***Critical path statistics of max-plus linear systems
with Gaussian noise***

Hook, J

2012

MIMS EPrint: **2012.64**

Manchester Institute for Mathematical Sciences

School of Mathematics

The University of Manchester

Critical path statistics of max-plus linear systems with Gaussian noise

James Hook - University of Manchester, UK

December 2, 2012

Abstract

The critical paths of a max-plus linear systems with noise are random variables. In this paper we introduce the *edge criticalities* which measure how often the critical paths traverse each edge in the precedence graph. We also present the *parallel path approximation,* a novel method for approximating these new statistics as well as the previously studied max-plus exponent. We show that for low amplitude noise the critical paths spend most of their time traversing the deterministic maximally weighted cycle and that as the noise amplitude is increased the critical paths become more random and their distribution over the edges in the precedence graph approaches a highly uniform measure of maximal entropy.

Introduction

Max-plus matrix multiplication corresponds exactly to finding the weight of the maximally weighted paths through a graph whose edge weights are determined by the matrices' coefficients[1]. Consider $K(N)$ the complete graph on N vertices. We say that a sequence of $K(N)$'s vertices, $\sigma = [\sigma(k)]_{k=1}^{n+1}$ is a path of length n from $\sigma(1)$ to $\sigma(n+1)$ formed from the n edges $[\sigma(k+1), \sigma(k)]_{k=1}^{n}$.

Now let $[A(k)]_{k=1}^{n}$ be a sequence of $N \times N$ max-plus matrices and $\Sigma_n(i,j)$ be the set of all paths from j to i of length n through $K(N)$, we have

$$[A(n) \otimes A(n-1) \otimes ... \otimes A(1)]_{i,j} = \max_{\sigma \in \Sigma_n(i,j)} \sum_{k=1}^{n} W(\sigma),$$

where $\otimes$ stands for max-plus matrix multiplication and the weight function W is given by

$$W(\sigma) = \sum_{k=1}^{n} A(k)_{\sigma(k+1),\sigma(k)}.$$

---

[1]Max-plus matrices are arrays of elements in $\mathbb{R} \cup \infty$. If A and B are $N \times N$ max-plus matrices then their product $C = A \otimes B$ is the $N \times N$ max-plus matrix defined by $c_{i,j} = \bigoplus_{k=1}^{n} a_{i,k} \otimes b_{k,j} = \max_{k=1}^{n} a_{i,k} + b_{k,j}$.

Deterministic autonomous max-plus linear systems exhibit turnpike behaviour characterised by Cuninghame-Green in [1]. We say that a problem of finding an optimal (i.e. maximally weighted) path has the turnpike property if there is some globally optimal path, called the turnpike, such that for any a and b and for sufficiently large t the optimal path from a to b of length t will move from a to the turnpike as quickly as possible then stay on it for as long as possible. The max-plus turnpike theorem says that for all i, j and sufficiently large n the maximally weighted path of length n from j to i will spend most of its steps traversing a critical cycle so that in the limit $n \to \infty$ the maximally weighted path spends almost all of its steps traversing the critical cycle and its weight divided by length will equal the average weight of the critical cycle, which we call the critical cycle mean.

In this paper we examine max-plus linear systems with Gaussian noise from this path centric viewpoint. The systems we consider consist of a fixed deterministic part plus a Gaussian distributed noise term which changes from one step to the next. This is a natural way to define a random max-plus linear system and reflects the standard engineering approach of including noise in an otherwise deterministic model. Under a standard irreducibility assumption we show that whilst there is no longer a predetermined critical cycle there is still turnpike-like behaviour where for all i, j and large n the maximally weighted path of length n from j to i will spend most of its steps traversing a critical path common to all (i, j) pairs. This common path is a random variable which will typically not exhibit the periodic behaviour of the deterministic system. In the deterministic setting it is possible for two critical paths to traverse the same critical cycle out of phase so that they are never at the same vertex at the same step, but in these stochastic systems we see that all the critical paths come together and coincide exactly for most of their steps. Hence we call this conflux turnpike behaviour.

This behaviour makes it possible for us to prove the existence of a new set of critical path statistics, the *edge criticalities* which measure how often the critical paths traverse each edge. These statistics are particularly interesting when viewed as a function of the noise amplitude where we see a transition from deterministic turnpike behaviour where the edge criticalities are concentrated on the critical cycle to a highly uniform distribution as the noise amplitude is increased.

Another critical path statistic is the *max-plus exponent*, which is the average edge weight of the critical path. This has been well studied and can be proved to exists for a very general class of max-plus linear system; see for example Heidergrott [2], Baccelli et.al. [3] and Merlet [4] where some interesting examples in which this limit does not exist are also discussed. The max-plus exponent is of particular importance as a performance indicator for queuing systems where it corresponds to the reciprocal of the throughput.

Analytic investigation of any of these statistics has only proved possible for very specially structured systems, see for example Baccelli and Hong [5]. One major problem is that since different paths share edges which they traverse at the same step their total weights are interdependent which makes statistical analysis difficult. In [6] Goverde et al. describe an efficient method for estimating the max-plus exponent with an expectation expression. In this paper we introduce the *parallel path approximation* (PPA) a new method for approximating both the max-plus exponent and the edge

criticalities. The idea is to simply ignore the complicated path weight interdependencies when computing the weight of the maximally weighted path. This approximation provides an accurate approximation of the edge criticalities and an upper bound on the max-plus exponent which is accurate for large highly interconnected systems.

This paper is organised as follows. In section 1 we introduce the specific type of matrix distribution we will be using as well as some other important notation for the remainder of the paper. In section 2 we define the critical path statistics and the notion of conflux turnpike behaviour, we prove that these statistics are well defined and that this behaviour will always be observed. In section 3 we define the PPA approximation and show how to compute its statistical properties in order to approximate those of the original max-plus system. Finally we present a suite of examples and pose a conjecture that states there is a class of systems for which the PPA max-plus exponent approximation is perfectly accurate in the limit of large systems.

1   System specification

Although it is possible to treat a slightly more general set up we will, for the sake of simplicity, restrict all results and examples to systems of the following formulation. Our approach is to take a fixed max-plus matrix β (which could be used to model a deterministic system) and add a time varying Gaussian noise term to it. The noise terms in each component of the matrix and at each stage in the system are mutually independent. The amplitude of the noise in each component of the matrix can be chosen arbitrarily through the choice of the matrix α.

Definition 1 *Let $\beta \in (\mathbb{R} \cup \{-\infty\})^{N \times N}$ be the matrix of deterministic weights, $\alpha \in \mathbb{R}_+^{N \times N}$ be the matrix of noise amplitudes and $[Z(k)]_{k=1}^\infty$ be an i.i.d. sequence of $N \times N$ matrices whose components are mutually independent $(0, 1)$ Gaussians which we call the noise terms. Define the sequence of max-plus matrices $[A(k)]_{k=1}^\infty$ by their (i, j) components*

$$A(k)_{i,j} = \beta_{i,j} + \alpha_{i,j} z(k)_{i,j}.$$

Further define the sequence of product matrices $[M(n)]_{n=1}^\infty$ by

$$M(n) = \bigotimes_{k=1}^n A(k) = A(n) \otimes A(n-1) \otimes ... \otimes A(1),$$

where $\otimes$ stands for max-plus multiplication.

Note that the randomness in this distribution only affects the exact value of the finite entries in the matrix and not the position of the $-\infty$ entries. We say that matrix distributions with this property have fixed support.

