Notes of a Numerical Analyst

Which is Smaller, $O(n^2)$ or $O(n^3)$?

NICK TREFETHEN FRS

An old dream is the "Fast Matrix Inverse", which would invert an $n \times n$ matrix in essentially $O(n^2)$ operations — $O(n^2 \log n)$, perhaps. Such a discovery would revolutionise computational science, as the FFT revolutionised signal processing with its $O(n \log n)$ operation count for an n-point discrete Fourier transform.

But despite the importance of the problem, nobody has ever found the FMI, nor proved that it cannot exist. Mostly we use the classical $O(n^3)$ algorithms. There are theoretical alternatives needing just $O(n^{2.37})$, but the constants are enormous.

I was discussing these matters with a colleague the other day who startled me by saying, "But computers already achieve $O(n^2)$! Just give it a try on your machine!"

I did that, and the result is shown in Figure 1. Sure enough, for small n, the shape looks like $O(n^2)$. A user working with $n < 1000$ might think that the FMI already exists and is running on their laptop. On the other hand for $n \gg 1000$ we see equally cleanly $O(n^3)$, as we learned in our numerical analysis courses.

One could discuss why these results look the way they do, but my interest is in the more basic question, what do they *mean?* Would it be fair to say "Yes, it's $O(n^3)$ in theory, but the bad running time doesn't kick in until n is quite large"?

For there is a paradox here: the computation would obviously be faster if there were no $O(n^2)$ component at all and the $O(n^3)$ kicked in right from the start. Or how about this: if the running times were longer by $2 \cdot 10^{-5}\, n$, the complexity would look beautifully like $O(n)$ for $n < 1000$, but of course that would not be a better algorithm.

Analogously, I've seen people assert that although exponential convergence is provably impossible for a certain problem, they've got a method that "converges exponentially down to any specified accuracy $\varepsilon > 0$". You can depend upon it, the exponential initial transients of such a method lie above a subexponential envelope.

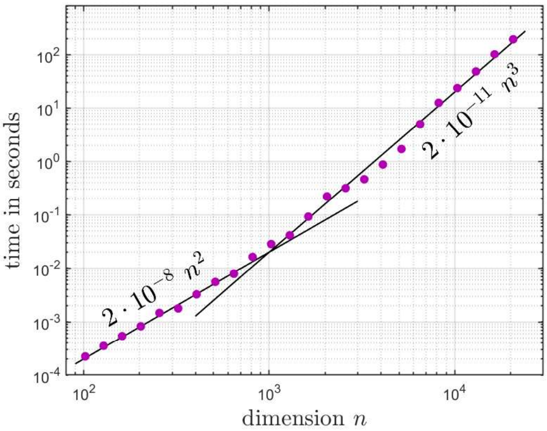

Figure 1. Inverting an $n \times n$ matrix on my laptop.

The disturbingly plausible idea that $O(n^2) + O(n^3)$ might be somehow faster than $O(n^3)$ alone reminds me of a moment in *Through the Looking-Glass.*

"It's a poor sort of memory that only works backwards," the Queen remarked.

"What sort of things do *you* remember best?" Alice ventured to ask.

"Oh, things that happened the week after next," the Queen replied in a careless tone. "For instance, now, . . . there's the King's Messenger. He's in prison now, being punished; and the trial doesn't even begin till next Wednesday: and of course the crime comes last of all."

"Suppose he never commits the crime?" said Alice.

"That would be all the better, wouldn't it?" the Queen said.

Nick Trefethen
After 26 years at Oxford, Trefethen has moved to Harvard University, where he is Professor of Applied Mathematics in Residence.