ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ
ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΥΧΑΙΟΠΟΙΗΜΕΝΕΣ ΜΕΘΟΔΟΙ ΠΡΟΣΕΓΓΙΣΗΣ ΠΟΛΛΑΠΛΑΣΙΑΣΜΟΥ ΠΙΝΑΚΩΝ ΚΑΙ ΠΑΡΑΓΟΝΤΟΠΟΊΗΣΗΣ ΙΔΙΑΖΟΥΣΩΝ ΤΙΜΏΝ

Διπλωματική Εργασία

Θεοχάρη Κατερίνα

Αριθμός Μητρώου: 09109102

Επιβλέπων Καθηγητής: Ψαρράκος Παναγιώτης

Αθήνα, 2020

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ
ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΥΧΑΙΟΠΟΙΗΜΕΝΕΣ ΜΕΘΟΔΟΙ ΠΡΟΣΕΓΓΙΣΗΣ ΠΟΛΛΑΠΛΑΣΙΑΣΜΟΥ ΠΙΝΑΚΩΝ ΚΑΙ ΠΑΡΑΓΟΝΤΟΠΟΊΗΣΗΣ ΙΔΙΑΖΟΥΣΩΝ ΤΙΜΏΝ

Διπλωματική Εργασία

Θεοχάρη Κατερίνα

Αριθμός Μητρώου: 09109102

Τριμελής Επιτροπή: Π. Στεφανέας, Επίκ. Καθηγητής Ε.Μ.Π.

Α. Χαραλαμπόπουλος, Καθηγητής Ε.Μ.Π.

Π. Ψαρράκος, Καθηγητής Ε.Μ.Π. (Επιβλέπων)

Αθήνα, 2020

ΠΕΡΙΕΧΟΜΕΝΑ

Ευχαριστίες

Με την παρούσα παράγραφο οφείλω να ευχαριστήσω όλους όσους συνέβαλαν στην εκπόνησή της διπλωματικής μου εργασίας.

Ιδιαίτερα ευχαριστώ τον επιβλέποντα καθηγητή μου, κ Ψαρράκο Παναγιώτη, για την πολύτιμη υποστήριξή του, τις παραγωγικές υποδείξεις του και το πολύ καλό κλίμα συνεργασίας που διαμόρφωσε συμβάλλοντας τα μέγιστα για την κατάρτιση της διπλωματικής μου εργασίας.

Ευχαριστώ επίσης ιδιαίτερα την οικογένεια μου για τη στήριξη , τη συμπαράσταση και την κατανόηση που έδειξαν καθ' όλη τη διάρκεια των σπουδών μου.

Ευχαριστώ θερμά τους συντρόφους μου που είναι πάντα εκεί, μου δίνουν δύναμη και με στηρίζουν σε κάθε μου προσπάθεια και σε κάθε βήμα της ζωής μου. Χωρίς αυτούς όλη αυτή η πορεία θα ήταν για εμένα σίγουρα πολύ πιο δύσκολη.

Τέλος θα ήθελα να ευχαριστήσω τον κύριο Α. Χαραλαμπόπουλο και τον κύριο Π. Στεφανέα που είχαν την ευγενή καλοσύνη να διαθέσουν τον χρόνο τους για να διαβάσουν το κείμενο της διπλωματικής μου και να συμμετάσχουν στην τριμελή επιτροπή αξιολόγησής της.

Συμβολισμοί

R , R^{mxn}	Τα σύνολα των πραγματικών αριθμών και mxn πινάκων αντίστοιχα
$\|\cdot\|_F$, $\|\cdot\|_2$	Οι νόρμα Frobenius και η Ευκλείδεια νόρμα αντίστοιχα
A^T	Ο ανάστροφος ενός πίνακα A
trace(A)	Το ίχνος ενός τετραγωνικού πίνακα A
rank(A)	Ο βαθμός ενός πίνακα A
Range(A) ή Image(A)	Η εικόνα ενός πίνακα A
diag$\{\alpha_1, \alpha_2, \ldots, \alpha_\kappa\}$	Ο διαγώνιος πίνακας με διαγώνια στοιχεία $\alpha_1, \alpha_2, \ldots, \alpha_\kappa$
span$\{x_1, x_2, \ldots, x_\kappa\}$	Η γραμμική θήκη των διανυσμάτων $x_1, x_2, \ldots, x_\kappa$

Περίληψη

Η συγγραφή της παρούσας διπλωματικής εργασίας έγινε στα πλαίσια του Προπτυχιακού Προγράμματος της Σχολής Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών του Εθνικού Μετσόβιου Πολυτεχνείου.

Σε πολλές εφαρμογές τα δεδομένα αναπαρίστανται με τη μορφή πολύ μεγάλων πινάκων. Η διαχείριση αυτού του όγκου των δεδομένων είναι απαγορευτική. Σκοπός της συγκεκριμένης διπλωματικής είναι η μελέτη και η εφαρμογή κάποιων αλγορίθμων που χρησιμοποιούν τη μέθοδο Monte Carlo για τη δημιουργία πολύ μικρότερων πινάκων πετυχαίνοντας πολύ καλές προσεγγίσεις του αρχικού προβλήματος και μικρό σφάλμα. Οι αλγόριθμοι αυτοί έχουν εφαρμογή σε πολλά προβλήματα της γραμμικής άλγεβρας και κάνουν πιο αποδοτική τη χρήση υπολογιστικών πόρων όπως ο χρόνος υπολογισμού , η μνήμη RAM και ο αριθμός περασμάτων πάνω από τα δεδομένα σε σχέση με άλλους αλγορίθμους

Οι μέθοδοι Monte Carlo είναι μια ευρεία κατηγορία υπολογιστικών αλγορίθμων που βασίζονται σε επαναλαμβανόμενη τυχαία δειγματοληψία για την επίτευξη αριθμητικών αποτελεσμάτων. Η λογική είναι ότι χρησιμοποιούν τυχαία στοιχεία για την επίλυση διάφορων προβλημάτων. Συχνά χρησιμοποιούνται σε φυσικά και μαθηματικά προβλήματα και είναι πολύ χρήσιμες όταν είναι δύσκολο ή αδύνατο να χρησιμοποιηθούν άλλες προσεγγίσεις.

Στο Κεφάλαιο 1 παρουσιάζουμε βασικούς ορισμούς και έννοιες της Γραμμικής Άλγεβρας και της Ανάλυσης Πινάκων οι οποίοι μας είναι απαραίτητοι για την κατανόηση των κεφαλαίων που ακολουθούν. Παρουσιάζουμε επίσης το μοντέλο pass efficient καθώς και δύο λήμματα δειγματοληψίας που θα χρησιμοποιήσουμε στους αλγορίθμους μας.

Στο Κεφάλαιο 2 παρουσιάζονται δύο αλγόριθμοι για το πρόβλημα του πολλαπλασιασμού πινάκων καθώς επίσης και η χρονική και χωρική πολυπλοκότητα τους. Αρχικά αναλύεται ο βασικός αλγόριθμος πολλαπλασιασμού πινάκων και στη συνέχεια ο αλγόριθμος ElementMatrix. Αναφέρεται επίσης και η αποτελεσματικότητά τους, σε σχέση με το σφάλμα που έχουν ως προς την Ευκλείδεια και τη Frobenius νόρμα, ανάλογα με το αν χρησιμοποιούνται αυθαίρετες ή σχεδόν βέλτιστες πιθανότητες.

Στο Κεφάλαιο 3 παρουσιάζονται δύο αλγόριθμοι που χρησιμοποιούν τη μέθοδο SVD (παραγοντοποίηση ιδιαζουσών τιμών) προκειμένου να προσεγγίσουν έναν πίνακα. Πιο συγκεκριμένα έχοντας έναν mxn πίνακα A οι αλγόριθμοι βρίσκουν έναν πίνακα A_k που είναι μια προσέγγιση χαμηλής τάξης του πίνακα A. Αρχικά αναλύεται ο προσεγγιστικός αλγόριθμος SVD γραμμικού χρόνου και στη συνέχεια ο προσεγγιστικός αλγόριθμος SVD σταθερού χρόνου καθώς και οι πολυπλοκότητες και τα σφάλματα που έχουν για την Ευκλείδεια και τη Frobenius νόρμα.

Στο Κεφάλαιο 4 παρουσιάζονται δύο αλγόριθμοι που υπολογίζουν τον πίνακα A' σαν μια προσέγγιση του πίνακα A. Σε πολλές εφαρμογές ένας mxn πίνακας A αποθηκεύεται στο δίσκο και είναι πολύ μεγάλος για να διαβαστεί από τη μνήμη RAM ή να κάνει πρακτικά υπολογισμούς σε υπεργραμμικό πολυωνυμικό χρόνο σε αυτήν. Επομένως ενδιαφερόμαστε για έναν πίνακα A' που υπολογίζεται εύκολα και είναι προσέγγιση του αρχικού πίνακα A. Αρχικά αναλύεται ο αλγόριθμος γραμμικού χρόνου CUR και στη συνέχεια ο αλγόριθμος σταθερού χρόνου CUR καθώς και οι πολυπλοκότητες τους και τα σφάλματα που έχουν για την Ευκλείδεια και τη Frobenius νόρμα.

Abstract

This dissertation was written as part of the Undergraduate Program of the school of Applied Mathematics and Natural Sciences of the National Technical University of Athens.

In many applications the data may be formulated as large matrices. Managing this volume of data is prohibitive. The motivation for this dissertation is to study and apply some algorithms that use the Monte Carlo method to create much smaller matrices, achieving very good approaches to the original problem and small error.

These algorithms are applicable to many linear algebra problems and make the use of computational resources such as computation time, RAM and the number of transitions over data more efficient than other algorithms.

Monte Carlo methods are a broad category of computational algorithms based on repeated random sampling to achieve arithmetic results. They use random elements to solve problems that may be predetermined. They are often used in natural and mathematical problems and are very useful when it is difficult or impossible to use other approaches.

In Chapter 1 we present basic definitions and concepts of Linear Algebra and Matrix Analysis which are necessary to understand the following chapters. We also present the pass efficient model as well as two sampling lemmas that we will use in our algorithms.

Chapter 2 presents two algorithms for the problem of matrix multiplication as well as their additional time and additional space. We first analyze the basic matrix multiplication algorithm and then the Element Wise Matrix algorithm. Their effectiveness is also reported in relation to the error in Spectral and Frobenius norm, depending on whether arbitrary or almost optimal probabilities are used.

Chapter 3 presents two algorithms that use the SVD method in order to approach a matrix. More specifically, having an mxn matrix *A* the algorithms find a matrix A_k which is a low rank approach of matrix *A.* First we analyze the approximate linear time SVD algorithm an then the approximate constant time SVD algorithm as well as the additional time and space and errors they have for the Spectral an Frobenius norm.

Chapter 4 presents two algorithms that calculate matrix A’ as an approximation of matrix A. In many applications, the data consist of (or may be naturally formulated as) an m×n matrix A which may be stored on disk but which is too large to be read into random access memory (RAM) or to practically perform superlinear polynomial time computations on it. Therefore, we are interested in an A´ matrix that is easily calculated and is an approximation of the original A matrix. First analyzed the CUR linear time algorithm and then the CUR fixed time algorithm as well as as the additional time and space and errors they have for the Spectral an Frobenius norm.

ΚΕΦΑΛΑΙΟ 1. ΒΑΣΙΚΟΙ ΟΡΙΣΜΟΙ ΚΑΙ ΕΝΝΟΙΕΣ

Σε αυτό το κεφάλαιο θα αναφέρουμε κάποια βασικά στοιχεία της γραμμικής άλγεβρας [9,11,13,3, 20,21] το μοντέλο pass efficient και δύο λήμματα που θα μας χρειαστούν παρακάτω στους αλγορίθμους μας.

1.1. ΣΤΟΙΧΕΙΑ ΓΡΑΜΜΙΚΗΣ ΑΛΓΕΒΡΑΣ

Για ένα διάνυσμα $x \in R^n$ θέτουμε $|x| = (\sum_{i=1}^{n}|x_i|^2)^{1/2}$ να είναι η Ευκλείδεια νόρμα. Για έναν πίνακα $A \in R^{mxn}$ θέτουμε $A^{(j)}$, $j=1,\ldots,n$, θέτοντας την j-οστή στήλη του A ως διάνυσμα στήλη και $A_{(i)}$ με $i = 1,\ldots,m$, θέτοντας την i-οστή γραμμή του A σαν διάνυσμα γραμμή. Επομένως αν το A_{ij} υποδηλώνει το στοιχείο που βρίσκεται στη θέση i,j του πίνακα A, τότε $A_{ij} = (A^{(j)})_i = (A_{(i)})_j$.

Η εικόνα ενός πίνακα $A \in R^{mxn}$ είναι

$$range(A) = \{y \in R^m : y = Ax \text{ για κάποια } x \in R^n\} = \text{span}(A^{(1)}, \ldots, A^{(n)}) \qquad (1.1.1)$$

Ο βαθμός του πίνακα A, rank(A), είναι η διάσταση της εικόνας του πίνακα, range(A), και είναι ίση με τον αριθμό των γραμμικά ανεξάρτητων στηλών του πίνακα A. Επίσης, είναι ίση με το βαθμό του πίνακα A^T, rank(A^T) και άρα είναι επίσης ίση με τον αριθμό και των γραμμικά ανεξάρτητων γραμμών του πίνακα A.

Ο μηδενικός χώρος του A είναι

$$\text{null}\{x \in R^n : Ax = 0\}. \qquad (1.1.2)$$

Ορίζουμε τις νόρμες πινάκων

$$\text{Frobenius νόρμα} \quad \|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2}\,, \qquad (1.1.3)$$

$$\text{Ευκλείδεια νόρμα } \|A\|_2 = sup_{x \in R^n, x \neq 0} \frac{|Ax|}{|x|}\,. \qquad (1.1.4)$$

Αν $Tr(A)$ είναι το ίχνος του πίνακα A το οποίο ισούται με το άθροισμα των διαγώνιων στοιχείων του A, τότε $\|A\|_F^2 = Tr(A^T A) = Tr(AA^T)$.

Οι δύο αυτές νόρμες συνδέονται μεταξύ τους με την εξής σχέση:

$$\|A\|_2 \leq \|A\|_F \leq \sqrt{n}\|A\|_2. \qquad (1.1.5)$$

Και οι δύο αυτές νόρμες παρέχουν μια πληροφορία για το μέγεθος του πίνακα A. Σημειώνουμε ότι αν $A \in R^{mxn}$ τότε υπάρχει $x \in R^n$ τέτοιο ώστε $|x| = 1$ και $A^T Ax = \|A\|_2^2 x$. Επίσης αν $\{x^1, x^2, \ldots, x^n\}$ είναι οποιαδήποτε βάση του R^n και $A \in R^{mxn}$, τότε $\|A\|_F^2 = \sum_{i=1}^{n}\left|Ax^i\right|^2$.

Αν $A \in R^{mxn}$, τότε υπάρχουν οι ορθομοναδιαίοι πίνακες $U = [u^1 u^2 \ldots u^m] \in R^{mxm}$ και $V = [v^1 v^2 \ldots v^m] \in R^{nxn}$ όπου $\{u^t\}_{t=1}^{m} \in R^m$ και $\{v^t\}_{t=1}^{n} \in R^n$ για τους οποίους ισχύει ότι

$$U^T AV = \Sigma = diag(\sigma_{1,\ldots,}\sigma_\rho), \qquad (1.1.6)$$

όπου $\Sigma \in R^{mxn}$, ρ=min{m,n}, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_\rho \geq 0$. Ισοδύναμα,

$$A = U\Sigma V^T.$$

Η παραπάνω γραφή του A με τους τρεις πίνακες U, V και Σ ονομάζεται παραγοντοποίηση ιδιαζουσών τιμών (singular value decomposition, SVD) του πίνακα A. Οι τιμές $\sigma_1, \sigma_2, \ldots, \sigma_\rho$ καλούνται ιδιάζουσες τιμές του πίνακα A. Τα διανύσματα στήλες u_i και v_i $(1 \leq i \leq min\{m, n\})$ λέγονται αριστερά και δεξιά ιδιάζοντα διανύσματα του A που αντιστοιχούν στην ιδιάζουσα τιμή σ_i.

Οι στήλες των πινάκων U και V ικανοποιούν τις σχέσεις $Av^i = \sigma_i u^i$ και $A^T u^i = \sigma_i v^i$. Στους συμμετρικούς πίνακες τα αριστερά και τα δεξιά ιδιάζοντα διανύσματα ταυτίζονται.

Από τις σχέσεις

$$AA^T = Udiag\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_{min\{n,m\}}^2\}U^T \tag{1.1.7}$$

και

$$A^T A = Vdiag\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_{min\{n,m\}}^2\}V^T, \tag{1.1.8}$$

συμπεραίνουμε ότι:

1) Οι μη μηδενικές ιδιάζουσες τιμές του A είναι οι τετραγωνικές ρίζες των μη μηδενικών ιδιοτιμών των (θετικά ημιορισμένων) πινάκων AA^T και $A^T A$.
2) Τα αριστερά ιδιάζοντα διανύσματα του A είναι τα ιδιοδιανύσματα του AA^T.
3) Τα δεξιά ιδιάζοντα διανύσματα του A είναι τα ιδιοδιανύσματα του $A^T A$.

Η SVD μπορεί να αποκαλύψει σημαντικές πληροφορίες σχετικά με τη δομή ενός πίνακα. Αν ορίσουμε το r έτσι ώστε $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots \geq \sigma_\rho = 0$, τότε rank($A$)=$r$, null($A$)=span($v^{r+1}, \ldots, v^\rho$) και range($A$)=span($u^1, \ldots, u^r$).

Ο πίνακας $U_r \in R^{mxr}$ είναι ο πίνακας που αποτελείται από τις πρώτες r στήλες του πίνακα U, ο πίνακας $V_r \in R^{rxn}$ αποτελείται από τις πρώτες r γραμμές του πίνακα V και ο πίνακας $\Sigma_r \in R^{rxr}$, υποδηλώνει κύριο r x r υποπίνακα του πίνακα Σ, τότε

$$A = U_r \Sigma_r V_r^T = \sum_{t=1}^r \sigma_t \, u^t v^{t^T}. \tag{1.1.9}$$

Δηλαδή, ο πίνακας A γράφεται ως άθροισμα r πινάκων βαθμού 1. Για τυχαίο φυσικό αριθμό $k < r$ θεωρούμε τον πίνακα

$$A_k = U_k \Sigma_k V_k^T = \sum_{t=1}^k \sigma_t \, u^t v^{t^T}. \tag{1.1.10}$$

Τότε $A_k = U_k U_k^T A = (\sum_{t=1}^k u^t u^{t^T})A$ και $A_k = AV_k V_k^T = A(\sum_{t=1}^k v^t v^{t^T})$, δηλαδή ο πίνακας A_k είναι η προβολή του πίνακα A στο διάστημα που αποτελείται από τα πρώτα k ιδιάζοντα διανύσματα του A.

Επιπλέον, η απόσταση του πίνακα A από οποιοδήποτε προσεγγιστικό πίνακα του A βαθμού k, (όπως φαίνεται και από τις δυο νόρμες ($\|\cdot\|_2, \|\cdot\|_F$) , ελαχιστοποιείται από τον A_k, δηλαδή

$$\min_{D \in R^{mxn}, rank(D) \leq k} \|A - D\|_2 = \|A - A_k\|_2 = \sigma_{k+1}(A) \tag{1.1.11}$$

και

$$\min_{D \in R^{mxn}, rank(D) \leq k} \|A - D\|_F^2 = \|A - A_k\|_F^2 = \sum_{t=k+1}^r \sigma_t^2 \, (A). \tag{1.1.12}$$

Έτσι ο πίνακας A_k που κατασκευάσαμε είναι η καλύτερη προσέγγιση του πίνακα $\boldsymbol{A}$ και για τα τις δύο νόρμες.

Γενικά μπορούμε να δούμε ότι $\|A\|_2 = \sigma_1$ και $\|A\|_F^2 = \sum_{i=1}^r \sigma_i^2$.

Από τη θεωρία διαταραχών των πινάκων γνωρίζουμε ότι το μέγεθος της διαφοράς μεταξύ δύο πινάκων μπορεί να χρησιμοποιηθεί σαν όριο της διαφοράς των ιδιαζουσών τιμών δύο πινάκων . Συγκεκριμένα αν $A, E \in R^{mxn}$, m $\geq n$ τότε

$$\max_{t:1\le t\le n} |\sigma_t(A+E)-\sigma_t(A)| \le \|E\|_2 \quad (1.1.13)$$

και

$$\sum_{k=1}^{n}(\sigma_k(A+E)-\sigma_k(A))^2 \le \|E\|_F^2. \quad (1.1.14)$$

Η τελευταία σχέση είναι γνωστή ως ανίσωση Hoffman-Wielandt.

1.2. Το μοντέλο Pass Efficient

Σε αυτό το κεφάλαιο, θα ορίσουμε ένα υπολογιστικό μοντέλο στο οποίο το υπολογιστικό κόστος είναι το πλήθος των φορών που περνάει πάνω από τα δεδομένα. Θα ορίσουμε επιπλέον τη χρονική και χωρική πολυπλοκότητα.

Το pass-efficient model βασίστηκε στην παρατήρηση ότι στους σύγχρονους υπολογιστές ο χώρος αποθήκευσης στο σκληρό δίσκο έχει αυξηθεί πολύ περισσότερο από τον αποθηκευτικό χώρο της RAM [4].

Έτσι κάποιος έχει τη δυνατότητα να αποθηκεύει μεγάλα ποσά δεδομένων αλλά δεν έχει τυχαία πρόσβαση σε αυτά. Από την άλλη όμως το να προσεγγίσεις αυτό το πρόβλημα με αλγόριθμους πολυωνυμικού ή γραμμικού χρόνου με μεγάλες σταθερές είναι απαγορευτικό.

Για να μοντελοποιήσουμε το φαινόμενο αυτό , θεωρούμε το pass-efficient model στο οποίο τα τρία πράγματα που μας ενδιαφέρουν είναι ο αριθμός των περασμάτων πάνω από τα δεδομένα, η χρονική και η χωρική πολυπλοκότητα [10]. Τα δεδομένα αποθηκεύονται σε έναν εξωτερικό χώρο στο δίσκο που αποτελείται από δεδομένα των οποίων το μέγεθος φράσσεται από μία σταθερά και επιτρέπεται μόνο η ανάγνωση τους. Η μόνη πρόσβαση που έχει ένας αλγόριθμος στα δεδομένα είναι μέσω ενός περάσματος από αυτά. Με τον όρο πέρασμα από τα δεδομένα εννοούμε μια διαδοχική ανάγνωση των δεδομένων εισόδου (input). Επιτρέπεται μόνο ένας σταθερός χρόνος επεξεργασίας ανά bit ανάγνωσης. Αυτή είναι μια πιο περιορισμένη έννοια της μετάβασης στα δεδομένα από ότι σε άλλα μοντέλα ροής δεδομένων [12,10,8].

Ειδικότερα το pass-efficient model απαιτεί σταθερό και όχι λογαριθμικό χρόνο για να διαβάσει τα δεδομένα εισόδου. Εκτός από τον εξωτερικό χώρο στο δίσκο για την αποθήκευση δεδομένων και έναν μικρό αριθμό περασμάτων πάνω από τα δεδομένα, ένας αλγόριθμος στο pass-efficient model επιτρέπεται να χρησιμοποιεί επιπλέον χώρο μνήμης RAM και πρόσθετο χρόνο υπολογισμού.

Ένας αλγόριθμος που δουλεύει με το μοντέλο αυτό θεωρείται pass-efficient όταν απαιτεί σταθερό αριθμό περασμάτων ανεξάρτητα από το μέγεθος εισόδου και χωρική και χρονική πολυπλοκότητα που είναι υπογραμμικά στο μήκος της ροής δεδομένων για να υπολογίσει μια περιγραφή της λύσης του, η οποία στη συνέχεια επιστρέφεται από τον αλγόριθμο. Περιγραφή της λύσης είναι είτε μια ρητή λύση (αν αυτό είναι εφικτό εντός της καθορισμένης χωρικής και χρονικής πολυπλοκότητας) είτε μια προσέγγιση της λύσης που μπορεί να υπολογιστεί στην επιθυμητή χωρική και χρονική πολυπλοκότητα και αυτό μπορεί να επεκταθεί σε μια ρητή λύση με την πρόσθετη δαπάνη ενός περάσματος πάνω από τα δεδομένα και γραμμική (ως προς το μέγεθος των δεδομένων εισόδου) χωρική και χρονική πολυπλοκότητα. Ανάλογα με την εφαρμογή αυτό το τελευταίο βήμα μπορεί να είναι απαραίτητο αλλά μπορεί και όχι.

Αν τα δεδομένα μας αναπαρίστανται ως ένας πίνακας mxn τότε η ροή δεδομένων έχει μήκος O(mn) και αν έχουμε έναν αλγόριθμο ο οποίος έχει γραμμική , ως προς τον αριθμό των δεδομένων, χωρική και χρονική πολυπλοκότητα ή στη διαστασιοποίηση των σημείων δεδομένων δηλαδή το O(m) ή το O(n) είναι υπογραμμικό στο μήκος της ροής δεδομένων και άρα είναι αποδοτικό.

Θα αναφερθούμε πρωτίστως στα μοντέλα που απαιτούν χωρική και χρονική πολυπλοκότητα που είναι είτε O(m+n) είτε σταθερή ως προς τα m και n.

Μπορεί να έχουμε αραιή αναπαράσταση των δεδομένων. Κάθε στοιχείο της ροής δεδομένων περιγράφεται από ένα ζεύγος (i,j) όπου τα στοιχεία στη ροή δεδομένων μπορεί να μην είναι αντιστοιχισμένα σε σχέση με τους δείκτες (i,j) και να πρέπει να παρουσιαστούν μόνο τα μη μηδενικά στοιχεία του πίνακα.

Αυτή η πολύ γενική μορφή είναι κατάλληλη για εφαρμογές όπου, για παράδειγμα, πολλοί άνθρωποι μπορεί να γράφουν τμήματα ενός πίνακα σε μια κεντρική βάση δεδομένων και όπου κανείς δεν μπορεί να κάνει παραδοχές σε σχέση με τους κανόνες (συρρίκνωση γραφής κλπ).

Αλγόριθμος Select .

Είσοδος : $\{\alpha_1, \ldots, \alpha_n\}$, $a_i \geq 0$
Έξοδος: i^*, a_{i^*}.
1. $D = 0$.
2. Για $i = 1$ μέχρι n,
(a) $D = D + a_i$.
(b) Με πιθανότητα a_i/D, $i^* = i$ και $a_{i^*} = a_i$.
3. Εμφάνισε i^*,a_{i^*}.

1.3. ΛΗΜΜΑΤΑ ΔΕΙΓΜΑΤΟΛΗΨΙΑΣ

Σε αυτή την ενότητα παρουσιάζονται δύο λήμματα δειγματοληψίας που θα χρησιμοποιηθούν από τους αλγόριθμους μας. Το πρώτο λήμμα δηλώνει ότι σε ένα πέρασμα πάνω από τα δεδομένα μπορεί κανείς να διαλέξει ένα στοιχείο σύμφωνα με ορισμένες κατανομές πιθανοτήτων

Λήμμα 1.4.1.

