An Improved Approximation Algorithm for the Column Subset Selection Problem

Authors: Christos Boutsidis(RPI)
Michael W. Mahoney(Stanford)
Petros Drineas(RPI)
Presenter: Chia-Chun Tsai
Department of Industrial and Systems Engineering
April, 29, 2010

Outline

- The Column Subset Selection Problem(CSSP)
  - Definition, complexity
  - Approximability framework
- Paper’s contribution on the CSSP
  - Novel randomized algorithm
  - Improved approximability results

The Column Subset Selection Problem

$$\underbrace{\begin{bmatrix} \\ \mathbf{A} \\ \\ \end{bmatrix}}_{\mathbf{m \times n}}, \mathbf{k} \rightarrow \mathbf{CSSP\ solver} \rightarrow \underbrace{\begin{bmatrix} \\ \mathbf{C} \\ \\ \end{bmatrix}}_{\mathbf{m \times k}}$$

$\mathbf{Input: matrix\ A \in R^{m \times n}, integer\ k < n}$

$\mathbf{Output: matrix\ C \in R^{m \times k}\ with\ k\ columns\ of\ A}$

Combinatorial Optimization Problem:

$$\min_C \left\| A - CC^+ A \right\|_\xi$$

$\xi = 2, F$ **stands for the spectral or the Frobenius norm**

Complexity of the CSSP

- NP-hardness of the CSSP is an open question.
- There are $\binom{n}{k}$ for the matrix C
- Optimal solution can be found in $\mathbf{O(n^k mn)}$ time.

Notation and Linear Algebra

SVD of a matrix $\mathbf{A} \in \mathbf{R}^{\mathbf{m} \times \mathbf{n}}$ **with rank(A)** $= \rho$

$$A = U_A \Sigma_A V_A^T = \begin{pmatrix} U_k & U_{\rho-k} \end{pmatrix} \begin{pmatrix} \Sigma_k & 0 \\ 0 & \Sigma_{\rho-k} \end{pmatrix} \begin{pmatrix} V_k^T \\ V_{\rho-k}^T \end{pmatrix}$$

Why SVD?

$$A_k = U_k \Sigma_k V_k^T = \arg\{ \min_{Y \in R^{m \times n}, rank(Y) \le k} \| A - Y \|_\xi \}$$

Pseudoinverse of C is $\mathbf{C}^+ = V_c \Sigma_c^+ U_c^T$

Why $\mathbf{C}^+\mathbf{A}$**?**

$$C^+ A = \arg\{ \min_{X \in R^{k \times n}} \| A - CX \|_\xi \}$$

Approximation Algorithm for the CSSP

SVD provides a lower bound for the CSSP:

$$\|A - A_k\|_\xi \le \|A - C_{opt} C_{opt}^{+} A\|_\xi$$

We seek algorithms with upper bounds of the from:

$$\|A - CC^{+} A\|_\xi \le p(k,n) \|A - A_k\|_\xi$$

p(k,n) is typically a polynomial function

Overview of best-existing and paper results

Spectral norm:

Gu & Eisenstat, 1996 Deterministic, $O(mn^2)$ time

$$\left\|A - CC^{+}A\right\|_2 \le O(k^{\frac{1}{2}}(n-k)^{\frac{1}{2}})\left\|A - A_k\right\|_2$$

Paper's Algorithm, 2009 Randomized, $O(mn^2)$ time

$$\left\|A - CC^{+}A\right\|_2 \le O(k^{\frac{3}{4}}(n-k)^{\frac{1}{4}})\left\|A - A_k\right\|_2$$

Frobenius norm:

Deshpande & Vempala, 2006 Randomized, $O(mnk)$ time

$$\left\|A - CC^{+}A\right\|_F \le \sqrt{(K+1)!}\left\|A - A_k\right\|_F$$

Paper's Algorithm, 2009 Randomized, $O(mn^2)$ time

$$\left\|A - CC^{+}A\right\|_F \le O(K\sqrt{\log K})\left\|A - A_k\right\|_F$$

Paper's novel randomized algorithm

$$\mathbf{A, K} \rightarrow \begin{matrix}\textbf{Randomized}\\ \textbf{Stage}\end{matrix} \rightarrow \hat{\mathbf{C}} \rightarrow \begin{matrix}\textbf{Deterministic}\\ \textbf{Stage}\end{matrix} \rightarrow \mathbf{C}$$

2 **stage column selection**

$$\textbf{Randomized Stage:}\ A \rightarrow \hat{C} \in R^{m\times c}, C = O(k \log k)$$

$$\textbf{Deterministic Stage:}\ \hat{C} \rightarrow C \in R^{m\times k}$$

Randomized Stage

$$\mathbf{A,k} \rightarrow \begin{array}{c}\textbf{Ramdomized}\\ \textbf{Stage}\end{array} \rightarrow \hat{\mathbf{C}}$$

RandomizedSampling:

1. You give a score to every column of A:

$$\mathbf{p_i} = \frac{\frac{1}{2}\left\|(V_k)_{(i)}\right\|_2^2}{\sum_{j=1}^{n}\left\|(V_k)_{(j)}\right\|_2^2} + \frac{\frac{1}{2}\left\|(\sum_{\rho-k} V_{\rho-k}^T)^{(i)}\right\|_2^2}{\sum_{j=1}^{n}\left\|(\sum_{\rho-k} V_{\rho-k}^T)^{(j)}\right\|_2^2}$$

2. You select the i-th column with probability $\min\{1,cp_i\}$

Strong aprroximation guarantee($\xi = \mathrm{F}$):

$$\left\|A - \hat{C}\hat{C}^{+}A\right\|_F \leq (1+\varepsilon)\left\|A - A_k\right\|_F$$

[Reference: Drineas, Mahoney, & Muthukrishnan, SIMAX 2008]

Deterministic Stage

Some more notation:

$C = O(k \log k)$

$S_1 \in R^{m \times c}$ **is a sampling matrix s.t** $\hat{C} = AS_1$

$D_1 \in R^{c \times c}$ **is a diagonal matrix s.t** $D_{ii} = 1/\sqrt{\min\{1, cp_i\}}$

RRQR factorization on $V_k^T S_1 D_1 \in R^{k \times c}$

$(V_k^T S_1 D_1)\prod = QR$

Extract k column after this RRQR factorization.

