau92, Chapter IV, Definition 1.3.4]) If $\vec{v} \in (\mathbf{R} \cup \{-\infty\})^n$ then the support of $\vec{v}$ is $sup(\vec{v}) = \{i \in \mathbf{N} \mid v_i \neq -\infty\}$.

If A is some $n \times n$ matrix, we can regard $sup(\vec{v})$ as a subset of the vertices of $\mathcal{G}(A)$. Any subset of vertices, S, can always be regarded as a subgraph by restriction of the edge relation in the obvious way.

- If p is a path in $\mathcal{G}(A)$ and S is a subgraph of $\mathcal{G}(A)$, then p lies entirely in S, denoted $p \subseteq S$, if p is a subgraph of S.

Lemma 4.4 *Let A be a square matrix in max-plus algebra and suppose that $A\vec{v}^T = \lambda\vec{v}^T$. The following statements hold.*

1. *If $i \in sup(\vec{v})$ then there exists a circuit, g, with $\omega(g) = \lambda$ and a path, p, from that circuit to i, such that $p, g \subseteq sup(\vec{v})$.*

2. *If $g \subseteq sup(\vec{v})$ is a circuit, then $\omega(g) \leq \lambda$.*

Proof: For the first part, suppose that $v_i \neq -\infty$. Since $\lambda + v_i = A_{i.}\vec{v}^T$, there must be an edge $j \to i$ of $\mathcal{G}(A)$ such that $v_j \neq -\infty$ and

$$\lambda + v_i = A_{ij} + v_j. \tag{14}$$

Repeating this argument for j we can construct a path of arbitrary length leading to i lying entirely in $sup(\vec{v})$. Since there are only finitely many vertices, the path must contain a circuit, g. By applying equation (14) to each vertex of this circuit and then adding the results, it is easy to see that $\omega(g) = \lambda$. This proves the first part.

For the second part, suppose that $g = i_1 \to \cdots \to i_m \to i_1$ is a circuit such that $g \subseteq sup(\vec{v})$. For each $1 \leq i < m$, it must be the case that $\lambda + v_{i+1} \geq A_{(i+1)i} + v_i$ and, for $i = m$, $\lambda + v_1 \geq A_{1m} + v_m$. By adding these inequalities, we see that $\lambda \geq \omega(g)$. This proves the second part.

QED

For more complete information about how the eigenvalues are distributed over the MSCSs of $\mathcal{G}(A)$ see Proposition 2.2.1 and Theorem 2.2.4 of [Gau92, Chapter IV]. The next result is not needed for the proof of the main theorem but it will be helpful later on. It is a trivial consequence of Lemma 4.4.

Lemma 4.5 *If A is a square matrix in max-plus algebra and $A\vec{v}^T = \lambda\vec{v}^T$ then,*

$$\lambda = \bigvee\{\omega(g) \mid g \text{ a circuit}, g \subseteq sup(\vec{v})\}.$$

In particular, any two eigenvectors with the same support have the same eigenvalue.

For a more precise characterization, see [Gau92, Chapter IV, Corollary 2.2.5]. We can now state the main result.

Theorem 4.1 *Let A be a square matrix in max-plus algebra. The following statements are equivalent.*

1. *Any matrix with the same pattern as A has a real eigenvector.*

2. *$\mathcal{G}(A)$ has one and only one non-trivial MSCS and this is the least element of $\rho(A)$.*

3. *$\mathcal{G}(A)$ has a circuit and every eigenvector of A is real.*

Proof: We first prove the equivalence of parts 1 and 2. Assume that part 1 holds and that $\mathcal{G}(A)$ does not satisfy part 2. Since A has an eigenvector, Lemma 4.4(1) implies that $\mathcal{G}(A)$ has a circuit. It follows from the assumption that $\mathcal{G}(A)$ must have a non-trivial MSCS, S, which is not the least element of $\rho(A)$. There is hence a vertex i such that there is no path in $\mathcal{G}(A)$ from any vertex of S to i. So far we have not said anything about the values of the annotations on the edges of $\mathcal{G}(A)$. We can certainly choose them so that maximum cycle mean of appears only on a circuit in S. For instance, we can pick some circuit in S and set the annotations on the edges in that circuit to 1, while setting all other annotations to 0. Let B be the matrix, having the same pattern as A, which results from this construction and let γ denote the maximum cycle mean of $\mathcal{G}(B)$. According to part 1, B has a real eigenvector, $\vec{v}$ such that $B\vec{v}^T = \lambda\vec{v}^T$. By Lemma 4.4(1), there exists a circuit, g, in $\mathcal{G}(B)$ with $\omega(g) = \lambda$ and, since $\vec{v}$ is real, a path from g to the vertex i. It follows that g must lie in an MSCS different from S. Since $\gamma \subseteq S$ was chosen as the circuit with maximum cycle mean, it must be that $\lambda < \gamma$. But, γ is the cycle mean of a circuit lying entirely in $\sup(\vec{v}) = \mathcal{G}(B)$, and so, by Lemma 4.4(2), $\lambda \geq \gamma$. This contradiction shows that part 1 implies part 2.