Θεωρούμε ότι τα στοιχεία $a_1, \ldots, a_n$>0 διαβάζονται σε ένα πέρασμα, δηλαδή μια διαδοχική ανάγνωση δεδομένων από τον αλγόριθμο select. Στη συνέχεια ο αλγόριθμος select απαιτεί χωρική πολυπλοκότητα O(1), δηλαδή σταθερή ως προς το n και επιστρέφει τυχαία ένα i* επιλεγμένο από την πιθανότητα

$$Pr[\, i^* = i] = a_i / \sum_{i'=1}^{n} a_i \,'$$

Απόδειξη.

Αρχικά απαιτείται η διατήρηση της επιλεγμένης τιμής και χωρικής πολυπλοκότητας O(1). Το υπόλοιπο της απόδειξης γίνεται με επαγωγή. Έπειτα διαβάζει το πρώτο στοιχείο a_1, i*=1, με πιθανότητα a_1 /a_1=1. Έστω $D_l = \sum_{i'}^{l} a_{i'}$ και υποθέτουμε ότι ο αλγόριθμος έχει διαβάσει τα πρώτα l στοιχεία $a_1,\ldots, a_l$ και έχει αποθηκεύσει το άθροισμα D_l και το επιλεγμένο i* έτσι ώστε $Pr[\, i^* = i] = a_i/D_l$. Όταν διαβάζει το a_{l+1} ο αλγόριθμος θετει $i^* = l + 1$ με πιθανότητα $\frac{a_{l+1}}{D_{l+1}}$ και αποθηκεύει την τελευταία τιμή του i* .Σε αυτό το σημείο έχουμε $Pr[\, i* = l + 1] = a_{l+1}/D_{l+1}$. Επιπλέον, για $i = 1,\ldots.,l$, $Pr[\, i* = i] = \frac{a_i}{D_l}(1 - \frac{a_{l+1}}{D_{l+1}}) = \frac{a_i}{D_{l+1}}$. Με επαγωγή παίρνουμε αυτό το αποτέλεσμα όταν $l + 1 = n$. ■

Σε ένα μόνο πέρασμα πάνω από τα δεδομένα αυτός ο αλγόριθμος μπορεί να τρέξει παράλληλα με O(s) μονάδες μνήμης και να επιστρέψει το s που είναι ανεξάρτητο του δείγματος $\iota_1^*, \ldots., \iota_s^*$ έτσι ώστε για κάθε i_t^*, t= 1,…,s, να έχουμε $Pr[\, i_t^* = i] = \frac{a_i}{\sum_{i'=1}^{n} a_{i'}}$.

Το επόμενο λήμμα είναι μια τροποποίηση του προηγούμενου λήμματος για την περίπτωση που κάποιος θέλει να επιλέξει από έναν πίνακα μια σειρά με κάποια πιθανότητα. Αυτό μπορεί να

υλοποιηθεί με χωρική και χρονική πολυπλοκότητα O(1). Είναι προφανές ότι με μια μικρή τροποποίηση εφαρμόζεται και για επιλογή στήλης.

Λήμμα 1.4.2.

Έστω ένας πίνακας $A \in R^{mxn}$ που διαβάζεται σε ένα πέρασμα δηλαδή μια διαδοχική ανάγνωση των δεδομένων από τον αλγόριθμο select. Τότε ο αλγόριθμος απαιτεί O(1) , δηλαδή σταθερό σε σχέση με τα *m* και *n* χώρο αποθήκευσης και επιστρέφει *i*, j** έτσι ώστε $Pr[\, i^* = i \wedge j^* = j] = A_{ij}^2 / \|A\|_F^2$ και έτσι $Pr[\, i^* = i] = \frac{|A_{(i)}|^2}{\|A\|_F^2}$.

Απόδειξη.

Όσο $A_{i^*,\, j^*}^2 > 0$ ο πρώτος ισχυρισμός ισχύει από το Λήμμα 1. Το δεύτερο αποδεικνύεται αφού

$$Pr[\, i^* = i] = \sum_{j=1}^{n} Pr[\, i^* = i \hat{} j^* = j] = \sum_{j=1}^{n} \frac{A_{ij}^2}{\|A\|_F^2} = \frac{|A_{(i)}|^2}{\|A\|_F^2}. \quad \blacksquare$$

Αλγόριθμοι όπως ο select που επιλέγουν στοιχεία από μια μεγάλη δεξαμενή δεδομένων των οποίων το μέγεθος είναι άγνωστο λέγονται reservoir algorithms [14].

ΚΕΦΑΛΑΙΟ 2. ΠΡΟΣΕΓΓΙΣΤΙΚΟΣ ΠΟΛΛΑΠΛΑΣΙΑΣΜΟΣ ΠΙΝΑΚΩΝ

2.1. Ο ΒΑΣΙΚΟΣ ΑΛΓΟΡΙΘΜΟΣ ΠΟΛΛΑΠΛΑΣΙΑΣΜΟΥ ΠΙΝΑΚΩΝ

Σε αυτήν την ενότητα περιγράφεται ο βασικός αλγόριθμος πολλαπλασιασμού πινάκων (BasicMatrixMultiplication) για την προσέγγιση του γινομένου δύο πινάκων.

2.1.1. Ο αλγόριθμος

Βασικός Αλγόριθμος Πολλαπλασιασμού Πινάκων

Είσοδος: $A \in R^{mxn}$, $B \in R^{nxp}$, $c \in Z^+$ τέτοιο ώστε $1 \leq c \leq n$ και $\{p_i\}_{i=1}^{n}$ τέτοια ώστε $p_i \geq 0$ και $\sum_{i=1}^{n} p_i = 1$.

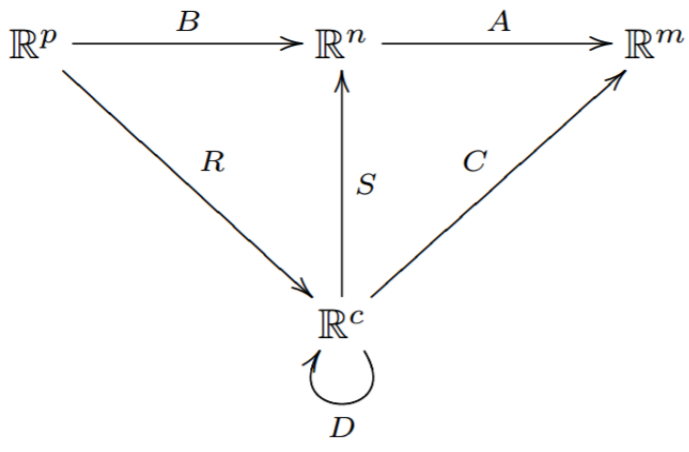
Έξοδος : $C \in R^{mxc}$ και $R \in R^{cxp}$.

1. Για t από 1 μέχρι c
   (a) Διάλεξε $i_t \in \{1,\ldots,n\}$ με πιθανότητα $\Pr[i_t = k] = p_k$ ανεξάρτητα και με επανάθεση.
   (b) Θέτουμε $C^{(t)} = A^{(i_t)}/\sqrt{cp_{i_t}}$ και $R_{(t)} = B_{(i_t)}/\sqrt{cp_{i_t}}$
2. Υπολόγισε C,R.

Για δύο πίνακες $A \in R^{mxn}$ και $B \in R^{nxp}$ το αποτέλεσμα του πολλαπλασιασμού πινάκων AB μπορεί να γραφτεί σαν άθροισμα από n πίνακες βαθμού 1

$$AB = \sum_{t=1}^{n} A^{(t)} B_{(t)}. \qquad (2.1.1)$$

Στην περίπτωση που ένας πολλαπλασιασμός πινάκων εκφράζεται με την παραπάνω σχέση, ένας απλός τυχαιοποιημένος αλγόριθμος προσεγγίζει το αποτέλεσμα του πολλαπλασιασμού πινάκων ως εξής: Παίρνει τυχαίο δείγμα με επανάθεση όρων στο άθροισμα c φορές σύμφωνα με μια κατανομή πιθανότητας $\{p_i\}_{i=1}^{n}$ κάνοντας κατάλληλη κανονικοποίηση στον κάθε όρο, και επιστρέφει το άθροισμα των κανονικοποιημένων όρων. Αν m=p=1 τότε $A^{(t)}$, $B_{(t)} \in R$ αποδεικνύεται κατά προφανή τρόπο ότι αυτή η δειγματοληψία παράγει έναν αμερόληπτο εκτιμητή για το άθροισμα. Όταν οι όροι του αθροίσματος είναι πίνακες βαθμού 1 αποδεικνύουμε ότι το αποτέλεσμα είναι παρόμοιο.



Σχήμα 1 . Διάγραμμα του βασικού αλγορίθμου πολλαπλασιασμού πινάκων.

Βασικός αλγόριθμος πολλαπλασιασμού πινάκων matlab

```
N=10;

n=randi([1000 2000],1,1);

m=randi([1000 2000],1,1);

p=randi([1000 2000],1,1);

A=randi(100,m,n);

B=randi(100,n,p);

CC=zeros(m,N);  % The initial CC.
RR=zeros(N,p);  % The initial RR.

for xi=1:S  % We repeat S times
the process.
            % Symbol xi is not
used elsewhere.

for i=1:m
 for j=1:n
  A2(i,j)=A(i,j)^2;
 end
end

AF=norm(A,'fro');
BF=norm(B,'fro');

for j=1:n
 sum=0;
 for i=1:m
   sum=A2(i,j)+sum;
 end
 P(j)=sum/AF^2;
end

sum=0;

for k=1:n
sum=sum+norm(A(:,k)*norm(B(k,:)))
;
end

for k=1:n
Pr(k)=(norm(A(:,k)*norm(B(k,:)))/
sum;
end

for i=1:n
 if P(i)>Pr(i)
 L(i)=i;
 end
end

L1=nonzeros(L);
x=randi(size(L1,1),1,N);  % The
vector of indices of columns.
COL=A(:,L1(x));
ROW=B(L1(x),:);

for i=1:N
 V(i)=P(L1(x(i)));
end

for t=1:N
 C(:,t)=COL(:,t)/sqrt(N*V(t));
 R(t,:)=ROW(t,:)/sqrt(N*V(t));
end



CC=CC+C;
RR=RR+R;
 end
 % Now, CC is the sum of all C's
     % and RR is the sum of all
R's.

CC=CC/S;  % Mean value of all
C's.
RR=RR/S;  % Mean value of all
R's.

Matrix_Product_Estimation=CC*RR;
% The estimation CC*RR of A*B.
Matrix_Product=A*B;
% The matrix product A*B.

Relative_Error=norm(Matrix_Produc
t-
Matrix_Product_Estimation,'fro')/
norm(Matrix_Product,'fro')  % The
relative error of the estimation.

toc  % The command tic-toc gives
the execution time
```

Ο αλγόριθμος αυτός δέχεται σαν είσοδο δύο πίνακες *A* και *B*, μία πιθανότητα $\{p_i\}_{i=1}^{n}$και έναν αριθμό *c* για τα ζεύγη στήλης-γραμμής που επιλέγει και επιστρέφει δύο πίνακες *C* και *R* έτσι ώστε το αποτέλεσμα του πολλαπλασιασμού *CR* να είναι μια προσέγγιση του *AB*.

Παρατηρούμε ότι

$$CR = \sum_{t=1}^{c} C^{(t)} R_{(t)} = \sum_{t=1}^{c} \frac{1}{cp_{i_t}} A^{(i_t)} B_{(i_t)}.$$

Ένα σημαντικό ζήτημα είναι η επιλογή των πιθανοτήτων $\{p_i\}_{i=1}^{n}$ και η κανονικοποίηση. Είναι εύκολα κατανοητό ότι ο όρος $\frac{1}{\sqrt{cp_{i_t}}}$ που χρησιμοποιείται στον αλγόριθμο καθιστά το *CR* αμερόληπτο εκτιμητή του *AB* (Λήμμα 2.1.1). Το λήμμα αυτό επίσης υπολογίζει τη διασπορά $Var[(CR)_{ij}]$ για τις πιθανότητες $\{p_i\}_{i=1}^{n}$. Στη συνέχεια, θα υπολογίσουμε την ποσότητα $E[\|AB - CR\|_F^2]$ και θα δούμε ότι οι πιθανότητες της μορφής $p_k = \frac{|A^{(k)}||B_{(k)}|}{N}$, *k*=1,…,*N*, είναι οι βέλτιστες γιατί ελαχιστοποιούν αυτή την ποσότητα (Λήμμα 2.1.2).

Αυτός ο τρόπος για την προσέγγιση του πολλαπλασιασμού πινάκων έχει αρκετά πλεονεκτήματα:

1) Είναι εννοιολογικά απλός και σε ορισμένες περιπτώσεις μπορεί να χρησιμοποιηθεί για να προσεγγίσει αποτελέσματα για περισσότερους από δύο πίνακες.
2) Δεδομένου ότι ο αλγόριθμος περιλαμβάνει πολλαπλασιασμό μικρότερων πινάκων μπορεί να χρησιμοποιήσει οποιονδήποτε αλγόριθμο από την βιβλιογραφία για να τον κάνει [9,19,22].
3) Η προσέγγιση αυτή δεν παραβιάζει την ακεραιότητα τον πινάκων.
4) Ο αλγόριθμος μπορεί εύκολα να υλοποιηθεί.

2.1.2. Υλοποίηση του χρόνου δειγματοληψίας και χρόνου λειτουργίας

Για την εφαρμογή του αλγορίθμου BasicMatrixMultiplication πρέπει να αποφασιστεί ποια θα είναι τα στοιχεία της εισόδου που θα επιλεγούν και από τα στοιχεία αυτά πρέπει στη συνέχεια να επιλεγεί ένα δείγμα. Στην περίπτωση ομοιόμορφης δειγματοληψίας μπορεί κανείς να αποφασίσει πριν φανεί η είσοδος ,ποια ζεύγη γραμμών-στηλών πρέπει να επιλεγούν. Στη συνέχεια ένα μόνο πέρασμα πάνω από τους πίνακες αρκεί για να επιλεγούν οι στήλες και οι γραμμές που μας ενδιαφέρουν και να κατασκευαστούν οι πίνακες *C* και *R*. Αυτό απαιτεί χρονική και χωρική πολυπλοκότητα O($c(m+p)$)

Θα δούμε παρακάτω ότι είναι χρήσιμο το δείγμα να επιλέγεται σύμφωνα με μία μη ομοιόμορφη κατανομή πιθανότητας που εξαρτάται από τα μήκη της στήλης και της σειράς.

Σε μια τέτοια περίπτωση προκειμένου να αποφασίσουμε ποια ζεύγη στηλών-γραμμών πρέπει να επιλεγούν αρκεί ένα πέρασμα από τους πίνακες με χρονική και χωρική πολυπλοκότητα O(n).

Στη συνέχεια σε ένα δεύτερο πέρασμα μπορούν να ληφθούν δείγματα από τις στήλες και τις σειρές που μας ενδιαφέρουν και να κατασκευαστούν και να αποθηκευτούν οι πίνακες *C* και *R*.

Αυτό απαιτεί χρονική και χωρική πολυπλοκότητα O($c(m+p)$).

Έτσι τόσο για ομοιόμορφη όσο και για μη ομοιόμορφη δειγματοληψία, είτε με ένα είτε με δύο περάσματα πάνω από τα δεδομένα η χωρική και χρονική πολυπλοκότητα O($c(m+n+p)$) επαρκεί για να γίνει η δειγματοληψία από τους πίνακες *A* και *B* και να κατασκευαστούν οι πίνακες *C* και *R*.

2.1.3. Ανάλυση του αλγορίθμου για αυθαίρετες πιθανότητες

Σε αυτή την ενότητα αποδεικνύουμε τα ανώτερα όρια για τη νόρμα $\|AB - CR\|_F^2$ όπου οι πίνακες *C* και *R* είναι οι πίνακες που επιστρέφει ο αλγόριθμος BasicMatrixMultiplication. Από την ανισότητα του Jensen που βάζει ένα όριο στο$\|AB - CR\|_F^2$ (κάτω από προϋποθέσεις) συνεπάγεται ένα όριο για

το $\|AB - CR\|_F$. Υπενθυμίζουμε ότι ένα όριο στο $\|AB - CR\|_F$ μας δίνει ένα όριο και για το $\|AB - CR\|_2$ δεδομένου ότι $\|AB - CR\|_2 \leq \|AB - CR\|_F$.

Το πρώτο λήμμα αποδεικνύει ότι το στοιχείο (i,j) της προσέγγισης είναι ίσο με το στοιχείο (i,j) του αρχικού γινομένου. Περιγράφει επίσης τη διακύμανση του στοιχείου (i,j).

Λήμμα 2.1.1.

Θεωρούμε ότι $A \in R^{mxn}$, $B \in R^{nxp}$ και $c \in Z^+$ τέτοιο ώστε $1 \leq c \leq n$, και $\{p_i\}_{i=1}^n$ με $p_i \geq 0$ και $\sum_{i=1}^n p_i = 1$. Κατασκευάζουμε τους πίνακες C και R με τον αλγόριθμο BasicMatrixMultiplication και παίρνουμε μια CR προσέγγιση του AB. Τότε

$$E[(CR)_{ij}] = (AB)_{ij} \quad \text{και} \quad Var[(CR)_{ij}] = \frac{1}{c}\sum_{k=1}^n \frac{A_{i_k}^2 B_{k_j}^2}{p_k} - \frac{1}{c}(AB)_{ij}^2.$$

Απόδειξη.

Έστω δεδομένα i,j. Για t =1,…,c, θέτουμε $X_t = (\frac{A^{(i_t)}B_{(i_t)}}{cp_{i_t}})_{ij} = \frac{A_{ii_t}B_{i_tj}}{cp_{i_t}}$.

Έτσι

$$E[X_t] = \sum_{k=1}^n p_k \frac{A_{ik}B_{kj}}{cp_k} = \frac{1}{c}(AB)_{ij} \quad \text{και} \quad E[X_t^2] = \sum_{k=1}^n \frac{A_{ik}^2 B_{kj}^2}{c^2 p_k}.$$

Αφού $(CR)_{ij} = \sum_{t=1}^c X_t$ έχουμε ότι $E[(CR)_{ij}] = \sum_{t=1}^c E[X_T] = (AB)_{ij}$. Επειδή $(CR)_{ij}$ είναι το άθροισμα από c ανεξάρτητες τυχαίες μεταβλητές $Var\left[(CR)_{ij}\right] = \sum_{t=1}^c Var[X_t]$. Επειδή $Var[X_t] = E[X_t^2] - E[X_t]^2$, έχουμε ότι

$$Var[X_t] = \sum_{k=1}^n \frac{A_{ik}^2 B_{kj}^2}{c^2 p_k} - \frac{1}{c^2}(AB)_{ij}^2. \qquad \blacksquare$$

Στο επόμενο λήμμα, βρίσκουμε ένα όριο για την ποσότητα $E[\|AB - CR\|_F^2]$ χρησιμοποιώντας το Λήμμα 2.1.1. Στη συνέχεια θα δούμε κατά πόσο η τιμή του σφάλματος εξαρτάται από τις πιθανότητες p_i.

Λήμμα 2.1.2.

Θεωρούμε ότι $A \in R^{mxn}$, $B \in R^{nxp}$, $c \in Z^+$ τέτοιο ώστε $1 \leq c \leq n$, και $\{p_i\}_{i=1}^n$ με $p_i \geq 0$ και $\sum_{i=1}^n p_i = 1$. Κατασκευάζουμε τους πίνακες C και R με τον αλγόριθμο BasicMatrixMultiplication και παίρνουμε μια CR προσέγγιση του AB. Τότε

$$E[\|AB - CR\|_F^2] = \sum_{k=1}^n \frac{\|A^{(k)}\|^2 \|B_{(k)}\|^2}{cp_k} - \frac{1}{c}\|AB\|_F^2. \qquad (2.1.2)$$

Επιπλέον, αν

$$p_k = \frac{|A^{(k)}||B_{(k)}|}{\sum_{k'=1}^n |A^{(k')}||B_{(k')}|}, \qquad (2.1.3)$$

τότε

$$E[\|AB - CR\|_F^2] = \frac{1}{c}(\sum_{k=1}^n \left|A^{(k)}\right| \left|B_{(k)}\right|)^2 - \frac{1}{c}\|AB\|_F^2. \qquad (2.1.4)$$

Απόδειξη.

Αρχικά έχουμε ότι

$$E[\|AB - CR\|_F^2 = \sum_{i=1}^m \sum_{j=1}^p E[(AB - CR)_{ij}^2] = \sum_{i=1}^m \sum_{j=1}^p Var[(CR)_{ij}].$$

Από το Λήμμα 2.1.1 συνεπάγεται ότι

$$E[\|AB - CR\|_F^2] = \frac{1}{c}\sum_{k=1}^n \frac{1}{p_k}\left(\sum_i A_{ik}^2\right)\left(\sum_j B_{kj}^2\right) - \frac{1}{c}\|AB\|_F^2$$

$$= \frac{1}{c}\sum_{k=1}^n \frac{1}{p_k}\left|A^{(k)}\right|^2 \left|B_{(k)}\right|^2 - \frac{1}{c}\|AB\|_F^2.$$

Αν $p_k = \frac{|A^{(k)}||B_{(k)}|}{\sum_{k'=1}^{n}|A^{(k')}||B_{(k')}|}$, τότε

$$E[\|AB - CR\|_F^2 = \frac{1}{c}(\sum_{k=1}^{n}\left|A^{(k)}\right|\left|B_{(k)}\right|)^2 - \frac{1}{c}\|AB\|_F^2.$$

Τέλος για να αποδείξουμε ότι η επιλογή αυτή του p_k ελαχιστοποιεί την ποσότητα $E[\|AB - CR\|_F^2]$ ορίζουμε τη συνάρτηση

$$f(p_1, \ldots\ldots, p_n) = \sum_{k=1}^{n}\frac{1}{p_k}\left|A^{(k)}\right|^2\left|B_{(k)}\right|^2,$$

η οποία δείχνει την εξάρτηση της ποσότητας $E[\|AB-CR\|_F^2]$ από το p_k. Για να ελαχιστοποιήσουμε την f καθώς $\sum_{k=1}^{n} p_k = 1$ εισάγουμε τον πολλαπλασιαστή Lagrange λ και ορίζουμε τη συνάρτηση

$$g(p_1, \ldots., p_n) = f(p_1, \ldots., p_n) + \lambda(\sum_{k=1}^{n} p_k - 1).$$

Έπειτα

$$0 = \frac{\partial g}{\partial p_i} = \frac{-1}{p_i^2}\left|A^{(i)}\right|^2\left|B_{(i)}\right|^2 + \lambda$$

και έτσι

$$p_i = \frac{|A^{(i)}||B_{(i)}|}{\sqrt{\lambda}} = \frac{|A^{(i)}||B_{(i)}|}{\sum_{i'=1}^{n}|A^{(i')}||B_{(i')}|},$$

όπου η δεύτερη ισότητα προκύπτει λύνοντας ως προς $\sqrt{\lambda}$ στο $\sum_{k=1}^{n-1} p_k = 1$. Το ότι αυτές οι πιθανότητες είναι οι βέλτιστες προκύπτει αφού $\frac{\partial^2 g}{\partial {p_i}^2} > 0$, για κάθε i έτσι ώστε $\left|A^{(i)}\right|^2\left|B_{(i)}\right|^2 > 0$. ■

2.1.4. Ανάλυση του αλγορίθμου για σχεδόν βέλτιστες πιθανότητες

Χρησιμοποιώντας το Λήμμα 2.1.2 και την ανισότητα Jensen μπορούν να βρεθούν άνω όρια για τις ποσότητες $E[\|AB - CR\|_F^2]$ και $E[\|AB - CR\|_F]$ για διάφορες πιθανότητες $\{p_i\}_{i=1}^{n}$. Θα περιορίσουμε την προσοχή μας σε δύο συγκεκριμένα σύνολα πιθανοτήτων . Είπαμε ότι οι πιθανότητες δειγματοληψίας $p_k = \frac{|A^{(k)}||B_{(k)}|}{\sum_{k'=1}^{n}|A^{(k')}||B_{(k')}|}$ είναι οι βέλτιστες αφού ελαχιστοποιούν την ποσότητα $E[\|AB - CR\|_F^2]$ το οποίο όπως δείχνει το Λήμμα 2.1.2 είναι ένα φυσικό μέτρο του σφάλματος. Θα δούμε ότι οι πιθανότητες $\{p_i\}_{i=1}^{n}$ είναι σχεδόν βέλτιστες αν $p_k \geq \frac{\beta|A^{(k)}||B_{(k)}|}{\sum_{k\prime=1}^{n}|A^{(k\prime)}||B_{(k\prime)}|}$ για $\beta \leq 1$. Τώρα θα αποδείξουμε ότι για σχεδόν βέλτιστες πιθανότητες έχουμε ανάλογα αποτελέσματα του Λήμματος 2.1.2.

Θεώρημα 2.1.1.

Θεωρούμε ότι $A \in R^{mxn}$, $B \in R^{nxp}$, $c \in Z^+$τέτοιο ώστε $1 \leq c \leq n$, *και* $\{p_i\}_{i=1}^{n}$ με $p_i \geq 0$ και $\sum_{i=1}^{n} p_i = 1$ και τέτοια ώστε για κάποια θετική σταθερά $\beta \leq 1$,

$$p_k \geq \frac{\beta|A^{(k)}||B_{(k)}|}{\sum_{k\prime=1}^{n}|A^{(k\prime)}||B_{(k\prime)}|}\,. \tag{2.1.5}$$

Κατασκευάζονται οι πίνακες C και R από τον αλγόριθμο BasicMatrixMultiplication και παίρνουμε το γινόμενο CR ως προσέγγιση του AB. Τότε

$$E[\|AB - CR\|_F^2] \leq \frac{1}{\beta c}\|A\|_F^2\|B\|_F^2. \tag{2.1.6}$$

Για $\delta \in (0,1)$ και $\eta = 1+\sqrt{(8/\beta)\log(1/\delta)}$, τότε με πιθανότητα τουλάχιστον 1-δ,

$$\|AB - CR\|_F^2 \leq \frac{\eta^2}{\beta c}\|A\|_F^2\|B\|_F^2 \tag{2.1.7}$$

Απόδειξη.

Στην ίδια λογική με το Λήμμα 2.1.2, αλλά χρησιμοποιώντας πιθανότητες που να ικανοποιούν τη σχέση (2.1.5), έχουμε

$$E[\|AB-CR\|_F^2] \le \frac{1}{c}\sum_{k=1}^{n}\frac{1}{p_k}\left|A^{(k)}\right|^2\left|B_{(k)}\right|^2$$

$$\le \frac{1}{bc}\left(\sum_{k=1}^{n}\left|A^{(k)}\right|\left|B_{(k)}\right|\right)^2$$

$$\le \frac{1}{bc}\|A\|_F^2\|B\|_F^2,$$

όπου η τελευταία ανισότητα προκύπτει από την ανισότητα του Cauchy-Schwartz.

Στη συνέχεια ορίζουμε τη σχέση

$$\varepsilon 2:\ \|AB-CR\|_F \le \frac{\eta}{\sqrt{\beta c}}\|A\|_F\|B\|_F. \tag{2.1.8}$$

Για να αποδείξουμε το υπόλοιπο θεώρημα αρκεί να αποδείξουμε ότι $Pr[\varepsilon 2] \ge 1-\delta$. Για να το κάνουμε αυτό θεωρούμε ότι οι πίνακες C και R αλλά και το γινόμενό τους $CR = \sum_{t=1}^{c}\frac{1}{cp_{i_t}}A^{i_t}B_{i_t}$, σχηματίζονται επιλέγοντας τυχαία c στοιχεία από το σύνολο $\{1,2,\ldots,n\}$ ανεξάρτητα και με επανατοποθέτηση.

