DAIR
DIAS Access to Institutional Repository

DIAS
Institiúid Ard-Léinn Bhaile Átha Cliath | Dublin Institute for Advanced Studies

Title	Large Deviations of Products of Random Topical Operators
Creators	Toomey, F.
Date	1999
Citation	Toomey, F. (1999) Large Deviations of Products of Random Topical Operators. (Preprint)
URL	https://dair.dias.ie/id/eprint/564/
DOI	DIAS-STP-99-04

DIAS-STP-99-04

LARGE DEVIATIONS OF PRODUCTS OF RANDOM TOPICAL OPERATORS

FERGAL TOOMEY
Dublin Institute for Advances Studies,
10 Burlington Road, Dublin 4, Ireland.
toomey@stp.dias.ie

Abstract

A topical operator on $\mathbb{R}^d$ is one which is isotone and homogeneous. Let $\{A(n) : n \geq 1\}$ be a sequence of i.i.d. random topical operators such that the projective radius of $A(n) \cdots A(1)$ is almost surely bounded for large n. If $\{x(n) : n \geq 1\}$ is a sequence of vectors given by $x(n) = A(n) \cdots A(1)x_0$, for some fixed initial condition x_0, then the sequence $\{x(n)/n : n \geq 1\}$ satisfies a weak large deviation principle. As corollaries of this result we obtain large deviation principles for products of certain random aperiodic max-plus and min-plus matrix operators, and for products of certain random aperiodic non-negative matrix operators.

1 Topical Operators

An operator $A : \mathbb{R}^d \to \mathbb{R}^d$ is *homogeneous* if it satisfies $A(x + a1) = Ax + a1$ for all $x \in \mathbb{R}^d$ and $a \in \mathbb{R}$, where 1 is the vector in $\mathbb{R}^d$ with all components equal to one. A is *isotone* if it satisfies $Ax \leq Ay$ whenever $x \leq y$ (the order here and throughput this paper is the product order on $\mathbb{R}^d$). An operator which is both homogeneous and isotone is called *topical*. This terminology was introduced by J. Gunawardena and M. Keane [GK95], who proposed the class of topical operators as a setting for the study of certain properties of discrete event systems. In this context, one considers recursive systems of equations of the form

$$x(n) = A(n)x(n-1), \qquad n = 1, 2, \ldots, \tag{1}$$

with the interpretation that $x(n) \in \mathbb{R}^d$ is a vector whose entries represent timing data: $x_i(n)$ is the time of the nth event of some type i, where d is the number of types of event which may occur. The operators $A(n) : \mathbb{R}^d \to \mathbb{R}^d$ determine the delays and synchronisation constraints present between events. Homogeneity of these operators reflects invariance of the system's dynamics under a shift in the origin of the time axis. Isotonicity of the $A(n)$'s implies that the system is monotonic, in the sense that if some events were to be artificially delayed, then all subsequent events would also be delayed, or at best they would occur no sooner than originally. For more information on topical operators and their application to discrete event systems see [GK95, Gun96] and references therein.

Well-known examples of topical operators include the max-plus and min-plus matrix operators, which are defined as follows. $A : \mathbb{R}^d \to \mathbb{R}^d$ is a *max-plus matrix operator* if it takes the form

$$(Ax)_i = \max_{j=1,\dots,d} A_{ij} + x_j, \qquad i = 1, \dots, d$$

for every $x \in \mathbb{R}^d$, where $\{A_{ij} : i, j = 1, \dots, d\}$ are elements of $\mathbb{R} \cup \{-\infty\}$. (We assume that each row of the matrix $\{A_{ij}\}$ has at least one entry different from $-\infty$, so that the image of $\mathbb{R}^d$ under A is contained in $\mathbb{R}^d$). A *min-plus matrix operator* is one which takes the form

$$(Ax)_i = \min_{j=1,\dots,d} A_{ij} + x_j, \qquad i = 1, \dots, d$$

for each $x \in \mathbb{R}^d$, where now $\{A_{ij} : i, j = 1, \dots, d\}$ are elements of $\mathbb{R} \cup \{+\infty\}$ (again with the caveat that each row of $\{A_{ij}\}$ has at least one finite entry). Matrix operators of these kinds arise in the theory of Markov decision processes and timed event graphs. A general reference is the book of F. Baccelli *et al.* [BCOQ92]. If we take a finite pointwise infimum of max-plus matrix operators, or a finite pointwise supremum of min-plus matrix operators, we obtain an operator which is again topical, known as a *min-max operator.* In the context of discrete event systems min-max operators were introduced and studied by G. J. Olsder [Ols91] and J. Gunawardena [Gun94].

Another interesting class of topical operators can be constructed from the isotone linear operators on the positive cone $\mathbb{R}^d_+$, in the following way [Gun96]. Let $\exp : \mathbb{R}^d \to \mathbb{R}^d_+$ be the componentwise exponential function and $\log : \mathbb{R}^d_+ \to \mathbb{R}^d$ the componentwise logarithm: $\exp(x)_i := \exp(x_i)$ and $\log(x)_i := \log(x_i)$. If $A : \mathbb{R}^d_+ \to \mathbb{R}^d_+$ is isotone and satisfies $A(ax) = aAx$ for all $x \in \mathbb{R}^d$ and $a \in \mathbb{R}_+$, then the operator $\tilde{A} : \mathbb{R}^d \to \mathbb{R}^d$ defined by $\tilde{A}x := \log(A\exp(x))$ is topical. A might be, for example, a non-negative matrix operator with at least one non-zero entry per row.