Now suppose part 2 holds. Let B be a matrix having the same pattern as A. We have to show that B has an eigenvector. Let λ be the maximum cycle mean of $\mathcal{G}(B)$ and consider the matrix $C = \lambda^{-1}B$ which is obtained from B by subtracting (in normal algebra) λ from each entry. Evidently, C has the same pattern as B and A and it is clear that $\mathcal{G}(C)$ is obtained from $\mathcal{G}(B)$ by subtracting λ from each annotation. It follows that any circuit of $\mathcal{G}(C)$ has a non-positive cycle mean and at least one circuit has cycle mean 0. Let k be a vertex on some circuit of maximum cycle mean. By hypothesis, there is a path from k to any vertex of $\mathcal{G}(C)$. Furthermore, since each circuit has non-positive weight, it is always possible to choose a path of maximum weight, which does not contain any circuits, from k to any vertex. Hence, C^+_{*k} exists and each $C^+_{ik} \in \mathbf{R}$. Furthermore, since k lies on a circuit of maximum cycle mean 0, $C^+_{kk} = 0$. It follows from Lemma 4.3 that $(C^+_{*k})^T$ is a real eigenvector of C of eigenvalue 0. Hence, $(C^+_{*k})^T$ must be a real eigenvector of B of eigenvalue λ. This shows that part 2 implies part 1 and hence that parts 1 and 2 are equivalent.

We now show the equivalence of parts 2 and 3. Suppose that part 2 holds. Since $\mathcal{G}(A)$ has a non-trivial MSCS, S, it must have a circuit. We need to show that every eigenvector of A is real. Let $\vec{v}$ be an eigenvector. By Lemma 4.4(1), there exists a circuit $g \subseteq sup(\vec{v})$. Clearly, $g \subseteq S$. But, since S is the least element of $\rho(A)$, there must be a path from any vertex in S to any other vertex, $j \in \mathcal{G}(A)$. Hence there is a path from a vertex in $sup(\vec{v})$ to j. But then it is easy to see that $j \in sup(\vec{v})$. Hence $\vec{v}$ must be real, as required.

Finally, suppose that part 3 holds and part 2 does not. Since A has a circuit there must be a non-trivial MSCS, S, of $\mathcal{G}(A)$ with the following two properties. Firstly, there is no non-trivial MSCS which is greater than S in the partial order $\rho(A)$. Secondly, there is a vertex $i \in \mathcal{G}(A)$ which cannot be reached by any path from S. Let λ be the maximum cycle mean of all circuits lying in S, let g be a circuit in S with cycle mean λ and let k be a vertex lying on g. Consider, as before, $B = \lambda^{-1}A$, and use the same notation for the corresponding S, g and k in $\mathcal{G}(B)$. Because of the choice of S, the only circuits on any path from k must lie entirely in S and must hence have non-positive weight. It follows that B^+_{*k} exists and $B^+_{kk} = 0$. Hence, by Lemma 4.3, $\vec{v} = (B^+_{*k})^T$ is an eigenvector of B of eigenvalue 0. Since there is no path in $\mathcal{G}(B)$ from k to i, $v_i = -\infty$. But, $\vec{v}$ is also an eigenvector of A of eigenvalue λ. Hence we have constructed an eigenvector of A which is not real. This contradiction shows that part 3 implies part 2 and hence that parts 2 and 3 are equivalent.

QED

Condition 2 of Theorem 4.1 is identical to the graph condition stated as part of property

P3′ in [Ols91, page 189], although the reader will have to add the words "non-trivial" before "strongly connected" in P3′ to bring it into line with the definitions given earlier. The equivalence between conditions 2 and 3 is related to the structural characterization of eigenvalues in Corollary 2.2.5 of [Gau92, Chapter IV].

We now want to compare the balance property with the irreducibility property which is widely used in max-plus algebra. First, we need to compare like with like.

Definition 4.4 *If A is a square matrix in max-plus algebra, then A is balanced if it satisfies any of the equivalent conditions in Theorem 4.1.*

If F is max-only and A is the corresponding matrix in max-plus algebra, then F is balanced if, and only if, A is balanced.

Definition 4.5 *If A is a square matrix in max-plus algebra then $\mu(A) = \bigvee\{\omega(g) \mid g \text{ any circuit}\}$ denotes the maximum cycle mean of A.*

If A has no circuits then, following the convention introduced before Proposition 3.1, $\mu(A) = -\infty$. However, if A is the matrix corresponding to some max-only function then $A_{i*} \neq \vec{c}(-\infty)$. It follows that A always has a circuit and hence $\mu(A) \neq -\infty$. There is a simple formula for $\mu(A)$ when A has dimension 2 which we shall make use of in the next section. If

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

where $a, b, c, d \in \mathbf{R} \cup \{-\infty\}$, then

$$\mu(A) = a \vee (b + c)/2 \vee d. \tag{15}$$

This quickly becomes less simple as the dimension increases!