Έστω ότι η ακολουθία των επιλεγμένων στοιχείων είναι $\{i_t\}_{t=1}^{c}$. Εξετάζουμε τη συνάρτηση

$$F(i_1,\ldots,i_c) = \|AB-CR\|_F. \tag{2.1.9}$$

Θα δείξουμε ότι η αλλαγή ενός i_t κάποια στιγμή δεν αλλάξει κατά πολύ την τιμή της F. Αυτό μας επιτρέπει να εφαρμόσουμε μια ανισότητα martingale. Τέλος, θεωρούμε ότι αλλάζουμε ένα από τα i_t σε $i_t^{'}$ ενώ διατηρούμε το άλλο i_t ίδιο.

Στη συνέχεια κατασκευάζουμε τους πίνακες C' και R'. Σημειώνουμε ότι ο πίνακας C' διαφέρει από τον πίνακα C σε μία μόνο στήλη και ο πίνακας R' διαφέρει από τον πίνακα R σε μία μόνο γραμμή. Επίσης,

$$\|CR-C'R'\|_F = \left\|\frac{A^{(i_t)}B_{(i_t)}}{cp_{i_t}} - \frac{A^{(i_t^{'})}B_{(i_t^{'})}}{cp_{i_t^{'}}}\right\|_F \tag{2.1.10}$$

$$\le \frac{1}{cp_{i_t}}\left\|A^{(i_t)}B_{(i_t)}\right\|_F + \frac{1}{cp_{i_t^{'}}}\left\|A^{(i_t^{'})}B_{(i_t^{'})}\right\|_F \tag{2.1.11}$$

$$= \frac{1}{cp_{i_t}}\left|A^{(i_t)}\right|\left|B_{(i_t)}\right| + \frac{1}{cp_{i_t^{'}}}\left|A^{(i_t^{'})}\right|\left|B_{(i_t^{'})}\right| \tag{2.1.12}$$

$$\le \frac{2}{c}\max_a \frac{\left|A^{(a)}\right|\left|B_{(a)}\right|}{P_a}. \tag{2.1.13}$$

Η ισότητα (2.1.10) προκύπτει από την κατασκευή και η (2.1.12) προκύπτει από τη σχέση $\|xy^T\|_F = |x||y|$ για $x \in R^n$ και $y \in R^n$ έτσι χρησιμοποιώντας τις πιθανότητες της σχέσης (2.1.5) και εφαρμόζοντας την ανισότητα Cauchy-Schwarz βλέπουμε ότι

$$\|CR-C'R'\|_F \le \frac{2}{\beta c}\sum_{k=1}^{n}\left|A^{(k)}\right|\left|B_{(k)}\right| \tag{2.1.14}$$

$$\le \frac{2}{\beta c}\|A\|_F\|B\|_F. \tag{2.1.15}$$

Στη συνέχεια χρησιμοποιώντας την τριγωνική ανισότητα έχουμε ότι

$$\|AB - CR\|_F \leq \|AB - C'R'\|_F + \|C'R' - CR\|_F$$

$$\leq \left\|AB - C^{'}R^{'}\right\|_F + \frac{2}{\beta c}\|A\|_F\|B\|_F. \quad (2.1.16)$$

Με παρόμοιο τρόπο προκύπτει η σχέση

$$\|AB - C'R'\|_F \leq \|AB - CR\|_F + \frac{2}{\beta c}\|A\|_F\|B\|_F. \quad (2.1.17)$$

Θέτω $\Delta = \frac{2}{\beta c}\|A\|_F\|B\|_F$ οπότε

$$|F(i_1, \ldots, i_k, \ldots, i_c) - F(i_1, \ldots, i'_k, \ldots, i_c)| \leq \Delta. \quad (2.1.18)$$

Θέτω $\gamma = \sqrt{2c\, log(\frac{1}{\delta})}\Delta$ και από την ανισότητα Hoeffding-Azuma έχουμε ότι

$$Pr[\,\|AB - CR\|_F \geq \frac{1}{\sqrt{\beta c}}\|A\|_F\|B\|_F + \gamma] \leq exp(\frac{-\gamma^2}{2c\Delta^2}) = \delta \quad (2.1.19)$$

και η απόδειξη ολοκληρώθηκε. ■

Μια άμεση συνέπεια του Θεωρήματος 2.1.1 είναι ότι επιλέγοντας αρκετά ζεύγη στηλών-γραμμών το σφάλμα της προσέγγισης του πολλαπλασιασμού πινάκων μπορεί να γίνει αυθαίρετα μικρό. Συγκεκριμένα αν $c \geq 1/\beta\varepsilon^2$ με τη χρήση της ανισότητας του Jensen έχουμε ότι

$$E[\|AB - CR\|_F] \leq \varepsilon\|A\|_F\|B\|_F \quad (2.1.20)$$

και αν $c \geq \eta^2/\beta\varepsilon^2$ τότε με πιθανότητα τουλάχιστον 1-δ, έχουμε ότι

$$\|AB - CR\|_F \leq \varepsilon\|A\|_F\|B\|_F. \quad (2.1.21)$$

Σε ορισμένες εφαρμογές [5,6] μπορεί κάποιος να ενδιαφέρεται για την εφαρμογή του Θεωρήματος 2.1.1 στην περίπτωση που B=A^T δηλαδή να ενδιαφέρεται για την προσέγγιση $\|AA^T - CC^T\|_F^2$.

Σε αυτή την περίπτωση οι βέλτιστες πιθανότητες είναι οι $p_k \geq \frac{\beta|A^{(k)}|^2}{\|A\|_F^2}$ για κάποια θετική σταθερά $\beta \leq 1$ και παρουσιάζουμε το επόμενο θεώρημα σαν συμπέρασμα του Θεωρήματος 2.1.1 .

Θεώρημα 2.1.2.

Έστω $A \in R^{mxn}$, $c \in Z^+$, $1 \leq c \leq n$, *και* $\{p_i\}_{i=1}^n$ τέτοιο ώστε $\sum_{i=1}^n p_i = 1$ και τέτοιο ώστε $p_k \geq \frac{\beta|A^{(k)}|^2}{\|A\|_F^2}$ για κάποια θετική σταθερά $\beta \leq 1$. Επιπλέον, έχουμε $\delta \in (0,1)$ και $\eta = 1 + \sqrt{\left(\frac{8}{\beta}\right)\log(\frac{1}{\delta})}$. Κατασκευάζεται ο πίνακας C με τον αλγόριθμο BasicMatrixMultiplication και παίρνουμε το CC^T σαν προσέγγιση του AA^T. Τότε

$$E[\|AA^T - CC^T\|_F] \leq \frac{1}{\sqrt{\beta c}}\|A\|_F^2, \quad (2.1.22)$$

και με πιθανότητα τουλάχιστον 1-δ, έχουμε

$$\|AA^T - CC^T\|_F \leq \frac{\eta}{\sqrt{\beta c}}\|A\|_F^2. \quad (2.1.23)$$

2.1.5 Παραδείγματα

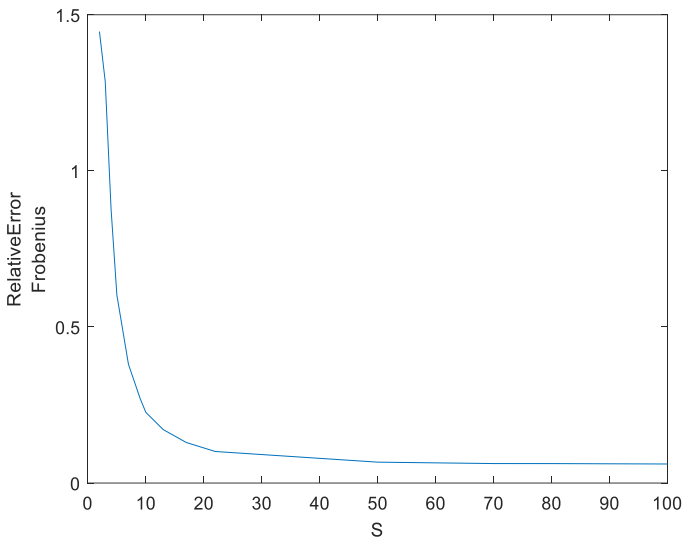
ΠΑΡΑΔΕΙΓΜΑ 1

c=10 (αριθμός στηλών που επιλέγονται), A =200x300, B=300x400

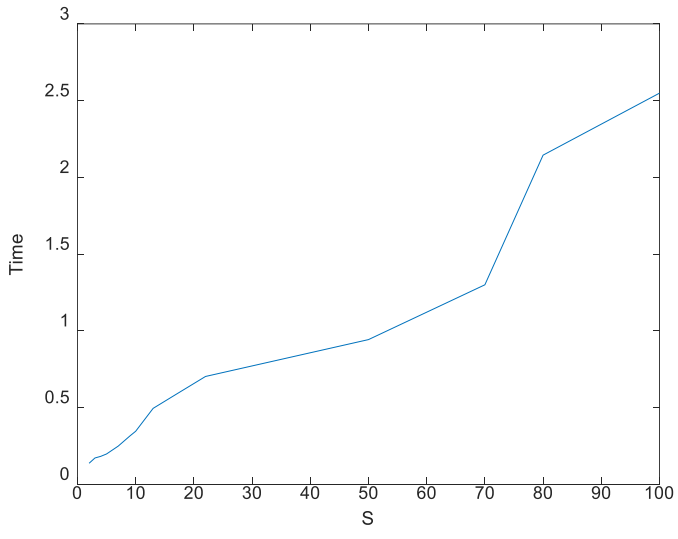
S (επαναλήψεις)	**Σχετικό σφάλμα** $\|AB - CR\|_F / \|AB\|_F$	**Σχετικό σφάλμα** $\|AB - CR\|_2 / \|AB\|_2$	**t (s)** (χρόνος)
2	1.4463	1.4455	0.0872
3	1.2854	1.2849	0.1144
4	0.8783	0.8780	0.1152
5	0.6027	0.6021	0.1261
7	0.3805	0.3796	0.1326
9	0.2717	0.2703	0.1490
10	0.2264	0.2247	0.1554
13	0.1712	0.1688	0.1760
17	0.1297	0.1256	0.2076
22	0.1010	0.0956	0.3327
50	0.0669	0.0516	0.4638
70	0.0624	0.0429	0.6337
80	0.0623	0.0418	0.6782
100	0.0610	0.0413	0.7685

Πίνακας 2.1.1.

Παρουσιάζονται τα για c=10 και A =200x300, B=300x400 τα σχετικά σφάλματα και ο χρόνος του αλγορίθμου.



Διάγραμμα1. S- σχετικό σφάλμα R.



Διάγραμμα 2. S-t.

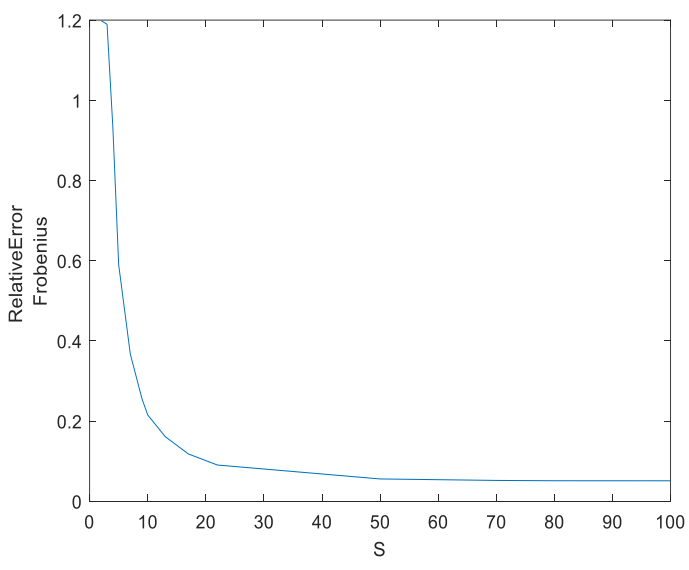
ΠΑΡΑΔΕΙΓΜΑ 2

c=80, A =200x300 , B=300x400

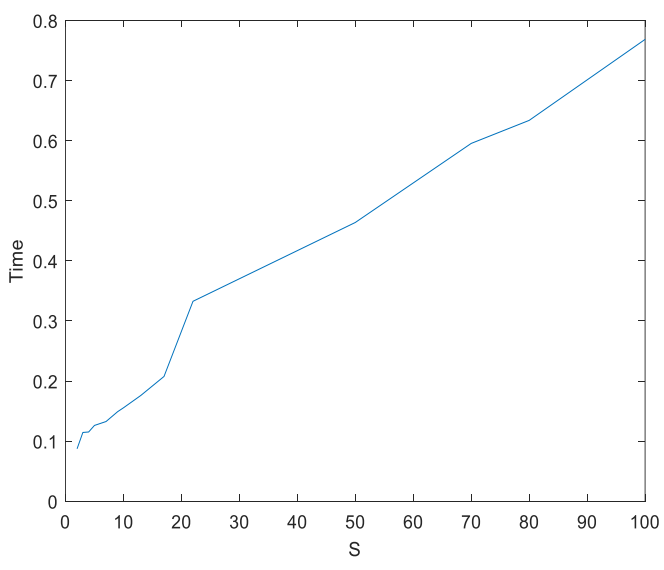
S (επαναλήψεις)	**Σχετικό σφάλμα** $\|AB - CR\|_F / \|AB\|_F$	**Σχετικό σφάλμα** $\|AB - CR\|_2 / \|AB\|_2$	**t (s)** (χρόνος)
2	1.1986	1.1932	0.1359
3	1.1896	1.1827	0.1704
4	0.7322	0.6483	0.1811
5	0.5899	0.5897	0.1967
7	0.3673	0.3669	0.2478
9	0.2560	0.2553	0.3137
10	0.2151	0.2143	0.3446
13	0.1612	0.1599	0.4939
17	0.1178	0.1157	0.5866
22	0.0902	0.0869	0.7018
30	0.0599	0.0522	0.8382
50	0.0555	0.0439	0.9421
70	0.0518	0.0358	1.2993
80	0.0510	0.0355	1.7327
100	0.0509	0.0349	2.5486

Πίνακας 2.1.2.

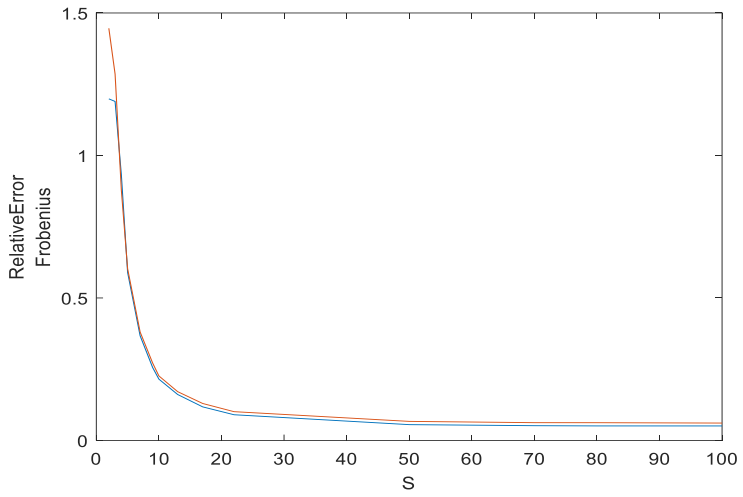
Παρουσιάζονται για c=80 και A =200x300, B=300x400 τα σχετικά σφάλματα και ο χρόνος του αλγορίθμου.



Διάγραμμα 3. S- σχετικό σφάλμα Frobenius.



Διάγραμμα 4. S-t.



Διάγραμμα 5. S- σχετικό σφάλμα Frobenius για c=10 και c=80.

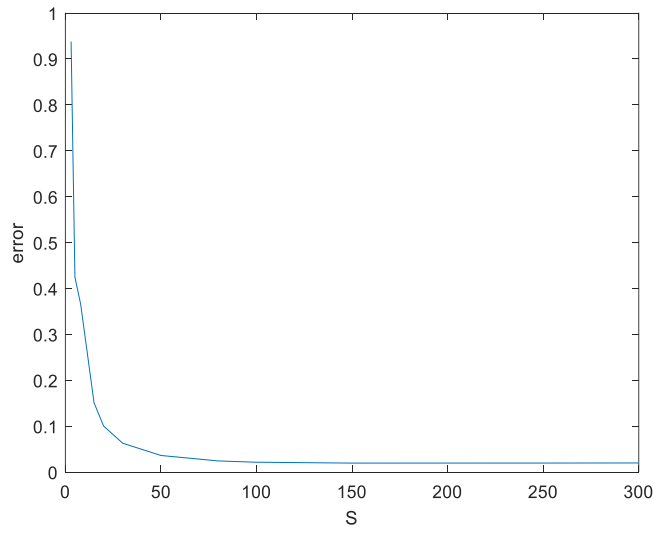
ΠΑΡΑΔΕΙΓΜΑ 3

c=500 , A =1.000x2.000 , B=2.000x2.500

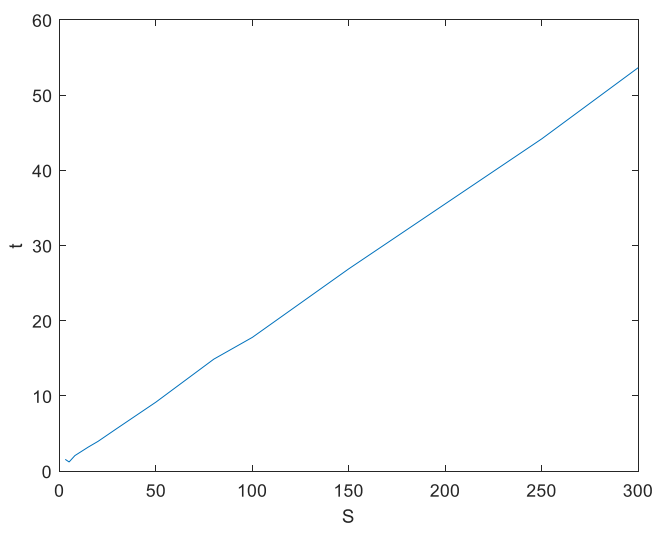
S (επαναλήψεις)	**Σχετικό σφάλμα** $\|AB - CR\|_F / \|AB\|_F$	**t (s)** (χρόνος)
3	0.9379	1.5649
5	0.4247	1.2188
8	0.3659	2.0672
15	0.1511	3.2100
20	0.1004	3.9604
30	0.0629	5.7120
50	0.0362	9.1775
80	0.0243	14.8879
100	0.0215	17.7876
150	0.0196	26.9059
250	0.0197	44.1804
300	0.0199	53.6439

Πίνακας 2.1.3.

Παρουσιάζονται τα για c=500 και A =1000x2000, B=2000x2500 τα σχετικά σφάλματα και ο χρόνος του αλγορίθμου.



Διάγραμμα 6. S- σχετικό σφάλμα Frobenius.



Διάγραμμα 7. S-t.

Στα παραδείγματα φαίνεται πως όσο αυξάνονται οι επαναλήψεις του αλγορίθμου, το σχετικό σφάλμα της Frobenius και της Ευκλείδειας νόρμας μειώνεται. Φαίνεται επίσης και από τα διαγράμματα ότι για μικρό αριθμό επαναλήψεων αυτή η μείωση είναι αρκετά αισθητή ενώ όσο αυξάνονται οι επαναλήψεις το σφάλμα τείνει να σταθεροποιηθεί. Ενώ λοιπόν στην αρχή είχαμε να πολλαπλασιάσουμε δυο πίνακες A και B με διαστάσεις 1.000x2.000 και 2.000x2.500 αντίστοιχα (Παράδειγμα 3) , τελικά οι πίνακες C και R που δημιουργήσαμε , είχανε διαστάσεις 1.000x500 και 500x2.500 . Κάνοντας δηλαδή πολύ λιγότερες πράξεις, ο αλγόριθμος προσεγγίζει αρκετά καλά τον πολλαπλασιασμό των αρχικών πινάκων.

2.2. ΑΛΓΟΡΙΘΜΟΣ ΠΟΛΛΑΠΛΑΣΙΑΣΜΟΥ ΠΙΝΑΚΩΝ ELEMENT WISE

Στην ενότητα αυτή θα περιγράψουμε τον αλγόριθμο ElementWiseMatrix για να προσεγγίσουμε το γινόμενο δύο πινάκων. Θα αναλύσουμε τον αλγόριθμο και τα όρια σφάλματος λαμβάνοντας υπόψιν τη Frobenius νόρμα και την Ευκλείδεια νόρμα.

Element Wise matrix multiplication algorithm

Είσοδος: $A \in R^{nxm}$, $B \in R^{nxp}$, $\{p_{ij}\}_{i,j=1}^{m,n}$ τέτοια ώστε

$0 \le p_{ij} \le 1$ και $\{q_{ij}\}_{i,j=1}^{n,p}$ τέτοιο ώστε $0 \le q_{ij} \le 1$.

Έξοδος: $S \in R^{mxn}$ και $R \in R^{nxp}$.

Αλγόριθμος:

1. Για i =1,…,m και j =1,…,n, ανεξάρτητα,

(a) θέτουμε

$$S_{ij} = \begin{cases} \frac{A_{ij}}{p_{ij}} & με\ πιθανότητα\ p_{ij} \\ 0 & αλλιώς \end{cases}.$$

2. Για i =1,…,n και j =1,…,p, ανεξάρτητα,

(a) θέτω

$$R_{ij} = \begin{cases} \frac{B_{ij}}{q_{ij}} & με\ πιθανότητα\ q_{ij} \\ 0 & αλλιώς \end{cases}$$

3.Return S, R.

Ο αλγόριθμος αυτός δέχεται σαν είσοδο δύο πίνακες $A \in R^{mxn}$ και $B \in R^{nxp}$, και δημιουργεί δύο άλλους πίνακες $S \in R^{mxn}$ και $R \in R^{nxp}$. Ο αλγόριθμος διατηρεί μερικά στοιχεία των πινάκων A και B κανονικοποιώντας με κατάλληλο τρόπο τα στοιχεία αυτά και μηδενίζοντας τα υπόλοιπα. Το γινόμενο SR των πινάκων που επιστρέφονται είναι μια προσέγγιση του γινομένου AB των αρχικών πινάκων. Να σημειώσουμε εδώ ότι από τη στιγμή που οι πίνακες S και R σχηματίζονται ανεξάρτητα ο ένας από τον άλλον, ο αλγόριθμος δεν διατηρεί τα αντίστοιχα στοιχεία γιατί κάτι τέτοιο θα δημιουργούσε εξάρτηση και θα έκανε πιο περίπλοκη την ανάλυση.

Ο αλγόριθμος ElementWiseMatrix Multiplication μπορεί να εφαρμοστεί για μη ομοιόμορφες πιθανότητες. Αυτός ο αλγόριθμος διαφέρει από τον βασικό αλγόριθμο πολλαπλασιασμού πινάκων (BasicMatrixMultiplication) στο ότι έχουμε έναν αναμενόμενο αριθμό στοιχείων άρα έχουμε μία αναμενόμενη χωρική και χρονική πολυπλοκότητα. Στον παρακάτω αλγόριθμο δίνουμε σε κάθε στοιχείο των πινάκων A και B μία πιθανότητα και δημιουργούμε έτσι δύο πίνακες πιθανοτήτων P και $P1$ αντίστοιχα. Από τον αρχικό πίνακα A διατηρούμε τα στοιχεία που έχουν μεγαλύτερη πιθανότητα από τη μέση τιμή του πίνακα των πιθανοτήτων P, στον οποίο όπως θα δούμε παρακάτω κάθε στοιχείο είναι ίσο με $p_{i,j} = min\{1, lA_{ij}^2/\|A\|_F^2\}$.

(Αντίστοιχα για τον πίνακα B διατηρούμε τα στοιχεία που έχουν μεγαλύτερη πιθανότητα από τον πίνακα $P1$, στον οποίο κάθε στοιχείο είναι ίσο με $q_{i,j} = \min\{1, l'B_{ij}^2/\|B\|_F^2\}$.

Τα στοιχεία που δεν διατηρούνται γίνονται μηδενικά.

Αλγόριθμος πολλαπλασιασμού πινάκων ElementWise matlab

```
m=100;
n=200;
p=300;
A=randi(100,m,n);
B=randi(100,n,p);
CC=zeros(m,n);  % The initial CC.
RR=zeros(n,p);  % The initial RR.
R=zeros(n,p);
C=zeros(m,n);
WW=zeros(p,n);
 tic
%for xi=1:100  % We repeat S
times the process.
            % Symbol xi is not
used elsewhere.

for i=1:m
 for j=1:n
  A2(i,j)=A(i,j)^2;
 end
end

 for i=1:n
 for j=1:p
  B2(i,j)=B(i,j)^2;
 end
end


AF=norm(A,'fro');
BF=norm(B,'fro');
l=1;

%pinakas S
sum=0;
for i=1:m
for j=1:n
   sum=A2(i,j)+sum;
end
end
 for i=1:m
     for j=1:n

P(i,j)=min(1,l*A2(i,j)/AF^2);
     end
 end
 x=mean(P);
 xx=mean(mean(P));
for i=1:m
    for j=1:n
        if P(i,j)>xx
            S(i,j)=A(i,j);
%/P(i,j);
        else
            S(i,j)=0;
        end
    end
end
%pinakas R

sum=0;
for i=1:n
for j=1:p
   sum=B2(i,j)+sum;
end
end
 for i=1:n
     for j=1:p

P1(i,j)=min(1,l*B2(i,j)/BF^2);
     end
 end
 x1=mean(P1);
 xx1=mean(mean(P1));
for i=1:n
    for j=1:p
        if P1(i,j)>xx1
            R(i,j)=B(i,j);
%/P1(i,j);
        else
            R(i,j)=0;
        end
    end
end

 Relative_ErrorF=norm(A*B-
S*R,'fro')/norm(A*B,'fro')
Relative_Error2=norm(A*B-
S*R)/norm(A*B)
```

Τα αποτελέσματα του αλγορίθμου για το σχετικό σφάλμα της Frobenius και της Ευκλείδιας νόρμας είναι :

Relative_ErrorF = 0.5612

Relative_Error2 = 0.5606

2.2.1. Ανάλυση του αλγορίθμου

Λήμμα 2.2.1.

Έστω $A \in R^{nxm}$, $B \in R^{nxp}$, $l, l' \in Z^+$ και έστω $p_{i,j} = min\{1, lA_{ij}^2/\|A\|_F^2\}$ και $q_{i,j} = min\{1, l'B_{ij}^2/\|B\|_F^2\}$.

Κατασκευάζουμε τους πίνακες S και R με τον αλγόριθμο πολλαπλασιασμού element wise και υπολογίζουμε το γινόμενο SR σαν μια προσέγγιση του γινομένου των αρχικών πινάκων A και B. Έπειτα, $\forall\, i, j$,

$$E\big[(SR)_{ij}\big] = (AB)_{ij},$$

$$Var\big[(SR)_{ij}\big] = \sum_{k=1}^{n} \frac{A_{ik}^2}{p_{ik}} \frac{B_{kj}^2}{q_{kj}} - \sum_{k=1}^{n} A_{ik}^2 B_{kj}^2,$$

$$E[\|AB - SR\|_F^2] \geq \frac{mpn}{ll'} \|A\|_F^2 \|B\|_F^2 - \sum_{k=1}^{n} \left|A^{(k)}\right|^2 \left|B_{(k)}\right|^2. \qquad (2.2.1)$$

Σημειώνουμε ότι τα l και l' διαλέγονται έτσι ώστε να μην διατηρούνται περισσότερα από l και l' στοιχεία των πινάκων A και B αντίστοιχα.

