Fast Computation of Low Rank Matrix Approximations

by Dimitris Achlioptas and Frank McSherry
Presentation: Boris Temkin

Motivation

Idea of Algorithm

Definitions

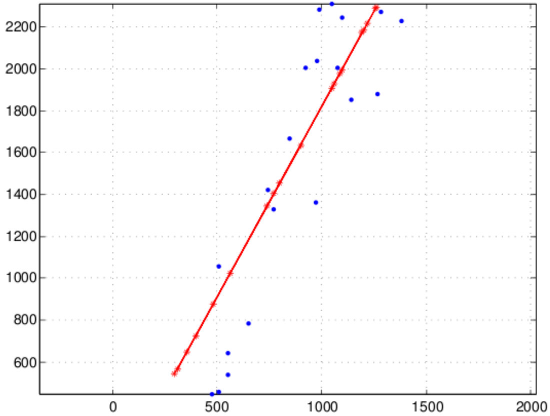
Quantization

Sparsification

Adaptive Non-Uniform Sampling

Practical considerations

Example



Truncated SVD

- $A = U\Sigma V^T$
- $A_k = U\Sigma_k V^T$

Optimal Low Rank Matrix Approximation

Given $m \times n$ real valued matrix A and integer k, find a $m \times n$ matrix A_k which minimizes $\|A - A_k\|$ over all matrices of rank k.

Optimal Low Rank Approximation Algorithms

- Truncated SVD
- Orthogonal Iteration
- Lanczos Iteration

Motivation

Used in areas

- Computer Vision
- Information Retrieval
- Machine Learning

Used for

- Correlation Extraction
- Noise Elimination

Near Optimal Approximation

Given $m \times n$ matrix A, integer k and $\delta > 0$, find $m \times n$ matrix B such that

- B_k is easy to compute
- $\|A - B_k\| \leqslant \|A - A_k\| + \delta$

Approaches

- Sparsification
- Quantization

Definitions

For any matrix M and integer k,

Norms

- Frobenius Norm

$$\|M\|_F = \sqrt{\sum_{i,j} M_{ij}^2}$$

- Second Norm

$$\|M\|_2 = \max_{\|x\|_2=1} \|Mx\|_2$$

Fact

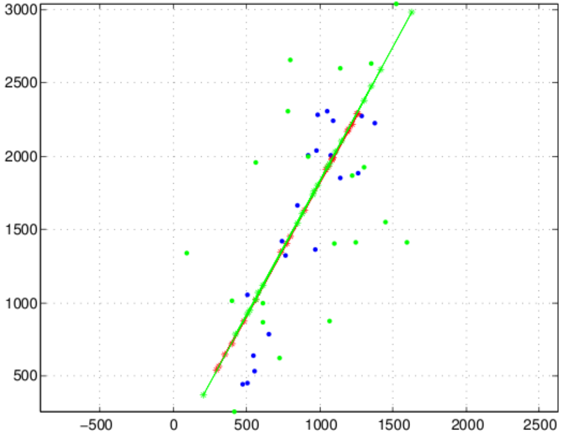
- $\|M_k\|_F \leqslant \sqrt{k}\|M\|_2$
- $\|M_k\|_2 = \|M\|_2$

Mental experement

Gaussian matrix

Let G be a matrix whose entries are independent Gaussian random variables with mean 0 and variance σ^2. If σ is not too big, then $\|A - (A+G)_k\| \approx \|A - A_k\|$

Example



Should it be Gaussian?

It is enough that G is a random matrix such that

- Entries are independent
- Mean of G_{ij} is zero
- Variance of G_{ij} is small

Example

Set $G_{ij} = \pm A_{ij}$ with equal probability, independently for all i, j.

