From *SIAM News*, Volume 43, Number 8, October 2010

Max-Plus Algebra: From Discrete-event Systems to Continuous Optimal Control Problems

By James Case

Discrete events come in different shapes and sizes. Earthquakes, hurricanes, and solar eclipses, occurring in necessarily finite numbers in a given time or place, certainly qualify. So do the arrival of planes at an airport, retirements of Supreme Court justices, and meteor showers in the night sky. Accurate records of such events have been kept for decades, centuries, or even (in a few cases) millennia. The events may be as random as the flooding of the Mississippi, or as regular as the ticking of a clock. They may be as subject to human control as the arrival of airplanes, or as far beyond it as the appearance of Halley's comet.

Because no single mathematical model—or type of model—can be expected to mimic all such phenomena, a variety of mathematical tools contribute to the analysis of discrete-event systems. A particularly interesting tool known as max-plus algebra permits a remarkably complete and concise treatment of certain problems in the design and control of such systems. See the box for an introduction to max-plus algebra.

Applications to Transportation Systems

Many of the applications of max-plus algebra are to transportation networks, such as highway systems, bus routes, and railway schedules. Figure 1 depicts a simple railway network served by four distinct trains, two traversing the middle loop and a single train serving each of the other two loops. A well-designed schedule for such a network places trains in stations simultaneously, so that passengers can change from one to another simply by walking across the platform between them. Dutch train schedules have long been designed with this in mind.

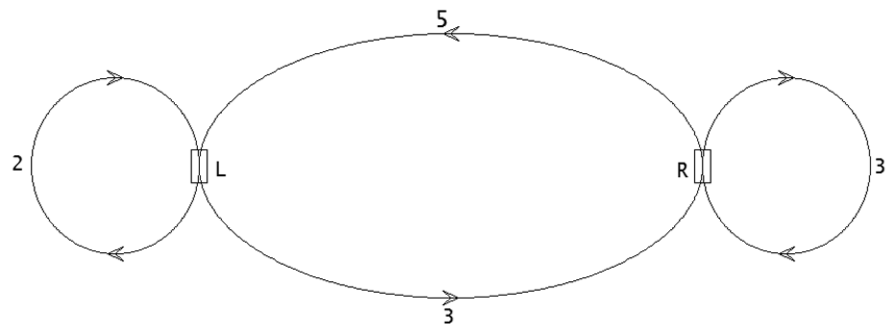

Figure 1. *A simple railway network. The numbers along the tracks represent transit times, and the boxes labeled L and R represent stations.*

In a simple model of the network shown in the figure, $x_S(k)$ denotes the instant at which the kth (simultaneous) departure from station S (either L or R) takes place; $a_{SS'}$ denotes the transit time between stations S and S', and δ the time two trains must stand side by side while passengers switch from one to the other. Then, because no train can leave a station before the connecting train has arrived, *and* the two have tarried side by side for δ minutes,

$$x_L(k+1) \geq \max\{x_L(k) + a_{LL} + \delta, x_R(k) + a_{RL} + \delta\}$$
$$x_R(k+1) \geq \max\{x_R(k) + a_{RR} + \delta, \ x_L(k) + a_{LR} + \delta\}.$$

If time is not to be wasted, the inequalities must be replaced by equalities, which can be further reduced to the vector–matrix form

$$\boldsymbol{x}(k+1) = A \otimes \boldsymbol{x}(k)$$

by writing $\boldsymbol{x}(k) = [x_L(k), x_R(k)]^T$ and absorbing the transfer time δ into the matrix $A = (a_{SS'})$ of transit times. The superscript T indicates matrix transposition. The preceding equation has the obvious solution $\boldsymbol{x}(k) = A^{\otimes k}\boldsymbol{x}(0)$, where $A^{\otimes k} = A \otimes A \otimes \ldots \otimes A$ to k factors. If $\boldsymbol{x}(0)$ is an eigenvector v of A, then $\boldsymbol{x}(k) = A^{\otimes k}v = \lambda^{\otimes k}v = k\lambda\mathbf{1}+v$, where $\mathbf{1}$ is a conformable vector of 1's. Finally, the fact that

$$\begin{bmatrix} 5 & 2 \\ 3 & 3 \end{bmatrix} \otimes \begin{bmatrix} 1-h \\ h \end{bmatrix} = 4 \otimes \begin{bmatrix} 1-h \\ h \end{bmatrix}$$

for any h implies that 4 is the eigenvalue of the transit-time matrix A corresponding to the depicted network and $v = [1,0]^T$ is a suitable (eigen)vector of original departure times.

In [1], Heidergott, Olsder, and van der Woude applied the foregoing technique to successively more detailed models of the Dutch passenger train network. The simplest of their models concerns deluxe intercity trains only. It consists of 19 lines serving 70 stations over 361 track segments represented by edges. There are 317 departure events, subject to 44 simultaneity constraints. A minimum of 112 trains is necessary to properly serve the network in question.