Απόδειξη.

Για κάθε k έχουμε ότι $S_{ik} = \frac{A_{ik}}{p_{ik}}$ με πιθανότητα p_{ik} και S_{ik}=0 με πιθανότητα 1-p_{ik} και $E[S_{ik}] = A_{ik}$,$E[R_{kj}] = B_{kj}$. Επίσης, επειδή οι πίνακες S και R δημιουργούνται ανεξάρτητα, έχουμε ότι

$$E[(SR)_{ij}] = E[\sum_{k=1}^{n} S_{ik}\, R_{kj}] = \sum_{k=1}^{n} E[S_{ik}]\, E\big[R_{kj}\big] = (AB)_{ij}$$

Αφού

$$Var[(SR)_{ij}] = E[(SR)_{ij}^2] - E[(SR)_{ij}]^2 \quad \text{και} \quad (SR)_{ij} = \sum_{k=1}^{n} S_{ik}\, R_{kj},$$

έχουμε ότι

$$\begin{aligned} Var[(SR)_{i,j}] &= \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} E[S_{ik_1} R_{k_1 j} S_{ik_2} R_{k_2 j}] - E[(SR)_{ij}]^2 \\ &= \sum_{k=1}^{n} E[S_{ik}^2] E[R_{kj}^2] + \sum_{k_1=1}^{n} \sum_{k_2 \neq k_1} E[S_{ik_1}] E[R_{k_1 j}] E[S_{ik_2}] E[R_{k_2 j}] - (AB)_{ij}^2 \\ &= \sum_{k=1}^{n} \frac{A_{ik}^2}{p_{ik}} \frac{B_{kj}^2}{q_{kj}} + \sum_{k_1=1}^{n} \sum_{k_2 \neq k_1} A_{ik_1} B_{k_1 j} A_{ik_2} B_{k_2 j} - (AB)_{ij}^2 \\ &= \sum_{k=1}^{n} \frac{A_{ik}^2}{p_{ik}} \frac{B_{kj}^2}{q_{kj}} - \sum_{k=1}^{n} A_{ik}^2 B_{kj}^2. \end{aligned}$$

Επίσης, αφού $E[\|AB - SR\|_F^2] = \sum_{i=1}^{m} \sum_{j=1}^{p} Var[(SR)_{ij}]$, και αφού για τις πιθανότητες p_{ij} και q_{ij} ισχύει ότι $\frac{1}{p_{ik}} \geq \frac{\|A\|_F^2}{lA_{ik}^2}$ και $\frac{1}{q_{kj}} \geq \frac{\|B\|_F^2}{l'B_{kj}^2}$, έχουμε ότι

$$E[\|AB - SR\|_F^2] = \sum_{i=1}^{m} \sum_{j=1}^{p} \sum_{k=1}^{n} \frac{A_{ik}^2}{p_{ik}} \frac{B_{kj}^2}{q_{kj}} - \sum_{i=1}^{m} \sum_{j=1}^{p} \sum_{k=1}^{n} A_{ik}^2\, B_{kj}^2$$

$$\geq \sum_{i,j=1}^{m,p} \sum_{k=1}^{n} \frac{\|A\|_F^2 \|B\|_F^2}{ll'} - \sum_{k=1}^{n} \left|A^{(k)}\right|^2 \left|B_{(k)}\right|^2. \qquad \blacksquare$$

Για να αποδείξουμε το Θεώρημα 2.2.2 το οποίο μας δίνει το όριο για το $\|AB - SR\|_2$, θα χρησιμοποιήσουμε το επόμενο θεώρημα.

Θεώρημα 2.2.1.

Έστω ένας *n*xn πίνακας *A*, και έστω $\hat{A}$ ένας τυχαίος πίνακας του οποίου οι καταχωρήσεις είναι ανεξάρτητες τυχαίες μεταβλητές έτσι ώστε $E[\hat{A}_{ij}] = A_{ij}$, $Var[\hat{A}_{ij}] \leq \sigma^2$ και

$$\left|\hat{A}_{i,j} - A_{i,j}\right| \leq \frac{\sigma\sqrt{2n}}{log^3(2n)} . \tag{2.2.2}$$

Για κάθε $n \geq 10$ με πιθανότητα τουλάχιστον 1-1/(2n) έχουμε ότι

$$\left\|A - \hat{A}\right\|_2 \leq 7\sigma\sqrt{2n} . \tag{2.2.3}$$

Πριν προχωρήσουμε στο κύριο αποτέλεσμα χρειάζεται να σταθούμε σε ένα τεχνικό ζήτημα. Η κατασκευή των πινάκων *S* και *R* με τον αλγόριθμο Element Wise μπορεί να θεωρηθεί σαν προσθήκη προσεκτικά κατασκευασμένων τυχαίων πινάκων *E* και *D* έτσι ώστε *S*=*A*+*E* και *R*=*B*+*D* [1, 2].

Θα δούμε παρακάτω ότι αν μπορέσουμε να φράξουμε τις ποσότητες $\|E\|_2$ και $\|D\|_2$ τότε μπορούμε να βρούμε εύκολα ένα όριο και για την ποσότητα $\|AB - CR\|_2$. Με δεδομένο ότι θα εφαρμόσουμε το Θεώρημα 2.2.1 για να πετύχουμε αυτά τα όρια θα πρέπει να ικανοποιείται η σχέση (2.2.2). Κάνοντας τη δειγματοληψία με μια μη ομοιόμορφη κατανομή πιθανότητας του Λήμματος 2.2.1 θα υπήρχε περίπτωση να παραβιαστεί αυτός ο κανόνας επειδή στην περίπτωση που διατηρηθεί ένα μικρό στοιχείο, το στοιχείο $S_{i,j} = A_{i,j}/p_{i,j}$ που προκύπτει, θα είναι πολύ μεγάλο. Για το λόγο αυτό τροποποιούμε λίγο τις πιθανότητες δειγματοληψίας [1] ώστε τα μικρά στοιχεία να διατηρούνται με μια ελαφρώς μεγαλύτερη πιθανότητα

$$p_{ij} = \begin{cases} min\{1, lA_{ij}^2/\|A\|_F^2\} & \text{αν } \left|A_{ij}\right| \geq \frac{\|A\|_F\, log^3(2n)}{\sqrt{2nl}} \\ min\{1, \frac{\sqrt{l}\,|A_{ij}|\, log^3(2n)}{\sqrt{2n\|A\|}_F}\} & \text{διαφορετικά} \end{cases} , \tag{2.2.4}$$

$$q_{ij} = \begin{cases} min\{1, l'B_{ij}^2/\|B\|_F^2\} & \text{αν } \left|B_{ij}\right| > \frac{\|B\|_F\, log^3(2n)}{\sqrt{2nl'}} \\ min\{1, \frac{\sqrt{l'}|B_{ij}|\, log^3(2n)}{\sqrt{2n}\|B\|_F}\} & \text{διαφορετικά} \end{cases} . \tag{2.2.5}$$

Θεώρημα 2.2.2.

Έστω δύο πίνακες $A \in R^{nxm}$ και $B \in R^{nxp}$ και παίρνουμε τις πιθανότητες p_{ij} , q_{ij} όπως τις ορίσαμε στις σχέσεις (2.2.4) και (2.2.5) αντίστοιχα με $l = l' \geq 1$.

Θεωρούμε ότι $l \leq \|A\|_F^2/\, max_{i,j}\, A_{ij}^2$ και $l \leq \|B\|_F^2/\, max_{i,j}\, B_{ij}^2$. Θεωρούμε επίσης ότι *m*=*n*=*p* και ότι το n είναι αρκετά μεγάλο έτσι ώστε $2n \geq log^6(2n)$. Κατασκευάζουμε τους πίνακες *S* και *R* με τον αλγόριθμο Element Wise και παίρνουμε το γινόμενο τον πινάκων *SR* σαν μια προσέγγιση του γινομένου AB. Έπειτα με πιθανότητα τουλάχιστον 1-1/n έχουμε ότι

$$\|AB - SR\|_2 \leq \left(20\sqrt{\frac{n}{l}} + \frac{100n}{l}\right)\|A\|_F\|B\|_F. \tag{2.2.6}$$

Απόδειξη.

Λόγω των παραδοχών για τα n και l ούτε το p_{ij} ούτε το q_{ij} δεν ξεπερνάει το *1* για κάποιο *i,j*. Έστω *E*=*S*-*A* και *D*=*R*-*B*. Τότε έχουμε

$$SR = (A+E)\,(B+D) = AB+EB+AD+ED. \tag{2.2.7}$$

Από την τριγωνική ανισότητα έχουμε ότι

$$\|AB - SR\|_2 \leq \|A\|_2\|D\|_2 + \|E\|_2\|B\|_2 + \|E\|_2\|D\|_2. \tag{2.2.8}$$

Προκειμένου να εφαρμόσουμε το Θεώρημα 2.2.1 για τα $\|E\|_2$ και $\|D\|_2$, πρώτα επαληθεύουμε ότι οι υποθέσεις του θεωρήματος ικανοποιούνται. Από την απόδειξη του Λήμματος 2.2.1, έχουμε ότι $E[S_{i,j}] = A_{i,j}$.

Στη συνέχεια, βλέπουμε ότι

$$Var[S_{i,j}] \leq E[S_{i,j}^2] = \frac{A_{i,j}^2}{p_{i,j}} \leq \frac{\|A\|_F^2}{l}$$

ανεξάρτητα αν το $|A_{i,j}|$ είναι μεγαλύτερο ή μικρότερο από το όριο. Με παρόμοιο τρόπο βρίσκουμε ότι $E[R_{i,j}] = B_{i,j}$ και $Var[D_{i,j}] \leq \frac{\|B\|_F^2}{l}$. Είναι εύκολο να δείξουμε ότι ανεξάρτητα από το αν το $|A_{i,j}|$ είναι μεγαλύτερο ή μικρότερο από το όριο και ανεξάρτητα από το αν το $S_{i,j} = 0$ ή $S_{i,j} = A_{i,j}/p_{i,j}$ ισχύει ότι

$$|A_{i,j} - S_{i,j}| \leq \frac{\|A\|_F \sqrt{2n}}{\sqrt{l}\, log^3(2n)}\,. \tag{2.2.9}$$

Παρόμοια βρίσκουμε ότι

$$|B_{i,j} - R_{i,j}| \leq \frac{\|B\|_F \sqrt{2n}}{\sqrt{l}\, log^3(2n)}\,. \tag{2.2.10}$$

Έτσι οι υποθέσεις του Θεωρήματος 2.2.1 επαληθεύονται και με πιθανότητα τουλάχιστον 1-1/2n ισχύουν τα επόμενα

$$\|E\|_2 \leq 7\|A\|_F \sqrt{2n}/\sqrt{l}\,, \tag{2.2.11}$$

$$\|D\|_2 \leq 7\|B\|_F \sqrt{2n}/\sqrt{l}\,. \tag{2.2.12}$$

Συνδυάζοντας τις σχέσεις (2.2.11) και (2.2.12) με τη σχέση (2.2.8) και αφού $\|\cdot\|_2 \leq \|\cdot\|_F$ έχουμε ότι

$$\begin{aligned}\|AB - SR\|_2 &\leq \|A\|_2\|D\|_2 + \|E\|_2\|B\|_2 + \|E\|_2\|D\|_2 \\ &\leq \frac{7\sqrt{2n}\|A\|_F\|B\|_F}{\sqrt{l}} + \frac{7\sqrt{2n}\|A\|_F\|B\|_F}{\sqrt{l}} + \frac{98n\|A\|_F\|B\|_F}{l} \\ &\leq (20\sqrt{n/l} + 100n/l)\|A\|_F\|B\|_F.\end{aligned}$$

■

Σημειώνουμε ότι εάν θέσουμε *l=cn* στο Θεώρημα 2.2.2, τότε η σχέση (2.2.6) γίνεται

$$\|AB - SR\|_2 \leq (\frac{20}{\sqrt{c}} + \frac{100}{c})\|A\|_F\|B\|_F = O(1/\sqrt{c})\|A\|_F\|B\|_F.$$

Η σύγκριση με τη σχέση (2.1.7) του Θεωρήματος 2.1.1 (αφού $\|\cdot\|_2 \leq \|\cdot\|_F$), παρατηρούμε ότι οι δύο αλγόριθμοι έχουν παρόμοιο σφάλμα ως προς την Ευκλείδεια νόρμα.

2.2.2. Πίνακας αποτελεσμάτων για διάφορες πιθανότητες

Χρειάζεται να δώσουμε έμφαση στο γεγονός ότι στην περίπτωση δειγματοληψίας με μη ομοιόμορφες πιθανότητες η δειγματοληψία μας μπορεί να θεωρηθεί ως ένας αλγόριθμος που κάνει δύο περάσματα πάνω από τα δεδομένα. Στο πρώτο πέρασμα ο αλγόριθμος διαβάζει τον πίνακα, τότε αποφασίζει ποιες στήλες και σειρές πρέπει να κρατήσει, και στη συνέχεια στο δεύτερο πέρασμα εξάγει αυτές τις στήλες και σειρές. Σε ορισμένες εφαρμογές όμως επιτρέπεται μόνο ένα πέρασμα και έτσι το να κάνει ο αλγόριθμος δύο περάσματα πάνω από τα δεδομένα είναι απαγορευτικό [8]. Στις περιπτώσεις αυτές μπορούμε να εκτελούμε ομοιόμορφη δειγματοληψία αν τα ζεύγη στήλης-σειράς είναι περίπου στο ίδιο μέγεθος . Για παράδειγμα τα $\left|A^{(k)}\right|$ $\left|B_{(k)}\right|$ να είναι κοντά στη μέση τιμή τους.

	$E[\|AB-CR\|_F] \leq$	$\|AB-CR\|_F \leq$	Σχόλια
$p_k \geq \frac{\beta\left\|A^{(k)}\right\|\left\|B_{(k)}\right\|}{\sum_{k'}\left\|A^{(k')}\right\|\left\|B_{(k')}\right\|}$	$\frac{1}{\sqrt{\beta c}}\|A\|_F\|B\|_F$	$\frac{\eta}{\sqrt{\beta c}}\|A\|_F\|B\|_F$	$\eta=1+\sqrt{\frac{8}{\beta}log(\frac{1}{6})}$
$p_k \geq \frac{\beta\left\|A^{(k)}\right\|^2}{\|A\|_F^2}$	$\frac{1}{\sqrt{\beta c}}\|A\|_F\|B\|_F$	$\frac{\eta}{\sqrt{\beta c}}\|A\|_F\|B\|_F$	$\eta=1+\frac{\|A\|_F}{\|B\|_F}M\sqrt{\frac{8}{\beta}log(\frac{1}{6})}$ $M=max_a\left\|\frac{B_{(a)}}{A^{(a)}}\right\|$
$p_k \geq \frac{\beta\left\|B^{(k)}\right\|^2}{\|B\|_F^2}$	$\frac{1}{\sqrt{\beta c}}\|A\|_F\|B\|_F$	$\frac{\eta}{\sqrt{\beta c}}\|A\|_F\|B\|_F$	$\eta=1+\frac{\|A\|_F}{\|B\|_F}M\sqrt{\frac{8}{\beta}log(\frac{1}{6})}$ $M=max_a\left\|\frac{A^{(a)}}{B_{(a)}}\right\|$
$p_k \geq \frac{\beta\left\|A^{(k)}\right\|}{\sum_{k'=1}^{n}\left\|A^{(k')}\right\|}$	$\frac{1}{\sqrt{\beta c}}\|A\|_F\sqrt{n}M$	$\frac{\eta}{\sqrt{\beta c}}\|A\|_F\sqrt{n}M$	$\eta=1+\sqrt{\frac{8}{\beta}log(\frac{1}{6})}$ $M=max_a\left\|B_{(a)}\right\|$
$p_k \geq \frac{\beta\left\|B_{(k)}\right\|}{\sum_{k'=1}^{n}\left\|B_{(k')}\right\|}$	$\frac{1}{\sqrt{\beta c}}\sqrt{n}M\|B\|_F$	$\frac{\eta}{\sqrt{\beta c}}\sqrt{n}M\|B\|_F$	$\eta=1+\sqrt{\frac{8}{\beta}log(\frac{1}{6})}$ $M=max_a\left\|A^{(a)}\right\|$
$p_k \geq \frac{\beta\left\|A^{(k)}\right\|\left\|B_{(k)}\right\|}{\|A\|_F\|B\|_F}$	$\frac{1}{\sqrt{\beta c}}\|A\|_F\|B\|_F$	$\frac{\eta}{\sqrt{\beta c}}\|A\|_F\|B\|_F$	$\eta=1+\sqrt{\frac{8}{\beta}log(\frac{1}{6})}$
$p_k=\frac{1}{n}$			

ΚΕΦΑΛΑΙΟ 3. ΥΠΟΛΟΓΙΣΜΟΣ ΜΙΑΣ LOW-RANK ΠΡΟΣΕΓΓΙΣΗΣ ΠΙΝΑΚΑ

Οι αλγόριθμοι που παρουσιάζεται σε αυτό το κεφάλαιο χρησιμεύουν στο να μπορούμε να κάνουμε υπολογισμούς σε μεγάλους πίνακες.

Υπάρχουν πολλές εφαρμογές στις οποίες έχουμε μια καλή προσέγγιση των δεδομένων χρησιμοποιώντας έναν πίνακα μικρού βαθμού. Τέτοιες εφαρμογές καλύπτουν ένα ευρύ φάσμα διαχείρισης δεδομένων και χωρίς να θέλουμε να επεκταθούμε παραπάνω αναφέρουμε ενδεικτικά δύο

Παραδείγματος χάρη η εφαρμογή LSI (Latent Semantic Indexing) είναι μία μέθοδος που αντιμετωπίζει προβλήματα που σχετίζονται με λεξιλογική συσχέτιση [16, 17, 25].

Ένα άλλο παράδειγμα είναι η εφαρμογή DNA Microarray Technology η οποία χρησιμοποιείται για να μελετήσουμε μια ποικιλία από βιολογικά προτσές [15, 19, 25].

3.1. Προσεγγιστικός αλγόριθμος SVD γραμμικού χρόνου

Linear time svd algorithm

Είσοδος: $A \in R^{mxn}, c, k \in Z^+$ τέτοια ώστε $1 \leq k \leq c \leq n$, $\{p_i\}_{i=1}^n$ έτσι ώστε $p_i \geq 0$ και $\sum_{i=1}^n p_i = 1$.

Έξοδος: $H_k \in R^{mxk}$ και $\sigma_t(C),\ t = 1,\ldots,k.$

1. Για *t=1* μέχρι *c*,

(a) Διάλεξε $i_t \in 1,\ldots,n$ με πιθανότητα $Pr[\,i_t = a] = p_a$, α=1,…,*n*.

(b) Έστω $C^{(t)} = \frac{A^{(i_t)}}{\sqrt{cp_{i_t}}}$.

2. Υπολόγισε τον πίνακα $C^T C$ και το SVD του $C^T C = \sum_{t=1}^c \sigma_t^2(C) y^t y^{t^T}$.

3. Υπολόγισε τον πίνακα $H_k = Cy^t/\sigma_t(C)$ για t=1,…,*k*.

Παίρνοντας έναν πίνακα $A \in R^{mxn}$ θέλουμε να προσεγγίσουμε τις k ιδιάζουσες τιμές και τα αντίστοιχα ιδιάζοντα διανύσματα του, σε έναν σταθερό αριθμό περασμάτων πάνω από τα δεδομένα με χωρική πολυπλοκότητα O(cm+c^2) και χρονική πολυπλοκότητα O($c^2m + c^3$).

Η λογική του αλγόριθμου SVD γραμμικού χρόνου είναι ότι διαλέγουμε c στήλες του πίνακα A και διαιρούμε την κάθε μια με τον κατάλληλο παράγοντα για να φτιάξουμε τον πίνακα $C \in R^{mxc}$και μετά υπολογίζουμε τις ιδιάζουσες τιμές και τα αντίστοιχα ιδιάζοντα διανύσματα του πίνακα C τα οποία, όπως θα δούμε και αργότερα, θα αποτελέσουν μια προσέγγιση των ιδιαζουσών τιμών και των αντίστοιχων αριστερών ιδιαζόντων διανυσμάτων του πίνακα A.

Αυτά υπολογίζονται με την εκτέλεση της SVD του πίνακα $C^T C$ για να υπολογίσουμε τα δεξιά ιδιάζοντα διανύσματα του πίνακα C και από αυτά έπειτα να υπολογίσουμε τα αριστερά ιδιάζοντα διανύσματα του C.

Ο αλγόριθμος που περιγράψαμε δέχεται σαν είσοδο έναν πίνακα A και επιστρέφει σαν έξοδο μια προσέγγιση των πρώτων k αριστερών ιδιαζουσών τιμών και των αντίστοιχων αριστερών ιδιαζόντων διανυσμάτων.

Σημειώνουμε ότι στην κατασκευή της SVD του πίνακα C, έχουμε $C = H\Sigma_C y^T$. Θα δείξουμε ότι αν οι πιθανότητες $\{p_i\}_{i=1}^n$ επιλεχθούν σωστά τότε τα αριστερά ιδιάζοντα διανύσματα του πίνακα C είναι με μεγάλη πιθανότητα προσεγγίσεις των αριστερών ιδιαζόντων διανυσμάτων του πίνακα A.

Παρουσιάζουμε παρακάτω τον αλγόριθμο clown στον οποίο αρχική εικόνα του clown αναπαρίσταται με τη μορφή ενός πίνακα A με διαστάσεις 200x320 και στην συνέχεια τον αναλύουμε . Σκοπός του αλγορίθμου μας σύμφωνα με τη λογική που περιγράψαμε παραπάνω είναι να προσεγγίσει την αρχική εικόνα άρα και τον αρχικό πίνακα. Θα δούμε παρακάτω ότι για κατάλληλο k , κατάλληλο c και κατάλληλες πιθανότητες ο πίνακας $H_k H_k^T A$ είναι μια καλή προσέγγιση του αρχικού πίνακα A. Θα δούμε αναλυτικά τα βήματα στην ανάλυση του αλγορίθμου.

Αλγόριθμος SVD γραμμικού χρόνου clown matlab

```
load clown
imagesc(X)
newmap1 = contrast(X);
colormap(newmap1)
AT=X'
A=X
b=0.99;
k=50;
c=300;
m=200;
n=320;
P=[];
Pr=[];
matrixH=zeros(m,k);
singval=zeros(c,c);

AF=norm(A,'fro');

for j=1:n
sum=norm(A(:,j))^2;
 P(j)=sum/AF^2;
end

 sum=0;
for w=1:n
 sum = sum + norm(A(:,w))^2;
%*norm(s.X(w,:));
end

for w=1:n
 Pr(w)=norm(A(:,w))^2/sum;
%*norm(s.X(w,:))
end

for i=1:n
 if P(i)>Pr(i)

L(i)=i;
 end
end
```

```
L1=nonzeros(L);
x=randi(size(L1,1),1,c);  % The
vector of indices of columns.
COL=X(:,L1(x));
for i=1:c
 V1(i)=P(L1(x(i)));
end
for t=1:c
 C(:,t)=COL(:,t)/sqrt(c*V1(t));
end
Z=C'*C;
[U,S,V]=svd(Z);
H=zeros(m,k);
for t=1:k
    H(:,t)=C*U(:,t)/sqrt(S(t,t));
end

AF=norm(A-
H*H'*A,'fro')/norm(A,'fro')

imagesc(H*H'*A)

norm(A-
H*H'*A,'fro')/norm(A,'fro')

spectralnorm=norm(A-
H*H'*A)/norm(A)
```

Έχοντας έναν πίνακα A mxn απαιτείται η διαχείριση mxn στοιχείων . Γνωρίζοντας όμως μια προσέγγιση του A , $A_k = s_1u_1v_1^T + s_2u_2v_2^T + \ldots + s_ku_kv_k^T$ απαιτείται η διαχείριση (m+n+1)k στοιχείων. Οπότε στο δικό μας παράδειγμα που έχουμε έναν πίνακα με διαστάσεις 200x320 θα είχαμε να διαχειριστούμε 64.000 στοιχεία ενώ εφαρμόζοντας το SVD για k=40 , έχουμε μια ικανοποιητική προσέγγιση που απαιτεί τη γνώση και τη διαχείριση 20.840 στοιχείων. Στα παραδείγματα της ενότητας 3.1.3 θα δούμε πιο αναλυτικά τα αποτελέσματα αυτά.

3.1.1. Ανάλυση της εκτέλεσης και του χρόνου λειτουργίας του αλγορίθμου

Υποθέτοντας ότι χρησιμοποιούνται οι σχεδόν βέλτιστες πιθανότητες δειγματοληψίας του Θεωρήματος 2.1.1 τότε στον αλγόριθμο οι πιθανότητες p_k μπορούν να χρησιμοποιηθούν για να επιλεγούν οι στήλες σε ένα πέρασμα χρησιμοποιώντας τον αλγόριθμο select [7] με χωρική και χρονική πολυπλοκότητα O(c).

Στη συνέχεια, με δεδομένα τα στοιχεία που επιλέχθηκαν, κατασκευάζεται ο πίνακας C σε ένα πρόσθετο πέρασμα και αυτό απαιτεί χωρική και χρονική πολυπλοκότητα O(mc).

Έχοντας τον πίνακα $C \in R^{mxc}$ ο υπολογισμός του C^TC απαιτεί χωρική πολυπλοκότητα O(mc) και χρονική πολυπλοκότητα $O(mc^2)$. Επίσης, ο υπολογισμός της SVD του C^TC απαιτεί χρονική πολυπλοκότητα $O(c^3)$.

Τέλος ο υπολογισμός του H_k απαιτεί πολλαπλασιασμούς k πινάκων άρα απαιτεί χωρική και χρονική πολυπλοκότητα O(mck). Έτσι συνολικά ο αλγόριθμος έχει χωρική πολυπλοκότητα $O(cm + c^2)$ και χρονική πολυπλοκότητα $O(c^2m + c^3)$

3.1.2. Ανάλυση του σταδίου δειγματοληψίας

Η προσέγγιση του πίνακα A από το $A_k = U_kU_k^TA$ προκαλεί σφάλμα $\|A - A_k\|_F^2 = \sum_{t=k+1}^{r} \sigma_t^2(A)$ και $\|A - A_k\|_2 = \sigma_{k+1}(A)$. Το A_k είναι η βέλτιστη προσέγγιση βαθμού k του πίνακα A σε σχέση με τις δύο νόρμες $\|.\|_F$ και $\|.\|_2$. Θα δείξουμε στη συνέχεια ότι εκτός από το σφάλμα αυτό , ο πίνακας $H_kH_K^TA$ έχει ένα σφάλμα που εξαρτάται από το $\|AA^T - CC^T\|_F$. Έπειτα χρησιμοποιώντας το Θεώρημα 2.1.1 θα δείξουμε ότι αυτό το επιπλέον σφάλμα εξαρτάται από το $\|A\|_F^2$. Αρχικά υποθέτουμε πως έχουμε ένα όριο για τη Frobenius νόρμα.

Θεώρημα 3.1.1.

Θεωρούμε έναν πίνακα $A \in R^{mxn}$ και έστω ότι ο πίνακας H_k που κατασκευάζεται από τον αλγόριθμο SVD γραμμικού χρόνου. Τότε

$$\left\|A - H_kH_k^TA\right\|_F^2 \leq \|A - A_k\|_F^2 + 2\sqrt{k}\|AA^T - CC^T\|_F.$$

Απόδειξη.