Definition 2 *Further define the associated precedence graph $G = \langle V, E \rangle$ with vertex set $V = \{1, 2, ..., N\}$ and an edge $(i, j) \in E$ whenever $\beta_{i,j} \neq -\infty$. Now define the set of all paths through G of length n from j to i*

$$\Sigma_n(i, j) = \{\sigma \in \{1, 2, ..., N\}^{n+1} : \sigma(1) = j, \sigma(n+1) = i, [\sigma(k+1), \sigma(k)] \in E \text{ for } k = 1, 2, ..., n\}.$$

For each edge $(i, j) \in E$ associate a weight sequence $[w(i, j)_k]_{k=1}^\infty$ given by

$$w(i, j)_k = A(k)_{i,j} = \beta_{i,j} + \alpha_{i,j} z(k)_{i,j},$$

and define the weight of a path $\sigma \in \Sigma_n(i, j)$ by

$$W(\sigma) = \sum_{k=1}^n w[\sigma(k+1), \sigma(k)]_k.$$

By assigning a sequence of weights to each edge we are able to express components of the product matrices in terms of maximally weighted paths through the precedence graph just as in the deterministic case.

$$M(n)_{i,j} = \max_{\sigma \in \Sigma_n(i,j)} W(\sigma).$$

Throughout this paper we will make the standard assumption that G has the following properties.

Definition 3 *We say that a graph $G = \langle V, E \rangle$ is* irreducible *if for all $i, j \in V$ there exists a path through G from j to i. We say that G is* aperiodic *if the greatest common divisor of the lengths of the cycles in G is equal to one.*

2   Confluxes and critical path statistics

In this section we define exactly what we mean by a conflux turnpike, roughly the idea is that there is always a maximally weighted (critical) path between each pair of vertices and that a conflux occurs when all of these paths come together and coincide exactly for at least one step. The turnpike is then this subpath where they coincide.

We also define the new critical path statistic, the edge criticality's which measure how often the critical paths traverse each edge.

Theorem 9 states that systems of the sort outlined in section 1 will almost surely exhibit conflux turnpike behaviour and shows how this then enables us to factorize our sequence of matrices in a useful way. Theorem 10 makes use of this factorisation to prove the existence of the edge criticalities as an almost sure limit independent of any initial conditions.

Definition 4 *For a sequence of matrices $[A(k)]_{k=1}^n$, which define the weight function W, a path $\sigma \in \Sigma_n(i, j)$ is (i, j)-critical if*

$$W(\sigma) = \max_{\sigma \in \Sigma_n(i,j)} W(\sigma),$$

and a path $\sigma \in \Sigma_n(i, j)$ is critical *if*

$$W(\sigma) = \max_{i',j'} \max_{\sigma \in \Sigma_n(i',j')} W(\sigma) = \max_{\sigma \in \Sigma_n} W(\sigma),$$

where Σ_n is the set of all paths of length n through G.

For systems of the sort outlined in section 1 the probability of any two different paths having the same weight is zero so that these maximums are attained uniquely and we are able to define the following.

Definition 5 *The (i,j)-critical path of length n, $\phi_n^{i,j}$ is the unique path of length n that is (i,j)-critical. The critical path of length n, ϕ_n is the unique path of length n that is critical.*

Definition 6 *The critical path statistics are then the max-plus exponent $\lambda \in \mathbb{R}$ and the edge criticalities $P \in \mathbb{R}^{N \times N}$ defined by*

$$\lambda = \lim_{n \to \infty} \frac{1}{n} W(\phi_n),$$

provided this limit exists and

$$P_{i,j} = \lim_{n \to \infty} \frac{1}{n} \#[(i,j) \in \phi_n],$$

provided this limit exists, where $\#[(i,j) \in \phi_n]$ is the number of times that the path ϕ_n traverses the edge (i,j).

So a sequence of matrices $[A(k)]_{k=1}^n$ defines a critical path ϕ_n, which as a function of the noise terms $[Z(k)]_{k=1}^n$ is a random variable. The max-plus exponent measures the growth rate of the critical paths weight and the edge criticalities measure what proportion of the critical paths steps traverse each edge. Note that in general there is no guarantee that these limits will exist. In theorem 10 we will show that for systems of the sort outlined in section 1 the two expressions converge almost surely to constants independent of any initial conditions, so that the max-plus exponent or edge criticalities of the (i,j)-critical paths are all the same and equal to those of the critical path.

Definition 7 *A sequence of max-plus matrices $[A(k)]_{k=1}^n$ is a conflux if the different (i,j)-critical paths of length n all come together and coincide for one or more steps. So if there exists $l_T > 0$ and k_C such that for all (i,j), (i',j')*

$$\phi_n^{i,j}(k) = \phi_n^{i',j'}(k),$$

for $k_C \le k \le k_C + l_T$. The section φ for which they coincide is called the turnpike which a path through G of length l_T. We say that $[A(k)]_{k=1}^n$ is a minimal conflux if it is a conflux but $[A(k)]_{k=1}^{n-1}$ is not.

It is easy to show that if the sequence $[A(k)]_{k=1}^n$ is a conflux then the product $M(n) = A(n) \otimes A(n-1) \otimes ... \otimes A(1)$ is a rank-1 matrix and that the sequence therefore has the memory loss property used by Mairesse [9] to prove many statistical results for the max-plus exponent. This characterisation of the condition in terms of critical paths will enable us to prove some new statistical properties for these systems, in particular the existence of the edge criticalities.

Example 8 *Define the sequence of i.i.d. 2×2 max-plus matrices $[A(n)]_{n=1}^{\infty}$ by*

$$A(n) = \begin{pmatrix} 1 + aZ_1(n) & aZ_2(n) \\ 3 + aZ_3(n) & aZ_4(n) \end{pmatrix},$$

where $a \in \mathbb{R}$ is the single noise amplitude parameter. Clearly this is of the form outlined in section 1 with G the complete graph on 2 vertices. Through Monte Carlo simulation we sample the (i,j)-critical paths of length 6 for $a = 0$ and $a = 1$, and approximate the critical path statistics. The results are displayed in figure 1.

For $a = 0$ the system exhibits deterministic turnpike behaviour where all critical paths spend most of their steps traversing the critical cycle $1 \mapsto 2 \mapsto 1$. The edge criticalities are concentrated on the critical cycle with $P_{1,1} = P_{2,2} = 0$ and $P_{1,2} = P_{2,1} = 0.5$. The max-plus exponent is equal to the critical cycle mean $\lambda = (3+0)/2$.

For $a = 1$ the system exhibits stochastic conflux turnpike behaviour where all the critical paths coincide between steps 3 and 5 so that in the terms of definition 7 we have $K_C = 3$, $l_T = 3$ and $\varphi = (1,1,2,1)$. Observe that unlike in the deterministic case the turnpike is random and all paths coincide in phase. The edge criticalities are more uniform than in the deterministic case, roughly a convex combination of the deterministic $a = 0$ edge criticalities and the measure of maximal entropy $P_{1,1} = P_{1,2} = P_{2,1} = P_{2,2} = 0.25$. The max-plus exponent is greater than in the deterministic $a = 0$ case.

