Using Tropical Maths to Model Ribosome Dynamics

James Hook and **Marianne Johnson**, University of Manchester

1 The mathematics of scheduling

If two activities must be performed consecutively then the time required to complete both is the sum of the individual times, but if they may be performed concurrently then the time required is the maximum of the individual times. For instance, imagine you are designing a train timetable. What is the earliest time at which a train can depart a given station? Certainly, it cannot leave until after it has arrived (despite what departure boards are prone to suggest!), but what other factors should be considered? Suppose that there are passengers arriving at the station on connecting trains from three other towns. We would therefore like to schedule our train to depart after these three trains have arrived. We don't need to worry about the order in which the three connecting trains arrive; we only care that the departure time of our train occurs after the *maximum* of the three arrival times. To give the passengers a chance to change platforms and board, we should also *add* on a fixed amount of time to this maximum. Of course, in reality, rail networks involve huge numbers of trains with more complicated systems of dependencies between them. This sort of timing-analysis will therefore yield large systems of non-linear equations involving sums and maxima.

2 Tropical maths

Roughly speaking, tropical (or max-plus) mathematics is any kind of maths which uses the set $\mathbb{R}_{\max}$ of real numbers together with the extra element $-\infty$, under the operations of addition (which we shall denote by $\otimes$) and maximum (which we shall denote $\oplus$). This algebraic structure is called the *tropical semiring*.[1] Semirings are algebraic structures similar to rings, but which suffer from the unfortunate property that their addition need not be invertible; the most familiar example of a semiring is the set of natural numbers. The 'tropical' nature of $\mathbb{R}_{\max}$ lies in the fact that usual addition of numbers plays the role of multiplication, whilst maximisation plays the role of addition.[2] So for example, $2 \oplus 3 = 3$ whilst $2 \otimes 3 = 5$.

In spite of this strange definition, the tropical semiring shares many of the properties of a field; the main difference being that there is no analogue of subtraction in tropical mathematics. Of course, this makes a big difference mathematically! To take a simple example, notice that the equation $3 \oplus x = \max(3, x) = 2$ has no solution tropically. Notice that instead tropical addition satisfies $x \oplus x = \max(x, x) = x$.

Returning to our train timetable problem we see that in tropical terms the equations become *linear*. Of course, this does not immediately solve the problem, as we have already seen that the most basic of linear equations may have no solution in tropical mathematics. However, this way of thinking allows us to represent our complicated system of equations in a compact form (namely, by using tropical matrices and vectors) and to analyse its behaviour using techniques from tropical maths.

3 A biological queue

As we have seen, tropical maths can be used to model certain types of queuing system. This modelling has proved useful in a wide variety of applications. One particularly novel application, which David Broomhead was involved in, is a tropical model of part of the process used by a cell to create proteins. We'll need to know a little biology to understand this properly.

Amino acids are the building blocks of proteins. Whenever a cell produces a protein it does so by first creating a chain of amino acids. This amino acid chain, or peptide, can then be folded up or combined with other peptides to produce the required protein. The properties of an amino acid chain are completely determined by the sequence of amino acids involved. In the human body there are 20 standard species of amino acid that are used to produce amino acid chains.

The order in which amino acids are combined is dictated by the genetic code. The cell begins by creating something called an mRNA strand, which is a single strand of nucleotides transcribed from DNA. The 'm' in mRNA stands for 'messenger', indicating that genetic information is encoded in and passed on via this strand. For simplicity, we can think of an mRNA strand as encoding a template or sequence of instructions which the cell uses to produce amino acid chains and hence proteins. Each instruction is of the form 'get a particular amino acid and add it to the chain' with a final instruction of 'stop making this chain and release it'. The nucleotides in the mRNA strand are grouped into triplets, which are called codons and each codon corresponds to a particular instruction, and hence in general, to a particular amino acid. A ribosome is a biological gadget which is able to travel along the mRNA strand decoding and performing these instructions.

