Can I use Strassen's Method when doing the matrix multiples in the HPL benchmark or for the Top500 run?

The normal matrix multination algorithm requires n3 + O(n2) multiplications and about the same number of additions. Strassen's algorithm reduces the total number of operations to O(n2.82) by recursively multiplying 2n × 2n matrices using seven n × n matrix multiplications. Thus using Strassen's Algorithm will distort the true execution rate. As a result we do not allow Strassen's Algorithm to be used for the TOP500 reporting. As a side note, in the "usual" matrix multiplication, we have an n2 error term. In Strassen's method, the error exponent p for npranges from 2-3.85 and the numerical error can be 10-100 times greater than that for standard multiplication.

There are general algorithms to solve $Ax=b$ which are asymptotically faster than $O(n^3)$ as $n \to \infty$. In fact $O(n^{2.82})$ is possible with Strassen's algorithm..

Article | Open access | Published: 05 October 2022

Discovering faster matrix multiplication algorithms with reinforcement learning

Alhussein Fawzi, Matej Balog, Aja Huang, Thomas Hubert, Bernardino Romera-Paredes, Mohammadamin Barekatain, Alexander Novikov, Francisco J. R. Ruiz, Julian Schrittwieser, Grzegorz Swirszcz, David Silver, Demis Hassabis & Pushmeet Kohli

Nature **610**, 47–53 (2022) | Cite this article

Even faster algorithms discovered with AI guided optimization

Abstract

Improving the efficiency of algorithms for fundamental computations can have a widespread impact, as it can affect the overall speed of a large amount of computations. Matrix multiplication is one such primitive task, occurring in many systems—from neural networks to scientific computing

Warning: Tradeoff between numerical stability and number of operations.

Matrix norms: $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$, $Ax \in \mathbb{R}^m$

$\cancel{m \times n}$

$$\|A\|_p = \max\{ \underbrace{\|Ax\|_p}_{\substack{\text{m-dim} \\ \text{vector norm}}} : \underbrace{\|x\|_p}_{\substack{\text{n-dim} \\ \text{vector norm.}}} = 1 \}$$

Three values of p that are interesting.

$p=2$ Euclindean distance $\|x\|_2 = \sqrt{|x_1|^2 + \cdots + |x_n|^2}$

$p=1$ } simpler to compute $\|x\|_1 = |x_1| + \cdots + |x_n|$

$p=\infty$ } $\|x\|_\infty = \max\{|x_1|, \ldots, |x_n|\}$

also extra credit

So what is $\|A\|_1$?

Let $x \in \mathbb{R}^n$ with $\|x\|_1 = 1$. What is $\|Ax\|_1$?

$Ax \in \mathbb{R}^m$ is a vector

$$\|Ax\|_1 = \sum_{i=1}^{m} |[Ax]_i|$$

for each component of that vector.

$$[Ax]_i = \sum_{j=1}^{n} A_{ij} x_j$$ definition of Matrix-vector mult.

$$\boxed{\|Ax\|_1} = \sum_{i=1}^{m} \left| \sum_{j=1}^{n} A_{ij} x_j \right| \leq \sum_{i=1}^{m} \sum_{j=1}^{n} |A_{ij}| |x_j|$$

$$= \sum_{j=1}^{n} \underbrace{\left(\sum_{i=1}^{m} |A_{ij}| \right)}_{\text{ignore this then what's left is } \|x\|_1 = 1} |x_j| \leq \sum_{j=1}^{n} M |x_j| = M\|x\|_1 = \boxed{M}.$$

where

$$M = \max\left\{ \sum_{i=1}^{m} |A_{ij}| \; : \; j = 1, \dots, n \right\}$$

so far.

$$\|A\|_1 = \max\left\{ \|Ax\|_1 \; : \; \|x\|_1 = 1 \right\} \leq M$$

In fact $\|A\|_1 = M$. To see this

Now let ℓ be the value of j such that $M = \sum_{i=1}^{m} |A_{i\ell}|$

Consider $x = e_\ell \in \mathbb{R}^n$ where

standard basis. $\rightarrow e_\ell = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ ℓ^{th} element is 1 and all others 0.

$\|e_\ell\| = |0| + |0| + \cdots |1| + |0| + \cdots |0| = 1$

ℓ^{th} term

$$Ae_\ell = \ell^{th} \text{ column of } A = \begin{bmatrix} A_{1\ell} \\ A_{2\ell} \\ \vdots \\ A_{m\ell} \end{bmatrix}$$

same

$$\|Ae_\ell\|_1 = \left\| \begin{bmatrix} A_{1\ell} \\ A_{2\ell} \\ \vdots \\ A_{m\ell} \end{bmatrix} \right\| = \sum_{i=1}^{m} |A_{i\ell}| = M.$$

Thus

$$M = \|Ae_\ell\|_1 \leq \underbrace{\max\left\{ \|Ax\|_1 \; : \; \|x\|_1 = 1 \right\}}_{\|A\|_1} \leq M$$

implies

$$\boxed{\|A\|_1 = \max\left\{ \sum_{i=1}^{m} |A_{ij}| \; : \; j = 1, \dots, n \right\}.}$$

Theorem 2.7.6 : Matrix ∞-norm and 1-norm

next time

$$\|\mathbf{A}\|_\infty = \max_{1\le i\le n} \sum_{j=1}^{n} |A_{ij}|, \qquad (2.7.9)$$

Same for square matrix

$$\boxed{\|\mathbf{A}\|_1 = \max_{1\le j\le n} \sum_{i=1}^{n} |A_{ij}|.} \qquad (2.7.10)$$

$\|A\|_2 = ?$

Next section is conditioning (again)

relative error in the output $\le$ κ relative error in the input

condition number

Condition number:

$$\lim_{\tilde{x}\to x} \dfrac{\left|\dfrac{f(\tilde{x}) - f(x)}{f(x)}\right|}{\left|\dfrac{\tilde{x}-x}{x}\right|} = \kappa_f(x)$$

Then approximately

relative error in x. bounded by $\frac{1}{2}\varepsilon_{mach}$.

$$\frac{|f(\tilde{x}) - f(x)|}{|f(x)|} \approx \kappa_f(x) \left|\underbrace{\frac{\tilde{x}-x}{x}}\right| \le \kappa_f(x)\ \frac{1}{2}\varepsilon_{mach}.$$

Same thing for solving $Ax=b$:

b input and $f(b) = A^{-1}b = x$ as output