When the noise amplitude is increased further the edge criticalities slowly converge to the measure of maximal entropy on the graph $P_{1,1} = P_{1,2} = P_{2,1} = P_{2,2} = 0.25$.
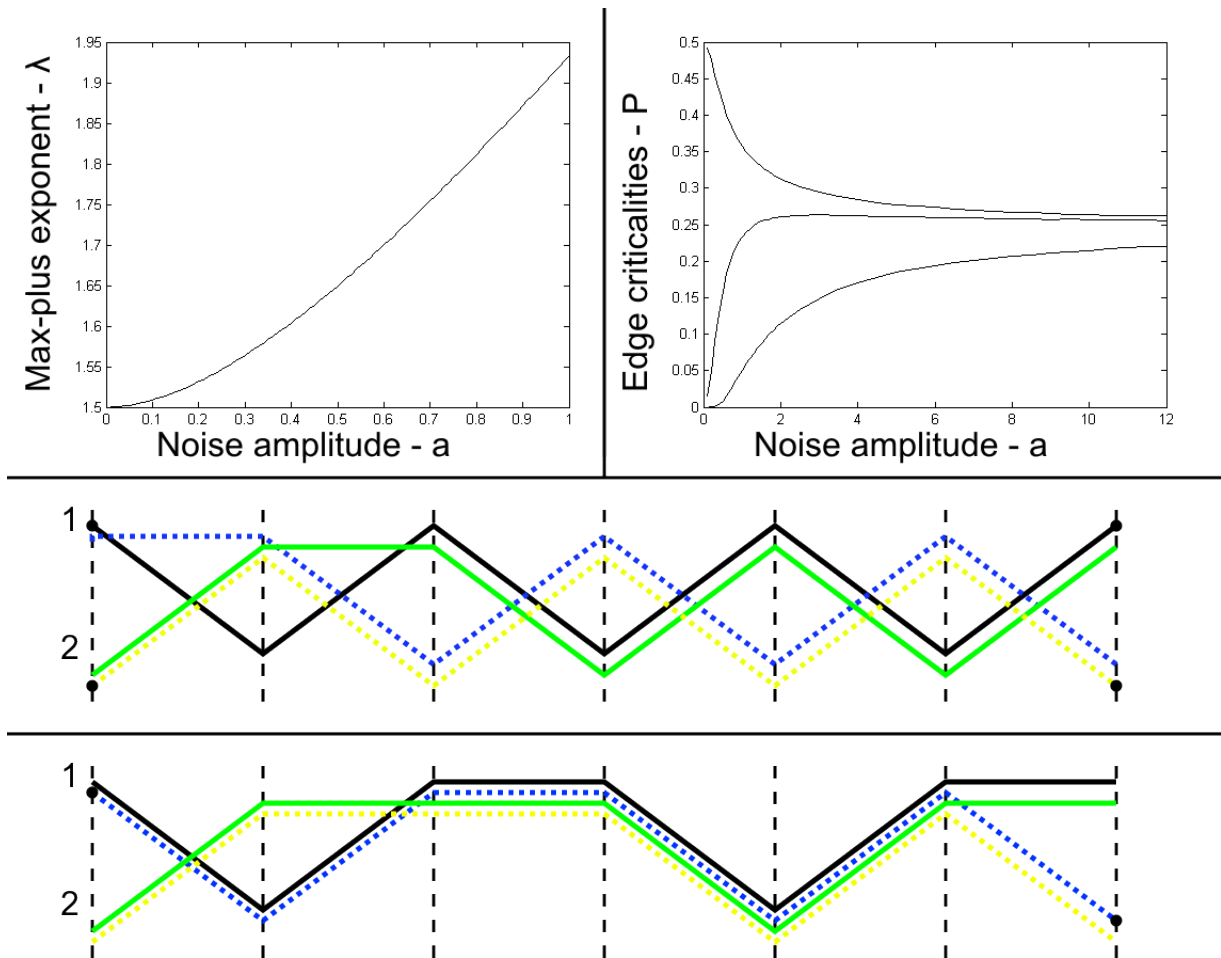
Figure 1: Top: critical path statistics vs noise amplitude, max-plus exponent (left), edge criticalities (right) with $P_{1,1}$, $P_{1,2} = P_{2,1}$ and $P_{2,2}$ bottom to top. Middle and bottom: (i,j)-critical paths of length 6 for $a = 0$ and $a = 1$ respectively with $\phi(1,1)_6$ black, $\phi(1,2)_6$ blue-dashed, $\phi(2,1)_6$ green and $\phi(2,2)_6$ yellow-dashed, Critical paths ϕ_6 are distinguished by black head and tail dots.

The following theorem tells us that for systems of the sort outlined in section 1 we will (with probability 1) observe conflux turnpike behaviour, and that as we look at longer and longer sequences of matrices the proportion of steps that each critical path spends on the turnpike tends to one.

Theorem 9 *Let $[A(k)]_{k=1}^n$ be a sequence of max-plus matrices of the sort outlined in section 1. Provided G is irreducible and aperiodic*

$$\lim_{n \to \infty} \mathbb{P}\{[A(k)]_{k=1}^n \text{ is a conflux }\} = 1.$$

And for all $\epsilon > 0$

$$\lim_{n \to \infty} \mathbb{P}\{|\frac{l_T}{n} - 1| \leq \epsilon\} = 1,$$

where l_T is the length of the turnpike.

Proof. The sequence $[A(k)]_{k=1}^n$ can be factorized into a sequence of minimal confluxes with a non-conflux remainder

$$[A(k)]_{k=1}^n = C^1, C^2, ..., C^{n_C}; A(t_{n_C} + 1), ..., A(n),$$

where each C^m is a minimal conflux of length $|C^m|$ and

$$t_m = \sum_{k=1}^m |C^k|.$$

So that $A(t_{m-1}+r) = C^m(r)$ for $r = 1, 2, .., |C^m|$. This factorisation can be performed inductively so that

$$[A(k)]_{k=1}^{n+1} = C^1, C^2, ..., C^{n_C}, C^{n_C+1};$$

if $A(t_{n_C} + 1), ..., A(n), A(n + 1) = C^{n_C+1}$ forms a conflux and

$$[A(k)]_{k=1}^{n+1} = C^1, C^2, ..., C^{n_C}; A(t_{n_C} + 1), ..., A(n), A(n + 1),$$

if not.

We will first show that the length of each of the minimal confluxes in this factorisation is exponentially bounded. W.l.o.g. $\alpha_{1,2} \neq 0$ and $\beta_{1,2} \neq -\infty$, irreducibility and aperiodicity implies that there exist n_1 and n_2 such that for all (i, j) there is a path of length n_1 from j to 2 and a path of length n_2 from 1 to i. Now suppose that the noise terms $[Z(k)]_{k=1}^{n_1+n_2+1}$ satisfy

$$z(n_1 + 1)_{1,2} \geq \frac{2(n_1 + n_2 + 1) + 1 - \beta_{1,2}}{\alpha_{1,2}},$$

and

$$\frac{-\beta_{i,j}}{\alpha_{i,j}} \leq z(k)_{i,j} \leq \frac{1 - \beta_{i,j}}{\alpha_{i,j}},$$

for all $(i, j, k) \neq (1, 2, n_1+1)$. Any path not traversing the edge $(1, 2)$ on the (n_1+1)th step will have weight $\leq n_1 + n_2 + 1$ whereas any path traversing the edge $(1, 2)$ on the $(n_1 + 1)$th step will have weight $\geq (n_1 + n_2 + 1) + 1$. Therefore all the critical paths will come together and coincide on this edge at this step and the sequence will be a conflux. Of course this is only one possible way for a conflux to occur, but it gives us the bound

$$\mathbb{P}\{|C| \geq n\} \leq (1 - \rho^{\frac{1}{n_1+n_2+1}})^{n-(n_1+n_2+1)},$$

for all $n \geq n_1 + n_2 + 1$, where ρ is the (non-zero) probability associated with the condition on the noise terms described above.