Our purpose in this paper is to study the large deviations of sequences $\{x(n) : n \geq 1\}$ which satisfy recursions of the form (1), in the case when

$\{A(n) : n \geq 1\}$ is a random sequence of i.i.d. topical operators. The approach we will take requires an assumption that the $A(n)$'s satisfy a certain range condition, which we now state.

Let t and b denote the *top* and *bottom* functions on $\mathbb{R}^d$:

$$\mathsf{t}[x] := \max_i \; x_i; \qquad \mathsf{b}[x] := \min_i \; x_i.$$

J. Gunawardena and M. Keane [GK95] show that $A : \mathbb{R}^d \to \mathbb{R}^d$ is topical if and only if it is non-expansive in t:

$$\mathsf{t}[Ax - Ay] \leq \mathsf{t}[x - y] \qquad \forall x, y; \tag{2}$$

and if and only if it is non-contractive in b:

$$\mathsf{b}[Ax - Ay] \geq \mathsf{b}[x - y] \qquad \forall x, y. \tag{3}$$

Together these inequalities imply that topical operators are non-expansive in the l_∞-norm on $\mathbb{R}^d$:

$$||Ax - Ay|| \leq ||x - y|| \qquad \forall x, y,$$

where $||x|| = \max_i \; |x_i| = \mathsf{t}[x] \vee (-\mathsf{b}[x])$. In fact, M. G. Crandall and L. Tartar have shown [CT80] that a homogeneous operator on $\mathbb{R}^d$ is isotone if and only if it is l_∞ non-expansive. Inequalities (2) and (3) also imply that topical operators are non-expansive in the *projective semi-norm* $|| \cdot ||_{\mathsf{P}}$ defined by

$$||x||_{\mathsf{P}} = \mathsf{t}[x] - \mathsf{b}[x].$$

We define the *projective radius* of a topical operator A to be the extended real number

$$\Pi[A] := \sup_{x \in \mathbb{R}^d} ||Ax||_{\mathsf{P}}.$$

Note that the projective radius of a translation operator, for example, is $+\infty$. The interest of projective radius is that, if A has finite projective radius, then there exists a vector $x \in \mathbb{R}^d$ and a scalar a such that $Ax = x + a1$ [BM96]: such a vector is sometimes called a *generalised fixed point* of A. Finite projective radius is not, however, a necessary condition for the existence of a generalised fixed point. More details and references on the fixed point properties of various types of topical operators can be found in [BCOQ92, GG98].

Turning to sequences of random operators, an important result is the following ergodic theorem due to F. Baccelli and J. Mairesse.

Theorem. [BM96] *Let $\{A(n) : n \geq 1\}$ be a stationary and ergodic sequence of random topical operators. If there is an integer N and a real number C such that*

$$\Pi[A(N) \cdots A(1)] \leq C$$

with positive probability, then there exists $\gamma \in \mathbb{R}$ such that

$$\lim_{n\to\infty} \frac{1}{n} A(n) \cdots A(1) x_0 = \gamma 1$$

almost surely, for every $x_0 \in \mathbb{R}^d$.

In this paper we study deviations from the behaviour described by this theorem, but under the following stronger assumption:

Assumption 1 *(a) $\{A(n) : n \geq 1\}$ is a random sequence of i.i.d. topical operators; (b) there exists an integer N and a real number C such that $\Pi[A(N) \cdots A(1)] \leq C$ almost surely.*

Our main result is that if this assumption holds for the sequence $\{A(n)\}$, and if $\{x(n) : n \geq 1\}$ is a sequence of vectors satisfying the recursive system (1) for some fixed initial condition, then the sequence $\{x(n)/n : n \geq 1\}$ satisfies a weak large deviation principle. The associated rate function is equal to $+\infty$ away from the line $x = a1$, $a \in \mathbb{R}$, so that at this scale the system's behaviour is effectively one-dimensional. This confinement is the consequence of part (b) of assumption 1. We also present some results to characterise the rate function, but explicit calculations turn out to be difficult in all but trivial cases. It is well known that calculation of the Lyapunouv exponent γ of the ergodic theorem is already a hard problem.

For the case of max-plus and min-plus matrix operators, our results extend previous work by F. Baccelli and T. Konstantopoulos [BK91], P. Glassermann and D. D. Yao [GY95], and C.-S. Chang [Cha96].

2 Large Deviations

Let $\{A(n) : n \geq 1\}$ be a sequence of random topical operators on $\mathbb{R}^d$. Given a fixed initial vector $x_0 \in \mathbb{R}^d$ we let $\{x(n; x_0) : n \geq 1\}$ be the sequence defined by

$$x(n;x_0) := A(n)A(n-1)\cdots A(1)x_0, \qquad n = 1,2,\ldots$$

It will be convenient to let $A(l,m)$ stand for the product of the $A(n)$'s from $n = l+1$ up to $n = m$, where $m > l \geq 0$:

$$A(l,m) := A(m)A(m-1)\cdots A(l+1).$$

We shall also use $x(l,m)$ to denote the vector $A(l,m)x_0$.

With the assumption that x_0 is a fixed, rather than random, initial condition, the non-expansive property of the $A(n)$'s ensures that the large deviations of the sequence $\{x(n;x_0)/n : n \geq 1\}$ are in fact independent of x_0. If y_0 is another fixed initial condition, then

$$\|x(n;x_0) - x(n;y_0)\| = \|A(0,n)x_0 - A(0,n)y_0\| \leq \|x_0 - y_0\|,$$

implying that, for any $\epsilon > 0$,

$$\mathbb{P}\Big(\|x(n;x_0) - x(n;y_0)\| > n\epsilon\Big) = 0$$

for n large enough. The sequences $\{x(n;x_0)/n\}$ and $\{x(n;y_0)/n\}$ are therefore exponentially equivalent ([DZ98], chapter 4), so that one satisfies a large deviation principle if and only if the other does, and with the same rate function. We set x_0 equal to the zero vector 0 and suppress the dependence of $x(n;x_0)$ on x_0 henceforth.