Lemma

Let A and N be any matrices and write $B = A + N$. Then

- $\|A - B_k\|_2 \leqslant \|A - A_k\|_2 + 2\|N_k\|_2$
- $\|A - B_k\|_F \leqslant \|A - A_k\|_F + \|N_k\|_F + 2\sqrt{\|N_k\|_F \|A_k\|_F}$

Proof.

$$\begin{aligned}
\|A - B_k\|_2 &\leqslant \|A - B\|_2 + \|B - B_k\|_2 \\
&\leqslant \|A - B\|_2 + \|B - A_k\|_2 \\
&\leqslant \|A - B\|_2 + \|B - A\|_2 + \|A - A_k\|_2 \\
&= \|A - A_k\| + 2\|A - B\|_2 \\
&= \|A - A_k\| + 2\|\underbrace{(A - B)_k}_{N_k}\|_2
\end{aligned}$$

□

Lemma

Let A and N be any matrices and write $B = A + N$. Then

- $\|A - B_k\|_2 \leqslant \|A - A_k\|_2 + \underbrace{2\|N_k\|_2}_{\delta}$
- $\|A - B_k\|_F \leqslant \|A - A_k\|_F + \underbrace{\|N_k\|_F + 2\sqrt{\|N_k\|_F\|A_k\|_F}}_{\delta}$

Theorem (w/o proof)

Given $m \times n$ matrix A such that $m \leqslant n$ and $(m+n) \geqslant 152$, fixed $\epsilon > 0$ and $\Theta > 0$. Let

$$K = \left(\frac{\log(1+\epsilon)}{2\log(m+n)}\right)^2 \times \sigma\sqrt{m+n}$$

Let B be a random matrix whose entries are independent random variables such that for all i,j

- $\mathbb{E}(B_{ij}) = A_{ij}$
- $\mathbb{V}ar(B_{ij}) \leqslant \sigma^2$
- *B_{ij} takes values on interval of length K*

Then

$$\mathbb{P}r\left[\|A-B\|_2 \geqslant 2(1+\epsilon+\Theta)\sigma\sqrt{m+n}\right] < 2\exp\left(-\frac{16\Theta^2}{\epsilon^4}(\log n)^4\right)$$

Quantization

Theorem (quantization)

Let A be any $m \times n$ matrix where $m \leqslant n$ and $b = \max_{ij} |A_{ij}|$.
Let B be a random $m \times n$ matrix whose entries are independently distributed as

$$B_{ij} = \begin{cases} +b & \text{with probability } \frac{1}{2} + \frac{A_{ij}}{2b} \\ -b & \text{with probability } \frac{1}{2} - \frac{A_{ij}}{2b} \end{cases}$$

Then, for large enough n,
with probability at least $1 - \exp(-19(\log n)^4)$

- $\|(A - B)_k\|_2 < 4b\sqrt{n}$
- $\|(A - B)_k\|_F < 4b\sqrt{kn}$

Proof.

Apply theorem with $\epsilon = 3/10$ and $\Theta = 1/10$.

- $\mathbb{E}[B_{ij}] = A_{ij}$, $\sigma = b$
- For $(m+n) = 3.08 + E9$, $K = 2.008b \Rightarrow K \geqslant 2b$.

□

Proof.

Apply theorem with $\epsilon = 3/10$ and $\Theta = 1/10$.

- $\mathbb{E}[B_{ij}] = A_{ij}$, $\sigma = b$
- For $(m+n) = 3.08 + E9$, $K = 2.008b \Rightarrow K \geqslant 2b$.

$$\mathbb{P}r\left[\|A - B\|_2 \geqslant \underbrace{2(1+\epsilon+\Theta)}_{\approx 3.9598/\sqrt{2}} \underbrace{\sigma}_{b} \underbrace{\sqrt{m+n}}_{\leqslant\sqrt{2n}}\right] < 2\exp\left(\frac{16\Theta^2}{\epsilon^4}(\log n)^4\right)$$

□

Proof.

Apply theorem with $\epsilon = 3/10$ and $\Theta = 1/10$.