So given a sequence $[A(k)]_{k=1}^\infty$ the finite sequences $[A(k)]_{k=1}^n$ can be factorized into sequences of minimal confluxes. As n grows, with probability one, the matrix sequences factorized into a longer and longer sequence of minimal confluxes which themselves form an i.i.d. random sequence $[C^m]_{m=1}^{n_C}$.

Now consider $\phi_n^{i,j}$, the (i, j)-critical path of length n for some sequence of matrices $[A(k)]_{k=1}^n$ which has a minimal conflux factorisation as described

$$[A(k)]_{k=1}^n = C^1, C^2, ..., C^{n_C}; A(t_{n_C} + 1), ..., A(n).$$

8

Since $\phi_n^{i,j}(t_{m-1}+1) = i'$ and $\phi_n^{i,j}(t_m) = j'$ for some i', j' the maximally weighted path from j to i of length n which accumulates weight according to $[A(k)]_{k=1}^n$ contains a subpath $\gamma_m^{i',j'}$ of length $|C^m|$ from j' to i' which accumulates weight according to the minimal conflux sequence C^m.

Therefore $\gamma_m^{i',j'}$ is the (i', j')-critical path of length $|C^m|$ for the matrix sequence C^m, since otherwise there would be a more weighted such path and we could use it to construct a new path of length n from j to i with greater weight than $\phi_n^{i,j}$ which would be a contradiction.

Therefore the factorisation of the matrix sequence gives rise to a factorisation of the (i,j)-critical paths

$$\phi_n^{i,j} = \gamma_1^{i'(1),j}, \gamma_2^{i'(2),i'(1)}, ..., \gamma_{n_C}^{i'(n_C),i'(n_C-1)}, \zeta^{i,i'(n_C)},$$

where the γ's are the subpaths corresponding to each minimal conflux sequence and ζ is a remainder term to complete the path.

Now since C^m is a conflux sequence and $\gamma_m^{i,j}$ is (i,j)-critical for C^m it follows that $\gamma_m^{i,j}$ traverses the turnpike φ^m of C^m.

Therefore $\gamma_m^{i'(m),i'(m-1)}$ will traverse φ^m and $\gamma_{m+1}^{i'(m+1),i'(m)}$ will traverse φ^{m+1} so that $\phi_n^{i,j}$ will contains a subpath from the beginning of φ^m to the beginning of φ^{m+1}, call it ψ^m - intuitively this is the second half of the subpath γ_m glued to the first half of the subpath γ_{m+1}. As before this path will be critical but most importantly it is independent of (i,j). We therefore have another path factorisation

$$\phi_n^{i,j} = \eta^j, \psi^1, \psi^2, ..., \psi^{n_C-1}, \theta^i,$$

where ψ^m is the critical path linking the turnpike of C^m to that of C^{m+1}, and η and θ are remainder terms to complete the path. Note that the only subpaths in this factorisation to depend on the conditions (i,j) are the first and last terms. The ψ subpaths are common to all (i,j) and therefore constitute the turnpike, as all the critical paths will coincide along this section.

Finally given a sequence $[A(k)]_{k=1}^n$ factorized in this way the total length of its turnpike l_T is equal to n minus the length of the two remainder terms η and θ. Each of these terms is smaller than some minimal conflux sequence, whose length we have shown to be exponentially bounded so that for all ϵ

$$\lim_{n \to \infty} \mathbb{P}\{|\frac{l_T}{n} - 1| \le \epsilon\} = 1.$$

■

The second theorem in this section uses the path factorisation introduced in the previous proof to show that the definition of the edge criticalities in terms of a limit exist almost surely.

Theorem 10 *Let $[A(k)]_{k=1}^\infty$ be a sequence of max-plus matrices of the sort outlined in section 1. Provided G is irreducible and aperiodic there exists $P \in \mathbb{R}^{N \times N}$ such that for all (i,j) and all $\epsilon > 0$*

$$\lim_{n \to \infty} \mathbb{P}\{|\frac{\#[(i',j') \in \phi_n^{i,j}]}{n} - P_{i',j'}| \le \epsilon\} = 1,$$

where $\#[(i', j') \in \phi_n^{i,j}]$ is the number of times that the (i, j)-critical path of length n traverses the edge (i', j').

Proof. We showed previously that the sequence of matrices $[A(k)]_{k=1}^n$ could be factorized into a sequence of minimal conflux sequences

$$[A(k)]_{k=1}^n = C^1, C^2, ..., C^{n_C}; A(t_{n_C} + 1), ..., A(n),$$

where each C^m is a minimal conflux of length $|C^m|$ and

$$t_m = \sum_{k=1}^m |C^k|.$$

We also showed how this factorisation gave us a factorisation of the (i, j)-critical paths

$$\phi_n^{i,j} = \eta^j, \psi^1, \psi^2, ..., \psi^{n_C-1}, \theta^i,$$

where ψ^m is the critical path from the beginning of C^m's turnpike to the beginning of C^{m+1}'s.

This factorisation can be used in the transition counting function

$$\#[(i', j') \in \phi_n^{i,j}] = \#[(i', j') \in \eta^j] + \sum_{m=1}^{n_C-1} \#[(i', j') \in \psi^m] + \#[(i', j') \in \theta^i],$$

where $\#[(i', j') \in \eta^j]$ is the number of times that the path η^j traverses the edge (i, j) and so on. Since the length of the remainder terms η and θ are exponentially bounded their contribution to this sum is also bounded. With probability one we have

$$\lim_{n \to \infty} \frac{\#[(i', j') \in \phi_n^{i,j}]}{n} = \lim_{n \to \infty} \frac{\sum_{m=1}^{n_C-1} \#[(i', j') \in \psi^m]}{n}.$$

From the proof of the previous theorem we know that

$$\lim_{n \to \infty} \frac{n_C}{n} = \bar{l}$$

exists and is non-zero so that

$$\lim_{n \to \infty} \frac{\#[(i', j') \in \phi_n^{i,j}]}{n} = \lim_{n_C \to \infty} \frac{\bar{l}}{n_C} \sum_{m=1}^{n_C-1} \#[(i', j') \in \psi^m].$$

Clearly each ψ^m has the same distribution and therefore this sample average is taken over identically distributed variables but we need to be a little careful as they are not independent.

The sequence of i.i.d. matrices $[A(k)]_{k=1}^n$ gives a sequence of i.i.d. minimal conflux sequences $[C^m]_{m=1}^{n_C}$ and the ψ subpaths are functions of successive pairs of these matrix sequences so that

$$\psi^m = F(C^m, C^{m+1}),$$

for some function F. The sequence of subpaths $[\psi^m]_{m=1}^{n_C-1}$ and their transition counts are therefore the the output of a hidden Markov model. The underlying Markov chain is a sequence of successive paris of i.i.d. variables

$$X(m) = (C^m, C^{m+1}) \mapsto X(m+1) = (C^{m+1}, C^{m+2}),$$

where $[C^m]_{m=1}^\infty$ is an i.i.d. sequence.