Let $\mathbb{M}_n$ be the law of $x(n)/n$, and let $\overline{m}$ and $\underline{m}$ be set functions defined on the Borel subsets of $\mathbb{R}^d$ by

$$\begin{aligned}\overline{m}[B] &:= \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{M}_n[B],\\ \underline{m}[B] &:= \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{M}_n[B].\end{aligned}$$

The *upper* and *lower deviation functions* $\overline{\mu}$ and $\underline{\mu}$ associated with the sequence $\{\mathbb{M}_n : n \geq 1\}$ are the maps from $\mathbb{R}^d$ into $[-\infty, 0]$ given by

$$\begin{aligned}\overline{\mu}(x) &:= \inf_{G \ni x} \overline{m}[G],\\ \underline{\mu}(x) &:= \inf_{G \ni x} \underline{m}[G],\end{aligned}$$

where the infima on the right-hand sides are taken over all open sets G containing the point x. As both $\overline{m}$ and $\underline{m}$ are increasing set functions these infima may in fact be taken over any base of Borel neighbourhoods of x. The properties of $\overline{\mu}$ and $\underline{\mu}$ are discussed in the review of J. T. Lewis and C.-F. Pfister [LP95]. They are upper semi-continuous and for all open sets G satisfy

$$\begin{aligned} \overline{m}[G] &\geq \sup_{x\in G} \overline{\mu}(x), \\ \underline{m}[G] &\geq \sup_{x\in G} \underline{\mu}(x). \end{aligned}$$

In addition, $\overline{\mu}$ satisfies

$$\overline{m}[K] \leq \sup_{x\in K} \overline{\mu}(x)$$

for all compact sets K.

The sequence $\{\mathbb{M}_n : n \geq 1\}$ satisfies a *weak large deviation principle* with rate function l if and only if l is lower semi-continuous and the inequalities

$$\begin{aligned} \overline{m}[K] &\leq -\inf_{x\in K} l(x), \\ \underline{m}[G] &\geq -\inf_{x\in G} l(x), \end{aligned}$$

hold for all compact sets K and open sets G. A neccessary and sufficient condition for the weak l.d.p. to hold with rate function l is that $\overline{\mu}$ and $\underline{\mu}$ should coincide and be equal to $-l$ throughout $\mathbb{R}^d$ [LP95]. $\{\mathbb{M}_n : n \geq 1\}$ satisfies a *large deviation principle* with rate function l if and only if it satisfies a weak l.d.p. in which the upper bound for compact sets extends to all closed subsets F of $\mathbb{R}^d$:

$$\overline{m}[F] \leq -\inf_{x\in F} l(x).$$

Lemma 1 below establishes that, if the $A(n)$'s are i.i.d., then $\overline{\mu}(x) = \underline{\mu}(x)$ on the line $x = a1$, $a \in \mathbb{R}$. The argument is based on the following observation: if A_1, A_2 are any pair of topical operators and a, b are any pair of real numbers, then

$$\begin{aligned} &\|A_1 A_2 \cdot 0 - (a+b)1\| \\ &\quad \leq \|A_1 A_2 \cdot 0 - A_1 \cdot a1\| + \|A_1 \cdot a1 - (a+b)1\| \\ &\quad \leq \|A_2 \cdot 0 - a1\| + \|A_1 \cdot 0 - b1\|. \end{aligned} \tag{4}$$

Let $B_r(a1)$ denote the l_∞-ball of radius r centred at $a1$. Since $x(n+m) = A(n, n+m)A(0,n)\cdot 0$, the inequality (4) implies that the sequence

$\{\mathbb{M}_n[B_r(a1)]\}$ is super-multiplicative:

$$\begin{aligned}
&\mathbb{M}_{n+m}[B_r(a1)] \\
&\quad = \mathbb{P}\Big(\|x(n+m) - (n+m)a1\| < (n+m)r\Big) \\
&\quad \geq \mathbb{P}\Big(\|x(n) - na1\| < nr,\ \|x(n, n+m) - ma1\| < mr\Big) \\
&\quad = \mathbb{M}_n[B_r(a1)]\mathbb{M}_m[B_r(a1)].
\end{aligned}$$

A variant of the standard sub-additivity lemma (see [Lan73, LPS94]) may now be used to show that $\overline{\mu}(a1) = \underline{\mu}(a1)$, and also that the resulting function $a \longmapsto \overline{\mu}(a1) = \underline{\mu}(a1)$ is concave.

Lemma 1 *Under part (a) of assumption 1, $\underline{\mu}(a1) = \overline{\mu}(a1)$ for each $a \in \mathbb{R}$.*

PROOF Fix $a \in \mathbb{R}$ and put $n = ps + q$, where $s > 0$, $p > 0$, and $0 \leq q < s$. We have from (4) that

$$\|x(n) - na1\| \leq \|x(ps, ps+q) - qa1\| + \|x(0, ps) - psa1\|,$$

and, continuing the expansion,

$$\begin{aligned}
\|x(n) - na1\| \quad \leq \quad & \|x(ps, ps+q) - qa1\| \\
& + \sum_{k=1}^{p} \|x((k-1)s, ks) - sa1\|. \qquad (5)
\end{aligned}$$