It follows immediately from Lemma 4.5 that any real eigenvector must have $\mu(A)$ for its eigenvalue. For balanced matrices, this is the only eigenvalue, as the following result shows.

Proposition 4.2 *Let A be a square matrix in max-plus algebra. If A is balanced then $\mu(A)$ is the unique eigenvalue of A.*

Proof: If A is balanced then, by Theorem 4.1(1), it has a real eigenvector. As we have just seen, the associated eigenvalue must be $\mu(A)$. Furthermore, by Theorem 4.1(3), every eigenvector of A is real. Hence $\mu(A)$ is the unique eigenvalue of A.

QED

What does this imply for irreducible matrices? Irreducibility has the same meaning in max-plus algebra as in classical linear algebra: a matrix A is irreducible if there is no permutation of its rows and columns which puts it into upper-triangular block form, [BCOQ92, Definition 2.13]. It is not difficult to see that this is equivalent to $\mathcal{G}(A)$ being strongly connected, [BCOQ92, Theorem 2.14]. In other words, there is a unique MSCS which is non-trivial. Hence, by Theorem 4.1(2), irreducibility implies balance. The converse is false: the matrix below

$$\begin{pmatrix} a & -\infty \\ c & -\infty \end{pmatrix},$$

where $a, c \in \mathbf{R}$, is balanced but not irreducible. Proposition 4.2 tells us that Theorem 3.23 of [BCOQ92], the analogue in max-plus algebra of the Perron-Frobenius theorem on the spectrum of an irreducible matrix, holds verbatim if the word "irreducible" is replaced by the word "balanced".

The disadvantage of irreducibility is that it is so firmly tied to linear algebra and graph theory that it is difficult to generalize it to systems with mixed constraints. The virtue of the balance condition is that it can be defined for any min-max function and, as we have seen in this section, we can give useful characterizations of it in cases of interest.

5 The cycle time formula

In this section we study the cycle time, $\chi_F(\vec{x})$, of a min-max function F. We recall that this is only defined when F is eventually periodic at $\vec{x}$. We begin with the case of a max-only function, F. Let A be the corresponding matrix in max-plus algebra and suppose that F has a fixed point where $\chi_F(\vec{x}) = \vec{x} + \vec{c}(h)$. It follows from Lemma 4.5 that $\chi_F(\vec{x}) = h = \mu(A)$. Hence the cycle time of any eventually fixed point is unique. What can we say about the cycle time at an eventually periodic point? The following result is presumably well-known to max-plus experts.

Lemma 5.1 *Let A be a square matrix in max-plus algebra. For any $k \geq 1$, $\mu(A^k) = k\mu(A)$.*

Proof: $\mathcal{G}(A)$ has a circuit if, and only if, $\mathcal{G}(A^k)$ has a circuit so we can assume that both $\mathcal{G}(A)$ and $\mathcal{G}(A^k)$ have circuits.

Let g be a circuit of maximum cycle mean in $\mathcal{G}(A)$ and let $n = |g|_\ell$. If we take any k consecutive edges on this circuit, say from vertex i to vertex j, then the sum of the annotations on these edges must be at most $A^k_{ji} \neq -\infty$. Hence, if we start from vertex i and run along k consecutive edges on the circuit to vertex j and keep on going, n times, we generate a circuit h of $\mathcal{G}(A^k)$ with $|h|_\ell = n$. By construction, $|h|_{\mathsf{w}} \geq nk\mu(A)$. Hence, $\omega(h) \geq k\mu(A)$ and so $\mu(A^k) \geq k\mu(A)$.

Now let g be a circuit of maximum cycle mean in $\mathcal{G}(A^k)$ and let $n = |g|_\ell$. We can decompose each edge on the circuit into a path of length k in $\mathcal{G}(A)$ such that the sum of the annotations on the path (in $\mathcal{G}(A)$) is equal to the annotation on the edge (in $\mathcal{G}(A^k)$). By abutting the paths end-to-end for each edge of g, we generate a circuit h in $\mathcal{G}(A)$ with $|h|_\ell = kn$. Then, $|h|_{\mathsf{w}} = n\mu(A^k)$. Hence $\omega(h) = \mu(A^k)/k$ and so $k\mu(A) \geq \mu(A^k)$.

QED

If F is a max-only function and A is the corresponding matrix in max-plus algebra then it is easy to see that the matrix corresponding to F^k is just A^k.

Proposition 5.1 *Let F be a max-only function. If $\vec{x}$ is an eventually periodic point of F then*

$$\chi_F(\vec{x}) = \mu(A),$$

where A is the matrix corresponding to F in max-plus algebra. In particular, the cycle time is independent of which eventually periodic point is chosen and depends only on F.

Proof: We have already seen that if $\vec{x}$ is a fixed point of F then $\chi_F(\vec{x}) = \mu(A)$.