From any state in this Markov chain it is possible to move to any other state in two moves by

$$(A, B) \mapsto (B, C) \mapsto (C, D).$$

It is also possible to start at one state and return to it in three moves by

$$(A, B) \mapsto (B, C) \mapsto (C, A) \mapsto (A, B).$$

The chain is therefore ergodic (irreducible and aperiodic). Finally since the length of each subpath ψ^m is bounded by the sum of the lengths of C^m and C^{m+1}, which are themselves exponentially bounded, the expectation of the edge counts of a subpath ψ^m are also bounded so that

$$P_{i',j'} = \bar{l}\mathbb{E}\#[(i', j') \in \psi^m]$$

exists and for all $\epsilon > 0$

$$\lim_{n \to \infty} \mathbb{P}\{|\frac{\#[(i', j') \in \phi_n^{i,j}]}{n} - P_{i',j'}| \le \epsilon\} = 1.$$

∎

3   Parallel path approximation

In this section we introduce the parallel path approximation (PPA) which is a simplification of the max-plus systems outlined in definition 1. Theorem 15 gives an exact formula for the statistical properties of the approximated system in the form of a convex optimisation problem. Lemma 13 is a technical result that supports the second part of theorem 15 which states that the PPA approximated max-plus exponent is greater than or equal to the true max-plus exponent, the PPA approximated exponent therefore provides a performance guarantee for queuing systems as its reciprocal gives a lower bound on the throughput.

Definition 11 *For a max-plus linear system*

$$M(n)_{i,j} = \max_{\sigma \in \Sigma_n(i,j)} W(\sigma).$$

define the parallel path approximation *by*

$$\widehat{M}(n)_{i,j} = \max_{\sigma \in \Sigma_n(i,j)} \widehat{W}(\sigma),$$

where the modified path weight function $\widehat{W}$ is given by

$$\widehat{W}(\sigma) = \sum_{k=1}^{n} \beta_{\sigma(k+1),\sigma(k)} + t_\sigma \sqrt{\sum_{k=1}^{n} \alpha^2_{\sigma(k+1),\sigma(k)}},$$

where $\{t_\sigma : \sigma \in \Sigma_n : n \in \mathbb{N}\}$ is a set of i.i.d. $(0,1)$-Gaussians, one for each and every path through G.

The idea is that since different paths through G share edges which they traverse on the same step their weights have very complicated interdependencies and this makes further analytic analysis extremely difficult. However the individual path weights are fairly simple

$$W(\sigma) = \sum_{k=1}^{n} \beta_{\sigma(k+1),\sigma(k)} + \alpha_{\sigma(k+1),\sigma(k)} z(k)_{\sigma(k+1),\sigma(k)}$$

The modified path wights are identically distributed to the original path weights but are mutually independent, the exact form of the PPA weight function is simply derived from the original weight function and basic properties of the Gaussian distribution.

Each modified path weight is a function of a unique independent Gaussian so that the modified weights are independent which greatly simplifies analysis. One interpretation is that we have assumed that no two paths between the same pair of vertices ever share an edge, hence parallel.

Definition 12 *In analogy to definition 4 define the* PPA critical path *of length n, $\widehat{\phi}_n \in \Sigma_n$ by*

$$\widetilde{W}(\widehat{\phi}_n) = \max_{\sigma \in \Sigma_n} \widehat{W}(\sigma),$$

where Σ_n is the set of all paths of length n through G. The PPA critical path statistics *are then the* PPA max-plus exponent *$\widehat{\lambda} \in \mathbb{R}$ defined by*

$$\widehat{\lambda} = \lim_{n \to \infty} \frac{1}{n} \widehat{W}(\widehat{\phi}_n),$$

and the PPA edge criticalities *$\widehat{P} \in \mathbb{R}^{N \times N}$ defined by*

$$\widehat{P}_{i,j} = \lim_{n \to \infty} \frac{1}{n} \#[(i,j) \in \widehat{\phi}_n],$$

where $\#[(i,j) \in \widehat{\phi}_n]$ is the number of times that the path $\widehat{\phi}_n$ traverses the edge (i,j).

Lemma 13 *A sequence of real random variables $[a_k]_{k=1}^n$ is said to be* associated *if the covariance*

$$cov[f(a_1, ..., a_n), g(a_1, ..., a_n)] \geq 0$$

for all f and $g : \mathbb{R}^n \to \mathbb{R}$ non-decreasing in each component [10].

Let $[a_k]_{k=1}^n$ be a sequence of associated real random variables and $[b_k]_{k=1}^n$ be a sequences of independent real random variables such that a_k is identically distributed to b_k for each k. We have

$$\mathbb{E}[\max_{k=1}^n a_k] \leq \mathbb{E}[\max_{k=1}^n b_k].$$

Proof. For any $i \in \{1, 2, ..., n\}$ $J \subset \{1, 2, ..., n\}$, with i not in J and for $t \in \mathbb{R}$ define the non-decreasing functions

$$f(\underline{a}) = \begin{cases} -1, & \text{if } a_i < t; \\ 0, & \text{otherwise.} \end{cases} \quad g(\underline{a}) = \begin{cases} -1, & \text{if } \max_{j \in J} a_j < t; \\ 0, & \text{otherwise.} \end{cases}$$

Substitution into the definition of associativity gives

$$\mathbb{P}[a_i < t | a_j < t \forall j \in J] \geq \mathbb{P}[a_i < t].$$

Now

$$\mathbb{P}[\max_{i=1}^{n} a_i < t] = \prod_{i=1}^{n} \mathbb{P}[a_i < t | a_j < t : j = 1, 2, ..., i-1],$$

so that repeated application of the above result using $J = \{1, 2, ..., i-1\}$ gives

$$\mathbb{P}[\max_{i=1}^{n} a_i < t] \geq \prod_{i=1}^{n} \mathbb{P}[a_i < t] = \mathbb{P}[\max_{i=1}^{n} b_i < t].$$

So that

$$\mathbb{E}[\max_{i=1}^{n} a_i < t] = \int_{0}^{\infty} 1 - \mathbb{P}[\max_{i=1}^{n} a_i < t] dt < \int_{0}^{\infty} 1 - \mathbb{P}[\max_{i=1}^{n} b_i < t] dt = \mathbb{E}[\max_{i=1}^{n} b_i < t].$$

∎

Before our man result we need a few more definitions.

Definition 14 *The simplex of G admissible edge criticalities is the simplex*

$$\Delta_G = \Big\{ x \in \mathbb{R}^{N \times N} : \sum_{i,j} x_{i,j} = 1 : \sum_{j} x_{i,j} = \sum_{j} x_{j,i} \Big\}.$$

For $x \in \Delta_G$ define $y \in \mathbb{R}^N$ by

$$y_i = \sum_{j} x_{i,j},$$

then the edge partition entropy *and* vertex partition entropy *are given by*

$$H^e(x) = -\sum_{e \in E} x_e \log x_e, \quad H^v(x) = -\sum_{i \in V} y_i \log y_i,$$

respectivly.

Theorem 15 *For a max-plus system of the sort outlined in section 1 the PPA max-plus exponent and edge criticalities defined in definition 11 are given by*

$$\widehat{\lambda} = \max_{x \in \Delta_G} \langle \beta, x \rangle + \sqrt{2 \langle \alpha^2, x \rangle [H^e(x) - H^v(x)]},$$

$$\widehat{P} = \arg \max_{x \in \Delta_G} \langle \beta, x \rangle + \sqrt{2 \langle \alpha^2, x \rangle [H^e(x) - H^v(x)]},$$

where $\alpha^2 \in \mathbb{R}^{N \times N}$ is the componentwise square of α. Furthermore the PPA max-plus exponent is greater than or equal to the max-plus exponent,

$$\widehat{\lambda} \geq \lambda.$$

13

Proof. For a path of length n, $\sigma \in \Sigma_n$ define $X(\sigma) \in \mathbb{R}^{N \times N}$ by $X(\sigma)_{i,j} = \frac{1}{n} \#[(i,j) \in \sigma]$ where $\#[(i,j) \in \sigma]$ is the number of times that the path σ traverses the edge (i,j).

