Numerical Accuracy and Sensitivity of Monte Carlo Matrix Multiplication

Ilse Ipsen

North Carolina State University
Raleigh, NC, USA

Joint work with: John Holodnak
Randomized Matrix Multiplication

Existing Work:

Cohen & Lewis 1997, 1999
Frieze, Kannan & Vempala 1998
Drineas & Kannan 2001
Sarlós 2006
Drineas, Kannan & Mahoney 2006
Belabbas & Wolfe 2008

Applications:

Importance sampling strategy for query matching

Overview:

Relative error due to randomization
Sensitivity to perturbations
Randomized Inner Product
[Drineas, Kannan & Mahoney 2006]

Input: real vectors $a = (a_1 \ldots a_n)^T$, $b = (b_1 \ldots b_n)^T$
probabilities $p_k > 0$, $\sum_{k=1}^n p_k = 1$
number $c$

Output: Approximation $X$ to $a^T b$
from $c$ randomly sampled element pairs $a_k, b_k$

$X = 0$
for $t = 1 : c$ do
    Sample $k_t$ from $\{1, \ldots, n\}$ with probability $p_{k_t}$
    independently and with replacement
    $X = X + \frac{a_{k_t} b_{k_t}}{c \ p_{k_t}}$
end for
Output of Randomized Inner Product

- Random variable $X_t \equiv \frac{a_k b_k}{c p_k}$

- $X_t$ takes on value $\frac{a_k b_k}{c p_k}$ with probability $p_k$

- Expected value ("average")

$$E [X_t] = \sum_{k=1}^{n} p_k \frac{a_k b_k}{c p_k} = \sum_{k=1}^{n} \frac{a_k b_k}{c} = \frac{a^T b}{c}$$

- Output $X = X_1 + \cdots + X_c$

$$E [X] = E [X_1] + \cdots + E [X_c] = \sum_{t=1}^{c} \frac{a^T b}{c} = a^T b$$

- Unbiased estimator: Expected value = exact value
Absolute Error due to Randomization

[Drineas, Kannan & Mahoney 2006]

- **Uniform probabilities:** $p_k = 1/n$, $1 \leq k \leq n$

For every $\delta > 0$ with probability at least $1 - \delta$

$$\left| X - a^T b \right| \leq n \max\{\|a\|_{\infty}, \|b\|_{\infty}\}^2 \sqrt{\frac{8 \ln(2/\delta)}{c}}$$

- **Identical products:** $a_k b_k = \gamma$, $1 \leq k \leq n$

$$X = \underbrace{\frac{n}{c} \gamma + \cdots + \frac{n}{c} \gamma}_{c} = n \gamma = a^T b$$

Randomized algorithm gives **exact result** for any $c$

Bound is too pessimistic
Relative Error due to Randomization

General probabilities: \( p_k > 0, \sum_k p_k = 1 \)

Deviation from identical products:

\[
\text{osc} \left( \frac{ab}{p} \right) \equiv \max_j \frac{a_j b_j}{p_j} - \min_k \frac{a_k b_k}{p_k}
\]

[Lynn & Timlake 1969, Deutsch & Zenger 1971]

For every \( \delta > 0 \) with probability at least \( 1 - \delta \)

\[
\left| \frac{X - a^T b}{a^T b} \right| \leq \frac{\text{osc} \left( \frac{ab}{p} \right)}{|a^T b|} \sqrt{\frac{\ln(2/\delta)}{2c}}
\]

For every \( \delta > 0 \) with probability at least \( 1 - \delta \)

\[
\left| \frac{X - a^T b}{a^T b} \right| \leq \frac{\text{osc} \left( \frac{ab}{p} \right)}{|a^T b|} \sqrt{\frac{\ln(2/\delta)}{2c}}
\]

\text{Condition} \quad \text{Algorithm}
Relative Error vs Bound

$a_k, b_k \ iid \ uniform \ [0, 1], \ n = 10^5, \ uniform \ probabilities$

Relative errors $\frac{|X - a^T b|}{|a^T b|}$ for every $c$

Bound with probability .99
Relative Error vs Bound

$a_k, b_k$ iid uniform $[-0.5, 0.5]$, $n = 10^5$, uniform probabilities

Relative errors $|X - a^T b|/|a^T b|$ for every $c$

Bound with probability .99
Randomized Matrix Multiplication

[Drineas, Kannan & Mahoney 2006]

\[ A = \begin{pmatrix} a_1 & \ldots & a_n \end{pmatrix} \quad B = \begin{pmatrix} b_1^T \\ \vdots \\ b_n^T \end{pmatrix} \]

Sum of outer products \( AB = a_1 b_1^T + \cdots + a_n b_n^T \)

- Random variable \( X_t = \frac{a_{kt} b_{kt}^T}{cp_{kt}} \)
- \( X_t \) takes on value \( \frac{a_k b_k^T}{cp_k} \) with probability \( p_k > 0 \)
- Output \( X = X_1 + \cdots + X_c \)
Normwise Error due to Randomization

$A$ is $m \times n$, $B$ is $n \times q$

\[
O_{ij} \equiv \max_k \frac{a_{ik} b_{kj}}{p_k} - \min_k \frac{a_{ik} b_{kj}}{p_k}
\]

for $1 \leq i \leq m$ and $1 \leq j \leq q$

For every $\delta > 0$ with probability at least $1 - \delta$

\[
\frac{\|X - AB\|}{\|AB\|} \leq \frac{\|O\|}{\|AB\|} \sqrt{\frac{\ln(2mq/\delta)}{2c}}
\]

in the $1$, $\infty$ and $F$ norms

Bound depends on dimensions of $A$ and $B$
Relative Error vs Bound

Elements of $A, B$ iid uniform $[0, 1]$, $m = 50$, $n = 1000$, $q = 80$ uniform probabilities