Let $z_s(k)$ denote the contribution coming from the kth block of size s:

$$z_s(k) := x((k-1)s, ks).$$

Then $\{z_s(k) : k \geq 1\}$ is a sequence of i.i.d. random variables and the law of $z_s(1)$ is $\mathbb{M}_s$. It follows from (5) that, for each $\epsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\Big(\|x(n) - na1\| < nr\Big) \\
&\quad \geq \mathbb{P}\Big(\|x(ps, ps+q) - qa1\| < n\epsilon,\ \sum_{k=1}^{p} \|z_s(k) - sa1\| < n(r-\epsilon)\Big) \\
&\quad \geq \mathbb{P}\Big(\|x(q) - qa1\| < n\epsilon\Big) \Big[\mathbb{P}\Big(\|z_s(1) - sa1\| < s(r-\epsilon)\Big)\Big]^p,
\end{aligned}$$

and therefore

$$\begin{aligned}
&\frac{1}{n}\log \mathbb{M}_n[B_r(a1)] \geq \\
&\qquad \frac{1}{n}\log \mathbb{P}\Big(\|x(q) - qa1\| < n\epsilon\Big) + \frac{p}{n}\log \mathbb{M}_s[B_{r-\epsilon}(a1)]. \qquad (6)
\end{aligned}$$

Now, since any finite collection of probability measures on $\mathbb{R}^d$ is tight, we can find a compact set $K \subset \mathbb{R}^d$ such that $x(q)$ falls in K with positive probability for each for each $q = 0, \dots, s-1$. Define

$$\alpha_s := \min_{0 \le q < s} \mathbb{P}(x(q) \in K).$$

Then $\alpha_s > 0$, and there exists $M < \infty$ such that for all $n \ge M$ and each $q = 0, \dots, s-1$,

$$\mathbb{P}\Big(\|x(q) - qa\mathbf{1}\| < n\epsilon\Big) \ge \mathbb{P}(x(q) \in K) \ge \alpha_s.$$

Returning to inequality (6), this yields

$$\begin{aligned} \frac{1}{n} \log \mathbb{M}_n[B_r(a\mathbf{1})] &\ge \frac{1}{n} \log \alpha_s + \frac{p}{n} \log \mathbb{M}_s[B_{r-\epsilon}(a\mathbf{1})] \\ &\ge \frac{1}{n} \log \alpha_s + \frac{1}{s} \log \mathbb{M}_s[B_{r-\epsilon}(a\mathbf{1})], \end{aligned}$$

with $\log \alpha_s > -\infty$. Taking first the liminf in n and then the limsup in s we obtain

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{M}_n[B_r(a\mathbf{1})] \ge \limsup_{s\to\infty} \frac{1}{s} \log \mathbb{M}_s[B_{r-\epsilon}(a\mathbf{1})],$$

or $\underline{m}[B_r(a\mathbf{1})] \ge \overline{m}[B_{r-\epsilon}(a\mathbf{1})]$. The statement of the lemma follows on taking infima over $r > \epsilon$, giving

$$\inf_{r>\epsilon} \underline{m}[B_r(a\mathbf{1})] \ge \overline{\mu}(a\mathbf{1}),$$

and then over $\epsilon > 0$, to get $\underline{\mu}(a\mathbf{1}) \ge \overline{\mu}(a\mathbf{1})$. ∎

Lemma 2 *The map $a \longmapsto \underline{\mu}(a\mathbf{1}) = \overline{\mu}(a\mathbf{1})$ resulting from lemma 1 is concave.*

PROOF For $a_1, a_2 \in \mathbb{R}$, we have from (4) that

$$\|x(2n) - n(a_1 + a_2)\mathbf{1}\| \le \|x(n) - na_1\mathbf{1}\| + \|x(n, 2n) - na_2\mathbf{1}\|,$$

which implies

$$\mathbb{M}_{2n}[B_r((a_1\mathbf{1} + a_2\mathbf{1})/2)] \ge \mathbb{M}_n[B_r(a_1\mathbf{1})]\, \mathbb{M}_n[B_r(a_2\mathbf{1})].$$

Therefore,

$$\begin{aligned} \overline{m}[B_r((a_1\mathbf{1} + a_2\mathbf{1})/2)] &\ge \liminf_{n\to\infty} \frac{1}{2n} \log \mathbb{M}_{2n}[B_r((a_1\mathbf{1} + a_2\mathbf{1})/2)] \\ &\ge \frac{1}{2}\, \underline{m}[B_r(a_1\mathbf{1})] + \frac{1}{2}\, \underline{m}[B_r(a_2\mathbf{1})], \end{aligned}$$

and taking infima over $r > 0$ we get

$$\overline{\mu}((a_1 1 + a_2 1)/2) \geq \frac{1}{2}\ \underline{\mu}(a_1 1) + \frac{1}{2}\ \underline{\mu}(a_2 1).$$

This inequality may be extended to cover all convex combinations $\lambda a_1 1 + (1-\lambda)a_2 1$, where λ is a dyadic rational in $[0,1]$, by iterating the above argument. The concavity of the map $a \mapsto \underline{\mu}(a1) = \overline{\mu}(a1)$ then follows from the fact that it is upper semi-continuous. ∎

Lemma 1 is enough to establish a weak large deviation principle if both parts of assumption 1 are satisfied.

Theorem 3 *Let both parts of assumption 1 hold. The sequence* $\{\mathbb{M}_n : n \geq 1\}$ *satisfies a weak large deviation principle with a convex rate function* l *which is equal to* $+\infty$ *on the set* $\{x : \|x\|_{\mathsf{P}} > 0\}$.