Figure 1: A cartoon of a ribosome queue. Once a ribosome (red) attaches itself to the mRNA strand, it begins to read the sequence of instructions provided by the codons (coloured triples) and creates an amino acid chain as it moves along the strand.

Figure 2: A coloured electron microscope photo of a ribosome queue. Ribosomes (large blueish dots) join the mRNA strand (red) at the top right of the picture and move along the U-shaped strand. In this picture you can see a few areas where the ribosomes are bunched up, held up by a ribosome waiting at a slow codon. As you look further along the strand you can clearly see the growing peptide chain (green) that is being constructed by each individual ribosome.

The process by which a ribosome attaches to the mRNA strand at one end and moves down it, scanning the mRNA codon sequence and constructing the desired amino acid chain as it goes, is called *mRNA translation* (see Figure 1 for a simplistic illustration). As the ribosome visits each codon in the mRNA it determines which amino acid is coded for this codon (illustrated by colour in Figure 1). At each codon the ribosome must wait until the correct amino acid is present. It then then adds that amino acid to the growing chain of amino acids that it is constructing and moves on. The ribosome may have to wait some time for the correct amino acid to be present. Typically a single strand of mRNA will be occupied by many different ribosomes at the same time. As the ribosomes move down the mRNA strand they can block each other causing traffic jams. This phenomenon will affect the time taken for each chain to be completed and thus affect the peptide production rate, which is a very important factor in the cell's metabolism. The simplistic illustration of mRNA translation shown in Figure 1 makes it clear that tropical mathematics naturally lends itself to modelling this process. In the electron microscope photo in Figure 2 you can clearly see the amino acid chain being constructed as the ribosomes move along the strand (just as we have indicated in our cartoon).

4 A tropical model of mRNA translation

David and his co-authors proposed the following tropical model for the dynamics of mRNA translation [1]. Let n be the length of the mRNA strand, in terms of the number of codons it consists of. Recall that the final codon in any strand is a stop codon, which encodes the instruction to stop making the current chain and release it. Now for each ribosome that translates the mRNA we define a vector of transition times. For the kth ribosome to join the mRNA the transition time vector is denoted $x(k)$ and for $i = 1, \ldots, n$ the ith component $x(k)_i$ records the time at which the kth ribosome arrives at the ith codon of the mRNA. The $(n+1)$th component $x(k)_{n+1}$ records the time at which the kth ribosome disconnects from the mRNA chain having successfully translated it.

When a ribosome arrives at a particular codon (excepting the final stop codon) it has to wait for a suitable amino acid to become available and then attach it to its growing peptide chain. The amount of time required for this to happen might be well modelled by an exponentially distributed random variable; however, for the sake of simplicity we will assume here that the total time required for this process is a constant that only depends on the species of amino acid coded for by the codon. Thus for $i = 1, \ldots, n-1$ we set $a(i)$ to be the time taken to add the amino acid associated with codon i.

It it not possible for one ribosome to overtake another on the mRNA strand. It is therefore possible that one ribosome, which has added the amino acid associated with its current codon, may have to wait some extra time before it can move onto the next codon if there is another ribosome in front of it.

The only remaining assumptions needed for the model are the initiation time and termination time. That is the time gap between the first codon of the mRNA being free and a ribosome binding onto it, which we will denote by b and the time taken for a ribosome to disconnect from the mRNA at the end of the chain once it has completed its translation. We will denote this second time by c.