- $\mathbb{E}[B_{ij}] = A_{ij}$, $\sigma = b$
- For $(m+n) = 3.08 + E9$, $K = 2.008b \Rightarrow K \geqslant 2b$.

$$\mathbb{P}r\left[\|A-B\|_2 \geqslant \underbrace{2(1+\epsilon+\Theta)\sigma\sqrt{m+n}}_{\leqslant 4b\sqrt{n}}\right] < \underbrace{2\exp\left(\frac{16\Theta^2}{\epsilon^4}(\log n)^4\right)}_{\leqslant \exp(-19(\log n)^4)}$$

□

Sparsification
(uniform version)

Theorem (uniform sampling)

Let A be any $m \times n$ matrix where $76 \leqslant m \leqslant n$ and $b = \max_{ij} |A_{ij}|$. For $p \geqslant (8 \log n)^4 / n$, let B be a random $m \times n$ matrix whose entries are independently distributed as

$$B_{ij} = \begin{cases} A_{ij}/p & \textit{with probability } p \\ 0 & \textit{Otherwise} \end{cases}$$

Then with probability at least $1 - exp(-19(\log n)^4)$

- $\|(A - B)_k\|_2 < 4b\sqrt{n/p}$
- $\|(A - B)_k\|_F < 4b\sqrt{kn/p}$

Proof.

Apply theorem with $\epsilon = 3/10$ and $\Theta = 1/10$.

- $\mathbb{E}[B_{ij}] = A_{ij}$
- $\mathbb{V}ar(B_{ij}) = \frac{(1-p)}{p} A_{ij}^2 \leqslant \frac{b^2}{p} = \sigma^2$

$$\mathbb{P}r \left[\|A - B\|_2 \geqslant \underbrace{2(1 + \epsilon + \Theta)\sigma\sqrt{m+n}}_{\leqslant 4b\sqrt{np}} \right] < \underbrace{2\exp\left(\frac{16\Theta^2}{\epsilon^4}(\log n)^4\right)}_{\leqslant \exp(-19(\log n)^4)}$$

□

Proof.

Apply theorem with $\epsilon = 3/10$ and $\Theta = 1/10$.

- $\mathbb{E}[B_{ij}] = A_{ij}$
- $\mathbb{V}ar(B_{ij}) = \frac{(1-p)}{p} A_{ij}^2 \leqslant \frac{b^2}{p} = \sigma^2$
- $$K = \left(\frac{\log(1+\epsilon)}{2\log(m+n)}\right)^2 \times \frac{b}{\sqrt{p}}\sqrt{m+n} \geqslant \frac{2b}{p} \iff$$ $$p \geqslant \left(\frac{2\sqrt{2}\log(m+n)}{\log(1+\epsilon)}\right)^4 \frac{1}{m+n}$$

$$\Bigg(\underbrace{\frac{2\sqrt{2}}{\log(1+3/10)}}_{\approx 7.4725}\Bigg)^4 \frac{(\log(m+n))^4}{m+n} \leqslant \frac{(8\log n)^4}{n} \leqslant p$$

Remark: $\log^4(a)/a$ decreasing for $a > 55$.

Sparsification
(non-uniform version)

May it be better?

Yes, by non-uniform sampling with probability $p_{ij} \leqslant p$ such that

- $\mathbb{E}[B_{ij}] = A_{ij}$
- $\mathbb{V}ar(B_{ij}) \leqslant \sigma^2 = \frac{b^2}{p}$
- B_{ij} taken from interval of length K.

Sparsification

(non-uniform version)

May it be better?

Yes, by non-uniform sampling with probability $p_{ij} \leqslant p$ such that

- $\mathbb{E}[B_{ij}] = A_{ij}$
- $\mathbb{V}ar(B_{ij}) \leqslant \sigma^2 = \frac{b^2}{p}$
- B_{ij} taken from interval of length K.

How to?

- Set $B_{ij} = A_{ij}/p_{ij}$ with probability $p_{ij} = p \times (A_{ij}/b)^2$
- $\mathbb{V}ar(B_{ij}) = \frac{1-p_{ij}}{p_{ij}} A_{ij}^2 = \frac{b^2}{p} - A_{ij}^2 \leqslant \sigma^2$
- Exected number of non-zero entries: $\sum_{ij} p_{ij} = p\|A\|_F^2/b^2$

Sparsification
(non-uniform version)

May it be better?