Υπενθύμιση: Για τους πίνακες X και Y ισχύουν τα εξής: $\|X\|_F^2 = Tr(X^TX)$, $Tr(X + Y) = Tr(X) + Tr(Y)$ και $H_k^TH_k = I_k$.

Επίσης, εκφράζουμε την ποσότητα $\left\|A - H_kH_k^TA\right\|_F^2$ ως εξής :

$$\begin{aligned}\left\|A - H_kH_k^TA\right\|_F^2 &= Tr((A - H_kH_k^TA)^T(A - H_kH_k^TA)) \\ &= Tr(A^TA - 2A^TH_kH_k^TA + A^TH_kH_k^TH_kH_k^TA) \\ &= Tr(A^TA) - Tr(A^TH_kH_k^TA)\end{aligned}$$

$$= \|A\|_F^2 - \|A^T H_k\|_F^2. \quad (3.1.1)$$

Στη συνέχεια συσχετίζουμε τις ποσότητες $\|A^T H_k\|_F^2$ και $\sum_{t=1}^{k} \sigma_t^2(C)$ ως εξής :

$$\begin{aligned} \left|\|A^T H_k\|_F^2 - \sum_{t=1}^{k} \sigma_t^2(C)\right| &\le \sqrt{k}\left(\sum_{t=1}^{k}\left(\left|A^T h^t\right|^2 - \sigma_t^2(C)\right)^2\right)^{1/2} \\ &= \sqrt{k}\left(\sum_{t=1}^{k}\left(\left|A^T h^t\right|^2 - \left|C^T h^t\right|^2\right)^2\right)^{1/2} \\ &= \sqrt{k}\left(\sum_{t=1}^{k}\left(h^{t^T}(AA^T - CC^T)h^t\right)^2\right)^{1/2} \\ &\le \sqrt{k}\|AA^T - CC^T\|_F. \end{aligned} \quad (3.1.2)$$

Η πρώτη ανισότητα προκύπτει από την ανισότητα Cauchy-Schwartz και η τελευταία ανισότητα γράφοντας AA^T και CC^T

Χρησιμοποιώντας την ανίσωση Hoffman-Wielandt (1.1.14) και τη σχέση $\sigma_t^2(X) = \sigma_t(XX^T)$ για τον πίνακα X συσχετίζουμε επίσης τις ποσότητες $\sum_{t=1}^{k} \sigma_t^2(C)$ και $\sum_{t=1}^{k} \sigma_t^2(A)$ ως εξής:

$$\begin{aligned} \left|\sum_{t=1}^{k} \sigma_t^2(C) - \sum_{t=1}^{k} \sigma_t^2(A)\right| &\le \sqrt{k}\left(\sum_{t=1}^{k}\left(\sigma_t^2(C) - \sigma_t^2(A)\right)^2\right)^{1/2} \\ &= \sqrt{k}\left(\sum_{t=1}^{k}\left(\sigma_t(CC^T) - \sigma_t(AA^T)\right)^2\right)^{1/2} \\ &\le \sqrt{k}\left(\sum_{t=1}^{m}\left(\sigma_t(CC^T) - \sigma_t(AA^T)\right)^2\right)^{1/2} \\ &\le \sqrt{k}\|CC^T - AA^T\|_F. \end{aligned} \quad (3.1.3)$$

Από τις σχέσεις (3.1.2) και (3.1.3) προκύπτει ότι

$$\left|\|A^T H_k\|_F^2 - \sum_{t=1}^{k} \sigma_t^2(A)\right| \le 2\sqrt{k}\|AA^T - CC^T\|_F \quad (3.1.4)$$

Από τις σχέσεις (3.1.1) και (3.1.4) προκύπτει το ζητούμενο. ■

Στη συνέχεια αποδεικνύουμε το ίδιο για την Ευκλείδεια νόρμα.

Θεώρημα 3.1.2.

Θεωρούμε έναν πίνακα $A \in R^{mxn}$ και έστω ότι ο πίνακας H_k κατασκευάζεται από τον αλγόριθμο SVD γραμμικού χρόνου. Τότε

$$\left\|A - H_k H_k^T A\right\|_2^2 \le \|A - A_k\|_2^2 + 2\|AA^T - CC^T\|_2.$$

Απόδειξη.

Έστω ότι $\mathcal{H}_k = range(H_k) = span(h^1, \ldots, h^k)$ και έστω ότι το $\mathcal{H}_{m-k}$ είναι το ορθογώνιο συμπλήρωμα του πίνακα $\mathcal{H}_k$. Έστω $x \in R^m$ και $x = ay + \beta z$, όπου $y \in \mathcal{H}_k$, $z \in \mathcal{H}_{m-k}$ και $a^2 + \beta^2 = 1$, τότε

$$\begin{aligned} \left\|A - H_k H_k^T A\right\|_2 &= \max_{x \in R^m, |x|=1} \left|x^T (A - H_k H_k^T A)\right|. \\ &= \max_{y \in \mathcal{H}_k, |y|=1, z \in \mathcal{H}_{m-k}, |z|=1, a^2+\beta^2=1} \left|(ay^T + \beta z^T)(A - H_k H_k^T A)\right| \\ &\le \max_{y \in \mathcal{H}_k, |y|=1} \left|y^T (A - H_k H_k^T A)\right| + \max_{z \in \mathcal{H}_{m-k}, |z|=1} \left|z^T (A - H_k H_k^T A)\right| \quad (3.1.5) \\ &= \max_{z \in \mathcal{H}_{m-k}, |z|=1} |z^T A|. \quad (3.1.6) \end{aligned}$$

Η ανίσωση (3.1.5) προέκυψε επειδή $\alpha, \beta \le 1$ και η ανίσωση (3.1.6) επειδή $y \in \mathcal{H}_k$, $z \in \mathcal{H}_{m-k}$.

Συνεχίζοντας έχουμε ότι

$$|z^T A|^2 = z^T CC^T z + z^T (AA^T - CC^T)z$$

$$\leq \sigma_{k+1}^2(C) + \|AA^T - CC^T\|_2 \quad (3.1.7)$$

$$\leq \sigma_{k+1}^2(A) + 2\|AA^T - CC^T\|_2 \quad (3.1.8)$$

$$= \|A - A_k\|_2^2 + 2\|AA^T - CC^T\|_2. \quad (3.1.9)$$

Η ανίσωση (3.1.7) προέκυψε επειδή το $\max_{z \in \mathcal{H}_{m-k}} |z^T C|$ ισχύει όταν το z είναι το $(k+1)$ αριστερό ιδιάζον διάνυσμα, η ανίσωση (3.1.8) διότι $\sigma_{k+1}^2(C) = \sigma_{k+1}(CC^T)$ και επειδή από τη σχέση (1.1.13) γνωρίζουμε ότι $\sigma_{k+1}^2(C) \leq \sigma_{k+1}(AA^T) + \|AA^T - CC^T\|_2$ και τέλος η σχέση (3.1.9) προέκυψε διότι $\|A - A_k\|_2 = \sigma_{k+1}(A)$.

Από τις σχέσεις (3.1.6) και (3.1.9) προκύπτει το ζητούμενο. ■

Το Θεώρημα 3.1.1 και το Θεώρημα 3.1.2 ισχύουν ανεξάρτητα από τις πιθανότητες $\{p_i\}_{i=1}^n$. Δεδομένου ότι $\|A - A_k\|_\xi$, ξ=2,*F*, είναι μια ιδιότητα του πίνακα *A*, η επιλογή των πιθανοτήτων δειγματοληψίας περιλαμβάνονται στο σφάλμα $\left\|A - H_k H_k^T A\right\|_\xi^2$, μόνο μέσω του όρου που περιλαμβάνεται στο πρόσθετο σφάλμα πέρα από τη βέλτιστη προσέγγιση βαθμού *k*, για παράδειγμα ο όρος $\|AA^T - CC^T\|_\xi$.

Αν και το πρόσθετο σφάλμα στο Θεώρημα 3.1.2 εξαρτάται από το $\|AA^T - CC^T\|_2$, παρατηρούμε ότι $\|AA^T - CC^T\|_2 \leq \|AA^T - CC^T\|_F$. Τη σχέση αυτή θα την χρησιμοποιήσουμε στην πορεία.

Σημειώνουμε επίσης ότι ο παράγοντας στο επιπρόσθετο σφάλμα είναι $2\sqrt{k}$ για την $\|\quad\|_F^2$ ενώ για την $\|\cdot\|_2$ είναι μόνο 2.

Στο επόμενο θεώρημα προσαρμόζουμε τις πιθανότητες δειγματοληψίας ώστε να είναι σχεδόν βέλτιστες. Έτσι διαλέγοντας αρκετές στήλες θα παρατηρήσουμε ότι το σφάλμα της προσέγγισης γίνεται αρκετά μικρό.

Θεώρημα 3.1.3.

Θεωρούμε έναν πίνακα $A \in R^{mxn}$ και έστω ότι ο πίνακας H_k κατασκευάζεται από τον αλγόριθμο SVD γραμμικού χρόνου διαλέγοντας c στήλες του πίνακα *A* με πιθανότητες $\{p_i\}_{i=1}^n$τέτοιες ώστε $p_i \geq \beta \left|A^{(i)}\right|^2 / \|A\|_F^2$, για έναν θετικό αριθμό $\beta \leq 1$ και έστω $\eta = 1 + \sqrt{(8/\beta)\, log(1/\delta)}$. Έστω επίσης ε>0.

Αν $c \geq 4k/\beta\varepsilon^2$ τότε

$$E[\left\|A - H_k H_k^T A\right\|_F^2] \leq \|A - A_k\|_F^2 + \varepsilon\|A\|_F^2 \quad (3.1.10)$$

και αν $c \geq 4k\eta^2/\beta\varepsilon^2$ τότε με πιθανότητα τουλάχιστον 1-δ έχουμε ότι

$$\left\|A - H_k H_k^T A\right\|_F^2 \leq \|A - A_k\|_F^2 + \varepsilon\|A\|_F^2. \quad (3.1.11)$$

Επιπλέον αν $c \geq 4/\beta\varepsilon^2$ τότε

$$E[\left\|A - H_k H_k^T A\right\|_2^2] \leq \|A - A_k\|_2^2 + \varepsilon\|A\|_F^2 \quad (3.1.12)$$

και αν $c \geq 4\eta^2/\beta\varepsilon^2$, τότε με πιθανότητα τουλάχιστον 1-δ

$$\left\|A - H_k H_k^T A\right\|_2^2 \leq \|A - A_k\|_2^2 + \varepsilon\|A\|_F^2. \quad (3.1.13)$$

Απόδειξη.

Συνδυάζοντας τα Θεωρήματα 3.1.1 και 3.1.2 με το Θεώρημα 2.1.1 έχουμε ότι

$$E[\left\|A - H_k H_k^T A\right\|_F^2] \leq \|A - A_k\|_F^2 + (\frac{4k}{\beta c})^{1/2}\|A\|_F^2, \quad (3.1.14)$$

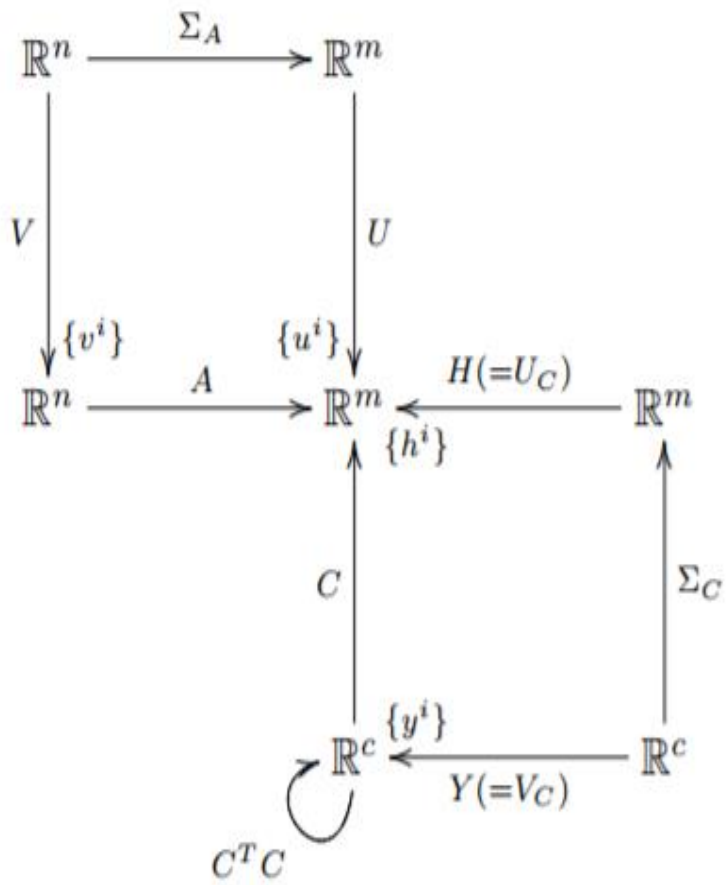
$$E[\left\|A - H_k H_k^T A\right\|_2^2] \leq \|A - A_k\|_2^2 + (\frac{4}{\beta c})^{1/2}\|A\|_F^2 \quad (3.1.15)$$

και με πιθανότητα τουλάχιστον 1-δ έχουμε ότι

$$\left\|A - H_k H_k^T A\right\|_F^2 \leq \|A - A_k\|_F^2 + (\frac{4\eta^2 k}{\beta c})^{1/2}\|A\|_F^2. \quad (3.1.16)$$

$$\left\|A - H_k H_k^T A\right\|_2^2 \leq \|A - A_k\|_2^2 + (\frac{4\eta^2}{\beta c})^{1/2}\|A\|_F^2. \quad (3.1.17)$$

Το θεώρημα αποδεικνύεται χρησιμοποιώντας κατάλληλη τιμή για το c. ■



Σχήμα 2. Διάγραμμα του αλγορίθμου *SVD γραμμικού χρόνου.*

Θα παρουσιάσουμε τον αλγόριθμο corn ο οποίος είναι αντίστοιχος του αλγορίθμου του clown . Και εδώ τα δεδομένα της εικόνας του καλαμποκιού αναπαρίστανται με έναν πίνακα ο οποίος τώρα έχει διαστάσεις 415x312. Αυτό απαιτεί τη διαχείριση 129480 στοιχείων. Θα δείξουμε πως και στο συγκεκριμένο παράδειγμα , όσο αυξάνουμε το *c*, δηλαδή των αριθμό των στηλών που επιλέγονται αλλά και το *k* δηλαδή των αριθμό των *k* πρώτων ιδιαζόντων διανυσμάτων που διατηρούμε αυξάνεται και η αποτελεσματικότητα της προσέγγισης . Ανάλογα πάλι με το k που θα επιλέξουμε τα στοιχεία που θα έχουμε να διαχειριστούμε θα είναι πολύ λιγότερα.

Αλγόριθμος SVD γραμμικού χρόνου corn matlab

```
corn_gray =
imread('corn.tif',3);
imshow(corn_gray)
I=im2double(corn_gray
);
AT=I';
A=I;
b=0.99;
k=8;
c=50;
m=415;
n=312;
P=[];
Pr=[];
matrixH=zeros(m,k);

%h=zeros(m,k);
%for tt=1:20
singval=zeros(c,c);

AF=norm(A,'fro');

for j=1:n
sum=norm(A(:,j))^2;
 P(j)=sum/AF^2;
end

 sum=0;
for w=1:n
 sum = sum +
norm(A(:,w))^2;
%*norm(s.X(w,:));
end

for w=1:n

Pr(w)=norm(A(:,w))^2/
sum;
%*norm(s.X(w,:))
end

for i=1:n
 if P(i)>=Pr(i)
 L(i)=i;
 end
end
```

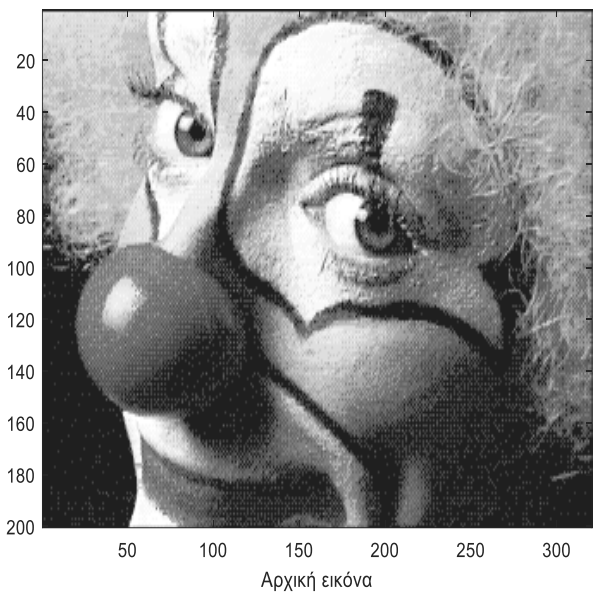
```
L1=nonzeros(L);
x=randi(size(L1,1),1,
c);  % The vector of
indices of columns.
COL=A(:,L1(x));
for i=1:c
 V1(i)=P(L1(x(i)));
end
for t=1:c

C(:,t)=COL(:,t)/sqrt(
c*V1(t));
end
Z=C'*C;
[U,S,V]=svd(Z);
H=zeros(m,k);
for t=1:k

H(:,t)=C*U(:,t)/sqrt(
S(t,t));
end
%matrixH=matrixH+H;
%singval=singval+S;
%end
%H=matrixH/20;
%S=singval/20;
AF=norm(A-
H*H'*A,'fro')/norm(A,
'fro');

imagesc(H*H'*A)
norm(A-
H*H'*A,'fro')/norm(A,
'fro')
spectralnorm=norm(A-
H*H'*A)/norm(A)
```

3.1.3. Παραδείγματα SVD

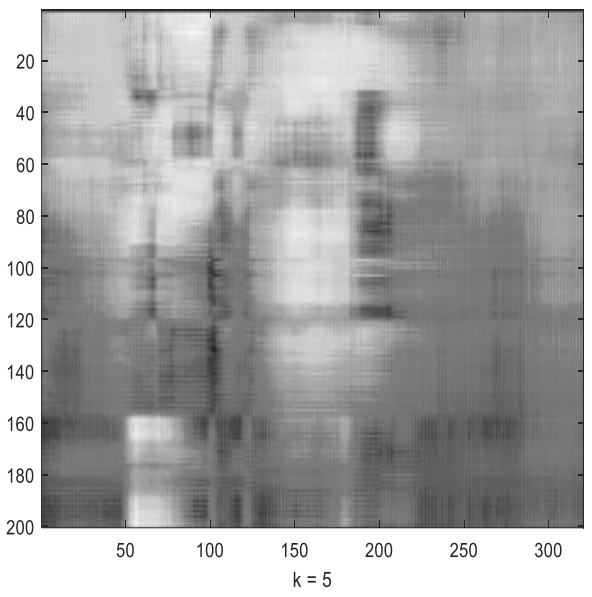
Στα παραδείγματα που ακολουθούν εφαρμόζουμε τους αλγορίθμους. Τα δεδομένα της εικόνας του clown και του corn όπως έχουμε αναφέρει αναπαρίστανται με έναν πίνακα. Δείχνουμε παρακάτω πως για τον ίδιο πίνακα και συνεπώς και την ίδια εικόνα όσο αυξάνουμε το c, δηλαδή των αριθμό των στηλών που επιλέγονται αλλά και το k δηλαδή των αριθμό των k πρώτων ιδιαζόντων διανυσμάτων που διατηρούμε αυξάνεται και η αποτελεσματικότητα της προσέγγισης. Σύμφωνα με τις πιθανότητες δειγματοληψίας παρατηρούμε ότι επιλέγονται οι στήλες που περιέχουν την περισσότερη πληροφορία. Η επιλογή των στηλών που χρησιμοποιεί ο αλγόριθμος για να φτιάξει τον πίνακα C γίνεται χωρίς να τηρείται κάποια συγκεκριμένη σειρά και έχοντας τη δυνατότητα να επιλέγει την ίδια στήλη όσες φορές θέλει. Παρόλα αυτά θα δούμε παρακάτω στα παραδείγματά μας ότι έχουμε μια αρκετά καλή προσέγγιση του αρχικού πίνακα.