These waiting times together with the no-overtaking rule give us enough information to formulate a tropical model of the ribosome dynamics. The $(k+1)$th ribosome will join the mRNA strand b seconds after the kth ribosome moves from the first to the second codon so that

$$x(k+1)_1 = b + x(k)_2 = b \otimes x(k)_2. \tag{1}$$

For $i = 2, \ldots, n$ the $(k+1)$th ribosome will arrive at the ith codon either immediately after it has included the $(i-1)$th amino acid (this happens if it has not been blocked by the previous ribosome) or immediately after the kth ribosome moves from the ith to the $(i+1)$th codon (which happens when it has been blocked by the previous ribosome). Thus

$$\begin{aligned} x(k+1)_i &= \max\{a(i-1) + x(k+1)_{i-1}, x(k)_{i+1}\} \\ &= a(i-1) \otimes x(k+1)_{i-1} \oplus x(k)_{i+1}. \end{aligned} \tag{2}$$

Finally the $(k+1)$th ribosome disconnects from the mRNA strand c seconds after it includes the nth amino acid so that

$$x(k+1)_{n+1} = c + x(k+1)_n = c \otimes x(k+1)_n. \tag{3}$$

Thus the vectors of times for the different ribosomes satisfy a tropical linear equation which relates the vector of times for the kth and $(k+1)$th ribosomes. Indeed, equations (1)–(3) can be written compactly in matrix form as

$$x(k+1) = [A \otimes x(k+1)] \oplus [B \otimes x(k)],$$

where the procedure for multiplying matrices and vectors tropically mimics the usual procedure, replacing the usual operations of muliplcation and addition by $\otimes$ and $\oplus$ respectively.

For example, suppose that we have a very short mRNA strand with only four codons, the first three of which encode instructions to add particular amino acids to the chain, followed by the fourth and final stop codon. Then the equations are given by

$$\begin{bmatrix} x(k+1)_1 \\ x(k+1)_2 \\ x(k+1)_3 \\ x(k+1)_4 \\ x(k+1)_5 \end{bmatrix} = \begin{bmatrix} -\infty & -\infty & -\infty & -\infty & -\infty \\ a(1) & -\infty & -\infty & -\infty & -\infty \\ -\infty & a(2) & -\infty & -\infty & -\infty \\ -\infty & -\infty & a(3) & -\infty & -\infty \\ -\infty & -\infty & -\infty & c & -\infty \end{bmatrix} \otimes \begin{bmatrix} x(k+1)_1 \\ x(k+1)_2 \\ x(k+1)_3 \\ x(k+1)_4 \\ x(k+1)_5 \end{bmatrix}$$
$$\oplus \begin{bmatrix} -\infty & b & -\infty & -\infty & -\infty \\ -\infty & -\infty & 0 & -\infty & -\infty \\ -\infty & -\infty & -\infty & 0 & -\infty \\ -\infty & -\infty & -\infty & -\infty & 0 \\ -\infty & -\infty & -\infty & -\infty & -\infty \end{bmatrix} \otimes \begin{bmatrix} x(k)_1 \\ x(k)_2 \\ x(k)_3 \\ x(k)_4 \\ x(k)_5 \end{bmatrix}.$$

Note that these matrix equations do not currently give us a useful formula for $x(k+1)$ in terms of $x(k)$ as the right-hand side of the equation depends on $x(k+1)$ as well as $x(k)$. The exact details of how one goes about rearranging this equation are a bit too involved to go through in detail here. However, we note that the method of solution in tropical algebra turns out to bear a remarkable similarity to that used when rearranging a conventional system of linear equations of the form $y = Ay + Bx$, where we find that $y = (I-A)^{-1}Bx$. Of course, in the tropical case, we do not have a satisfactory notion of subtraction, let alone matrix inverse, so it is clear that the method of solution cannot be exactly the same. It turns out that the correct analogue of $(I-A)^{-1}$ is the so-called Kleene star of A, which roughly speaking is defined by $A^\star = I \oplus A \oplus A^{\otimes 2} \oplus \dots$. The Kleene star of A can be viewed as the tropical version of the Taylor series for $(I-A)^{-1}$. Thus we obtain the equation $x(k+1) = A^\star \otimes B \otimes x(k)$, where $A^* \otimes B$ is the matrix

$$\begin{bmatrix} -\infty & b & -\infty & -\infty & -\infty \\ -\infty & b+a(1) & 0 & -\infty & -\infty \\ -\infty & b+a(1)+a(2) & a(2) & 0 & -\infty \\ -\infty & b+a(1)+a(2)+a(3) & a(2)+a(3) & a(3) & 0 \\ -\infty & b+a(1)+a(2)+a(3)+c & a(2)+a(3)+c & a(3)+c & c \end{bmatrix}$$