Along with the deluxe intercity trains, the Dutch passenger system employs two other types: Local trains cover shorter distances at low speed, stopping at every station; express trains travel faster and stop at fewer stations than the locals, but are slower and make more stops than the deluxe trains. A model including all three types would involve at least 440 trains and a transit-time matrix A with thousands of rows and columns. Existing numerical methods are expected to prove equal to the task, if and when the requisite data become available.

Optimal Control Problems

Somewhat surprisingly, max-plus algebra has proved to be applicable to the solution of continuous-time, continuous-state optimal control problems of the form

$$\text{Maximize } J = \int_0^T L[x(t),u(t)]dt$$
$$\text{Subject to } \dot{x}(t) = f[x(t),u(t)], \text{etc.}$$

The "etc." refers to certain ancillary constraints on the state vectors $x(t)$, the control vectors $u(t)$, the terminal state $x(T)$, and various combinations thereof. Pontryagin's famous maximum principle characterizes the optimal pairs $[x^*(t), u^*(t)]$, $0 \leq t \leq T$ for which the objective functional J assumes its maximum value, while the complete solution of an optimal problem requires the further specification of a closed-loop optimal control $u^*(x)$ that "synthesizes" the optimal pairs in the sense that $u^*(x^*(t)) = u^*(t)$.

In favorable circumstances, the optimal closed-loop control $u^*(x)$ can be obtained directly—without recourse to Pontryagin's principle—by solving the Hamilton–Jacobi PDE

$$\max_u \{L(x,u) + \langle f(x,u), V_x \rangle\} = 0,$$

subject to boundary conditions imposed by the ancillary constraints, for the unknown value function $V(x)$. Here $\langle\cdot,\cdot\rangle$ denotes the ordinary inner product of vectors in $\mathbb{R}^n$, and $V(\xi)$ is the number obtained by integrating J along the solution of $\dot{x} = f(x,u^*(x))$ for which $x(0) = \xi$.

Hamilton–Jacobi equations have inhabited the intersection of physics and the calculus of variations for generations. Until the middle of the 20th century, attention focused on classical (continuously differentiable) solutions. Since then, motivated by developments in fluid mechanics, the theory of viscosity solutions of PDEs has emerged. These need not be classical solutions, as discontinuities in the gradient of the solution—and sometimes in the solution itself—are ubiquitous.

In practice, Hamilton–Jacobi equations can seldom be solved in closed form, unless $f(x,u)$ is of the linear form $Ax + Bu$ and $L(x,u)$ assumes the quadratic form $\frac{1}{2}\langle x, Qx\rangle - \frac{1}{2}\langle u, Ru\rangle$. Here A, B, Q, and R denote real matrices; Q and R must be symmetric and positive semidefinite, and R must be positive definite. The characteristics of the Hamilton–Jacobi equation are then linear ODEs with constant coefficients, which can in principle be solved in closed form. Max-plus algebra has been used, in recent years, to expand the class of problems for which this direct method is effective. It is effective for both finite- and infinite-horizon problems.

In either case, it is fruitful to define vector spaces over $\mathbb{R}_{\max}$, introduce the concept of semiconvexity for $\mathbb{R}_{\max}$-valued functions of several real variables, and observe that various classes of such functions indeed form vector spaces over $\mathbb{R}_{\max}$. The spaces of interest are typically complete, and a fairly comprehensive theory—including a duality theory—of "complete max-plus vector spaces" $\chi = (\chi,\oplus,\otimes)$ has been developed.

For infinite-time-horizon problems, in which $T = \infty$, it is useful to define a family of operators on the space of semiconvex functions $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}_{\max}$ as follows:

$$S_\tau\varphi(x) =$$
$$\int_0^\tau \left\{L(x(s)) - (\gamma^2/2)|u(s)|^2\right\}ds + \varphi(P_\tau x),$$

where $P_\tau x$ refers to the point subgroup associated with the ODE $\dot{x} = f(x, u^*(x))$. The collection of operators P_τ would form a group if $u^*(x)$ were, say, continuously differentiable, which is not generally the case in optimal control theory. Optimal closed-loop controls are often mildly discontinuous, and optimal paths frequently merge. Thus, we restrict attention to the semigroup $\{P_\tau\}_{\tau>0}$ of operators on $\mathbb{R}^n$. The related operators $\{S_\tau\}_{\tau>0}$ on the max-plus vector space χ are of interest both because they inherit from $\{P_\tau\}$ the semigroup properties $S_\sigma \circ S_\tau = S_{\sigma+\tau}$ and $S_0 = I$ and because the value function $V(x)$ that solves the Hamilton–Jacobi equation is a fixed point of each S_τ. Lastly, these operators can be shown to be "max-plus linear" in the sense that

$$S_\tau[a \otimes \varphi \oplus b \otimes \psi] = a \otimes S_\tau[\varphi] \oplus b \otimes S_\tau[\psi].$$

If max-plus vector spaces are in some sense spanned by bases, the operators S_τ can thus be associated with (presumably infinite) matrices likely to possess eigenvalues and eigenvectors.