[Reference: Pan, LAA 2000, Gu & Eisentstat, SISC 1996]

Intuition for 2-stgae algorithm

- The key idea:
  - Select k linearly independent columns from A
- Randomized Stage: The sampling scores are biased toward outlier columns
  - Leverage scores in the diagnostic regression analysis domain
  - Resistance scores in the graph sparsification domain [Srivastava & Spielman, STOC 2008]
- Deterministic Stage: RRQR factorizations were explicitly designed to identify sets of linearly independent columns

Proof step by step

$S \in R^{n\times k}$ **denotes a sampling operator such that**

$$C = AS$$

The proof :

$$\left\|A\text{-}CC^{+}A\right\|_2 = \left\|A - AS(AS)^{+}A\right\|_2$$

$$\leq \left\|A - AS(A_k S)^{+}A_k\right\|_2$$

$$= \left\|(A_k + A_{\rho-k}) - (A_k + A_{\rho-k})S(A_k S)^{+}A_k\right\|_2$$

$$\leq \left\|A_k - (A_k S)(A_k S)^{+}A_k\right\|_2 + \left\|A_{\rho\text{-}k}\right\|_2 + \left\|A_{\rho-k}S(A_k S)^{+}A_k\right\|_2$$

$$\|\mathbf{A}-\mathbf{C}\mathbf{C}^{+}\mathbf{A}\|_2 \le \underbrace{\|\mathbf{A_k}-(\mathbf{A_k}\mathbf{S})(\mathbf{A_k}\mathbf{S})^{+}\mathbf{A_k}\|_2}_{1} + \underbrace{\|\mathbf{A}_{\rho-\mathbf{k}}\|_2}_{2} + \underbrace{\|\mathbf{A}_{\rho-\mathbf{k}}\mathbf{S}(\mathbf{A_k}\mathbf{S})^{+}\mathbf{A_k}\|_2}_{3}$$

1. $\|A_k-(A_kS)(A_kS)^{+}A_k\|_2 = \|A_k - I\,A_k\|_2 = 0$

(it is non-trivial to prove that $(\mathbf{A_k}S)(A_kS)^{+} = I$)

2. $\|A_{\rho\text{-}k}\|_2 = \|A - A_k\|_2$

(This trivial comes from the SVD of A)

3. $\|A_{\rho\text{-}k}S(A_kS)^{+}A_k\|_2 \le \|A_{\rho-k}S\|_2 \|(A_kS)^{+}A_k\|_2$

(Standard property for matrix norms)

$\|A_{\rho\text{-}k}S\|_2 \le O((n-k)^{\frac{1}{4}})\|A-A_k\|$

(proof based on [Rudelson & Vershynin, JACM 2007])

$\|(A_kS)^{+}A_k\|_2 \le O(\sqrt{k(k\log k-1)})$

(proof based on [Pan, LAA 2000])

Past work : Algorithm to bound $\|A - CC^{+} A\|_2$

Deterministic algorithms

Develope by the numerical linear algebra people

(G. Golub, G.W Stewart,..)

Best Result: The algorithm in [Gu & Eisenstat, SISC1996]

runs in $O(mn^2)$ time and gaurantees that

$$\|A - CC^{+} A\|_2 \leq O(\sqrt{1 + k(n-k)})\|A - A_k\|_2$$

Deep connections with the so-called Rank Revealing QR

(RRQR) factorization.

Past work : Algorithm to bound $\left\| A - CC^{+} A \right\|_F$

Randomized algorithms

Develope by the theoritical computer science people

(A. Frieze, R. Kannan, S. Vempala,..)

Best Result: The algorithm in [Deshpande & Vempala, SODA 2006] runs in O(mnk + kn) time and gaurantees that with high probability

$$\left\| A\text{-}CC^{+} A \right\|_F \leq \sqrt{(k+1)!} \left\| A - A_k \right\|_F$$

Algorithm idea : Assign a probability to each column of A, and then sample the columns of A based on these probability.

Conclusions: $\|A\text{-}CC^{+}A\|_{\xi} \leq p(k,n)\|A - A_k\|_{\xi}$

- Best Algorithm so far

	Reference	p(k,n)	Time
$\xi = 2$	[Gu,Eisenstat,96]	$O(k^{\frac{1}{2}}(n-k)^{\frac{1}{2}})$	$O(mn^2)$
$\xi = F$	[Desphande,Vempala,06]	$\sqrt{(k+1)!}$	$O(mnk)$

- Paper's Algorithm results

	Reference	p(k,n)	Time
$\xi = 2$	[This paper,09]	$O(k^{\frac{3}{4}}(n-k)^{\frac{1}{4}})$	$O(mn^2)$
$\xi = F$	[This paper,09]	$O(k\sqrt{\log k})$	$O(mn^2)$

Thanks