5 Why is this model useful?

On the whole, the model outlined above gives a good representation of ribosome dynamics, and can be viewed as an improvement on several widely used models which came before it. First of all, it takes into account the fact that different codons may have different waiting times associated with them and that therefore the nature of the *sequence* of codons to be translated is of critical importance. Secondly, many of the previous models of ribosome dynamics use a random sequential update rule, whereby at each step the movement of a randomly chosen ribosome is updated according to some fixed rule. Of course, this is not particularly realistic, as in general ribosomes will behave reasonably independently of one another, unless one is holding another up. The tropical model allows for the fact that multiple ribosomes can move simultaneously.

Understanding processes such as translation forms an important aspect of systems biology. The tropical model allows one to predict protein production rates and codon occupation densities.

6 Where can I find out more about tropical maths and how it is used?

Tropical or max-plus linear algebra has been an active area of study since the mid 1970s. Much of the initial interest and motivation in this area of mathematics can be seen in connection to the modelling of discrete event systems typically arising in areas involving allocation of resources, scheduling and queueing theory. Indeed, as we have seen, tropical matrices are particularly amenable to the modelling of queueing problems. Today there are a wealth of results concerning the dynamics of such systems, including descriptions of periodic regimes and asymptotic behaviour (see [2] for details). Such results required the development of tropical analogues of several ideas from classical linear algebra such as spectral theory and linear independence (see [3] for a comprehensive treatment).

Tropical mathematics has attracted a considerable amount of attention amongst the mathematical community in recent years, spurred on by applications in a wide range of research and application areas including optimisation and scheduling, semigroup theory, algebraic geometry, economics and numerical linear algebra to name but a few. Many of these new applications take a slightly different point of view, by considering the tropical situation as a shadow (or logarithmic limit) of a classical algebraic variety, by means of a particular valuation. The resulting tropical variety shares many features with the original variety, but has a pleasing combinatorial structure that is often easier to reason with (see [4] for a survey of results in tropical geometry).

7 A bit about David and tropical maths

David Broomhead became interested in tropical mathematics through his role as the head of a large interdisciplinary research project in Manchester called CICADA back in 2007. David was instrumental in setting up a weekly reading group on tropical maths, which has since evolved into a thriving research group. One of the great achievements of this group was to get pure and applied mathematicians (such as ourselves), as well as biologists and computer scientists talking to each other and working together. Without David this simply would not have happened.

Notes

1. Depending upon the particular application, some authors prefer to use the natural numbers, integers or rationals rather than real numbers. Some work with minimum instead of maximum; if this applies to you, then please read this article standing on your head!
2. The origin of the word 'tropical' in a mathematical context stems from the involvement of Brazilian mathematician and computer scientist Imre Simon in the late 1980s and early 1990s. Before this time, research in this area was referred to using the more descriptive adjective 'max-plus'.

References

1 Brackley, C.A., Broomhead, D.S., Romano, M.C. and Thiel, M. (2012) A max-plus model of ribosome dynamics during mRNA translation, *J. Theoret. Biol.*, vol. 303, pp. 128–140.

2 Heidergott, B., Oldser, G.J. and van der Woude, J. (2006) *Max Plus at Work. Modeling and Analysis of Synchronized Systems: A Course on Max-plus Algebra and Its Applications*, Princeton Series in Applied Mathematics, Princeton University Press, Princeton, NJ.

3 Butkovic, P. (2010) *Max-Linear Systems: Theory and Algorithms*, Springer Monographs in Mathematics, Springer-Verlag London, Ltd, London.

4 Maclagan, D. (2012) *Introduction to Tropical Algebraic Geometry*, Contemporary Mathematics, 580, American Mathematics Society, Providence, RI, pp. 1–19.