The lack of additive inverses in max-plus algebra precludes the existence of anything as tractable as an orthonormal basis, but countably infinite bases—akin to Schauder bases in Banach spaces—have been found for the max-plus vector spaces of interest in control theory. Truncation of such bases produces finite-dimensional subspaces of max-plus function spaces, onto which the max-plus linear operators S_τ can be projected. And because these projections correspond to finite max-plus matrices, they are fully susceptible to the eigenmethods developed for discrete-event systems!

All this and more is described in the monograph [2], which treats finite- as well as infinite-time-horizon control problems, along with a few differential games and H_∞ estimation problems. The analysis is technical, as one might expect, particularly as it applies to the errors inherent in any truncation process. Yet the underlying ideas are well explained, and the book is commendably readable. It is carefully annotated, with a bibliography of more than a hundred items. Together with [1], it furnishes a welcome introduction to an exciting new branch of applied mathematics.

References

[1] B. Heidergott, G.J. Olsder, and J. van der Woude, *Max Plus at Work*, Princeton University Press, Princeton, NJ, 2006.
[2] W. McEneaney, *Max-Plus Methods for Nonlinear Control and Estimation*, Birkhäuser, Berlin, 2006.

James Case writes from Baltimore, Maryland.

A Max-Plus Algebra Primer

Max-plus algebra is predicated on the observation that the binary operations $\oplus$ and $\otimes$, defined by $x \oplus y = \max(x,y)$ and $x \otimes y = x + y$, turn $[-\infty,\infty)$ into a computer-friendly algebra. The element $e = -\infty$ serves as an "additive identity" in this algebra, because $x \oplus e = e \oplus x = \max(-\infty, x) = x$ in all cases; $\varepsilon = 0$ serves as a "multiplicative identity" on $(-\infty,\infty)$ for similar reasons. Indeed, the latter interval forms an abelian group under $\otimes$, and $[-\infty,\infty)$ forms—because of the lack of additive inverses—an abelian semigroup under $\oplus$. Finally, $\otimes$ is distributive over $\oplus$, and $e \otimes x = x \otimes e = e = -\infty$ for all x. Thus, the set $\mathbb{R}_{max} = (\mathbb{R}_{max}, \otimes, \oplus, e, \varepsilon)$ consisting of $[-\infty,\infty)$, together with the operations $\oplus$, $\otimes$, and distinguished elements e, ε, forms an algebra that is almost a field. The term "semifield" is becoming popular.

Most applications of max-plus algebra involve matrices of elements of $\mathbb{R}_{max}$, which can be manipulated in much the same way as real or complex matrices. In particular, $A \oplus B$ and $A \otimes B$ are defined componentwise, as follows: $[A \oplus B]_{ij} = a_{ij} \oplus b_{ij} = \max(a_{ij}, b_{ij})$ and $[A \otimes B]_{ik} = \bigoplus_{1 \le j \le n} (a_{ij} \otimes b_{jk}) = \max_j \{a_{ij} + b_{jk}\}$. There is an $n \times m$ null matrix, which has e's (aka negative infinities) in all positions, and an $n \times n$ identity matrix E, which differs from a null matrix only in that the e's on the main diagonal are replaced by ε's (aka zeros).

Perhaps the most important result concerning such matrices is an analogue of the Perron–Frobenius theorem of ordinary linear algebra. It asserts that any square irreducible matrix A with elements in $\mathbb{R}_{max}$ has a unique eigenvalue λ that satisfies $A \otimes v = \lambda \otimes v$ for some conformable eigenvector v with one or more finite elements also in $\mathbb{R}_{max}$. Moreover, λ is finite and can be computed by means of a simple formula.

To explain the eigenvalue formula—as well as the meaning of irreducibility in the max-plus context—it is necessary to describe the communication graph $G(A)$ associated with an $n \times n$ matrix A. $G(A)$ is a directed graph on the set $N = \{1, 2, \ldots, n\}$ of nodes in which an edge runs from i to j if and only if $a_{ji} \neq \varepsilon = 0$.* A path of length m in $G(A)$ is an ordered $(m+1)$-tuple of distinct nodes, each successive pair of which is joined by an appropriately directed edge. A circuit is a path that closes on itself. A is called irreducible if and only if it is possible to reach any node of $G(A)$ from any other node without traversing an edge in the wrong direction.

The quantity a_{ji} is called the weight of the edge from i to j, and the sum of the weights of all the edges in a path (or circuit) p in $G(A)$ is called the weight of p. This is denoted $|p|_w$; the length of p is $|p|_1$. The average edge weight of a path (or circuit) p is then $|p|_w / |p|_1$. Finally, $\lambda = \lambda(A) = \max_\gamma |\gamma|_w / |\gamma|_1$, where γ ranges over all circuits in $G(A)$. Other important theorems concern the numerical computation of eigenvalues and eigenvectors for large matrices.

*Such apparent reversal of the subscripts i and j is standard in the max-plus literature.