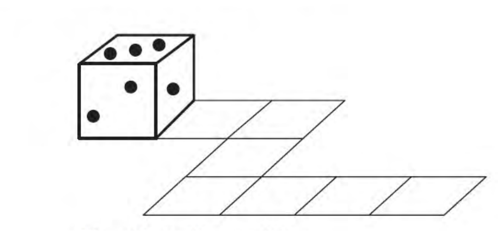
Notes of a Numerical Analyst

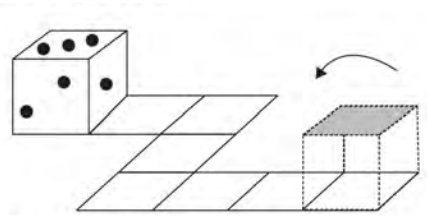
From Dice to Adjoints

NICK TREFETHEN FRS

David Acheson has a wonderful puzzle involving a dice rolling along a track [1]. When it gets to the end, what number will be on top?



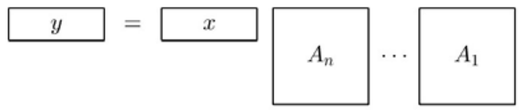
If you try to solve this in your head, it will drive you crazy. It's just too complicated. Eventually you'll give up and look in the kitchen drawer to see if you've got an old dice hanging around to help you out. But then Acheson reveals his elegant trick. Just run the problem backwards!



Imagine starting from the end, with the grey face on top, and tracking the position of that face step by step backwards to the start position. This is easy in your head. At the start, the grey face is facing to the left, opposite the single dot. So it must be the face with 6 dots.

The trick of running it backwards turns out to be at the heart of many things. What makes this problem hard in forward mode is that there are 24 states of the dice — 6 possible numbers on top, 4 rotations. What makes it easy in reverse mode is that we don't care about the rotations, just the position of the grey face. Actually, you could solve it forward via easy 6-state simulations: but you'd have to do six of them, not just one. One forward simulation tells you where the 3-face ends up, another where the 2-face ends up, and so on. After six runs (or three, exploiting symmetry), you'll have solved the problem. But that's nowhere near as slick as one run in reverse.

For an analogy from linear algebra, think of the 6×24 matrix y resulting from a product of a sequence of 24×24 matrices $A_1, \ldots, A_n$ finally times a 6×24 matrix x:



If you work from right to left multiplying the matrices in the usual order of composition, it's a sequence of big square matrix products, but if you start from the vector x and work from left to right, all the products are small rectangular ones.

Rolling dice is not a major problem of computational science, but the computation of derivatives is. The same switch from forward to reverse mode is what made the technique of Automatic Differentiation take off late in the last century [3]. More recently the idea has grown even more conspicuous in the technique of *backpropagation* for training neural networks. These are all ideas related to the distinction between an operator and its adjoint [2].

FURTHER READING

[1] D. Acheson, *The Spirit of Mathematics*, Oxford University Press, 2023.
[2] M. B. Giles and N. A. Pierce, An introduction to the adjoint approach to design, *Flow, Turbulence, and Combustion*, 65 (2000), 393–415.
[3] A. Griewank and A. Walther, *Evaluating Derivatives*, SIAM, 2008.

Nick Trefethen
Trefethen is Professor of Applied Mathematics in Residence at Harvard University.