Now suppose that $\vec{x}$ is a periodic point of F of period k and that $F^k(\vec{x}) = \vec{x} + \vec{c}(h)$. Then $\vec{x}$ is a fixed point of F^k and $\chi_{F^k}(\vec{x}) = h = k\chi_F(\vec{x})$. But then,

$$\begin{array}{rcll} \chi_{F^k}(\vec{x}) & = & \mu(A^k) & \text{by the first part} \\ & = & k\mu(A) & \text{by Lemma 5.1.} \end{array}$$

Hence, $\chi_F(\vec{x}) = \mu(A)$. The last assertion is clear.

QED

There is an important distinction between $\chi_F(\vec{x})$ and $\mu(A)$. The latter is always defined. The former only has meaning when we know that F is eventually periodic at $\vec{x}$. It is easy to calculate $\mu(A)$ in principle but this is not enough to tell us whether F has an eventually periodic point. As we saw in Theorem 3.3, this need not be the case.

We now want to extend this result to min-max functions. It will be convenient to first introduce some dual notation.

- If B is a square matrix in min-plus algebra then $\eta(B) = \bigwedge\{\omega(g) \mid g \text{ any circuit}\}$ denotes the minimum cycle mean of B.

For the remainder of this section we shall use the letter A for max-plus matrices and the letter B for min-plus matrices. Also, if $\vec{v}, \vec{w} \in (\mathbf{R} \cup \{-\infty\})^n$, then $\vec{v} \geq \vec{w}$ if $v_i \geq w_i$ for each $1 \leq i \leq n$.

Now let F be a min-max function of dimension n. As we saw in (5), F can be placed in conjunctive form so that, for $1 \leq k \leq n$,

$$F_k(\vec{x}) = (A^k_{11} + x_1 \vee \cdots \vee A^k_{1n} + x_n) \wedge \cdots \wedge (A^k_{\ell(k)1} + x_1 \vee \cdots \vee A^k_{\ell(k)n} + x_n), \qquad (16)$$

where $A^k_{ij} \in \mathbf{R} \cup \{-\infty\}$. Here $\ell(k)$ is the number of max-only expressions in the component F_k. We can now associate a max-plus matrix A to F by choosing, for the k-th row of the matrix, one of the $\ell(k)$ max-only expressions in (16): $A_{kj} = A^k_{i_k j}$ where $1 \leq i_k \leq \ell(k)$ specifies which max-only expression is chosen in row k

Definition 5.1 *The matrix A constructed in this way is called a max-only projection of F. A set of max-only projections is the collection of all such matrices from a single conjunctive form for F such as (16). If the conjunctive form is also normal, then the corresponding matrices are called normal max-only projections. projections.*

It follows from the discussion surrounding Theorem 2.1 that if A is any max-only projection of F then there exists some normal max-only projection A' such that $A'_{i*} \leq A_{i*}$ for each row

i. By virtue of Theorem 2.1 the sets of normal max-only and normal min-only projections of F are uniquely defined. In practice it is often more convenient to work with whatever set of projections is easiest to construct instead of doing the additional work necessary to find the set of normal projections. It is, however, of theoretical interest to know that there is a way to associate a unique canonical set of projections to any min-max function. Note that sets of projections can be quite large: the function (16) has a set of $\prod_{1\leq i\leq n}\ell(i)$ distinct max-only projections. Let us consider the following example and work out sets of max-only and min-only projections:

$$\begin{array}{rcl} F_1(x_1,x_2) & = & (a+x_1 \vee b+x_2)\wedge c+x_1 \\ F_2(x_1,x_2) & = & (t+x_1\wedge u+x_2) \end{array}$$

where $a,b,c,t,u\in\mathbf{R}$. F_1 is already in conjunctive form while F_2 is in disjunctive form. We first put F_2 into conjunctive form using the algorithm discussed in §2:

$$F_2(x_1,x_2) = (t+x_1\vee -\infty+x_2)\wedge(-\infty+x_1\vee u+x_2)$$

and then read off a set of max-only projections:

$$\{\left(\begin{array}{cc} a & b \\ t & -\infty \end{array}\right), \left(\begin{array}{cc} a & b \\ -\infty & u \end{array}\right), \left(\begin{array}{cc} c & -\infty \\ t & -\infty \end{array}\right), \left(\begin{array}{cc} c & -\infty \\ -\infty & u \end{array}\right)\}.$$

If $a<c$ then F_1 is in conjunctive normal form and the set of projections above is also a set of normal max-only projections. If $a\geq c$ then a set of normal projections consists of only the first two matrices. Dually, we can put F_1 into disjunctive form—an exercise which we have already performed in (8):

$$F_1(x_1,x_2) = ((a\wedge c)+x_1\wedge +\infty+x_2)\vee(c+x_1\wedge b+x_2)$$

and then read off a set of min-only projections:

$$\{\left(\begin{array}{cc} a\wedge c & +\infty \\ t & u \end{array}\right), \left(\begin{array}{cc} c & b \\ t & u \end{array}\right)\}.$$

We hope this example has clarified these important constructs.

The proof of the following lemma is very similar to that of Lemma 4.4 and can safely be left to the reader. See also Lemma 1.3.9 of [Gau92, Chapter IV].