Yes, by non-uniform sampling with probability $p_{ij} \leqslant p$ such that

- $\mathbb{E}[B_{ij}] = A_{ij}$
- $\mathbb{V}ar(B_{ij}) \leqslant \sigma^2 = \frac{b^2}{p}$
- B_{ij} taken from interval of length K.

How to?

- Set $B_{ij} = A_{ij}/p_{ij}$ with probability $p_{ij} = p \times (A_{ij}/b)^2$
- $\mathbb{V}ar(B_{ij}) = \frac{1-p_{ij}}{p_{ij}} A_{ij}^2 = \frac{b^2}{p} - A_{ij}^2 \leqslant \sigma^2$
- Exected number of non-zero entries: $\sum_{ij} p_{ij} = p\|A\|_F^2/b^2$

Problem

$B_{ij} = A_{ij}/p_{ij} = \frac{b^2}{pA_{ij}}$ may violate range constraint! (i.e. $2b/p$)

Sparsification

(non-uniform version)

Theorem (non-uniform sampling)

Let A be any $m \times n$ matrix where $76 \leqslant m \leqslant n$ and $b = \max_{ij} |A_{ij}|$. For any $p > 0$, define $\tau_{ij} = p(A_{ij}/b)^2$ and let

$$p_{ij} = \max\{\tau_{ij}, \sqrt{\tau_{ij} \times (8 \log n)^4/n}\}$$

Let B be a random $m \times n$ matrix whose entries are independently distributed as

$$B_{ij} = \begin{cases} A_{ij}/p_{ij} & \textit{with probability } p_{ij} \\ 0 & \textit{Otherwise} \end{cases}$$

Then with probability at least $1 - exp(-19(\log n)^4)$

- $\|(A - B)_k\|_2 < 4b\sqrt{n/p}$
- $\|(A - B)_k\|_F < 4b\sqrt{kn/p}$

Expected number for non-zero entries

- Uniform Version: pmn
- Non-Uniform Version: $pmn \times \text{Avg}(A_{ij}/b)^2 + mn \times (8 \log n)^4/n$ (As $p_{ij} \leqslant \tau_{ij} + (8 \log(n))^4/n$)

Adaptive Non-Uniform Sampling in a Single Pass

Sample(s, n)

Let Q be empty priority queue and let $Z = 0$
for all entry A_{ij} **do**
    $Z \leftarrow Z + A_{ij}^2$
    Select $r_{ij} \in_R [0, 1]$
    $k_{ij} \leftarrow \max\{sA_{ij}^2/r_{ij}, sA_{ij}^2/r_{ij}^2 \times (8 \log n)^4/n\}$
    Insert A_{ij} in Q with key k_{ij}
    Remove from Q all elements with key smaller than Z
end for
return Q

Adaptive Non-Uniform Sampling in a Single Pass

Lemma

Let A be $m \times n$ matrix where $76 \leqslant m \leqslant n$. For every $s > 0$, Sample(s, n) yields a matrix B such that

- *With probability at least $1 - \exp(-19 \log(n)^4)$, the matrix $N = A - B$ satisfies*

  $$\|N_k\|_2 \leqslant 4\sqrt{n/s} \times \|A\|_F \qquad \text{and} \qquad \|N_k\|_F \leqslant 4\sqrt{kn/s} \times \|A\|_F$$

- *The expected number of non-zero entries in B is bounded by $s + m(8 \log n)^4$*

Proof.

- Set $p = sb^2/\|A\|_F^2$ then $\tau_{ij} = s(A_{ij}/\|A\|_F)^2$
- $A_{i,j}$ is in Q if and only if

$$sA_{ij}^2/r_{ij} \geqslant \|A\|_F^2 \qquad \text{or} \qquad sA_{ij}^2/r_{ij}^2 \times (8\log(n))^4/n \geqslant \|A\|_F^2$$

- Which equivalent to $r_{ij} \leqslant p_{ij}$, but r_{ij} choosen i.i.d.

□

Practical considerations

- Combining Sampling and Quantization
- Computing Optimal Low Rank Approximations