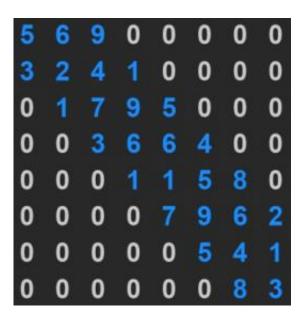# Unraveling the Matrix

July 29 2010, by Larry Hardesty

In a banded matrix, all the nonzero entries cluster around the diagonal. Graphic: Christine Daniloff

A new way of analyzing grids of numbers known as matrices could improve signal-processing applications and data-compression schemes.

Among the most common tools in electrical engineering and computer science are rectangular grids of numbers known as matrices. The numbers in a matrix can represent data: The rows, for instance, could represent temperature, air pressure and humidity, and the columns could represent different locations where those three measurements were taken. But matrices can also represent mathematical equations. If the

expressions t + 2p + 3h and 4t + 5p + 6h described two different mathematical operations involving temperature, pressure and humidity measurements, they could be represented as a matrix with two rows, [1 2 3] and [4 5 6]. Multiplying the two matrices together means performing both mathematical operations on every column of the data matrix and entering the results in a new matrix. In many time-sensitive engineering applications, multiplying matrices can give quick but good approximations of much more complicated calculations.

In a paper published in the July 13 issue of *Proceedings of the National Academy of Science*, MIT math professor Gilbert Strang describes a new way to split certain types of matrices into simpler matrices. The result could have implications for software that processes video or audio data, for compression software that squeezes down digital files so that they take up less space, or even for systems that control mechanical devices.

Strang's analysis applies to so-called banded matrices. Most of the numbers in a banded matrix are zeroes; the only exceptions fall along diagonal bands, at or near the central diagonal of the matrix. This may sound like an esoteric property, but it often has practical implications. Some applications that process video or audio signals, for instance, use banded matrices in which each band represents a different time slice of the signal. By analyzing local properties of the signal, the application could, for instance, sharpen frames of video, or look for redundant information that can be removed to save memory or bandwidth.

Working backwards

Since most of the entries in a banded matrix — maybe 99 percent, Strang says — are zero, multiplying it by another matrix is a very efficient procedure: You can ignore all the zero entries. After a signal has been processed, however, it has to be converted back into its original form. That requires multiplying it by the "inverse" of the processing

matrix: If multiplying matrix A by matrix B yields matrix C, multiplying C by the inverse of B yields A.

But the fact that a matrix is banded doesn't mean that its inverse is. In fact, Strang says, the inverse of a banded matrix is almost always "full," meaning that almost all of its entries are nonzero. In a signal-processing application, all the speed advantages offered by banded matrices would be lost if restoring the signal required multiplying it by a full matrix. So engineers are interested in banded matrices with banded inverses, but which matrices those are is by no means obvious.

In his *PNAS* paper, Strang describes a new technique for breaking a banded matrix up into simpler matrices — matrices with fewer bands. It's easy to tell whether these simpler matrices have banded inverses, and if they do, their combination will, too. Strang's technique thus allows engineers to determine whether some promising new signal-processing techniques will, in fact, be practical.

Faster than Fourier?

One of the most common digital-signal-processing techniques is the [discrete Fourier transform](link) (DFT), which breaks a signal into its component frequencies and can be represented as a matrix. Although the [matrix](link) for the Fourier transform is full, Strang says, "the great fact about the Fourier transform is that it happens to be possible, even though it's full, to multiply fast and to invert it fast. That's part of what makes Fourier wonderful." Nonetheless, for some signal-processing applications, banded matrices could prove more efficient than the Fourier transform. If only parts of the signal are interesting, the bands provide a way to home in on them and ignore the rest. "Fourier transform looks at the whole signal at once," Strang says. "And that's not always great, because often the signal is boring for 99 percent of the time."

Richard Brualdi, the emeritus UWF Beckwith Bascom Professor of Mathematics at the University of Wisconsin-Madison, points out that a mathematical conjecture that Strang presents in the paper has already been proven by three other groups of researchers. "It's a very interesting theorem," says Brualdi. "It's already generated a couple of papers, and it'll probably generate some more." Brualdi points out that large data sets, such as those generated by gene sequencing, medical imaging, or weather monitoring, often yield matrices with regular structures. Bandedness is one type of structure, but there are others, and Brualdi expects other mathematicians to apply techniques like Strang's to other types of structured matrices. "Whether or not those things will work, I really don't know," Brualdi says. "But Gil's already said that he's going to look at a different structure in a future paper."

More information: www.pnas.org/content/107/28/12413.abstract

Provided by Massachusetts Institute of Technology