Lemma 5.2 *Let A be an $n\times n$ matrix in max-plus algebra and suppose that $A\vec{v}\geq\lambda\vec{v}$ for some $\vec{v}\in(\mathbf{R}\cup\{-\infty\})^n$ such that $\vec{v}\neq\vec{c}(-\infty)$. Then $\mu(A)\geq\lambda$.*

We can now state the main result of this section.

Theorem 5.1 *Let F be a min-max function of dimension n. If $\vec{x}$ is a fixed point of F then*

$$\bigvee_{B\in Q}\eta(B) = \chi_F(\vec{x}) = \bigwedge_{A\in P}\mu(A) \qquad (17)$$

where P and Q are any sets of max-only and min-only projections, respectively, of F.

Proof: Suppose that $F(\vec{x}) = \vec{x}+\vec{c}(h)$ so that $\chi_F(\vec{x}) = h$. Let F be expressed in the conjunctive form from which P arises and let us use the same notation as in (16) above. For each component F_k, let i_k be the index of some max-only expression for which $F_k(\vec{x}) = A^k_{i_k *}\vec{x}^T$. Such an expression must clearly exist for each $1 \leq k \leq n$. It then follows that, for any $1 \leq i \leq \ell(k)$,

$$A^k_{i*}\vec{x}^T \geq A^k_{i_k *}\vec{x}^T \tag{18}$$

Let $A_1 \in P$ be the max-only projection constructed by choosing the max-only expression i_k in row k. Two conclusions can now be drawn. Firstly, since $F(\vec{x}) = \vec{x} + \vec{c}(h)$, it follows that $A_1\vec{x}^T = h\vec{x}^T$. Hence $\vec{x}$ is a real eigenvector of A_1 and it follows from Lemma 4.5 that $h = \mu(A_1)$. Secondly, it follows from (18) that if $A_2 \in P$ is any max-only projection from the same set, then $A_2\vec{x}^T \geq h\vec{x}^T$. But then Lemma 5.2 tells us that $\mu(A_2) \geq h$. Hence,

$$h = \bigwedge_{A\in P} \mu(A).$$

The other assertion is clearly the dual statement.

QED

Corollary 5.1 *Let F be a min-max function. The cycle time of F at an eventually periodic point is independent of which eventually periodic point is chosen and depends only on F.*

Proof: Suppose that $\vec{x}$ and $\vec{y}$ are periodic points of F such that $F^k(\vec{x}) = \vec{x} + \vec{c}(r)$ and $F^l(\vec{y}) = \vec{y} + \vec{c}(s)$. Then, by (9), $\vec{x}$ and $\vec{y}$ are both fixed points of F^{kl} with $F^{kl}(\vec{x}) = \vec{x} + \vec{c}(lr)$ and $F^{kl}(\vec{y}) = \vec{y} + \vec{c}(ks)$. It follows that $lr = \chi_{F^{kl}}(\vec{x})$ and $ks = \chi_{F^{kl}}(\vec{y})$. But by Theorem 5.1 applied to F^{kl}, we see that $lr = ks$. Hence $\chi_F(\vec{x}) = r/k = s/l = \chi_F(\vec{y})$.

QED

With this result in place, we can dispense with the dependence of the cycle time on the chosen point in $\mathbf{R}^n$.

Definition 5.2 *Let F be any min-max function. If F is eventually periodic somewhere, then the cycle time of F, $\chi(F)$, is its cycle time at any eventually periodic point. $\chi(F)$ is undefined if F is not eventually periodic somewhere.*

Theorem 5.1 should be thought of as the generalization to min-max functions of the Perron-Frobenius theorem for max-plus matrices discussed in §4, [BCOQ92, Theorem 3.23]. The hypothesis of irreducibility, or balance, is replaced by the weaker hypothesis that F has a fixed point. Note that we do not have to use normal projections to calculate the cycle time.

The conclusion of Theorem 5.1 is relatively weaker than that of Proposition 5.1. For instance, if a min-max function has a periodic point of period $k > 1$ but is not known to have a fixed point, then Theorem 5.1 cannot be used to calculate $\chi(F)$ in terms of some set of projections of F. We could, of course, use the projections of F^k but this is usually very inconvenient. However, if F is max-only then Proposition 5.1 gives us a formula for $\chi(F)$ no matter what the period. The missing ingredient in Theorem 5.1 is some analogue of Lemma 5.1, which seems difficult to find. Alternatively, we could hope that later research will confirm the connection between periodicity and fixed points which is suggested by Theorem 3.3. In that case we would know that Theorem 5.1 always gives the right answer.

By way of illustration of the results in this paper we reconsider some of the work of [Ols91] and [Ols93]. Suppose first that F is a min-max function of dimension $n+m$ such that F_i is max-only for $1 \leq i \leq n$. (If F has max-only components which are distributed differently, we can always renumber the variables to bring it into this form.) We do not care, for the moment, about the syntactic structure of F_i when $n+1 \leq i \leq n+m$. Express F in conjunctive form and let us use the same notation as in (16). Let A be the $n \times n$ matrix such that $A_{ij} = A^i_{ij}$ for $1 \leq i, j \leq n$. Note that A is not necessarily the matrix of a min-max function because, for instance, A_{i*} could be $\vec{c}(-\infty)$ for some row i.