Παράδειγμα SVD 1

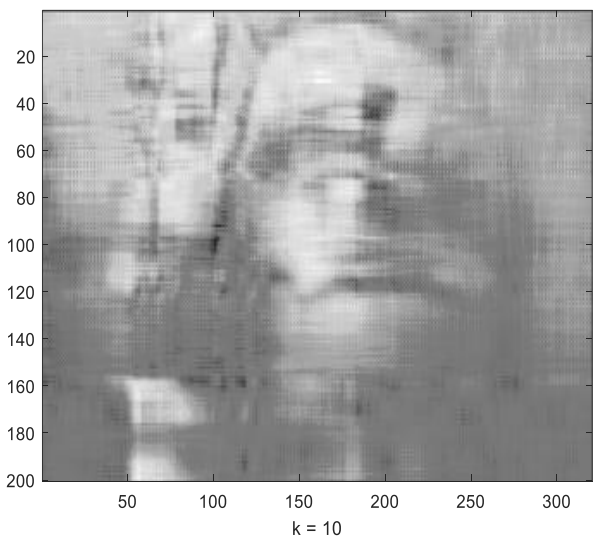
c=50



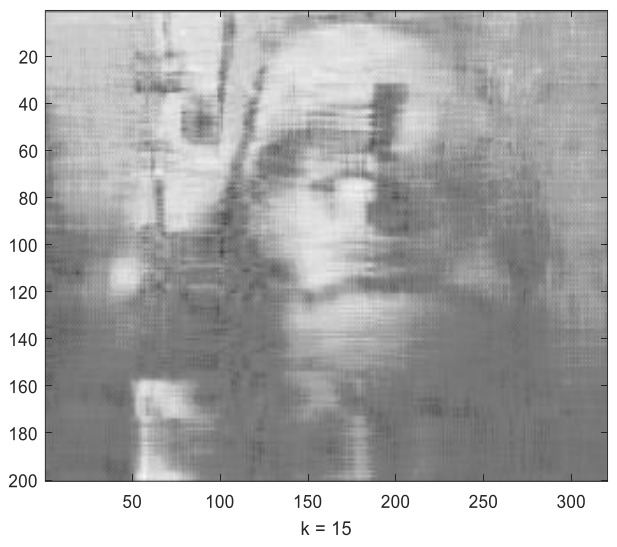
Αρχική εικόνα



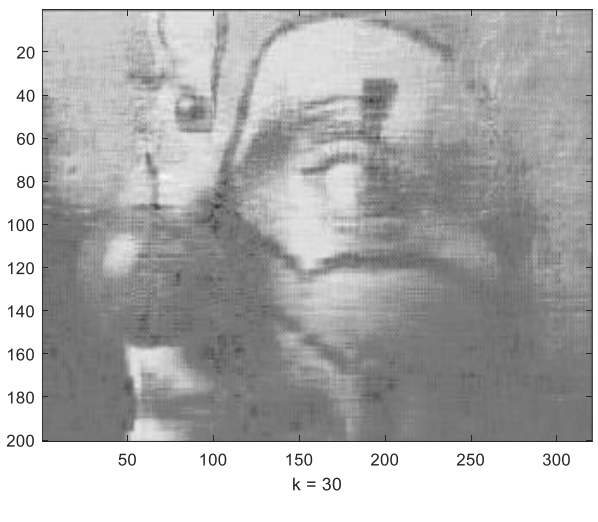
k = 5



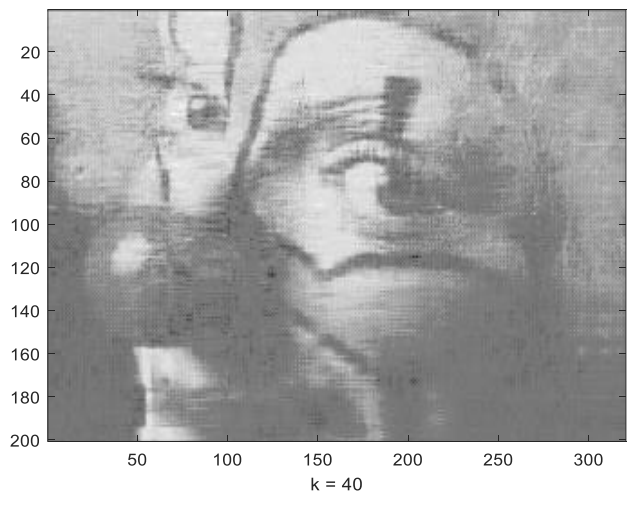
k = 10



k = 15



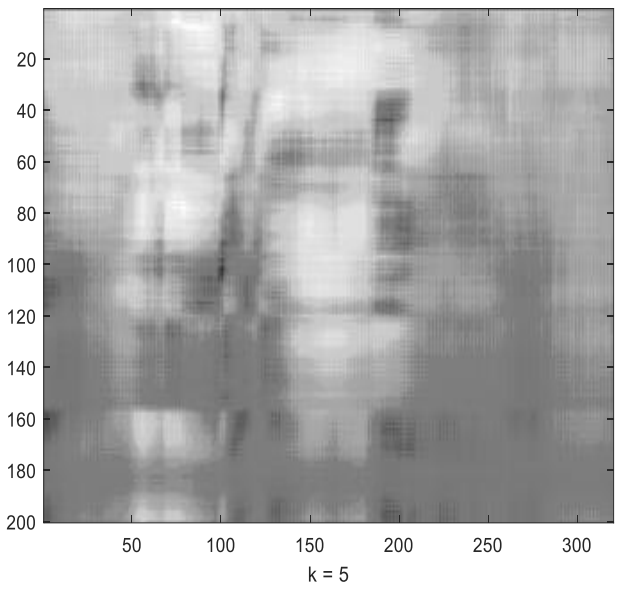
k = 30

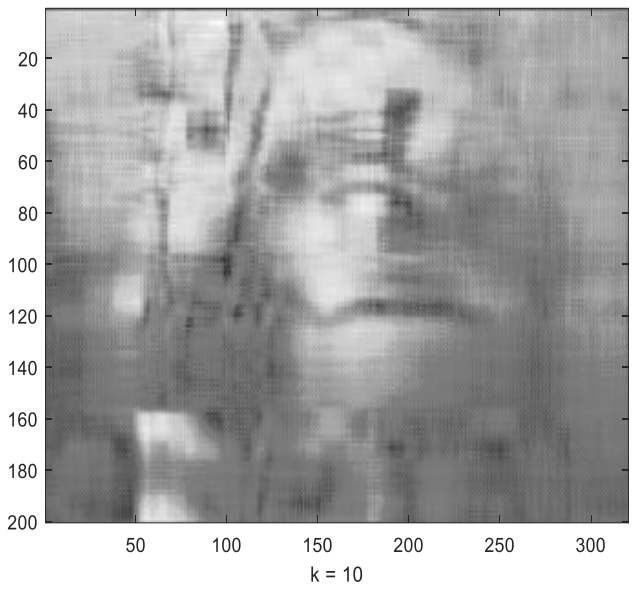


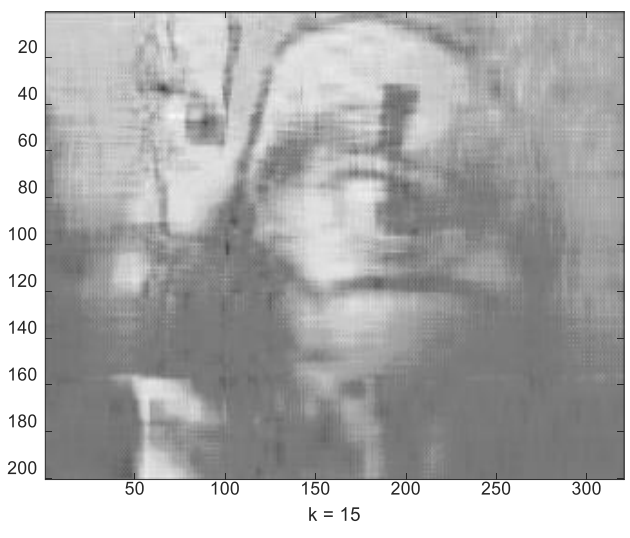
k = 40

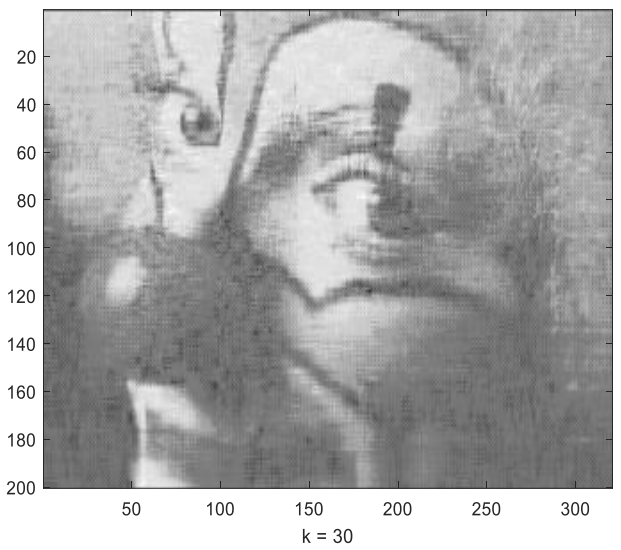
Παράδειγμα SVD 2

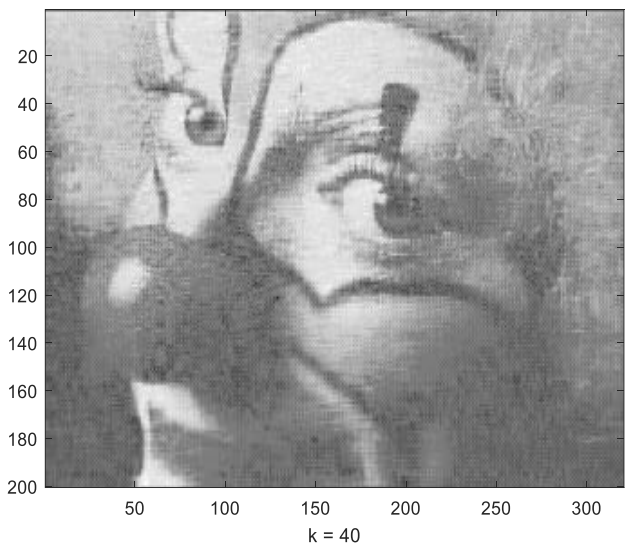
c=150

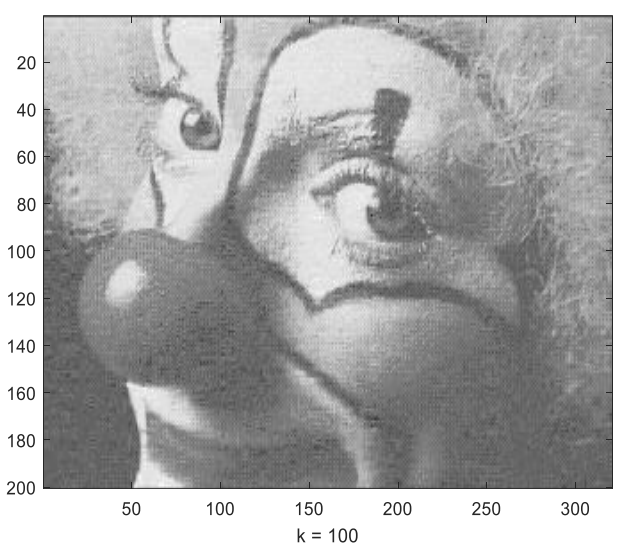












c=50

k	Σχετικό σφάλμα $\left\|A - H_k H_k^T A\right\|_F / \|A\|_F$	Σχετικό σφάλμα $\left\|A - H_k H_k^T A\right\|_2 / \|A\|_2$
5	0.3265	0.1671
10	0.2782	0.1053
15	0.2386	0.0867
30	0.2054	0.0630
40	0.1822	0.0625
45	0.1787	0.0620
50	0.1703	0.0600

Πίνακας 3.1.1.

c =150

k	Σχετικό σφάλμα $\left\|A - H_k H_k^T A\right\|_F / \|A\|_F$	Σχετικό σφάλμα $\left\|A - H_k H_k^T A\right\|_2 / \|A\|_2$
5	0.2792	0.1498
10	0.2544	0.0899
15	0.2199	0.0645
30	0.1661	0.0436
40	0.1469	0.0427
80	0.1057	0.0297
100	0.0929	0.0228

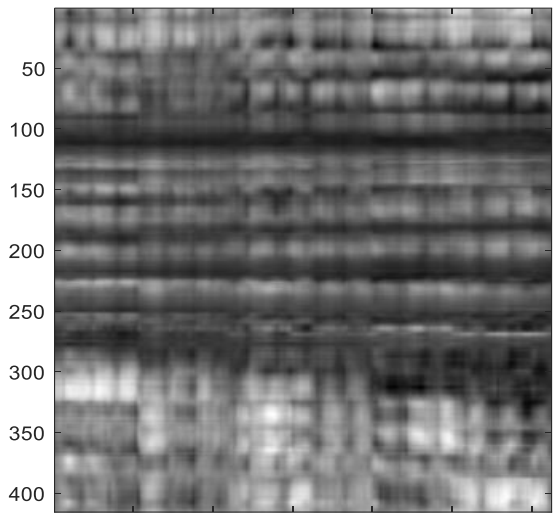
Πίνακας 3.1.2.

Παρατηρούμε ότι στον συγκεκριμένο αλγόριθμο η Ευκλείδεια νόρμα δίνει αρκετά καλύτερο αποτέλεσμα σε σχέση με τη νόρμα Frobenius. Επιπλέον, φαίνεται ότι κρατώντας το k σταθερό και αυξάνοντας το c, δηλαδή των αριθμό των στηλών του αρχικού πίνακα A που επιλέγονται, μειώνεται το σφάλμα και συνεπώς η προσέγγιση γίνεται καλύτερη. Όπως αναφέραμε και στο Θεώρημα 3.1.3 χρησιμοποιώντας τις κατάλληλες πιθανότητες δειγματοληψίας $p_i \geq \beta \left|A^{(i)}\right|^2 / \|A\|_F^2$ για έναν θετικό αριθμό $\beta \leq 1$ και $\eta = 1 + \sqrt{(8/\beta)\, log(1/\delta)}$ τότε το σφάλμα μου είναι αρκετά μικρό και πετυχαίνω μια αρκετά καλή προσέγγιση. Κάτι που δεν είναι εύκολο να διακρίνω στις εικόνες αλλά φαίνεται καθαρά στους πίνακες των αποτελεσμάτων 3.1.1 και 3.1.2 είναι ότι κρατώντας σταθερό το k και αυξάνοντας το c μειώνεται αισθητά το σφάλμα.

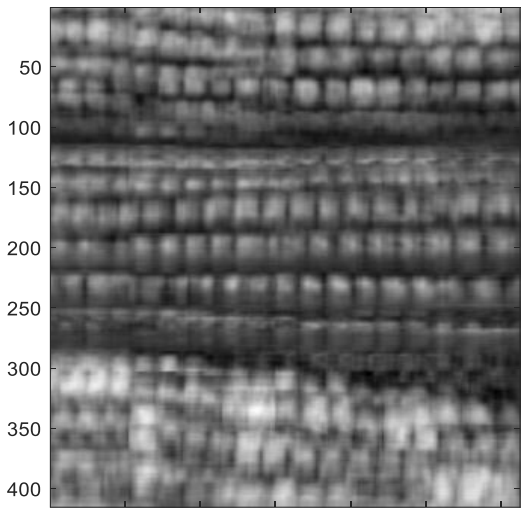
ΠΑΡΑΔΕΙΓΜΑ SVD 3

c=50

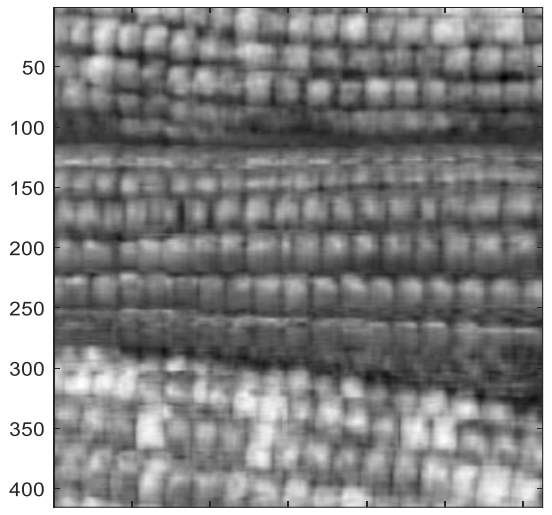
αρχική εικόνα



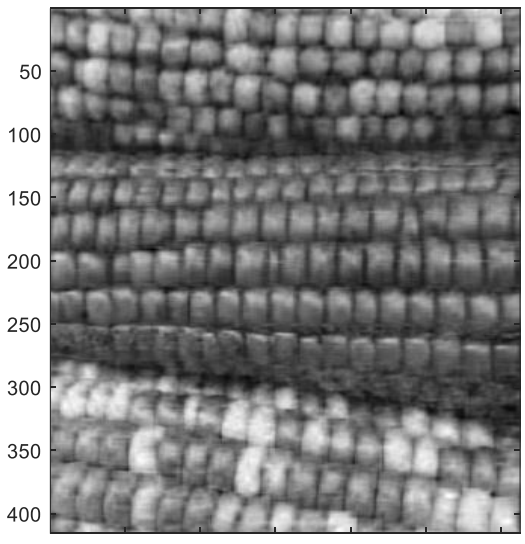
k=5



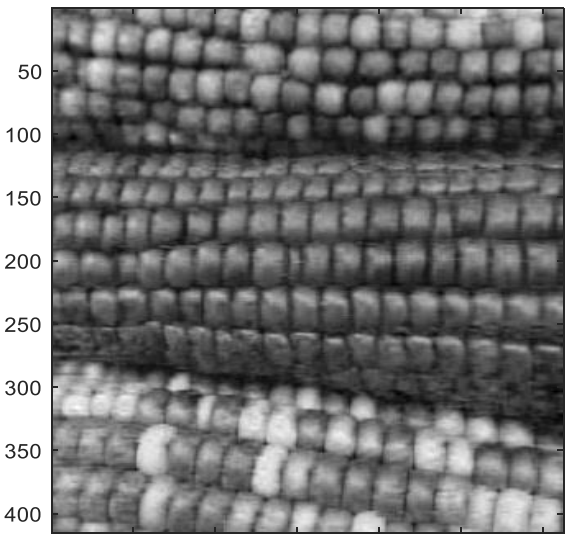
k=10



k=15



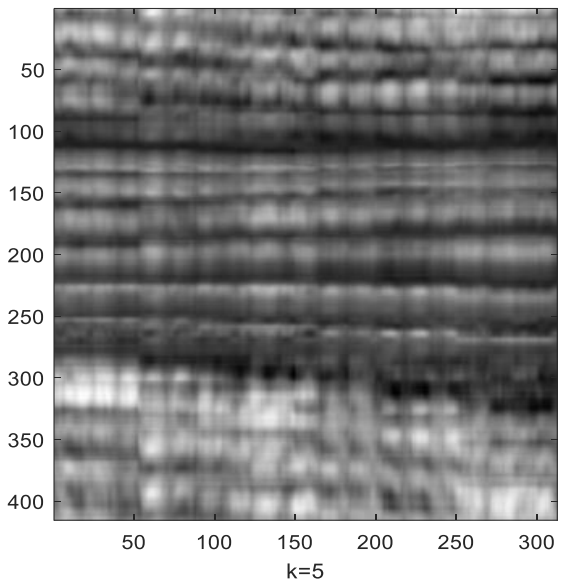
k=30

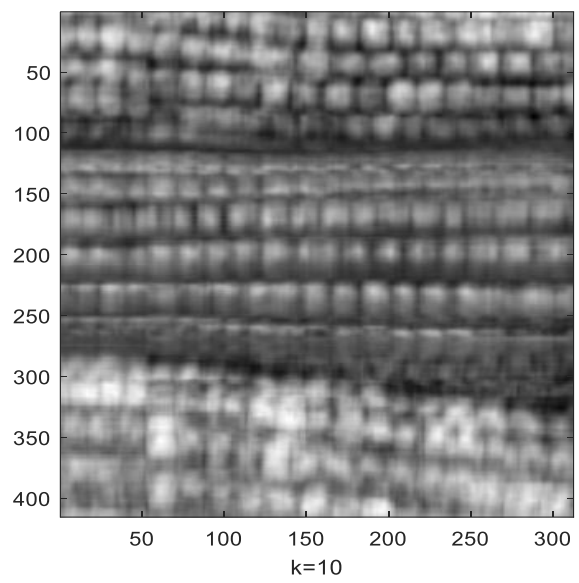


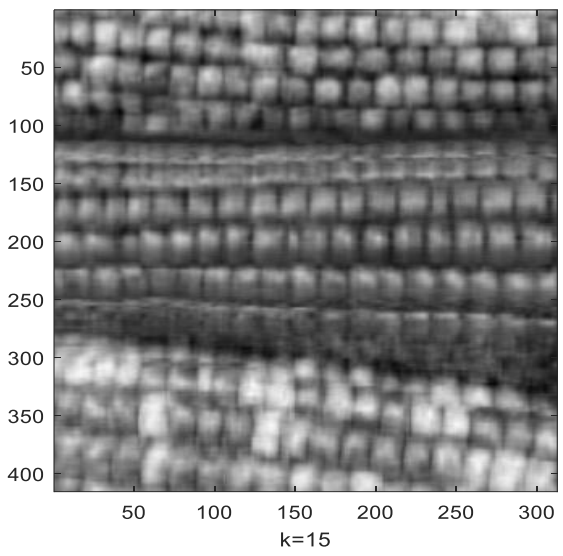
k=40

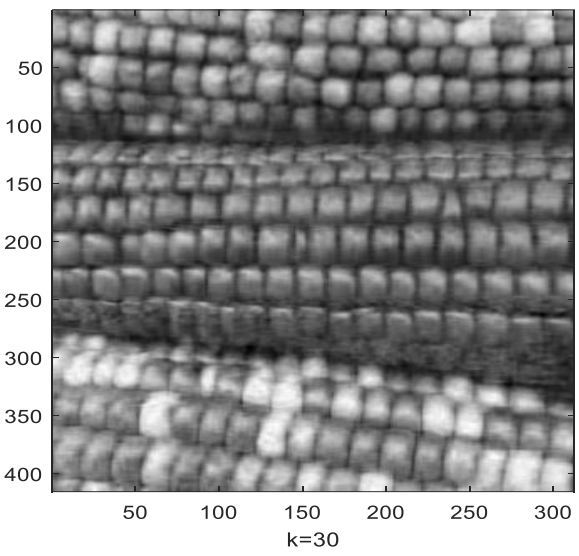
ΠΑΡΑΔΕΙΓΜΑ SVD 4

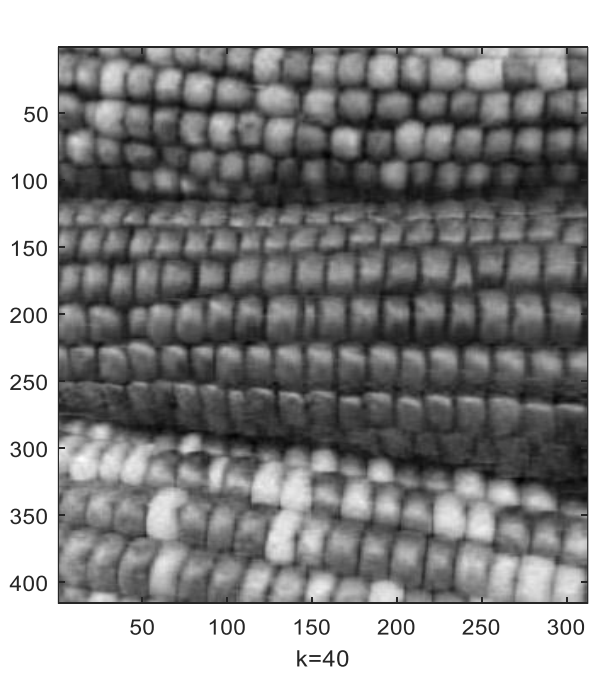
c=150

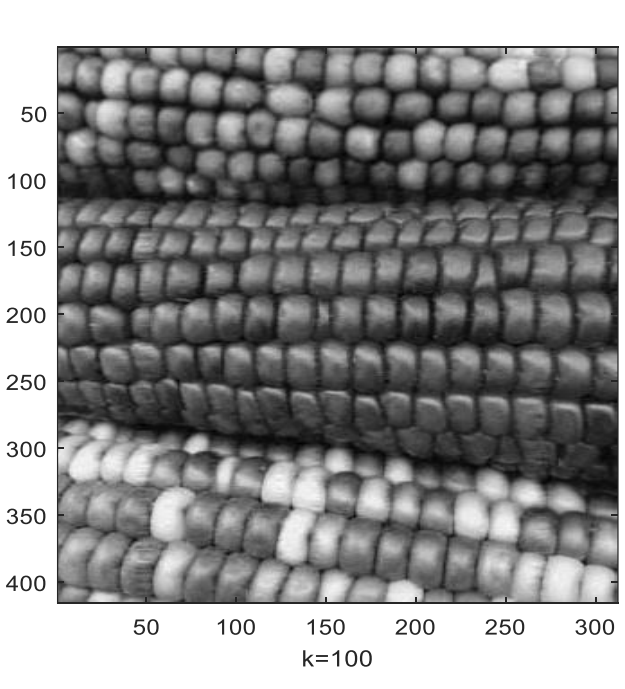












c= 50

k	Σχετικό σφάλμα $\left\|A - H_k H_k^T A\right\|_F / \|A\|_F$	Σχετικό σφάλμα $\left\|A - H_k H_k^T A\right\|_2 / \|A\|_2$
5	0.3013	0.1243
10	0.2404	0.0879
15	0.1978	0.0695
30	0.1381	0.0552
40	0.1181	0.0422

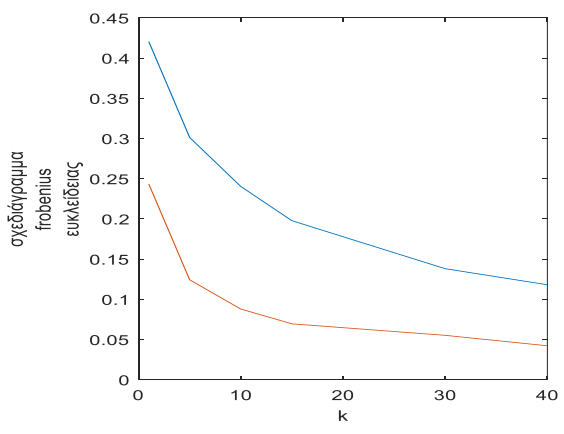
Πίνακας 3.1.3

c=150

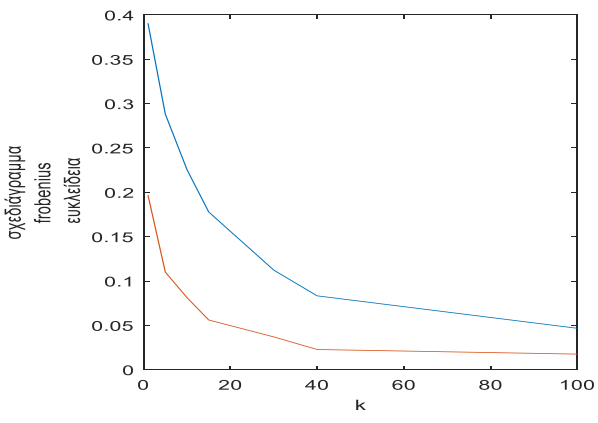
k	Σχετικό σφάλμα $\left\|A - H_k H_k^T A\right\|_F / \|A\|_F$	Σχετικό σφάλμα $\left\|A - H_k H_k^T A\right\|_2 / \|A\|_2$
5	0.2882	0.1101
10	0.2256	0.0813
15	0.1779	0.0561
30	0.1123	0.0370
40	0.0833	0.0228
100	0.0467	0.0176

Πίνακας 3.1.4

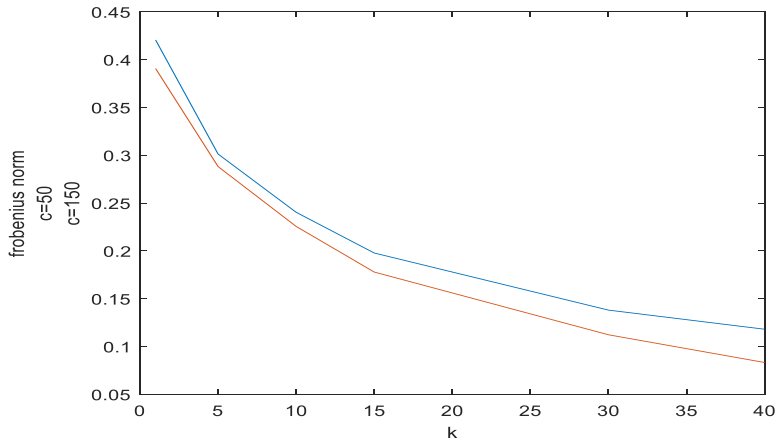
Παρατηρούμε και στους πίνακες ότι το σφάλμα μικραίνει καθώς αυξάνουμε το k .Εδώ φαίνεται καλύτερα και ότι ,για σταθερό k, το σφάλμα μικραίνει καθώς αυξάνουμε το c. Αυτή είναι μια πληροφορία που ίσως δυσκολευόμαστε να διακρίνουμε στις εικόνες για αυτό παρουσιάζουμε αναλυτικά τα αποτελέσματα του αλγορίθμου μας . Παρακάτω φαίνονται οι γραφικές παραστάσεις για c=50 και c=150 οι οποίες μας επιβεβαιώνουν ακριβώς αυτό το αποτέλεσμα. Παρατηρούμε επίσης όπως και στον αλγόριθμο BasicMatrixMultiplication ότι από μία τιμή του k και μετά το σφάλμα αρχίζει και σταθεροποιείται.



Διάγραμμα 8



Διάγραμμα 9



Διάγραμμα 10

3.2. ΠΡΟΣΕΓΓΙΣΤΙΚΟΣ ΑΛΓΟΡΙΘΜΟΣ SVD ΣΤΑΘΕΡΟΥ ΧΡΟΝΟΥ

Αλγόριθμος SVD σταθερού χρόνου

Είσοδος: $A \in R^{mxn}$, $c, w, k \in Z^+$ τέτοια ώστε $1 \le w \le m$, $1 \le c \le n$ και $1 \le k \le min(w, c)$ και $\{pi\}_{i=1}^{n}$τέτοιο ώστε $p_i \ge 0$ και $\sum_{i=1}^{n} p_i = 1$

Έξοδος: $\sigma_t(w)$, $t = 1,....l$ και μια περιγραφή του $\bar{H}_l \in R^{mxl}$

1. *Για* t=1 μέχρι c
   (*a*) Διάλεξε $i_t \in 1,\ldots,n$ με $\Pr[i_t = a] = p_a$, α=1,...,n και αποθήκευσε $\{(i_t, p_{jt}): t = 1,\ldots c\}$
   (*b*) $C^{(t)} = A^{(i_t)}/\sqrt{cp_{i_t}}$.
2. *Διάλεξε* $\{q_j\}_{j=1}^{m}$ τέτοιο ώστε $q_j = \left|C_{(j)}\right|^2/\|C\|_F^2$
3. *Για* t=1 μέχρι w,
   (*a*) Διάλεξε $j_t \in 1,\ldots.m$ με $\Pr[j_t = a] = q_a$, α=1,....m.
   (*b*) $W_{(t)} = C_{(j_t)}/\sqrt{wq_{j_t}}$
4. *Υπολόγισε το* $W^T W$ και το SVD του. $W^T W = \sum_{t=1}^{c} \sigma_t^2(W)\, z^t z^{t^T}$
5. *Υπολόγισε τις ιδιάζουσες τιμές* $\{\sigma_t(W)\}_{t=1}^{l}$ και τα αντίστοιχα ιδιάζοντα διανύσματα $\{z^t\}_{t=1}^{l}$

3.2.1. Ο αλγόριθμος

Με δεδομένο έναν πίνακα $A \in R^{mxn}$ θέλουμε να προσεγγίσουμε τις μεγαλύτερες k ιδιάζουσες τιμές και τα αντίστοιχα ιδιάζοντα διανύσματα σε έναν σταθερό αριθμό περασμάτων πάνω από τα δεδομένα και χωρική και χρονική πολυπλοκότητα O(1) ανεξάρτητες από τα m και n.

Η στρατηγική του αλγορίθμου είναι ότι διαλέγει c στήλες του πίνακα A και έπειτα ανακατασκευάζει διαιρώντας με τον κατάλληλο παράγοντα για να φτιάξει τον πίνακα $C \in R^{mxc}$. Έπειτα υπολογίζει τις προσεγγίσεις των ιδιαζουσών τιμών και των αντίστοιχων αριστερών ιδιαζόντων διανυσμάτων του πίνακα C τα οποία είναι προσεγγίσεις των αντίστοιχων του πίνακα A.

Στον αλγόριθμο SVD γραμμικού χρόνου, ο υπολογισμός των αριστερών ιδιαζόντων διανυσμάτων απαιτούσε χωρική και χρονική πολυπλοκότητα γραμμική στο $m+n$ (αν c σταθερό).

Με τον αλγόριθμο SVD σταθερού χρόνου, για να πετύχουμε σταθερή χωρική και χρονική πολυπλοκότητα O(1) γίνεται δειγματοληψία ξανά επιλέγοντας w γραμμές του πίνακα C για να κατασκευάσουμε τον πίνακα $W \in R^{wxc}$.

Στη συνέχεια υπολογίζεται το SVD του $W^T W$. Έστω $W^T W = Z\Sigma_{W^T W} Z^T = Z\Sigma_W^2 Z^T$. Οι ιδιάζουσες τιμές και τα αντίστοιχα ιδιάζοντα διανύσματα που υπολογίζονται είναι με μεγάλη πιθανότητα

προσεγγίσεις των ιδιαζουσών τιμών και των ιδιαζόντων διανυσμάτων του $C^T C$ και επομένως προσεγγίσεις των ιδιαζουσών τιμών και των δεξιά ιδιαζόντων διανυσμάτων του πίνακα C.

Σημειώνουμε ότι χρησιμοποιεί απλά τον αλγόριθμο SVD γραμμικού χρόνου για να προσεγγίσει τα δεξιά ιδιάζοντα διανύσματα του πίνακα C με τυχαία δειγματοληψία γραμμών του C.