For an open cover $\Delta_G = \bigcup_{\gamma \in C} U_\gamma$,

$$\widehat{\lambda} = \max_{\gamma \in C} \lim_{n \to \infty} \underbrace{\max_{\sigma \in \Sigma_n : X(\sigma) \in U_\gamma} \frac{1}{n} \widehat{W}(\sigma)},$$

where the underbraced term is greater than or equal to

$$\min_{x \in U_\gamma} \langle \beta, x \rangle + \sqrt{\min_{x \in U_\gamma} \langle \alpha^2, x \rangle} \big[\lim_{n \to \infty} \max_{\sigma \in \Sigma_n : X(\sigma) \in U_\gamma} \frac{t_\sigma}{\sqrt{n}} \big]$$

and less than or equal to

$$\max_{x \in U_\gamma} \langle \beta, x \rangle + \sqrt{\max_{x \in U_\gamma} \langle \alpha^2, x \rangle} \big[\lim_{n \to \infty} \max_{\sigma \in \Sigma_n : X(\sigma) \in U_\gamma} \frac{t_\sigma}{\sqrt{n}} \big].$$

We now apply an extreme value result from [11]. Let $[t(n)]_{n=1}^\infty$ be a sequence of i.i.d. $(0,1)$-Gaussians. For a measure-1 set of such sequences

$$\lim_{n \to \infty} \frac{\max_{k=1}^n t(k)}{\sqrt{2 \log n}} = 1.$$

Therefore

$$\lim_{n \to \infty} \max_{\sigma \in \Sigma_G^n : X(\sigma) \in U_\gamma} \frac{t_\sigma}{\sqrt{n}} = \sqrt{2 \big(\lim_{n \to \infty} \frac{1}{n} \log \#_n[\sigma \in U_\gamma] \big)},$$

where $\#_n[\sigma \in U_\gamma]$ is the number of paths $\sigma \in \Sigma_n$ with $X(\sigma) \in U_\gamma$.

This limit can be calculated using the large deviation principal from [12],

$$\lim_{n \to \infty} \frac{1}{n} \log \#_n[\sigma \in U_\gamma] = \max_{x \in \Delta_G \cap U_\gamma} H^e(x) - H^v(x).$$

Finally taking the limit $\max_{\gamma \in C} \|U_\gamma\| \to 0$ gives

$$\widehat{\lambda} = \max_{x \in \Delta_G} \langle \beta, x \rangle + \sqrt{2 \langle \alpha^2, x \rangle [H^e(x) - H^v(x)]}.$$

The performance guarantee inequality follows from lemma 13. Since the weight of each path is itself a non-decreasing function of the independent noise terms it follows that the path weights are associated. Also by construction the approximated path weights are independent with $W(\sigma)$ and $\widehat{W}(\sigma)$ identically distributed for each σ so that

$$\mathbb{E}[\max_{\sigma \in \Sigma_n} W(\sigma)] \leq \mathbb{E}[\max_{\sigma \in \Sigma_n} \widehat{W}(\sigma)].$$

Finally since the limits λ and $\widehat{\lambda}$ exist with probability-1 they are equal to their expectations and

$$\lambda = \lim_{n \to \infty} \frac{1}{n} \mathbb{E}[\max_{\sigma \in \Sigma_n} W(\sigma)] \leq \lim_{n \to \infty} \frac{1}{n} \mathbb{E}[\max_{\sigma \in \Sigma_n} \widehat{W}(\sigma)] = \widehat{\lambda}.$$

■

The formula for the PPA critical statistics gives some insight into the factors determining the maximally weighted path in the original max-plus linear system. Each path's weight has a deterministic and random part and the critical path is the path that maximizes the sum of these parts. Consider paths σ with $X(\sigma)$ in a small ball containing $X(c^*)$ where c^* is the critical cycle in the deterministic system, these path's weights will have large deterministic part but there will be relatively few of them and they will be unable to freely explore the whole graph in order to have a large random part. On the other hand paths with $X(\sigma)$ in a small ball in the center of Δ_G will not have a large deterministic part but there will be many of them and they will be able to freely explore the whole graph so that there will be some paths in this ball with large random part. It is this balance that determines the critical statistics of the true max-plus system as well as the PPA system.

Example 16 *Taking the max-plus linear system from example 8 we represent $X(\sigma) \in \mathbb{R}^{2 \times 2}$ by $x \in \mathbb{R}^4$ with*

$$x = [X(\sigma)_{1,1}, X(\sigma)_{1,2}, X(\sigma)_{2,1}, X(\sigma)_{2,2}].$$

Δ_G is then the triangle with vertices $(1, 0, 0, 0)$, $(0, \frac{1}{2}, \frac{1}{2}, 0)$ and $(0, 0, 0, 1)$. The PPA Statistics are given by

$$\widehat{\lambda} = \max_{x \in \Delta_G} \langle \begin{pmatrix} 1 \\ 0 \\ 3 \\ 0 \end{pmatrix}, x \rangle + a\sqrt{2[H^e(x) - H^v(x)]},$$

with $\widehat{P}$ the $\arg\max$ of the above, where

$$H^e(x) = -x_1 \log x_1 - x_2 \log x_2 - x_3 \log x_3 - x_4 \log x_4,$$

and

$$H^v(x) = -(x_1 + x_2) \log(x_1 + x_2) - (x_3 + x_4) \log(x_3 + x_4).$$

We solve this optimisation problem to obtain $\widehat{\lambda}$ and $\widehat{P}$ using gradient ascent. The results are displayed in figure 2.
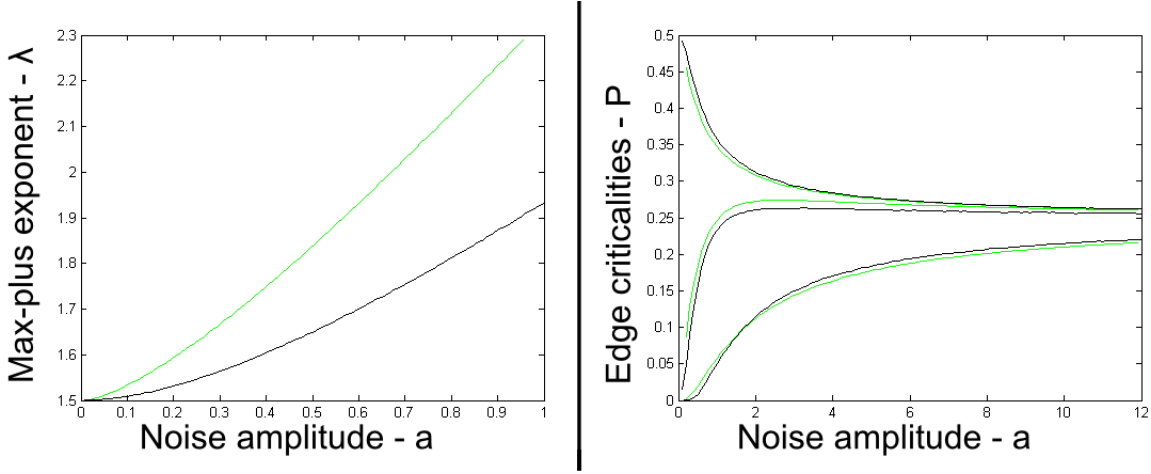
Figure 2: Critical path statistics and PPA vs noise amplitude, max-plus exponent (left), edge criticalities (right) with $P_{1,1}$, $P_{1,2} = P_{2,1}$ and $P_{2,2}$ top to bottom in black and approximations $\widehat{\lambda}$ and $\widehat{P}$ in green.