Lemma 5.3 *With the assumptions above, if F has a fixed point then $\chi(F) \geq \mu(A)$.*

Proof: Let P be the set of max-only projections corresponding to the chosen conjunctive form. Suppose that $A_2 \in P$. It is easy to see that A must appear as the top, left-hand $n \times n$ sub-matrix of A_2. In particular, any circuit of $\mathcal{G}(A)$ can be considered a circuit of $\mathcal{G}(A_2)$ with the same cycle mean. Hence $\mu(A_2) \geq \mu(A)$. This holds for any $A_2 \in P$. Now suppose that F has a fixed point. It follows from (17) that $\chi(F) \geq \mu(A)$.

QED

Now suppose that, in addition to the assumptions above, F_i is min-only for $n+1 \leq i \leq n+m$. We may call such a function separated, since different constraints never appear in the same expression. Place F in disjunctive form and let Q be the corresponding set of min-only projections. We can, in similar fashion, identify an $m \times m$ matrix, B, in min-plus algebra, which corresponds to the bottom, right-hand $m \times m$ sub-matrix of any min-only projection in Q. It follows from Lemma 5.3 and its dual that if F has a fixed point then

$$\mu(A) \leq \chi(F) \leq \eta(B). \tag{19}$$

This is identical to the necessary condition of [Ols91, Theorem 2.1] but does not require such strong assumptions: the proof in [Ols91] relies on the irreducibility of A and B. The fact that, in the presence of these same assumptions, (19) is also sufficient for the existence of a fixed point—this is the main result of [Ols91]—is beyond the scope of this paper.

In [Ols93] more detailed information is presented about separated functions in dimension 2. The only case of interest is example (13) which we looked at in §3. Theorem 16 of [Ols93] states that, under the assumptions made in §3, if $a < (b+c)/2 < d$ then for any $\vec{x} \in \mathbf{R}^2$, there exists $K \geq 0$ such that, for all $k \geq K$,

$$F^{k+2}(x_1, x_2) = F^k(x_1, x_2) + \vec{c}(b+c).$$

(The constant is stated to be $(b+c)/2$ in [Ols93] but this is a misprint.) In our language, F is eventually periodic everywhere with period at most 2 and cycle time $(b+c)/2$. Now we have already seen that Proposition 3.1 implies that this function has a fixed point if, and only if, $a \leq d$. Furthermore, by Theorem 3.3, F is eventually periodic everywhere, with period at most 2, if, and only if, the same conditions hold. Note that the values of b and c play no role so far. Finally, it is easy to see that

$$\{ \begin{pmatrix} a & b \\ c & -\infty \end{pmatrix}, \begin{pmatrix} a & b \\ -\infty & d \end{pmatrix} \}$$

is a set of max-only projections of (13). It follows from Theorem 5.1 and (17) that, when $a \leq d$,

$$\chi(F) = (a \vee (b+c)/2) \wedge (a \vee d) = (a \vee (b+c)/2) \wedge d.$$

In particular, if $a < (b+c)/2 < d$, then $\chi(F) = (b+c)/2$ which demonstrates Theorem 16 of [Ols93].

6 Conclusion

The theory of min-max functions is in its infancy and has some way to go before reaching the level of maturity of max-plus algebra. What we have tried to do in this paper is to set up the basic definitions, give a thorough treatment of the 2-dimensional case, establish the connections with max-plus algebra and find a simple generalization of the fundamental formula of the maximum cycle mean. There are many open problems. Perhaps the most interesting and difficult of these is to extend the results on eventual periodicity in dimension 2 to higher dimensions. We believe that the picture we have sketched in §3 will be useful to anyone contemplating an attack on that problem. Even if the details of the high-dimensional case turn out to be more subtle—painful experience has taught us the wisdom of not making conjectures—it will give the prospective attacker something to aim at. The work in §4 suggests that some of the well-known concepts and results of max-plus algebra should be reconsidered in the light of new results on min-max functions. Finally, we believe that the projections which we introduced in §5 will prove to be a key tool in analysing the deeper properties of min-max functions. Indeed, recent work has already demonstrated their significance, [Gun94a, Gun94b].

References

[BCGZ84] R. E. Burkard, R. A. Cuninghame-Green, and U. Zimmermann, editors. *Algebraic and Combinatorial Methods in Operations Research*, volume 19 of *Annals of Discrete Mathematics*. North-Holland, 1984.

[BCOQ92] F. Baccelli, G. Cohen, G. J. Olsder, and J.-P. Quadrat. *Synchronization and Linearity*. Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, 1992.

[BO93] J. G. Braker and G. J. Olsder. The power algorithm in max algebra. *Linear Algebra and its Applications*, 182:67–89, 1993.