Αλγόριθμος SVD σταθερού χρόνου matlab

```
A=randi(100,200,320);
b=0.7;
k=100;
c=150;
w=100
m=200;
n=320;
P=[];
Pr=[];
matrixH=zeros(k,w);
%h=zeros(m,k);
%for tt=1:20
singval=zeros(c,c);


AF=norm(A,'fro');

 [U,S,V]=svd(A);

for j=1:n
sum=norm(A(:,j))^2;
 P(j)=sum/AF^2;
end

 sum=0;
for j=1:n
 sum = sum + norm(A(:,j))^2;
%*norm(s.X(w,:));
end

for j=1:n
 Pr(j)=norm(A(:,j))^2/sum;
%*norm(s.X(w,:))
end

for i=1:n
 if P(i)>Pr(i)
 L(i)=i;
 end
end

L1=nonzeros(L);
x=randi(size(L1,1),1,c);  % The
vector of indices of columns.
COL=A(:,L1(x));
for i=1:c
 V1(i)=P(L1(x(i)));
end
for t=1:c
 C(:,t)=COL(:,t)/sqrt(c*V1(t));
end

% kataskeuh w

CF=norm(A,'fro');

for i=1:m
sum=norm(C(i,:))^2;
 P1(i)=sum/CF^2;
end
```

```
sum=0;
for i=1:m
 sum = sum + norm(C(i,:))^2;
%*norm(s.X(w,:));
end
for i=1:m
 Pr1(i)=b*norm(C(i,:))^2/sum;
%*norm(s.X(w,:))
end

for i=1:m
 if P1(i)>Pr1(i)
   L2(i)=i;
 end
end

l1=nonzeros(L2);
x1=randi(size(l1,1),1,w);  % The
vector of indices of columns.
W1=C(l1(x1),:);
for i=1:w
 V2(i)=P(l1(x1(i)));
end
for t=1:w
 W(t,:)=W1(t,:)/sqrt(w*V2(t));
end

[U1,S1,V1]=svd(W);

sum=0;
for i=1:10
    for j=1:10
Error(i,j)=norm(S(i,j)-
S1(i,j),'fro')/norm(S(i,j),'fro')
 sum=sum+1;
end
end
```

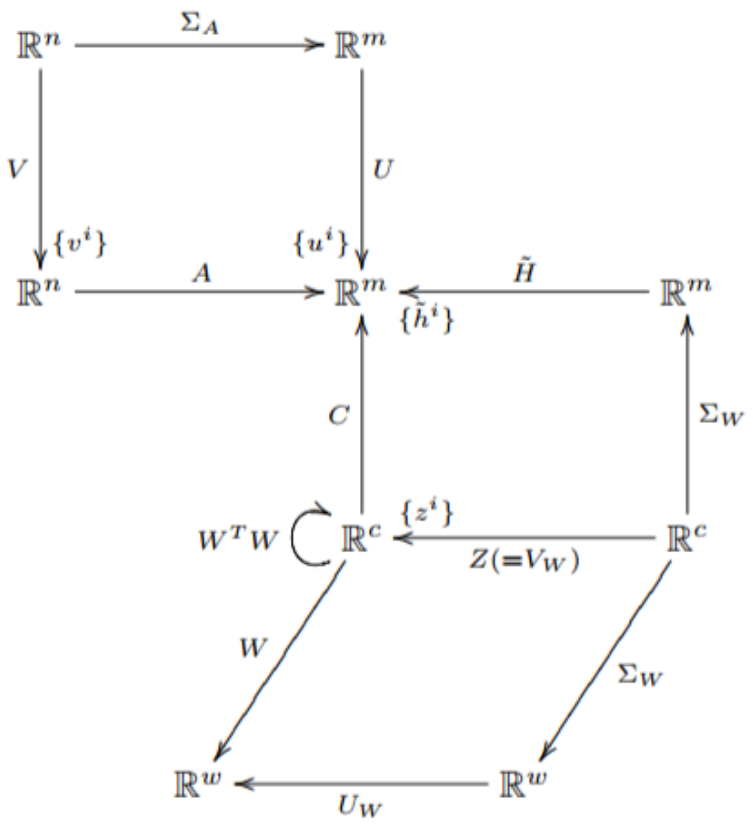
Στον παρακάτω πίνακα παρουσιάζεται το σχετικό σφάλμα των δέκα πρώτων ιδιαζουσών τιμών του πίνακα *A* (πίνακας *S*) και των δέκα πρώτων ιδιαζουσών τιμών του πίνακα *W* (πίνακας *S1*). Ο πίνακας *A* έχει διαστάσεις 200x320 ενώ ο πίνακας *W* έχει διαστάσεις 100x150. Παρόλο που έχουμε κάνει δύο δειγματοληψίες και έχουμε μειώσει αρκετά τις διαστάσεις του αρχικού πίνακα, παρατηρούμε ότι το σχετικό σφάλμα των ιδιαζουσών τιμών είναι αρκετά μικρό , το οποίο σημαίνει ότι έχουμε μια αρκετά καλή προσέγγιση.

Πίνακας Error

norm(S(i,j)- S1(i,j),'fro')/norm(S(i,j),'fro')

0.256	0	0	0	0	0	0	0	0	0
0	1.354	0	0	0	0	0	0	0	0
0	0	1.162	0	0	0	0	0	0	0
0	0	0	1.180	0	0	0	0	0	0
0	0	0	0	1.082	0	0	0	0	0
0	0	0	0	0	1.054	0	0	0	0
0	0	0	0	0	0	0.993	0	0	0
0	0	0	0	0	0	0	0.941	0	0
0	0	0	0	0	0	0	0	0.905	0
0	0	0	0	0	0	0	0	0	0.850

Πίνακας 3.1.5



Σχήμα 3. Διάγραμμα του αλγορίθμου SVD σταθερού χρόνου.

3.2.2. Ανάλυση της εκτέλεσης και του χρόνου λειτουργίας

Υποθέτοντας ότι χρησιμοποιούνται οι βέλτιστες πιθανότητες δειγματοληψίας (Θεώρημα 2.1.1) , στον αλγόριθμο SVD σταθερού χρόνου, οι πιθανότητες p_k μπορούν να χρησιμοποιηθούν για να επιλεγούν οι στήλες που θα επιλέξουμε σε ένα πέρασμα το οποίο απαιτεί χρονική και χωρική πολυπλοκότητα O(c) χρησιμοποιώντας τον αλγόριθμο select.

Έχοντας ως δεδομένες τις στήλες του πίνακα A που θα επιλεχθούν δεν κατασκευάζει ακριβώς τον πίνακα C αλλά κάνει και μια δεύτερη δειγματοληψία στην οποία διαλέγει w γραμμές του πίνακα C με πιθανότητες $\{q_i\}_{i=1}^{m}$ προκειμένου να κατασκευάσει τον πίνακα W.

Το πετυχαίνει αυτό κάνοντας ένα δεύτερο πέρασμα πάνω από τα δεδομένα το οποίο απαιτεί χωρική και χρονική πολυπλοκότητα O(w) χρησιμοποιώντας ξανά τον αλγόριθμο select [7]. Έπειτα σε ένα τρίτο πέρασμα κατασκευάζεται ο πίνακας W το οποίο απαιτεί χωρική και χρονική πολυπλοκότητα O(cw).

Στη συνέχεια έχοντας τον πίνακα W, υπολογίζει το W^TW με χωρική πολυπλοκότητα O(cw) και χρονική πολυπλοκότητα O(c^2w) και τέλος υπολογίζει το SVD του W^TWμε χρονική πολυπλοκότητα O(c^3). Οι ιδιάζουσες τιμές και τα αντίστοιχα ιδιάζοντα διανύσματα που υπολογίζονται μπορούν να επιστραφούν σαν μια προσέγγιση της λύσης.

Η συνολική χρονική πολυπλοκότητα του αλγορίθμου είναι O(c^3+cw^2) . Αυτό είναι σταθερό αν τα w και c είναι σταθερά. Για να υπολογίσει ακριβώς τον πίνακα $\tilde{H_k}$ χρειάζονται k πολλαπλασιασμοί το οποίο απαιτεί άλλο ένα πέρασμα πάνω από τα δεδομένα και χωρική και χρονική πολυπλοκότητα O(mck).

3.2.3. Χρήσιμα λήμματα

Υπενθυμίζουμε ότι η SVD του $W^TW \in R^{cxc}$ είναι

$$W^TW = \sum_{t=1}^{c} \sigma_t^2(W) z^t z^{t^T} = Z\Sigma_W^2 Z^T, \quad (3.2.1)$$

όπου $Z \in R^{cxc}$. Ορίζουμε $Z_{\alpha,\beta} \in R^{cx(\beta-\alpha+1)}$ να είναι ο πίνακας του οποίου οι στήλες είναι τα α μέσω των β ιδιαζόντων διανυσμάτων του W^TW . Τότε

$$\tilde{H_l} = CZ_{1,l}T, \quad (3.2.2)$$

όπου ο πίνακας $T \in R^{lxl}$ είναι ένας διαγώνιος πίνακας με στοιχεία $T_{tt} = 1/\sigma_t(W)$. Στη συνέχεια ορίζουμε την SVD του πίνακα $\tilde{H_l}$ να είναι

$$\tilde{H_l} = B_l \Sigma_{\tilde{H_l}} D_l^T \quad (3.2.3)$$

και θέτουμε τον πίνακα $\Delta \in R^{lxl}$ να είναι

$$\Delta = TZ_{1,l}^T(C^TC - W^TW)Z_{1,l}T. \quad (3.2.4)$$

Σε αυτή την ενότητα παρουσιάζουμε τέσσερα λήμματα που θα μας βοηθήσουν στην απόδειξη του Θεωρήματος 3.2.1 παρακάτω.

Λήμμα 3.2.1.

Για ξ=2,F, και για κάθε ε>0,

$$\left\|A - \widetilde{H_l}\widetilde{H_l^T}A\right\|_\xi^2 \le \left(1 + \frac{\varepsilon}{100}\right)\left\|A - B_lB_l^TA\right\|_\xi^2 + (1 + \frac{\varepsilon}{100})\left\|B_lB_l^T - \widetilde{H_l}\widetilde{H_l^T}\right\|_\xi^2 \|A\|_\xi^2.$$

Απόδειξη.

$$\left\|A-\widetilde{H_l}\widetilde{H_l^T}A\right\|_\xi^2 \le (\|A-B_lB_l^TA\|_\xi + \left\|B_lB_l^T-\widetilde{H_l}\widetilde{H_l^T}\right\|_\xi \|A\|_\xi)^2. \quad \blacksquare$$

Το λήμμα αποδείχθηκε αφού $(\alpha+\beta)^2 \le (1+\varepsilon)\alpha^2 + (1+1/\varepsilon)\beta^2$ για όλα τα ε≥0.

Στη συνέχεια παρόλο που τα διανύσματα δεν είναι στη γενική τους μορφή ορθοκανονικά , θα περίμενε κανείς από την κατασκευή τους ότι αν ο πίνακας W^TW είναι κοντά στον πίνακα C^TC τότε με μεγάλη πιθανότητα θα είναι προσεγγιστικά ορθοκανονικά.

Το Λήμμα 3.2.2 δηλώνει ότι το Δ που ορίζεται στη σχέση (3.2.4) χαρακτηρίζει πόσο μακριά είναι το H_l από το να έχουμε ορθοκανονικές στήλες και δείχνει ότι το σφάλμα που έχουμε οριοθετείται από μια απλή συνάρτηση του γ και το σφάλμα από το δεύτερο στάδιο της δειγματοληψίας.

Λήμμα 3.2.2.

$$\widetilde{H_l^T}\widetilde{H}_l = I_l + \Delta$$

Επιπλέον, για ξ=2,F,

$$\|\Delta\|_\xi \le \frac{1}{\gamma\|W\|_F^2}\|C^TC - W^TW\|_\xi$$

Απόδειξη.

Υπενθυμίζουμε ότι $\widetilde{H}_l = CZ_{1,l}T$ και ότι $T^TZ_{1,l}^TW^TWZ_{1,l}T = I_l$, οπότε

$$\left\|\widetilde{H_l^T}\widetilde{H}_l - I_l\right\|_\xi = \left\|T^TZ_{1,l}^TC^TCZ_{1,l}T - T^TZ_{1,l}^TW^TWZ_{1,l}T\right\|_\xi \quad (3.2.5)$$

$$= \left\|T^TZ_{1,l}^T(C^TC - W^TW)Z_{1,l}\right\|_\xi. \quad (3.2.6)$$

Από τις ιδιότητες των δύο νορμών και ιδιαίτερα από

$$\|AB\|_\xi \le \|A\|_2\|B\|_\xi, \quad (3.2.7)$$

$$\|AB\|_\xi \le \|A\|_2\|B\|_2. \quad (3.2.8)$$

Για ξ=2,F, έχουμε ότι

$$\left\|\widetilde{H_l^T}\widetilde{H}_l - I_l\right\|_\xi \le \left\|T^TZ_{1,l}^T\right\|_2\|C^TC - W^TW\|_\xi\left\|Z_{1,l}T\right\|_2 \quad (3.2.9)$$

$$\le \|T\|_2^2\|C^TC - W^TW\|_\xi \quad (3.2.10)$$

$$\le \max_{t=1,..,l}(1/\sigma_t^2(W))\left\|C^TC - W^TW\right\|_\xi. \quad (3.2.11)$$

Αφού $\left\|Z_{1,l}\right\|_2 = 1$. Το λήμμα αποδείχθηκε αφού $\sigma_t^2(W) \ge \gamma\|W\|_F^2$ για όλα τα t=1,...,l, από τον ορισμό του l. ■

Στη συνέχεια θεωρούμε τον δεύτερο όρο του Λήμματος 3.2.1, $\left\|B_lB_l^T - \widetilde{H_l}\widetilde{H_l^T}\right\|_\xi^2$ και δείχνουμε ότι μπορεί να σχετίζεται με το $\|\Delta\|_\xi$.

Λήμμα 3.2.3.

Για ξ=2,F,

$$\left\|B_lB_l^T - \widetilde{H_l}\widetilde{H_l^T}\right\|_\xi = \left\|B_l(I_l - \Sigma_{\widetilde{H_l}}^2)B_l^T\right\|_\xi$$

$$= \left\|I_l - \Sigma_{\widetilde{H_l}}^2\right\|_\xi$$

$$= \left\| D_l\left(I_l - \Sigma^2_{\widetilde{H_l}}\right)D_l^T \right\|_\xi$$

$$= \left\| I_l - \widetilde{H_l^T}\widetilde{H_l} \right\|_\xi . \qquad \blacksquare$$

Το Λήμμα 3.2.4 αφορά την ειδική περίπτωση στην οποία οι πιθανότητες που εισάγονται στον αλγόριθμο είναι οι βέλτιστες, όπως συμβαίνει και στο Θεώρημα 3.2.5 που θα δούμε παρακάτω.

Λήμμα 3.2.4.

Έστω ένας πίνακας $A \in R^{mxn}$ορίζουμε μια προσέγγιση του πίνακα $\tilde{H_l}$ η οποία κατασκευάζεται τον αλγόριθμο SVD σταθερού χρόνου επιλέγοντας c στήλες του πίνακα A με πιθανότητες $\{p_i\}_{i=1}^n$ και w γραμμές του πίνακα C με πιθανότητες $\{q_j\}_{j=1}^m$ όπου $p_i = Pr[\, i_t = i] = \left|A^{(i)}\right|^2 / \|A\|_F^2$ και $q_j = Pr[j_t = 1] = \left|C_{(j)}\right|^2 / \|C\|_F^2$. Τότε

$$\|W\|_F = \|C\|_F = \|A\|_F.$$

Απόδειξη.

Αν $p_i = \left|A^{(i)}\right|^2 / \|A\|_F^2$, τότε έχουμε ότι $\|C\|_F^2 = \sum_{t=1}^c \left|C^{(t)}\right|^2 = \sum_{t=1}^c \frac{\left|A^{(i_t)}\right|^2}{c p_{i_t}} = \|A\|_F^2$.

Παρόμοια, αν $q_j = \left|C_{(j)}\right|^2 / \|C\|_F^2$, τότε έχουμε ότι $\|W\|_F^2 = \sum_{t=1}^w \left|W_{(t)}\right|^2 = \sum_{t=1}^w \frac{\left|C_{(i_t)}\right|^2}{w q_{i_t}} = \|C\|_F^2$. Το λήμμα αποδείχθηκε. ■

Θεώρημα 3.2.1.

Έστω ένας πίνακας $A \in R^{mxn}$ ορίζουμε μια προσέγγιση του πίνακα $\tilde{H_l}$ η οποία κατασκευάζεται τον αλγόριθμο SVD σταθερού χρόνου επιλέγοντας c στήλες του πίνακα A με πιθανότητες $\{p_i\}_{i=1}^n$ και w γραμμές του πίνακα C με πιθανότητες $\{q_j\}_{j=1}^m$ όπου $p_i = \left|A^{(i)}\right|^2 / \|A\|_F^2$ και $q_j = \left|C_{(j)}\right|^2 / \|C\|_F^2$.

Έστω η=1+$\sqrt{8\, log(\, 2/\delta)}$ και ε>0. Αν το σφάλμα της Frobenius νόρμας είναι το επιθυμητό και εφόσον ο αλγόριθμος τρέχει με $\gamma = \varepsilon/100k$ τότε διαλέγοντας $c = \Omega(k^2\eta^2/\varepsilon^4)$ στήλες του A και $w = \Omega(k^2\eta^2/\varepsilon^4)$ σειρές του C έχουμε με πιθανότητα τουλάχιστον 1-δ,

$$\left\| A - \widetilde{H_l}\widetilde{H_l^T}A \right\|_F^2 \leq \|A - A_k\|_F^2 + \varepsilon\|A\|_F^2. \qquad (3.2.12)$$

Αν το σφάλμα της Ευκλείδειας νόρμας είναι το επιθυμητό και εφόσον ο αλγόριθμος τρέχει με $\gamma = \varepsilon/100$ τότε διαλέγοντας $c = \Omega(\eta^2/\varepsilon^4)$ στήλες του A και $w = \Omega(k^2\eta^2/\varepsilon^4)$ γραμμές του C έχουμε με πιθανότητα τουλάχιστον 1-δ,

$$\left\| A - \widetilde{H_l}\widetilde{H_l^T}A \right\|_2^2 \leq \|A - A_k\|_2^2 + \varepsilon\|A\|_F^2. \qquad (3.2.13)$$

Απόδειξη.

Στη συνέχεια θα χρησιμοποιήσουμε κάποια λήμματα τα οποία θα μας βοηθήσουν στην απόδειξη του Θεωρήματος 3.2.1.

Ξεκινάμε πρώτα αποδεικνύοντας τη σχέση (3.2.12).

Λήμμα 3.2.5.

$$\left\| A - B_l B_l^T A \right\|_F^2 = \|A\|_F^2 - \left\| B_l^T A \right\|_F^2.$$

Απόδειξη.

$$\left\|A - B_l B_l^T A\right\|_F^2 = Tr((A - B_l B_l^T A)^T (A - B_l B_l^T A))$$

$$= Tr(A^T A - A^T B_l B_l^T A).$$ ■

Έπειτα, θέλουμε να θεωρήσουμε ένα κάτω όριο για την ποσότητα $\left\|B_l^T A\right\|_F^2$,όσον αφορά τις ιδιάζουσες τιμές του πίνακα *W*. Αυτό το κάνουμε σε διάφορα βήματα.

Αρχικά συσχετίζουμε το με το $\left\|B_l^T A\right\|_F^2$ με το $\left\|H_l^T A\right\|_F^2$. Σημειώνουμε ότι η παραδοχή $\|\Delta\|_F < 1$ γίνεται αφού στο Θεώρημα 3.2.1 λαμβάνονται πολλές γραμμές και στήλες. Αν δεν κάνουμε αυτή τη παραδοχή η δειγματοληψία θα είναι πιο δύσκολη.

Λήμμα 3.2.6.

Αν $\|\Delta\|_F < 1$, τότε

$$\left\|B_l^T A\right\|_F^2 \geq (1 - \|\Delta\|_F)\left\|\widetilde{H_l^T} A\right\|_F^2.$$

Απόδειξη.

Αφού $\widetilde{H_l}^T = B_l \Sigma_{\widetilde{H}_l} D_l^T$, ισχύει

$$\left\|\widetilde{H_l^T} A\right\|_F^2 = \left\|\Sigma_{\widetilde{H}_l} B_l^T A\right\|_F^2 \leq \left\|\Sigma_{\widetilde{H}_l}\right\|_2^2 \left\|B_l^T A\right\|_F^2 = \left\|\widetilde{H_l^T}\widetilde{H}_l\right\|_2^2 \left\|B_l^T A\right\|_F^2 \qquad (3.2.14)$$

χρησιμοποιώντας τη σχέση (3.2.7).

Από την τριγωνική ανισότητα έχουμε ότι

$$\left\|\widetilde{H_l^T}\widetilde{H}_l\right\|_2 \geq \left|\|I_l\|_2 - \left\|\widetilde{H_l^T}\widetilde{H}_l - I_l\right\|_2\right| = |1 - \|\Delta\|_2| \,. \qquad (3.2.15)$$

Το λήμμα αποδείχθηκε αφού $\|\Delta\|_2 \leq \|\Delta\|_F < 1$ και παρατηρούμε ότι $1 + x \leq 1/(1 - x)$ για όλα τα $x \leq 1$. ■

Στη συνέχεια, συσχετίζουμε την ποσότητα $\left\|\widetilde{H_l^T} A\right\|_F^2$ με την ποσότητα $\left\|\widetilde{H_l^T} C\right\|_F^2$.

Λήμμα 3.2.7.

$$\left\|\widetilde{H_l^T} A\right\|_F^2 \geq \left\|\widetilde{H_l^T} C\right\|_F^2 - (k + \sqrt{k}\|\Delta\|_F)\|AA^T - CC^T\|_F.$$

Απόδειξη.

Αφού $\left\|\widetilde{H_l^T} A\right\|_F^2 = Tr(\widetilde{H_l^T} AA^T \widetilde{H_l^T})$, έχουμε ότι

$$\left\|\widetilde{H_l^T} A\right\|_F^2 = Tr\left(\widetilde{H_l^T} CC^T \widetilde{H_l^T}\right) + Tr(\widetilde{H_l^T}(AA^T - CC^T)\widetilde{H_l^T})$$

$$\geq \left\|\widetilde{H_l^T} C\right\|_F^2 - \|AA^T - CC^T\|_2 \left\|\widetilde{H_l}\right\|_F^2$$

όπου από την ανισότητα έχουμε ότι

$$\left|Tr(\widetilde{H_l^T}(AA^T - CC^T)\widetilde{H_l^T})\right| \leq \sum_t \left|\left(\widetilde{H_l^T}\right)_{(t)} (AA^T - CC^T)(\widetilde{H_l})^{(t)}\right|$$

$$\leq \|AA^T - CC^T\|_2 \left\|\widetilde{H_l}\right\|_F^2.$$

Το λήμμα αποδείχθηκε αφού $\|\square\|_2 \leq \|\square\|_F$, και αφού

$$\left\|\widetilde{H_l}\right\|_F^2 = \sum_{t=1}^{l}\left|\tilde{h}^{t^T} h^t\right| = \sum_{t=1}^{l} 1 + \Delta_{tt} \leq k + \sqrt{k}\,\|\Delta\|_F. \qquad \blacksquare$$

Έπειτα συσχετίζουμε τις ποσότητες $\left\|\widetilde{H}_l^T C\right\|_F^2$ και $\sum_{t=1}^{l}\sigma_t^2(W)$.

Λήμμα 3.2.8.

$$\left\|\widetilde{H}_l^T C\right\|_F^2 \geq \sum_{t=1}^{l}\sigma_t^2(W) - \frac{2}{\sqrt{\gamma}}\|CC^T - W^T W\|.$$

Απόδειξη.

Αφού $\left\|\widetilde{H}_l^T C\right\|_F^2 = \left\|C^T \widetilde{H}_l\right\|_F^2 = \left\|C^T C Z_{1,l} T\right\|_F^2$, έχουμε ότι

$$\left\|\widetilde{H}_l^T C\right\|_F^2 \geq (\left\|W^T W Z_{1,l} T\right\|_F - \left\|(C^T C - W^T W) Z_{1,l} T\right\|_F)^2$$

$$\geq ((\sum_{t=1}^{l}\sigma_t^2(W)^{\frac{1}{2}} - \frac{1}{\sqrt{\gamma\|W\|_F}}\|C^T C - W^T W\|_F)^2,$$

όπου η δεύτερη ανισότητα χρησιμοποιεί τη σχέση $\|XZ\|_F \leq \|X\|_F$ για κάθε πίνακα X αν ο πίνακας Z έχει ορθοκανονικές στήλες. Πολλαπλασιάζοντας το δεξί μέλος και αγνοώντας κάποιους όρους το λήμμα αποδεικνύεται, αφού $(\sum_{t=1}^{l}\sigma_t^2(W))^{\frac{1}{2}}/\|W\|_F \leq 1$. ■

Συνδυάζοντας τα Λήμματα 3.2.6 , 3.2.7 και 3.2.8 έχουμε το επιθυμητό όριο για την ποσότητα $\left\|B_l^T A\right\|_F^2$ σε σχέση με τις ιδιάζουσες τιμές του πίνακα W.

Τέλος, χρησιμοποιούμε τη θεωρία διαταραχών πίνακα για να συσχετίσουμε τις ποσότητες $\sum_{t=1}^{l}\sigma_t^2(W)$ και $\sum_{t=1}^{k}\sigma_t^2(A)$.

Λήμμα 3.2.9.

$$\sum_{t=1}^{l}\sigma_t^2(W) \geq \sum_{t=1}^{k}\sigma_t^2(A) - \sqrt{k}\|AA^T - CC^T\|_F - \sqrt{k}\|C^T C - W^T W\|_F - (k-l)\gamma\|W\|_F^2.$$

Απόδειξη.

Από την ανισότητα Hoffman-Wielandt βλέπουμε ότι

$$\left|\sum_{\mathrm{t=1}}^{\mathrm{k}}(\sigma_t^2(C) - \sigma_t^2(A))\right| \leq \sqrt{k}(\sum_{t=1}^{k}(\sigma_t^2(C) - \sigma_t^2(A))^2)^{1/2}$$

$$\leq \sqrt{k}(\sum_{t=1}^{k}(\sigma_t(CC^T) - \sigma_t(AA^T))^2)^{1/2}$$

$$\leq \sqrt{k}\|AA^T - CC^T\|_F \qquad (3.2.16)$$

και με παρόμοιο τρόπο έχουμε ότι

$$\left|\sum_{\mathrm{t=1}}^{\mathrm{k}}(\sigma_t^2(W) - \sigma_t^2(C))\right| \leq \sqrt{k}(\sum_{t=1}^{k}(\sigma_t^2(W) - \sigma_t^2(C))^2)^{1/2}$$

$$\leq \sqrt{k}(\sum_{t=1}^{k}(\sigma_t(WW^T) - \sigma_t(CC^T))^2)^{1/2}$$

$$\leq \sqrt{k}\|C^T C - W^T W\|_F. \qquad (3.2.17)$$

Συνδυάζοντας τις σχέσεις (3.2.16) και (3.2.17), βλέπουμε ότι

$$\left|\sum_{t=1}^{k}\sigma_t^2(W) - \sigma_t^2(A)\right| \leq \sqrt{k}\|AA^T - CC^T\|_F + \sqrt{k}\|C^T C - W^T W\|_F. \qquad (3.2.18)$$

Αφού $\sigma_t^2(W) < \gamma\|W\|_F^2$ για όλα τα $t = l+1, \dots, k$, έχουμε τη σχέση

$$\sum_{t=l+1}^{k} \sigma_t^2(W) \leq (k-l)\gamma\|W\|_F^2$$

η οποία μαζί με τη σχέση (3.2.18) αποδεικνύουν το λήμμα. ■

Τώρα συνδέουμε αυτά τα αποτελέσματα προκειμένου να αποδείξουμε τη σχέση (3.2.12).