Example 17 *For an irreducible aperiodic graph G define the homogenous noise max-plus linear system on G by setting $\beta_{i,j} = 0$ if (i,j) is an edge in G and $-\infty$ otherwise and setting $\alpha_{i,j} = 1$ for all (i,j).*

The PPA statistics of these systems are determined by

$$\widehat{\lambda} = \max_{x \in \Delta_G} \sqrt{2[H^e(x) - H^v(x)]},$$

which is equal to $\sqrt{2\log \rho}$ where ρ is the classical Perron root *of G's adjacency matrix $\mathcal{A}$. The maximum is attained by the measure of maximal entropy P^*, defined by*

$$P_{i,j}^* = \mathcal{A}_{i,j}\frac{u_i v_j}{\rho},$$

where u is defined by $u\mathcal{A} = \rho u$ with $\|u\|_{l1} = 1$, and v by $\mathcal{A}v = \rho v$ with $\|v\|_{l1} = 1$.

Define χ_n the greedy path of length n by setting $\chi_n(1) = 1$ and then choosing the maximally weighted edge at each step inductively so that

$$a_{\chi_n(k+1),\chi_n(k)}(k) = \max_i a_{i,\chi_n(k)}(k)$$

for $k = 2, ..., n$. Since ϕ_n is the maximally weighted path of length n we have

$$\lim_{n \to \infty} \frac{1}{n}W(\chi_n) = \widetilde{\lambda} \le \lambda \le \widehat{\lambda}.$$

Now consider the homogeneous noise max-plus linear system on $K(N)$, the complete graph on N vertices. The PPA statistics are given by $\widehat{\lambda} = \sqrt{2\log N}$ and $P_{i,j} = N^{-2}$ for all (i,j). The greedy path's max-plus exponent is

$$\widetilde{\lambda} = \lim_{n \to \infty} \frac{1}{n}W(\chi_n) = \mathbb{E}_{k=1}^N t(k),$$

16

where $[t(k)]_{k=1}^{\infty}$ is a sequence of i.i.d. $(0,1)-Gaussians$. Applying the same extreme value result we used in the proof of theorem 15 gives,

$$\lim_{N\to\infty}\frac{\widetilde{\lambda}_N}{\lambda_N} = \lim_{N\to\infty}\frac{\lambda_N}{\lambda_N} = 1.$$

So that the proportional error in the approximation $\lambda \approx \widehat{\lambda}$ goes to zero in the limit of a large highly interconnected system.

Figure 3 shows the convergence of the ratio of the true exponent (which we approximate through simulation) to the PPA exponent for Gaussian noise systems on $K(N)$. We also make the same plot for systems with uniformly distributed noise terms and exponentially distributed noise terms. The PPA exponents for the uniform case are given by

$$\widehat{\lambda}_N = 1 - \frac{1}{eN},$$

where $e = 2.718...$ and for the exponential case the PPA exponents satisfy

$$\widehat{\lambda}_N = \log N + \log(1 + e\widehat{\lambda}_N),$$

and we can find them by iterating the RHS as it is a contraction mapping.

Example 18 *To demonstrate this phenomena in a more robust setting we generate a family of randomly configured homogeneous noise max-plus linear systems on Erdos Renyi graphs over a range of parameters and compare their PPA max-plus exponent to a Monte Carlo approximation of their true max-plus exponent. The results are displayed in figure 4. Observe that the true max-plus exponent also appears to be a function of the Perron root and that the proportional error decreases with the size of the root.*

Example 19 *Since the Perron root of the adjacency matrix is roughly the average degree of a vertex in the graph we might hope that having a large Perron root would guarantee accuracy of the PPA exponent, as a large Perron root would mean we had a large highly interconnected system. However this is not always the case as we have the following counter example.*

For some $m, N \in \mathbb{N}$ let $G = \langle V, E \rangle$ be a graph with $V = V_1 \cup V_2 \cup ... \cup V_m$ where for $i = 1, 3, 4, ..., m$ $V_i = \{v(i)\}$ contains a single vertex and

$$V_2 = \{v(2,j) : j = 1, 2, ..., N^m\},$$

contains N^m vertices. Let $E = E_1 \cup E_2 \cup ... \cup E_m$ where

$$E_1 = \{[v(2,j), v(1)] : j = 1, 2, ..., N^m\},$$

contains an edge from vertex $v(1)$ to every vertex in V_2 and where

$$E_2 = \{[v(3), v(2,j)] : j = 1, 2, ..., N^m\},$$

contains an edge from every vertex in V_2 to vertex $v(3)$ and where $E_i = \{(v(i+1), v(i)\}$ contains an edge from $v(i)$ to $v(i+1)$ for $i = 3, 4, ..., m$ where we identify $v(m+1)$ with $v(m)$. See figure 3.

This graph's adjacency matrix has Perron root $\rho = N$. Every path through G is quasi periodic, visiting $V_1, V_2, ..., V_m, V_1...$ and the only possible choice in how to move through G is which of the N^m routes from $v(1)$ to $v(3)$ to pick in each cycle. The average weight of the maximally weighted path, the max-plus exponent is therefore

$$\lambda = \frac{1}{m}(\mathbb{E}[\max_{j=1}^{N^m} t_j] + \mathbb{E}[t']),$$

where the t_j are i.i.d. $(0, \sqrt{2})$ Gaussians and t' is an $(0, \sqrt{m-2})$ Gaussian so that

$$\lambda = \frac{\sqrt{2}\sqrt{2 \log N^m} + 0}{m} = 2\sqrt{\frac{\log N}{m}}.$$

Thus for fixed N and large m we have a graph with Perron root N but arbitrarily small max-plus exponent.
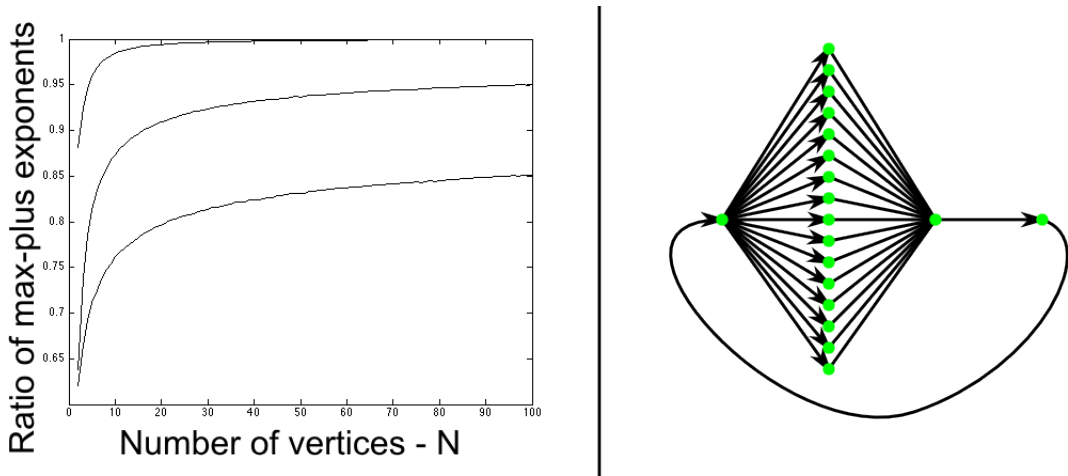

Figure 3: Left: ratio of max-plus exponent to PPA exponent $\lambda/\widehat{\lambda}$ vs graphs size for homogenous noise systems on complete graphs with uniform noise terms (top), Gaussian noise terms (middle) and exponential noise terms (bottom). Right: counter example graph for $N = 2$, $m = 4$.