PROOF For $\epsilon > 0$ and $n \geq N$,

$$\begin{aligned}\mathbb{P}\Big(\|x(n)\|_{\mathsf{P}} > n\epsilon\Big) &= \mathbb{P}\Big(\|A(n-N,n)x(n-N)\|_{\mathsf{P}} > n\epsilon\Big) \\ &\leq \mathbb{P}\Big(\Pi[A(n-N,n)] > n\epsilon\Big) \\ &= \mathbb{P}\Big(\Pi[A(0,N)] > n\epsilon\Big),\end{aligned}$$

which is zero for n sufficiently large. Therefore $\underline{\mu}(x) = \overline{\mu}(x) = -\infty$ for each x with $\|x\|_{\mathsf{P}} > 0$. Combining this with lemma 1 we have $\underline{\mu} = \overline{\mu}$ everywhere, and the resulting rate function $l = -\overline{\mu} = -\underline{\mu}$ is convex by lemma 2. ∎

The next two lemmas are directed towards proving that the rate function l is the convex dual of the scaled cumulant generating function λ of the sequence $\{x(n)\}$. For $\theta \in \mathbb{R}^d$, let $\{\mathbb{M}_n^\theta\}$ be the sequence of measures defined by

$$\mathbb{M}_n^\theta[B] := \int_B e^{n\langle\theta,\ x\rangle}\mathbb{M}_n[dx].$$

$\{\mathbb{M}_n^\theta : n \geq 1\}$ satisfies a weak large deviation principle with rate function l^θ given by $l^\theta(x) = l(x) - \langle\theta,\ x\rangle$ [LP95]. Let λ_n be the cumulant generating function of $x(n)$ (automatically proper, convex, and l.s.c.):

$$\lambda_n(\theta) := \log \mathbb{M}_n^\theta[\mathbb{R}^d] = \log \mathbb{E}e^{\langle\theta,\ x(n)\rangle}.$$

Lemma 4 shows that the limit $\lambda(\theta) := \lim_{n\to\infty} \lambda_n(\theta)/n$ exists, and lemma 5 shows that λ is the convex dual of l.

Lemma 4 also gives two expressions for λ which may be of use in approximating it. To state these, we define $\psi_n : \mathbb{R}^2 \to \mathbb{R}$ to be the map

$$\psi_n(\phi_1, \phi_2) := \log \mathbb{E} \exp\Big(\phi_1 \mathsf{t}[x(n)] + \phi_2 \mathsf{b}[x(n)]\Big).$$

For $\theta \in \mathbb{R}^d$ we let θ_+ represent the sum of the positive components of θ and θ_- the sum of the negative components.

Lemma 4 *Under assumption 1, $\lambda(\theta)$ exists for each $\theta \in \mathbb{R}^d$, and is given by*

$$\lambda(\theta) = \sup_{n\geq 1} \frac{1}{n}\psi_n(\theta_-, \theta_+)$$

and

$$\lambda(\theta) = \inf_{n\geq 1} \frac{1}{n}\psi_n(\theta_+, \theta_-).$$

PROOF For n, $m \geq 1$,

$$\begin{aligned} x(n+m) &= A(n, n+m)A(0,n)\mathsf{0} \\ &\leq A(n,n+m)(\mathsf{t}[A(0,n)\mathsf{0}]1) = A(n,n+m)\mathsf{0} + \mathsf{t}[A(0,n)\mathsf{0}]1, \end{aligned}$$

and similarly

$$x(n+m) \geq A(n,n+m)\mathsf{0} + \mathsf{b}[A(0,n)\mathsf{0}]1.$$

These yield the inequalities

$$\mathsf{t}[x(n+m)] \leq \mathsf{t}[x(n,n+m)] + \mathsf{t}[x(n)], \tag{7}$$

$$\mathsf{b}[x(n+m)] \geq \mathsf{b}[x(n,n+m)] + \mathsf{b}[x(n)], \tag{8}$$

which together imply that the sequence $\{\psi_n(\theta_-, \theta_+) : n \geq 1\}$ is super-additive:

$$\psi_{n+m}(\theta_-, \theta_+) \geq \psi_n(\theta_-, \theta_+) + \psi_m(\theta_-, \theta_+)$$

for all n, $m \geq 1$. Now ψ_n is the cumulant generating function of the pair of random variables $(\mathsf{t}[x(n)], \mathsf{b}[x(n)])$, and as such cannot take the value $-\infty$. It follows that the limit

$$\lim_{n\to\infty} \frac{1}{n}\psi_n(\theta_-, \theta_+) = \sup_{n\geq 1} \frac{1}{n}\psi_n(\theta_-, \theta_+)$$

exists for all θ.

Clearly $\lambda_n(\theta) \geq \psi_n(\theta_-, \theta_+)$ for all n. But for $n \geq N$, $x(n)$ satisfies $\mathsf{t}[x(n)] - \mathsf{b}[x(n)] \leq C$, so that $\mathsf{t}[x(n)] - x_i(n) \leq C$ and $x_i(n) - \mathsf{b}[x(n)] \leq C$ for each i. Therefore

$$\langle\theta,\ x(n)\rangle \leq \theta_+ \mathsf{b}[x(n)] + \theta_+ C + \theta_- \mathsf{t}[x(n)] - \theta_- C,$$

implying that

$$\lambda_n(\theta) \leq \psi_n(\theta_-, \theta_+) + \|\theta\|_1 C,$$

and

$$\lim_{n\to\infty} \frac{1}{n}\lambda_n(\theta) = \lim_{n\to\infty} \frac{1}{n}\psi_n(\theta_-, \theta_+).$$

