Advanced Scientific Computing Research
Applied Mathematics
FY 2007 Accomplishment

"Implementing an Interior Point Method for Linear Programs on a CPU-GPU System"

Dianne P. O'Leary, Jin Hyuk Jung
University of Maryland
College Park, Maryland

Summary

Hidden inside your desktop or laptop computer is a very powerful parallel processor, the graphics processing unit (GPU). This hardware is dedicated to rendering images on your screen, and its design was driven by the demands of the gaming industry. These processors are freely-available and potentially powerful computational engines that can be used with the host central processing unit (CPU) to solve numerical problems. We present a CPU-GPU algorithm for solving linear programming problems using interior point methods. Comparisons with a CPU implementation demonstrate that we can improve performance by using the GPU for sufficiently large problems. Since GPU architectures and programming languages are rapidly evolving, we expect that GPUs will be an increasingly attractive tool for matrix computation in the future.

Hidden inside your desktop or laptop computer is a very powerful parallel processor, the graphics processing unit (GPU). This hardware is dedicated to rendering images on your screen, and its design was driven by the demands of the gaming industry. These processors are freely-available and potentially powerful computational engines that can be used with the host central processing unit (CPU) to solve numerical problems. We present a CPU-GPU algorithm for solving linear programming problems using interior point methods. Comparisons with a CPU implementation demonstrate that we can improve performance by using the GPU for sufficiently large problems. Since GPU architectures and programming languages are rapidly evolving, we expect that GPUs will be an increasingly attractive tool for matrix computation in the future.

Originally, GPUs were much slower than CPUs and had very limited programmability. Now they show superior performance on some applications, and their speed is increasing at a rate faster than Moore's law predictions for CPUs.

The key to using GPUs for matrix computations lies in mapping matrix operations to shapes that the GPU can define. We use a rectangular-packed format to store lower triangular matrices, as shown in Figure 1. This makes computations such as the product of a matrix times its transpose easy to perform. Using this storage scheme, we can use the GPU to very quickly form the product of a matrix with its transpose and then compute the lower triangular factor of the resulting matrix. These operations form the core of the interior point method for solving linear programming problems. By comparing our implementations with a CPU implementation, we demonstrated that we can improve performance by using the GPU for sufficiently large problems.

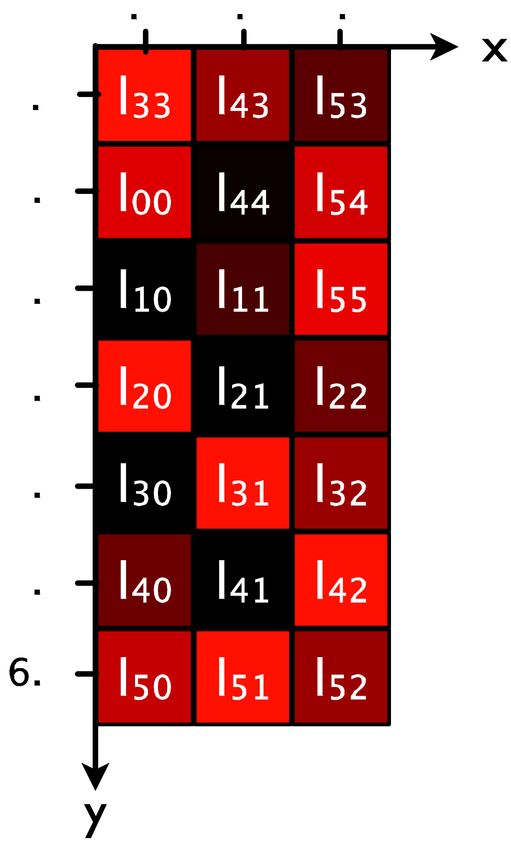

Figure 1. Storage of a lower triangular matrix in a texture using a packed storage scheme of Gunnels and Gustavson, with values stored as intensities of red. The 6×6 matrix is stored in a 3×7 texture, with the entries arranged as indicated.

For further information on this subject contact:

Dr. Dianne P. O'Leary
University of Maryland
oleary@cs.umd.edu
Phone: 301-405-2678
Or
Dr. Anil Deane
Applied Mathematics Research Program
Office of Advanced Scientific Computing
Phone: 301-903-1465
deane@ascr.doe.gov