Conjecture 20 *There is a class of graphs (e.g. Erdos Renyi graphs with parameters in some range) for which the PPA exponent is an accurate approximation of the true max-plus exponent. The relative error between the two exponents converges to zero as we consider large graphs in this class.*

Furthermore if we consider homogenous noise max-plus linear systems on this same class of graphs but with different noise term distributions (e.g. uniform, exponential,...) then the above statement still holds and the accuracy of PPA will be superior for distributions with lighter tails.

Example 21 *We generate a further randomly configured proportional noise max-plus linear system by constructing an Erdos Renyi graph with $N = 10$ and $p =*

0.5 *and then assign each edge* (i, j) *an edge weighting parameter* $\delta_{i,j}$ *which is itself an independent mean-1 exponential. The edge's weight sequence is then defined by* $w[(i, j), k] = \delta_{i,j}(1 + aZ_{i,j}[k])$ *where* $a \in \mathbb{R}$ *is the single noise parameter and* $[Z_{i,j}(k)]_{k=1}^{n}$ *is an independent sequence of i.i.d.* $(0, 1) - Gaussians.*

Through Monte Carlo simulation we approximate the max-plus exponent λ *of this system and compare it to the PPA exponent* $\widehat{\lambda}$ *computed using gradient ascent. The results are displayed in figure 4.*
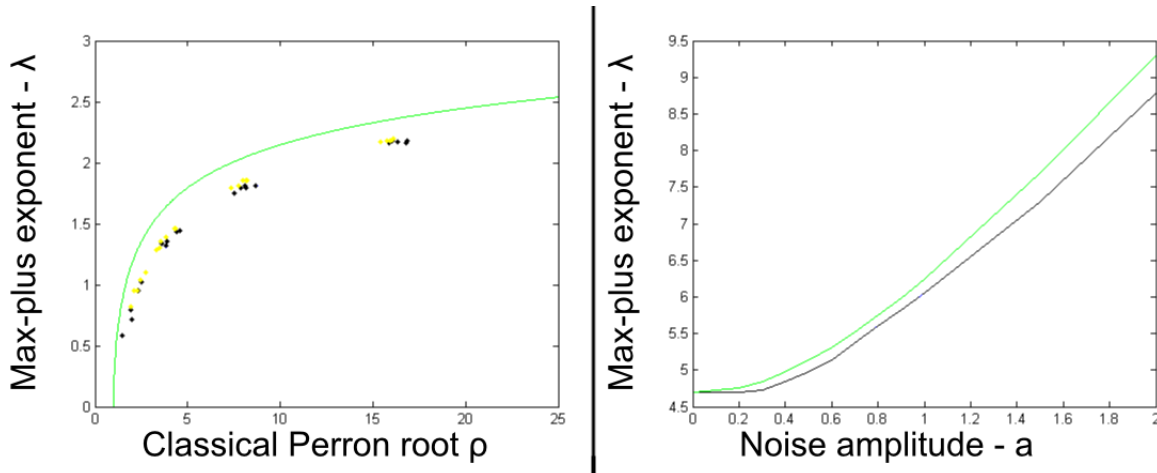

Figure 4: Left: max-plus exponent vs classical Perron root for randomly configured homogeneous noise systems with PPA approximation in green, for $p = 0.5$ there are five samples of each of $N = 4, 8, 16$ and 32 clustered left to right and in black, for $p = 0.25$ there are five samples of each of $N = 8, 16, 32$ and 64 clustered left to right in yellow. Right: max-plus exponent vs noise amplitude for proportional noise system with PPA approximation in green.

4   Conclusion

In this paper we introduced the notion of conflux turnpike behaviour, the edge criticalities and the PPA approximation. Theorems 9, 10 and 15 only apply to systems with Gaussian noise as outlined in section 1. Can these results be extended to a broader class of stochastic max-plus linear system?

The max-plus exponent has received considerable attention as it is equal to the reciprocal of the throughput of a queuing system. Are there useful practical interpretations of the edge criticalities? For instance should edges with high criticality be prioritised to have their weight reduced somehow in order to reduce the exponent? Can the PPA critical statistics be used to analytically optimise the constrained design of a queuing system?

We have shown that the PPA approximation is accurate for homogeneous systems on large complete graphs and large Erdos Renyi graphs. These homogeneous systems (introduced in example 17) should be of theoretical interest as they are the simplest

class of stochastic max-plus linear systems with interesting graph topology. Answering conjecture 20 would give us a sub class of systems where the max-plus exponent depends only on the Perron root of the associated adjacency matrix, adding to the existing links between max-plus matrix algebra and classical Perron Frobenius theory.

References

[1] R. A. Cuninghame-Green. (1962). Describing Industrial Processes with Interference and Approximating Their Steady-State Behaviour. *Oper. Res. Quarterly.* Vol. 13, pp 95-100

[2] B. Heidergott. (2007). Max-Plus Linear Stochastic Systems and Perturbation Analysis. *Springer.*

[3] F. Baccelli, G. Cohen, G. J. Olsder and J. P. Quadrat. (1992). Synchronisation and linearity. *Wiley.*

[4] G. Merlet. (2008). Cycle time of stochastic max-plus linear systems. *Electronic Journal of Probability.* pp 322-340

[5] F. Baccelli and D. Hong. (2000). Analytic Expansions of Max-Plus Lyapunov Exponents. *Annals of Applied Probability.* Vol. 10, pp. 779-827

[6] R. Goverde, B. Heidergott and G. Merlet.(2011). A coupling approach to estimating the Lyapunov exponent of stochastic max-plus linear systems. *European Journal of Operations Research.* Vol. 210, pp 249-257

[7] I. Necoara, E.C. Kerrigan, B. De Schutter, and T.J.J. van den Boom, Finite-horizon min-max control of max-plus-linear systems, IEEE Transactions on Automatic Control, vol. 52, no. 6, pp. 10881093, June 2007.

[8] T Muranta .(1989). Petri nets: Properties, analysis and applications. *Proceedings of the IEEE* Vol. 77, pp 541-580

[9] Jean Mairesse. (1997). Products of Irreducible Random Matrices in the (max, +) Algebra. Advances in Applied Probability, Vol. 29, No. 2, pp. 444-477

[10] A. Oliveira and P. Eduardo. (2012). Asymptotics for Associated Random Variables *Springer.*

[11] Y. Mittal and D. Ylvisaker. (1967). Strong Laws for the Maxima of Stationary Gaussian Processes. *Annals of probability* Vol. 4, pp 357-371.

[12] Large Deviations Techniques and Applications Series: Stochastic Modelling and Applied Probability, Vol. 38 Dembo, Amir, Zeitouni, Ofer 2nd ed. 1998. 2nd printing, 2010, XVI, 396 p.