To prove the second identity for λ we first note that the argument just given also establishes that

$$\psi_n(\theta_+, \theta_-) - \psi_n(\theta_-, \theta_+) \leq 2C\|\theta_1\|$$

for all $n \geq N$. Therefore $\psi_n(\theta_+, \theta_-)/n$ converges to $\lambda(\theta)$ as $n \to \infty$. Furthermore, inequalities (7) and (8) imply that $\{\psi_n(\theta_+, \theta_-)\}$ is a subadditive sequence: for all n, $m \geq 1$,

$$\psi_{n+m}(\theta_+, \theta_-) \leq \psi_n(\theta_+, \theta_-) + \psi_m(\theta_+, \theta_-);$$

hence

$$\lim_{n\to\infty} \frac{1}{n}\psi_n(\theta_+, \theta_-) = \inf_{n\geq 1} \frac{1}{n}\psi_n(\theta_+, \theta_-). \quad \blacksquare$$

Recall that the convex dual of l is the function $l^* : \mathbb{R}^d \to [-\infty, +\infty]$ given by

$$l^*(\theta) := \sup_{x\in\mathbb{R}^d} \{\langle\theta,\ x\rangle - l(x)\}.$$

Lemma 5 *Under assumption 1, λ is the convex dual of l.*

PROOF The sequence of measures $\{\mathbb{M}_n^\theta\}$ satisfies a weak large deviation principle with rate function $l(x) - \langle\theta,\ x\rangle$. Since $\mathbb{R}^d$ is an open set, the large deviation lower bound gives us

$$\lambda(\theta) = \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{M}_n^\theta[\mathbb{R}^d] \geq -\inf_x\{l(x) - \langle\theta,\ x\rangle\},$$

or $\lambda(\theta) \geq l^*(\theta)$. To prove the opposite inequality, let $z_m(n)$ be the random variable $x((n-1)m, nm)$, and let $y_m(n)$ denote the sum

$$y_m(n) := \frac{1}{nm} \sum_{k=1}^{n} z_m(k).$$

For fixed m, the sequence $\{z_m(n) : n \geq 1\}$ is i.i.d. Let $\mathbb{L}_n^{(m)}$ be the law of $y_m(n)$; by Cramér's theorem, the sequence $\{\mathbb{L}_n^{(m)}, n \geq 1\}$ satisfies a large deviation principle with convex rate function l_m equal to

$$l_m(x) = \sup_\theta \{\langle \theta,\ x \rangle - \lambda_m(\theta/m)\}.$$

Since both l_m and λ_m are proper convex l.s.c. functions, we also have

$$\lambda_m(\theta) = \sup_x \{\langle m\theta,\ x \rangle - l_m(x)\}.$$

Now for n, $m \geq N$ the two inequalities (7) and (8) imply that

$$\|x(n+m) - x(n, n+m) - x(n)\| \leq C.$$

Setting $n = pm$ with $m \geq N$ and applying this result repeatedly yields

$$\|x(n) - ny_m(p)\| = \|x(n) - \sum_{k=1}^{p} z_m(k)\| \leq mC.$$

Therefore

$$\begin{aligned} \mathbb{M}_n[B_{r+C/m}(x)] &= \mathbb{P}\Big(\|x(n) - nx\| < nr + pC\Big) \\ &\geq \mathbb{P}\Big(\|ny_m(p) - nx\| < nr\Big) \\ &= \mathbb{L}_p^{(m)}[B_r(x)]. \end{aligned}$$

Taking logs, dividing by n, and letting $p \to \infty$, this becomes

$$\overline{m}[B_{r+C/m}(x)] \geq \liminf_{p\to\infty} \frac{1}{pm} \log \mathbb{L}_p^{(m)}[B_r(x)],$$

and taking infima over $r > 0$ we get:

$$\inf_{r>C/m} \overline{m}[B_r(x)] \geq -\frac{1}{m}\ l_m(x).$$

Now fix $r' = C'/m > C/m$, and let $\bar{B}_{r'}(x)$ be the closed l_∞-ball of radius r' centred at x. Applying the large deviation upper bound for the set $\bar{B}_{r'}(x)$ produces

$$-\inf_{y\in\bar{B}_{r'}(x)} l(y) \geq \overline{m}[B_{r'}(x)] \geq -\frac{1}{m} l_m(x).$$

Hence

$$\begin{aligned}\sup_{y\in\bar{B}_{r'}(x)}\{\langle\theta,\ y\rangle - l(y)\} &\geq \langle\theta,\ x\rangle - \frac{1}{m}\|\theta\|C' - \inf_{y\in\bar{B}_{r'}(x)} l(y)\\ &\geq \langle\theta,\ x\rangle - \frac{1}{m}(\|\theta\|C' + l_m(x)),\end{aligned}$$

and taking the supremum over x on both sides:

$$l^*(\theta) \geq \frac{1}{m}(\lambda_m(\theta) - \|\theta\|C').$$

The upper bound $\lambda(\theta) \leq l^*(\theta)$ is now obtained by letting $m \to \infty$. ∎

If λ is finite in a neighbourhood of the origin then the sequence $\{\mathbb{M}_n : n \geq 1\}$ is *exponentially tight*: there exists a sequence of compact sets $\{K_n : n \geq 1\}$ such that

$$\limsup_{n\to\infty} \overline{m}[\mathbb{R}^d \setminus K_n] = -\infty.$$

Under exponential tightness the weak l.d.p. for $\{\mathbb{M}_n : n \geq 1\}$ extends to a full l.d.p. [LP95]

Theorem 6 *Let assumption 1 hold. The rate function l of theorem 3 is the convex dual of λ. If λ is finite in a neighbourhood of the origin, then the sequence $\{\mathbb{M}_n\}$ satisfies a large deviation principle with rate function l.*

PROOF λ exists by lemma 4, and by lemma 5 it is the convex dual of l. Since l is a proper convex l.s.c. function, it follows that $l = l^{**} = \lambda^*$. If λ is finite in a neighbourhood of the origin then the sequence $\{\mathbb{M}_n\}$ is exponentially tight and the l.d.p. follows. ∎

Note that since $\lambda(\theta) \leq \psi_n(\theta_+, \theta_-)/n$ for all n one needs only that $\psi_n(\theta_+, \theta_-)$ be finite in a neighbourhood of the origin, for any n, in order to establish the l.d.p.