[Bur90] S. M. Burns. *Performance Analysis and Optimization of Asynchronous Circuits.* PhD thesis, California Institute of Technology, 1990.

[CDQV85] 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.

[CG62] R. A. Cuninghame-Green. Describing industrial processes with interference and approximating their steady-state behaviour. *Operational Research Quarterly*, 13(1):95–100, 1962.

[CG79] R. A. Cuninghame-Green. *Minimax Algebra*, volume 166 of *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, 1979.

[CG91] R. A. Cuninghame-Green. Minimax algebra and applications. *Fuzzy Sets and Systems*, 41:251–267, 1991.

[CKR84] Z.-Q. Cao, K. H. Kim, and F. W. Roush. *Incline algebra and applications*. Mathematics and its Applications. Ellis-Horwood, 1984.

[CMQV89] G. Cohen, P. Moller, J.-P. Quadrat, and M. Viot. Algebraic tools for the performance evaluation of discrete event systems. *Proceedings of the IEEE*, 77(1):39–58, 1989.

[Dev89] R. L. Devaney. *An Introduction to Chaotic Dynamical Systems*. Addison-Wesley Publishing Company, Inc., 1989.

[Gal86] J. H. Gallier. *Logic for Computer Science*. Computer Science and Technology Series. Harper and Row, 1986.

[Gau92] S. Gaubert. *Théorie des Systèmes Linéaires dans les dioides*. PhD thesis, L'école Nationale Supérieure des mines de Paris, 1992.

[GM] M. Gondran and M. Minoux. Linear algebra in dioids: a survey of recent results. Appears in [BCGZ84].

[Gun93a] J. Gunawardena. Min-max functions, Part I. Technical Report STAN-CS-93-1474, Department of Computer Science, Stanford University, May 1993.

[Gun93b] J. Gunawardena. Periodic behaviour in timed systems with $\{\mathsf{AND}, \mathsf{OR}\}$ causality. Part I: systems of dimension 1 and 2. Technical Report STAN-CS-93-1462, Department of Computer Science, Stanford University, February 1993.

[Gun93c] J. Gunawardena. Timing analysis of digital circuits and the theory of min-max functions. In *TAU'93, ACM International Workshop on Timing Issues in the Specification and Synthesis of Digital Systems*, September 1993.

[Gun94a] J. Gunawardena. Cycle times and fixed points of min-max functions. To appear in the Proceedings of the *11th International Conference on Analysis and Optimization of Systems*, Sophia-Antipolis, France, June 1994.

[Gun94b] J. Gunawardena. A dynamic approach to timed behaviour. Draft, February 1994.

[Ho89] Y. C. Ho, editor. *Special issue on Dynamics of Discrete Event Systems*. Proceedings of the IEEE, 77(1), January 1989.

[Joh88] C. R. Johnson. Combinatorial matrix analysis: an overview. *Linear Algebra and its Applications*, 107:3–15, 1988.

[Kar78] R. M. Karp. A characterization of the minimum cycle mean in a digraph. *Discrete Mathematics*, 23:309–311, 1978.

[Mas87] V. P. Maslov. On a new principle of superposition for optimization problems. *Russian Math Surveys*, 42(3):43–54, 1987.

[MS92] V. P. Maslov and S. N. Samborskiĭ, editors. *Idempotent Analysis*, volume 13 of *Advances in Soviet Mathematics*. American Mathematical Society, 1992.

[Ols91] G. J. Olsder. Eigenvalues of dynamic max-min systems. *Discrete Event Dynamic Systems*, 1:177–207, 1991.

[Ols93] G. J. Olsder. Analysis of min-max systems. Rapports de Recherche 1904, INRIA, May 1993.

[Rei68] R. Reiter. Scheduling Parallel Computations. *Journal of the ACM*, 15(4):590–599, 1968.

[RH80] 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.

[Zim81] U. Zimmermann. *Linear and Combinatorial Optimization in Orderd Algebraic Structures*, volume 10 of *Annals of Discrete Mathematics*. North-Holland, 1981.

A Normal form

Proof (of Theorem 2.1): Suppose that f and g are min-max expressions of n variables which are both in conjunctive normal form:

$$f \equiv f_1 \wedge \cdots \wedge f_l, \qquad g \equiv g_1 \wedge \cdots \wedge g_m, \tag{20}$$

where f_i and g_i are max-only expressions in conjunctive normal form, as in (4), and $f_i \not\leq f_j$ and $g_i \not\leq g_j$ for $i \neq j$. Suppose that $f = g$. We are required to prove that, up to re-ordering of the conjunctions in g or f, $f \equiv g$.