Relative errors $\|X - AB\|_1/\|AB\|_1$ for every $c$

Bound with probability .99
Multiplying Rank One Matrices

\[ A = fa^T \quad B = bd^T \quad AB = \begin{pmatrix} a^T b \end{pmatrix} fd^T \]

inner product

Random variable \( X_t = \frac{a_{kt} b_{kt}}{c_{p_{kt}}} fd^T \)

For every \( \delta > 0 \) with probability at least \( 1 - \delta \)

\[
\frac{\| X - AB \|}{\| AB \|} \leq \frac{\text{osc} \left( \frac{ab}{p} \right)}{|a^T b|} \sqrt{\frac{\ln(2 mq / \delta)}{2c}}
\]

Condition

Algorithm

in the 1, \( \infty \) and \( F \) norms

Same condition number as inner product
but bound must hold for all \( mq \) elements of \( X \)
Error due to Randomization

Our bounds with probability .99

- Capture worst case error
- Informative even for small matrix dimensions
- Tight for inner products where all products are identical
- Recognize rank one matrices

How to pick good probabilities $p_k$:

- Minimize variance (importance sampling)
  
  [Drineas, Kannan & Mahoney 2006]
- Minimize $\|O\|$
Sensitivity of Randomized Inner Product

- **Exact inputs:** $a, b$  
  Desired result: $a^T b$

- **Randomized algorithm**
  - Fix $c$, fix probabilities $p_k$
  - Output from some run: 
    $$X = \sum_{t=1}^{c} \frac{a_{kt}b_{kt}}{cp_{kt}}$$

- **Perturbed inputs:** $\hat{a}, \hat{b}$
  - Same $c$, same probabilities $p_k$
  - Output from some run: 
    $$\hat{X} = \sum_{t=1}^{c} \frac{\hat{a}_{it}\hat{b}_{it}}{cp_{it}}$$

- **What to compare?**
  - $\hat{X}$ and $a^T b$: No info about sensitivity of algorithm
  - $\hat{X}$ from some run, and $X$ from another run: Too pessimistic
  - $\hat{X}$ and $X$ from same run
Sensitivity Bound: Numerator

- Relative perturbations
  \[ \hat{a}_k = a_k (1 + \alpha_k) \quad \hat{b}_k = b_k (1 + \beta_k) \quad |\alpha_k|, |\beta_k| \leq \epsilon \]

- Outputs from same run
  \[ X = \sum_{t=1}^{c} \frac{a_{kt} b_{kt}}{cp_{kt}} \quad \hat{X} = \sum_{t=1}^{c} \frac{\hat{a}_{kt} \hat{b}_{kt}}{cp_{kt}} \]

- For every \( \delta > 0 \) with probability at least \( 1 - \delta \)
  \[ |\hat{X} - X| \leq 3 \left[ |a|^T |b| + \text{osc} \left( \frac{|ab|}{p} \right) \sqrt{\frac{\ln(2/\delta)}{2c}} \right] \epsilon \]
Sensitivity Bound

\[
|\hat{X} - X| \leq 3 \left[ |a|^T |b| + \text{osc} \left( \frac{|ab|}{p} \right) \sqrt{\frac{\ln(2/\delta)}{2c}} \right] \frac{|X|}{|\epsilon|}
\]

Difficulties:
- Denominator $|X|$ unknown, can take on $O(n^c)$ different values
- Bound $|X|$ in terms of $|a^T b|$? Too pessimistic.
- Bound $|X|$ in terms of $\min_{k_1, \ldots, k_c} \left| \sum_{t=1}^c \frac{a_{k_t} b_{k_t}}{c p_{k_t}} \right|$? Too pessimistic. Too unwieldy.
Low Sensitivity

\( a_k, b_k \) iid uniform \([0, 1]\), \( n = 10^5 \), \( \epsilon = 10^{-8} \), \( c = 10^3 \)

uniform probabilities

Relative errors \( |\hat{X} - X|/|X| \) over 1000 runs

Sensitivity bound with probability .99

Bound almost constant \( \Rightarrow \) low sensitivity to perturbations
High Sensitivity

$a_k, b_k$ iid uniform $[-.5, .5]$, $n = 10^5$, $\epsilon = 10^{-8}$, $c = 10^3$

uniform probabilities

Relative errors $|\hat{X} - X|/|X|$ over 1000 runs

Sensitivity bound with probability .99

Bound oscillates $\Rightarrow$ high sensitivity to perturbations
Interpretation of Sensitivity Bound

Assumptions:

\( \hat{X} \) and \( X \) from same run

Same \( c \), same probabilities \( p \)

Relative perturbations \( \leq \epsilon \)

With probability at least \( 1 - \delta \)

\[
\frac{|\hat{X} - X|}{|X|} \leq \frac{\text{constant}(a, b, p, c, \delta)}{|X|} \epsilon
\]

Two factors influence sensitivity:

\[
\text{constant}(a, b, p, c, \delta) = O \left( |a|^T |b| + \text{osc} \left( \frac{|ab|}{p} \right) \right)
\]

“Variance” of \(|X|\) \hspace{1cm} \(|X|\) has \( O(n^c) \) different values
Randomized algorithm for matrix multiplication from [Drineas, Kannan & Mahoney 2006]

Relative error due to randomization
- Tighter bounds, apply to all probabilities
- Predictive even for small matrix dimensions

Sensitivity of randomized inner product
- Number of different outputs is exponential: \( \mathcal{O}(n^c) \)
- Capture variation across all of these outputs