3 Matrix Operators

From the results of the last section we may deduce for the l.d.p. for certain classes of the matrix operators introduced in section 1.

Lemma 7 *A max-plus matrix operator has finite projective radius if and only if each of its columns has all entries finite, or all entries equal to $-\infty$. Similarly a min-plus operator has finite projective radius if and only if each of its columns has all entries finite or all entries equal to $+\infty$.*

PROOF Suppose that A is a max-plus matrix operator with all matrix entries finite, and let x be any vector in $\mathbb{R}^d$. For a given value of $i \in [1, d]$ let $J(i)$ be the value of j which maximises $A_{ij} + x_j$. Then

$$\mathsf{t}[Ax] = \max_{i,j} A_{ij} + x_j = \max_i A_{iJ(i)} + x_{J(i)}$$

and

$$\mathsf{b}[Ax] = \min_k \max_l A_{kl} + x_l \geq \min_k A_{kJ(i)} + x_{J(i)}.$$

Therefore

$$\|Ax\|_{\mathsf{P}} \leq \max_i (A_{iJ(i)} - \min_k A_{kJ(i)}) \leq \max_{i,j} A_{ij} - \min_{kl} A_{kl}.$$

This gives a finite upper bound on $\|Ax\|_{\mathsf{P}}$, independent of x. Next, if one or more columns of A are identically equal to $-\infty$ then the projective radius of A is equal to that of the matrix obtained by deleting these columns. If the remaining entries are all finite then so is $\Pi[A]$. (Recall we assume that each row of A has at least one finite entry, so that A cannot be identically equal to $-\infty$).

Now suppose that for some column j we have A_{ij} finite and $A_{kj} = -\infty$. Then for any x

$$\|Ax\|_{\mathsf{P}} \geq (Ax)_i - (Ax)_k \geq A_{ij} + x_j - \max_{l \neq j} (A_{kl} + x_l),$$

and since x_j can be made arbitrarily large it follows that $\Pi[A] = +\infty$.

The proof for min-plus matrices is similar. ∎

In particular, if A is an *aperiodic* matrix operator, then there exists $N < \infty$ such A^N has all entries finite, and therefore finite projective radius. Turning to random sequences of matrix operators, we say that $\{A(n) : n \geq 1\}$ *has fixed structure* if, for each i, j, $A_{ij}(n)$ equal to $-\infty$ for all n with probability one or zero.

Assumption 2 *$\{A(n) : n \geq 1\}$ is a random sequence of i.i.d. aperiodic max-plus matrix operators, with fixed structure. In addition, the finite components of $A(1)$ take values in a bounded subset of $\mathbb{R}$.*

As in section 2, $\mathbb{M}_n$ denotes the law of $x(n)/n$, where

$$x(n) = A(n) \cdots A(1)x_0 \qquad n = 1, 2, \ldots,$$

with x_0 fixed.

Theorem 8 *Let assumption 2 hold. Then the sequence $\{\mathbb{M}_n\}$ satisfies a large deviation principle with convex rate function l equal to λ^*.*

PROOF Let $N < \infty$ be such that the matrix $(A(1))^N$ has all entries finite. Note that, when taking the product of matrices $A(1)$ and $A(2)$, the positions of the finite entries in $A(2)A(1)$ depend only on which entries of $A(1)$ and $A(2)$ are finite (and not on the values of these entries: this is similar to the situation with the zeros of products of positive matrices under the standard algebra). Since each matrix $A(k)$ has its finite entries in the same positions (due to the fixed structure assumption), it follows that $A(0, N) = A(N) \cdots A(1)$ has all entries finite, and therefore finite projective radius. Under the second part of assumption 2 the entries of $A(0, N)$ actually take values in a bounded subset of $\mathbb{R}$, implying that $\Pi[A(0, N)]$ is almost surely bounded. It follows from theorems 3 and 6 that the sequence $\{\mathbb{M}_n\}$ satisfies a weak l.d.p. with rate function λ^*. Assumption 2 also implies that for each n, $x(n)$ is a bounded random variable, so that the functions ψ_n of lemma 4 are finite throughout $\mathbb{R}^2$. Therefore so is $\lambda(\theta)$ and the l.d.p. holds for $\{\mathbb{M}_n\}$. ∎

The l.d.p. for a certain class of non-negative matrix operators on the positive cone $\mathbb{R}^d_+$ can be proved along similar lines. Recall that if $A : \mathbb{R}^d_+ \to \mathbb{R}^d_+$ is a non-negative matrix operator having at least one non-zero entry per row, then $\tilde{A} : \mathbb{R}^d \to \mathbb{R}^d$ is the topical operator defined by $\tilde{A}x := \log(A \exp(x))$.