We shall consider first the situation in which $l = 1$. We then have

$$f = g_1 \wedge \cdots \wedge g_m, \tag{21}$$

where $f \equiv a_1 + x_1 \vee \cdots \vee a_n + x_n$. We claim that, in this case, $f = g_i$ for some $1 \leq i \leq m$. The proof of this is by induction on n, the number of variables. If $n = 1$ then (21) becomes

$$a_1 + x_1 = b_1 + x_1 \wedge \cdots \wedge b_m + x_1$$

where $m \leq n$ and the claim follows easily. Now assume that the claim is true for $n < k$, where $k > 1$ and consider (21) when $n = k$. There exists some $a_i \neq -\infty$, since any min-max expression has at least one variable. Assume, without loss of generality, that this is a_k. We can then write (21) as

$$h \vee a_k + x_k = (h_1 \vee b_1 + x_k) \wedge \cdots \wedge (h_m \vee b_m + x_k), \tag{22}$$

where $b_i \in \mathbf{R} \cup \{-\infty\}$ and h, h_i are max-only expressions of dimension $k - 1$.

Choose arbitrary values for the variables other than x_k. By letting x_k become sufficiently small, (22) becomes

$$h = h_1 \wedge \cdots \wedge h_m.$$

This holds for any values of the variables $x_1, \cdots, x_{k-1}$. Hence we may apply the inductive hypothesis. We deduce that h equals one of the h_i and $h \leq h_i$ for all $1 \leq i \leq m$. It follows that, we can, by suitably re-ordering, rewrite (22) as

$$h \vee a_k + x_k = (h \vee b_1 + x_k) \wedge \cdots \wedge (h \vee b_r + x_k) \wedge (h_{r+1} \vee b_{r+1} + x_k) \wedge \cdots \wedge (h_m \vee b_m + x_k), \quad (23)$$

where $h < h_i$ for $r+1 \leq i \leq m$. (The notation $f < g$ for min-max functions, means that $f \leq g$ but $f \neq g$. That is to say, there exists some point $\vec{x}$ such that $f(\vec{x}) < g(\vec{x})$. However, it does not rule out the possibility that there are other points, $\vec{y}$, where $f(\vec{y}) = g(\vec{y})$.) By letting x_k become sufficiently large in (23) we see that $b_i \in \mathbf{R}$ for all $1 \leq i \leq m$ and (23) becomes

$$a_k + x_k = b_1 + x_k \wedge \cdots \wedge b_m + x_k.$$

Hence, a_k equals one of the b_i and $a_k \leq b_i$ for all $1 \leq i \leq m$. If $a_k = b_i$ for some $1 \leq i \leq r$, then we are done. So suppose that $a_k < b_i$ for each $1 \leq i \leq r$. We shall derive a contradiction.

We can, by re-ordering once again, rewrite (23) in the form

$$h \vee a_k + x_k = (h \vee b_1 + x_k) \wedge \cdots \wedge (h \vee b_r + x_k) \wedge \cdots \wedge (h_s \vee a_k + x_k) \wedge \cdots \wedge (h_m \vee a_k + x_k), \quad (24)$$

where $1 \leq r < s \leq m$, $a_k < b_i$ for $1 \leq i < s$ and $h < h_i$ for $r < i \leq m$.

As before, choose arbitrary values for the variables other than x_k. Choose x_k so that $h = a_k + x_k$, which we may always do. Under these conditions, we may simplify (24) so that

$$h = b_1 + x_k \wedge \cdots \wedge b_r + x_k \wedge h_s \wedge \cdots \wedge h_m, \quad (25)$$

where the terms $h_i \vee b_i + x_k$ for $r < i < s$ do not contribute because $b_i > a_k$. We can now substitute $x_k = h - a_k$ and rewrite (25) as

$$h = (b_1 \wedge \cdots \wedge b_r) - a_k + h \wedge h_s \wedge \cdots \wedge h_m.$$

Because this holds for any values of the variables $x_1, \cdots, x_{k-1}$, we may apply the inductive hypothesis once again. Since $h < h_i$ for $s \leq i \leq m$ and $a_k < b_i$ for $1 \leq i \leq r$, it must be the case that $b_1 \wedge \cdots \wedge b_r = a_k$. But then $a_k = b_i$ for some $1 \leq i \leq r$, which contradicts the constraints of (24). Hence $f = g_i$ for some $1 \leq i \leq m$ and the claim follows by induction.

We can now return to the general case, (20). Since $f = g$ by hypothesis, we deduce that

$$(f_1 \vee g_1) \wedge \cdots \wedge (f_l \vee g_1) = f \vee g_1 = g \vee g_1 = g_1.$$

Since $f_i \vee g_1$ is max-only, we may apply the first part above to deduce that $g_1 = f_p \vee g_1$ for some $1 \leq p \leq l$. Hence, $f_p \leq g_1$. But we can apply the same technique to f_p to deduce that $g_q \leq f_p$ for some $1 \leq q \leq m$. Hence, $g_q \leq g_1$ and so, by the requirements of normal form, $q = 1$ and $f_p = g_1$. Proceeding in this way, we can show that $l = m$ and each g_i is equal to one, and only one, f_j. We have already seen, by Lemma 2.1, that conjunctive normal form for max-only expressions is unique. Hence it follows that, by suitably re-ordering the g_i or the f_j, $f \equiv g$.

QED