Lemma 9 *Let $A : \mathbb{R}^d_+ \to \mathbb{R}^d_+$ be a non-negative matrix operator, with at least one non-zero entry per row. Then $\tilde{A}$ has finite projective radius if and only if each column of A has all entries greater than zero, or all entries equal to zero.*

PROOF Assume that A has no matrix entries equal to zero, and let x be any vector in $\mathbb{R}^d$. For a given value of i let $J(i)$ be the value of j which maximises $A_{ij}e^{x_j}$. Then

$$\mathsf{t}[\tilde{A}x] = \max_i \log\Big(\sum_j A_{ij}e^{x_j}\Big) \leq \max_i \log(dA_{iJ(i)}e^{x_{J(i)}})$$

and

$$\mathsf{b}[\tilde{A}x] = \min_k \log\Big(\sum_l A_{kl}e^{x_l}\Big) \geq \min_k \log(A_{kJ(i)}e^{x_{J(i)}}).$$

Therefore

$$\begin{aligned} \|\tilde{A}x\|_{\mathsf{P}} &\leq \max_i\Big(\log(dA_{iJ(i)}) - \min_k \log A_{kJ(i)}\Big) \\ &\leq \log d + \max_{i,j} \log A_{ij} - \min_{k,l} \log A_{kl}, \end{aligned}$$

a finite upper bound which is independent of x. As in the max-plus matrix case one can now observe that if A has all entries of one column equal to zero, then the projective radius of $\tilde{A}$ is equal to that of the operator obtained by excluding this column. This proves the 'if' part of the lemma.

On the other hand if column j of A has a non-zero entry A_{ij} and a zero entry A_{kj} then for any $x \in \mathbb{R}^d$

$$\|\tilde{A}x\|_{\mathsf{P}} \geq \log A_{ij} + x_j - \log\Big(\sum_{l\neq j} A_{il}e^{x_l}\Big).$$

Since x_j can be made arbitrarily large it follows that $\Pi[\tilde{A}] = +\infty$. ∎

The assumption analagous to assumption 2 is therefore the following. A *fixed structure* sequence of random non-negative matrices $\{A(n)\}$ will be one in which, for each i, j, $A_{ij}(n)$ is zero for all n with probability either one or zero.

Assumption 3 $\{A(n) : n \geq 1\}$ *is a random sequence of i.i.d. non-negative aperiodic matrix operators, with fixed structure. In addition, the non-zero entries of* $A(1)$ *take values in a compact subset of the positive real line.*

We continue to let $x(n)$ denote the vector $A(n)\cdots A(1)x_0$, for a fixed $x_0 \in \mathbb{R}^d_+$, but we now take $\mathbb{M}_n$ to be the law of $(\log x(n))/n$.

Theorem 10 *Under assumption 3, the sequence* $\{\mathbb{M}_n : n \geq 1\}$ *satisfies a large deviation principle with rate function* l *equal to* λ^**, where* $\lambda : \mathbb{R}^d \to \mathbb{R}$ *is given by*

$$\lambda(\theta) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E} \prod_{i=1}^{d} (x_i(n))^{\theta_i}.$$

PROOF Letting $\tilde{x}(n) = \log x(n)$ we find that $\tilde{x}(n)$ satisfies

$$\tilde{x}(n) = \tilde{A}(n) \cdots \tilde{A}(1)\tilde{x_0}.$$

The remainder of the proof parallels the proof of theorem 8, and is omitted. ∎

Acknowledgement. During the preparation of this work I benefitted from several useful conversations with John Lewis.

References

[BCOQ92] F. Baccelli, G. Cohen, G. J. Olsder, and J.-P. Quadrat. *Synchronization and Linearity.* Wiley, 1992.

[BK91] F. Baccelli and T. Konstantopoulos. Estimates of cycle times in stochastic Petri nets. In *Applied Stochastic Analysis*, volume 177 of *Lecture Notes in Control and Information Sci.*, pages 1–20. Springer, New York, 1991.

[BM96] F. Baccelli and J. Mairesse. Ergodic theorems for stochastic operators and discrete event networks. In J. Gunawardena, editor, *Idempotency.* Cambridge University Press, 1996.

[Cha96] C.-S. Chang. On the exponentiality of stochastic linear systems under the max-plus algebra. *IEEE Trans. Automatic Control*, 41:1182–1188, 1996.

[CT80] M. G. Crandall and L. Tartar. Some relations between non-expansive and order preserving maps. *Proc. Amer. Math. Soc.*, 78(3):385–390, 1980.

[DZ98] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications.* Springer-Verlag, New York, 1998.

[GG98] S. Gaubert and J. Gunawardena. A non-linear hierarchy for discrete event dynamical systems. In *Proceedings of the 4th Workshop on Discrete Event Systems, Cagliari, Italy*, 1998.

[GK95] J. Gunawardena and M. Keane. On the existence of cycle times for some nonexpansive maps. Technical Report HPL-BRIMS-95-003, HP Laboratories, 1995.

[Gun94] J. Gunawardena. Min-max functions. *Discrete Event Dynamic Systems*, 4:377–406, 1994.

[Gun96] J. Gunawardena. An introduction to idempotency. In J. Gunawardena, editor, *Idempotency*. Cambridge University Press, 1996.

[GY95] P. Glasserman and D. D. Yao. Stochastic vector difference equations with stationary coefficients. *J. Appl. Prob.*, 32:851–866, 1995.

[Lan73] O. E. Lanford. Entropy and equilibrium states in classical statistical mechanics. In *Lecture Notes in Phys.*, volume 20, pages 1–113. Springer, Berlin, 1973.

[LP95] J. T. Lewis and C.-E. Pfister. Thermodynamic probability theory: some aspects of large deviations. *Russian Math. Surveys*, 50(2):279–317, 1995.

[LPS94] J. T. Lewis, C.-E. Pfister, and W. G. Sullivan. Entropy, concentration of probability, and conditional limit theorems. *Markov Processes and Related Fields*, 1:319–386, 1994.

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