Έστω $E_{AA^T} = AA^T - CC^T$ και $E_{C^TC} = C^TC - W^TW$.

Αρχικά θεωρούμε ένα κάτω όριο για την ποσότητα $\left\|B_l^T A\right\|_F^2$. Συνδυάζοντας τα Λήμματα 3.2.6 και 3.2.7 και απορρίπτοντας κάποιους όρους έχουμε ότι

$$\left\|B_l^T A\right\|_F^2 \geq \left\|\widetilde{H}_l^T C\right\|_F^2 - \|\Delta\|_F\left\|\widetilde{H}_l^T C\right\|_F^2 - (k+\sqrt{k}\|\Delta\|_F)\left\|E_{AA^T}\right\|_F.$$

Η σχέση αυτή σε συνδυασμό με τα Λήμματα 3.2.8 και 3.2.9 , απορρίπτοντας κάποιους όρους μας δίνει

$$\left\|B_l^T A\right\|_F^2 \geq \sum_{t=1}^{k}\sigma_t^2(A) - \left(k+\sqrt{k}\right)\left\|E_{AA^T}\right\|_F - \left(\sqrt{k}+\frac{2}{\sqrt{\gamma}}\right)\left\|E_{C^TC}\right\|_F - \|\Delta\|_F\sum_{t=1}^{k}\sigma_t^2(A) - \sqrt{k}\,\|\Delta\|_F\left\|E_{AA^T}\right\|_F - (k-l)\gamma\|W\|_F^2 . \quad (3.2.19)$$

Έτσι από το Λήμμα 3.2.5 αμέσως έχουμε ένα άνω όριο για την ποσότητα $\left\|A - B_l B_l^T\right\|_F^2$.

$$\left\|A - B_l B_l^T\right\|_F^2 \leq \|A - A_k\|_F^2 + \left(k+\sqrt{k}\right)\left\|E_{AA^T}\right\|_F + \left(\sqrt{k}+\frac{2}{\sqrt{\gamma}}\right)\left\|E_{C^TC}\right\|_F + \|\Delta\|_F\sum_{t=1}^{k}\sigma_t^2(A) + \sqrt{k}\,\|\Delta\|_F\left\|E_{AA^T}\right\|_F + (k-l)\gamma\|W\|_F^2 .$$

Από τα Λήμμα 3.2.1 και 3.2.3 προκύπτει ότι

$$\left\|A - \widetilde{H}_l\widetilde{H}_l^T\right\|_F^2 \geq \left(1+\frac{\varepsilon}{100}\right)\left\|A - B_l B_l^T\right\|_F^2 + \left(1+\frac{\varepsilon}{100}\right)\|\Delta\|_F^2\|A\|_F^2. \quad (3.2.20)$$

Υπενθυμίζουμε ότι $\gamma = \frac{\varepsilon}{100\kappa}$, $\sum_{t=1}^{k}\sigma_t^2(A) \leq \|A\|_F^2$, $\|\Delta\|_F \leq \left\|E_{C^TC}\right\|_F/\gamma\|W\|_F^2$, από το Λήμμα 3.2.2 και ότι $\|W\|_F = \|C\|_F = \|A\|_F$ από το Λήμμα 3.2.4. Η σχέση (3.2.12) αποδεικνύεται λαμβάνοντας υπόψη τις σχέσεις (3.2.19) και (3.2.20), χρησιμοποιώντας τις κατάλληλες πιθανότητες που αναφέρονται στο θεώρημα και $c, w = \Omega(\frac{k^2\eta^2}{\varepsilon^4})$.

ΚΕΦΑΛΑΙΟ 4. ΥΠΟΛΟΓΙΣΜΟΣ ΜΙΑΣ 'ΣΥΜΠΙΕΣΜΕΝΗΣ' ΠΡΟΣΕΓΓΙΣΤΙΚΗΣ ΔΙΑΣΠΑΣΗΣ ΠΙΝΑΚΑ

Σε πολλές εφαρμογές ένας mxn πίνακας A αποθηκεύεται στο δίσκο και είναι πολύ μεγάλος για να διαβαστεί από τη μνήμη RAM ή να κάνει πρακτικά υπολογισμούς σε υπεργραμμικό πολυωνυμικό χρόνο σε αυτήν. Επομένως ενδιαφερόμαστε για έναν πίνακα A' που υπολογίζεται εύκολα και είναι προσέγγιση του αρχικού πίνακα A.

Έστω c και r δύο αριθμοί θετικοί, συνήθως σταθεροί και ακέραιοι και έστω ότι για κάθε πίνακα X ορίζουμε δύο νόρμες $\|X\|_F$ και $\|X\|_2$ που είναι η Frobenius και η Ευκλείδεια νόρμα αντίστοιχα, όπως τις έχουμε ορίσει παραπάνω.

Θα παρουσιάσουμε δύο αλγορίθμους που υπολογίζουν τον πίνακα A' σαν μια προσέγγιση του πίνακα A. Ο πίνακας A' έχει τις ακόλουθες ιδιότητες:

i) $A'=CUR$ όπου ο C είναι ένας πίνακας mxc που αποτελείται από c τυχαία επιλεγμένες στήλες του A, ο R είναι ένας rxn πίνακς που αποτελείται από r τυχαία επιλεγμένες σειρές του πίνακα A και ο U είναι ένας cxr πίνακας ο οποίος κατασκευάζεται από τους πίνακες C και R.

ii) Οι πίνακες C, U και R μπορούν να ορισθούν μετά από έναν μικρό σταθερό αριθμό περασμάτων (2 ή 3 για τους αλγόριθμους που παρουσιάζουμε εδώ) όλου του πίνακα από το δίσκο.

iii) Ο πίνακας U κατασκευάζεται χρησιμοποιώντας στη RAM επιπλέον χώρο και χρόνο $O(m+n)$ για τον αλγόριθμο γραμμικού χρόνου CUR και $O(1)$ για τον αλγόριθμο σταθερού χρόνου CUR .

iv) Για κάθε ε>0 και για κάθε k τέτοιο ώστε $1 \leq k \leq rank(A)$, μπορούμε να επιλέξουμε τα c και r έτσι ώστε με μεγάλη πιθανότητα ο πίνακας A' να ικανοποιεί τη σχέση

$$\|A - A'\|_2 \leq \min_{D:rank(D)\leq k} \|A - D\|_2 + \varepsilon\|A\|_F$$

και έτσι μπορούμε να επιλέξουμε τα c και r έτσι ώστε $\|A - A'\|_2 \leq \varepsilon\|A\|_F$.

v) Για κάθε ε>0 και για κάθε k τέτοιο ώστε $1 \leq k \leq rank(A)$, μπορούμε να επιλέξουμε τα c και r έτσι ώστε με μεγάλη πιθανότητα ο πίνακας A' να ικανοποιεί τη σχέση

$$\|A - A'\|_F \leq \min_{D:rank(A)\leq k} \|A - D\|_F + \varepsilon\|A\|_F.$$

4.1. ΑΛΓΟΡΙΘΜΟΣ CUR ΓΡΑΜΜΙΚΟΥ ΧΡΟΝΟΥ

Σε αυτή την ενότητα παρουσιάζουμε τον αλγόριθμο linear time CUR ο οποίος υπολογίζει μια κατά προσέγγιση διάσπαση CUR του πίνακα $A \in R^{mxn}$ έχοντας χρονική και χωρική πολυπλοκότητα γραμμική (ως προς τα m και n).

4.1.1. Ο αλγόριθμος

Αλγόριθμος CUR γραμμικού χρόνου

Είσοδος: $A \in R^{mxn}$, r, c, $k \in Z^+$ τέτοιο ώστε $1 \le r \le m$, $1 \le c \le n$ και $1 \le \kappa \le \min(r,c)$, $\{p_i\}_{i=1}^m$ τέτοιο ώστε $p_i \ge 0$ και $\sum_{i=1}^m p_i = 1$ και $\{q_j\}_{j=1}^n$ τέτοιο ώστε $q_j \ge 0$ και $\sum_{j=1}^n q_j = 1$.

Έξοδος : $C \in R^{mxc}$, $U \in R^{cxr}$, $R \in R^{rxn}$

1. Για t=1 to c

(a) Διάλεξε $j_t \in \{1,...,n\}$ με $\Pr[j_t = a] = q_a$, α=1,…,n.

(b) $C^{(t)} = A^{(j_t)} / \sqrt{cq_{j_t}}$

2. Υπολόγισε το $C^T C$ και το SVD του; $C^T C = \sum_{t=1}^c \sigma_t^2(C) y^t y^{t^T}$

3. Αν $\sigma_k(C) = 0$, τότε $k = \max\{k' : \sigma_{k'}(C) \ne 0\}$.

4. Για t=1 μέχρι r

(a) Διάλεξε $i_t \in \{1,...,m\}$ με $\Pr[i_t = a] = p_a$, α=1,..,m.

(b) $R_{(t)} = A_{(i_t)} / \sqrt{rp_{i_t}}$

(c) $\Psi_{(t)} = C_{(i_t)} / \sqrt{rp_{i_t}}$

5. Έστω $\Phi = \sum_{t=1}^k \frac{1}{\sigma_t^2(C)} y^t y^{t^T}$ και $U = \Phi\Psi^T$.

6. Υπολόγισε C , U και R

Έστω ένας πίνακας $A \in R^{mxn}$ θέλουμε να υπολογίσουμε με εύκολο τρόπο έναν πίνακα A' ο οποίος να γράφεται ως $A' = CUR \in R^{mxn}$ και ο οποίος ικανοποιεί τις πέντε ιδιότητες που αναφέραμε προηγουμένως.

Ο αλγόριθμος το πετυχαίνει αυτό κατασκευάζοντας αρχικά έναν πίνακα $C \in R^{mxc}$ χρησιμοποιώντας ένα τυχαία επιλεγμένο υποσύνολο c στηλών του πίνακα A. Οι στήλες επιλέγονται σε c ανεξάρτητες, ταυτόσημες δοκιμές όπου σε κάθε δοκιμή η α στήλη του πίνακα A επιλέγεται με πιθανότητα q_a και κάθε στήλη που επιλέγεται διαιρείται με το $1/\sqrt{cq_a}$ πριν από την καταχώρηση στον πίνακα C.

Ο αλγόριθμος στη συνέχεια κατασκευάζει έναν πίνακα $R \in R^{rxn}$ επανασχηματίζοντας ένα τυχαία επιλεγμένο υποσύνολο r σειρών του πίνακα A. Οι σειρές επιλέγονται σε r ανεξάρτητες ,ταυτόσημες δοκιμές όπου σε κάθε δοκιμή η α σειρά του πίνακα A επιλέγεται με πιθανότητα p_a και κάθε σειρά που επιλέγεται διαιρείται με το $1/\sqrt{rp_a}$ πριν από την καταχώρηση στον πίνακα R.

Χρησιμοποιώντας τις ίδιες τυχαία επιλεγμένες σειρές για την κατασκευή του πίνακα R, ο αλγόριθμος κατασκευάζει επίσης έναν πίνακα Ψ από τον C με τον ίδιο τρόπο.

Έτσι $\Psi \in R^{rxc}$και $\Psi_{i,j} = A_{it_1 jt_2}/\sqrt{crp_{it_1} q_{jt_2}}$όπου το i_{t_1} είναι ένα στοιχείο του συνόλου $\{1,\ldots,m\}$ που επιλέγεται στην t_1 προσπάθεια επιλογής γραμμής και το j_{t_2} είναι στοιχείο του συνόλου $\{1,\ldots,n\}$ που επιλέγεται στην t_2 προσπάθεια επιλογής στήλης.

Ας ορίσουμε τον πίνακα $S_C \in R^{nxc}$ να είναι ένας πίνακας με στοιχεία 0 και 1, όπου $(S_C)_{i,j} = 1$ αν η i στήλη του πίνακα A επιλεγεί στη j ανεξάρτητη τυχαία δοκιμή και $S_{i,j} = 0$ διαφορετικά.

Ας ορίσουμε επίσης τον πίνακα $D_C \in R^{cxc}$να είναι ένας διαγώνιος πίνακας με $(D_C)_{tt} = \frac{1}{\sqrt{cp_{i_t}}}$ όπου το i_t είναι ένα στοιχείο του συνόλου $\{1,\ldots,m\}$ το οποίο επιλέγεται στην t δοκιμή δειγματοληψίας.

Τέλος, για τη δειγματοληψία των σειρών, ορίζουμε με παρόμοιο τρόπο ορίζουμε τους πίνακες $S_R \in R^{rxm}$ και $D_R \in R^{rxr}$.

Ισχύουν οι σχέσεις

$$C = AS_C D_C \quad \text{και} \quad R = D_R S_R A. \tag{4.1.1}$$

Ο πίνακας $S_C D_C$μετασχηματίζει τον πίνακα A αλλάζοντας τις επιλεγμένες στήλες για να δημιουργηθεί ο πίνακας C και ο πίνακας $D_R S_R$ μετασχηματίζει τον πίνακα A αλλάζοντας τις επιλεγμένες σειρές για να δημιουργηθεί ο πίνακας R.

Έτσι έχουμε ότι

$$\Psi = D_R S_R C = D_R S_R A S_C D_C. \tag{4.1.2}$$

Έχοντας έναν πίνακα C ο αλγόριθμος υπολογίζει τις k ιδιάζουσες τιμές $\sigma_t^2(C)$, t=1,..,k, και τα αντίστοιχα ιδιάζοντα διανύσματα y^t, t=1, …., k, του πίνακα $C^T C$.

Υπενθυμίζουμε ότι αυτά είναι τα τετράγωνα των ιδιαζουσών τιμών και των αντίστοιχων δεξιών ιδιαζόντων διανυσμάτων του πίνακα C. Χρησιμοποιώντας αυτές τις ποσότητες, ένας πίνακας $\Phi \in R^{cxc}$ μπορεί να ορισθεί ως

$$\Phi = \sum_{t=1}^{k} \frac{1}{\sigma_t^2(C)} y^t y^{t^T} \tag{4.1.3}$$

και ο πίνακας $U \in R^{cxr}$ κατασκευάζεται ως $U = \Phi\Psi^T$.

Ο πίνακας A προσεγγίζεται από έναν πίνακα A' =CUR όπου ο C είναι ένας mxc πίνακας που αποτελείται από τις c τυχαία επιλεγμένες στήλες του πίνακα A και ο R είναι ένας rxn πίνακας που αποτελείται από τις r τυχαία επιλεγμένες σειρές του πίνακα A και ο $U = \Phi\Psi^T$είναι ένας cxr πίνακας που κατασκευάζεται από τους πίνακες C και R.

Όπως θα δούμε αν οι πιθανότητες με τις οποίες επιλέγονται οι σειρές και οι στήλες είναι οι κατάλληλες, τότε το c και το r μπορεί να επιλεχθούν να είναι σταθερές (ανεξάρτητες από τα m και n αλλά εξαρτώμενες από τα k και ε).

Πριν αποδείξουμε το θεώρημα χρειάζεται να δώσουμε μια βασική ιδέα στο γιατί αν οι πιθανότητες δειγματοληψίας μας ικανοποιούν συγκεκριμένες συνθήκες, τότε το γινόμενο CUR που υπολογίζεται από τον αλγόριθμο μας δίνει καλή προσέγγιση του πίνακα A ικανοποιώντας και τις σχέσεις (i)-(iv) που αναφέραμε παραπάνω. Θέτουμε $h^t = Cy^t/\sigma_t(C)$ να είναι τα αριστερά ιδιάζοντα διανύσματα του πίνακα C. Επίσης αν $H_k = (h^1\ h^2\ \ldots\ h^k) \in R^{mxk}$ τότε $H_k H_k^t A$ είναι η προβολή του A στην περιοχή που εκτείνεται στα ιδιάζοντα διανύσματα του πίνακα C. Αν οι πιθανότητες δειγματοληψίας $\{q_j\}_{j=1}^n$, ικανοποιούν συγκεκριμένες συνθήκες τότε τα k μεγαλύτερα από τα h^t είναι προσεγγίσεις των k αριστερών ιδιαζόντων διανυσμάτων του πίνακα A.

Ο πίνακας $H_k H_k^T A$ είναι μια προσέγγιση του πίνακα A. Κάτι που θα μπορούσε να μας απασχολήσει είναι αν με αυτή τη προσέγγιση ικανοποιούνται οι πέντε ιδιότητες που αναφέραμε παραπάνω. Αυτό ακριβώς είναι που κάνει ο πίνακας CUR.

Ισχύουν οι σχέσεις

$$\widetilde{H_k^T} = H_k^T (D_R S_R)^T \quad \text{και} \quad \widetilde{A} = D_R S_R A. \tag{4.1.4}$$

Το Λήμμα 4.1.1 όπως θα δούμε παρακάτω δείχνει ότι

$$\begin{aligned} CUR &= H_k H_k^T (D_R S_R)^T D_R S_R A \\ &= H_k \widetilde{H_k^T} \widetilde{A}. \end{aligned} \tag{4.1.5}$$

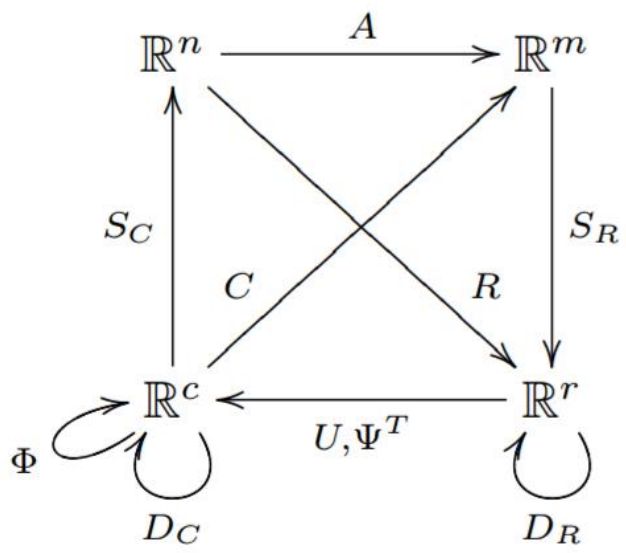
Προκειμένου να βρούμε ένα όριο για την ποσότητα $\|A - CUR\|_\xi$, ξ=2,F, αρχικά εφαρμόζουμε την τριγωνική ανισότητα

$$\|A - CUR\|_\xi \leq \left\|A - H_k H_k^T A\right\|_\xi + \left\|H_k H_k^T A - CUR\right\|_\xi. \tag{4.1.6}$$

Το πρώτο μέλος της σχέσης (4.1.6) μπορούμε να το φράξουμε χρησιμοποιώντας τα αποτελέσματα της SVD [3.11] εάν οι πιθανότητες δειγματοληψίας στήλης πληρούν ορισμένες πιθανότητες.

Για $H_k^T H_k = I_k$ θα δούμε παρακάτω ότι το Λήμμα 4.1.2 δείχνει ότι

$$\begin{aligned} \left\|H_k H_k^T A - CUR\right\|_F &= \left\|H_k^T A - H_k^T (D_R S_R)^T D_R S_R A\right\|_F \\ &= \left\|H_k^T A - \widetilde{H_k^T} \widetilde{A}\right\|_F. \end{aligned} \tag{4.1.7}$$



Σχήμα 4. Διάγραμμα του αλγορίθμου CUR γραμμικού χρόνου.

4.1.2. Ανάλυση της εκτέλεσης και του χρόνου λειτουργίας

Στον αλγόριθμο CUR γραμμικού χρόνου οι πιθανότητες $\{p_i\}_{i=1}^m$ και $\{q_j\}_{j=1}^n$, εφόσον ορισθούν κατάλληλα μπορούν να υπολογιστούν σε ένα πέρασμα και O(c+r) χωρική πολυπλοκότητα συν το χρόνο που χρησιμοποιεί ο αλγόριθμος select.

Έχοντας δεδομένα τα στοιχεία που θα επιλεχθούν, ο πίνακας C για να κατασκευαστεί χρειάζεται ένα επιπλέον πέρασμα. Αυτό απαιτεί χωρική και χρονική πολυπλοκότητα O(mc). Παρόμοια ο πίνακας R μπορεί να κατασκευαστεί στο ίδιο πέρασμα το οποίο απαιτεί χωρική και χρονική πολυπλοκότητα O(nr).

Έχοντας τον πίνακα $C \in R^{mxc}$, ο υπολογισμός του $C^T C$ απαιτεί O(mc) χωρική πολυπλοκότητα και O (mc^2) χρονική πολυπλοκότητα. Ο υπολογισμός της SVD του $C^T C$ απαιτεί χρονική πολυπλοκότητα O (c^3).

Ο πίνακας Ψ μπορεί να κατασκευαστεί στο ίδιο δεύτερο πέρασμα , διαλέγοντας τις ίδιες r γραμμές του πίνακα C, οι οποίες χρησιμοποιούνται για να κατασκευαστεί ο πίνακας R από τον A. Αυτό απαιτεί χωρική και χρονική πολυπλοκότητα O (cr). Ο πίνακας Φ υπολογίζεται εύκολα χρησιμοποιώντας χρονική πολυπλοκότητα $O(c^2k)$. Έτσι δεδομένου ότι τα c, k και r είναι σταθερές απαιτείται συνολικά O$(m+n)$ χωρική και χρονική πολυπλοκότητα.

Αλγόριθμος CUR γραμμικού χρόνου

```
load clown
imagesc(X)
newmap1 = contrast(X);
colormap(newmap1)
AT=X'
A=X;
b=0.99;
k=150;
c=300;
r=190;
m=200;
n=320;
P=[];
Pr=[];


A2F=norm(A','fro');
AF=norm(A,'fro');


%PINAKAS C
for j=1:n
sum=norm(A(:,j))^2;
 P(j)=sum/AF^2;
end

 sum=0;
for w=1:n
 sum = sum + norm(A(:,w))^2;
%*norm(s.X(w,:));
end

for w=1:n
 Pr(w)=norm(A(:,w))^2/sum;
%*norm(s.X(w,:))
end

for i=1:n
 if P(i)>Pr(i)

L(i)=i;
 end
end
L1=nonzeros(L);
x=randi(size(L1,1),1,c);  % Thw
vector of indices of columns.
COL=X(:,L1(x));
for i=1:c
 V1(i)=P(L1(x(i)));
end
for t=1:c
 C(:,t)=COL(:,t)/sqrt(c*V1(t));
end
Z=C'*C;
[U,S,V]=svd(Z);
CF=norm(C,'fro');
%PINAKAS R
A2=A';
for j=1:m
sum1=norm(A2(:,j))^2;
 P1(j)=sum1/A2F^2;
end

 sum1=0;
for w=1:m
 sum1 = sum1 + norm(A2(:,w))^2;
%*norm(s.X(w,:));
end

for w=1:m
 Pr1(w)=norm(A2(:,w))^2/sum1;
%*norm(s.X(w,:))
end

for i=1:m
 if P1(i)>Pr1(i)

L11(i)=i;
 end
end
L2=nonzeros(L11);
x1=randi(size(L2,1),1,r);  % The
vector of indices of columns.
COL1=A2(:,L2(x1));
for i=1:r
 V2(i)=P1(L2(x1(i)));
end
for t=1:r
 RR(:,t)=COL1(:,t)/sqrt(r*V2(t));
 Q(t,:)=C(t,:)/sqrt(r*V2(t));
end
R=RR';
for i=1:k
    f=U(:,i)*V(i,:)/sqrt(S(i,i));
end
U1=f*Q';
errorF=norm(A-
C*U1*R,'fro')/norm(A,'fro')
```

4.1.3. Ανάλυση του σταδίου δειγματοληψίας

Αρχικά παραθέτουμε δύο πολύ χρήσιμα λήμματα.

Λήμμα 4.1.1.

$$CUR = H_k\tilde{H_k^T}\tilde{A}.$$

Απόδειξη.

Σημειώνουμε ότι η SVD του πίνακα C είναι $C^TC = \sum_{t=1}^{c}\sigma_t^2(C)y^t y^{t^T}$ και ο πίνακας $\Psi = D_R S_R C$. Επίσης, $U = \Phi\Psi^T$, όπου $\Phi = \sum_{t=1}^{k}\frac{1}{\sigma_t^2(C)}y^t y^{t^T}$.

Έτσι έχουμε ότι

$$CUR = C(\sum_{t=1}^{k}\frac{1}{\sigma_t^2(C)}y^t y^{t^T})C^T(D_R S_R)^T R$$

$$= (\sum_{t_1}\sigma_{t_1}(C)h^{t_1}y^{t_1{}^T})\,(\sum_{t_2=1}^{k}\frac{1}{\sigma_{t_2}^2(C)}y^{t_2}y^{t_2{}^T})\,(\sum_{t_3}\sigma_{t_3}(C)y^{t_3}h^{t_3{}^T})\,(D_R S_R)^T R$$

$$= (\sum_{t=1}^{k}h^t h^{t^T})\,(D_R S_R)^T R.$$

Το λήμμα αποδείχθηκε αφού $\sum_{t=1}^{k}h^t h^{t^T} = H_k H_k^T$ και $R = D_R S_R A$ και από τις σχέσεις (4.1.4). ■

Λήμμα 4.1.2.

$$\left\|H_k H_k^T A - CUR\right\|_F = \left\|H_k^T A - \tilde{H_k^T}\tilde{A}\right\|_F.$$

Απόδειξη.

Από το Λήμμα 4.1.1 γνωρίζουμε ότι $CUR = H_k\tilde{H_k^T}\tilde{A}$. Ορίζουμε τον πίνακα $\Omega \in R^{kxn}$ως εξής

$$\Omega = H_k^T A - H_k^T(D_R S_R)^T D_R S_R A = H_k^T A - \tilde{H_k^T}\tilde{A}$$

και σημειώνουμε ότι

$$\left\|H_k H_k^T A - H_k\tilde{H_k^T}\tilde{A}\right\|_F^2 = \|H_k\Omega\|_F^2 = Tr(\Omega^T H_k^T H_k\Omega).$$

Το λήμμα αποδείχθηκε αφού $H_k H_k^T = I_k$ και $Tr(\Omega^T\Omega) = \|\Omega\|_F^2$. ■

Στη συνέχεια παρουσιάζουμε το βασικό θεώρημα σχετικά με τον αλγόριθμο LinearTimeCUR. Σημειώνουμε ότι οι πιθανότητες δειγματοληψίας που χρησιμοποιούμε είναι οι βέλτιστες.

Θεώρημα 4.1.1.

Έστω ένας πίνακας $A \in R^{mxn}$ και έστω C, U και R οι πίνακες που κατασκευάζονται από τον αλγόριθμο CUR γραμμικού χρόνου επιλέγοντας c στήλες του πίνακα A με πιθανότητες $\{q_j\}_{j=1}^{n}$ και r γραμμές του πίνακα A με πιθανότητες $\{p_i\}_{i=1}^{m}$. Θεωρούμε ότι $p_i = \left|A_{(i)}\right|^2/\|A\|_F^2$ και $q_j = \left|A^{(j)}\right|^2/\|A\|_F^2$. Τότε

$$E[\|A - CUR\|_F] \le \|A - A_k\|_F + ((\frac{4k}{c})^{1/4} + (\frac{k}{r})^{1/2})\|A\|_F, \quad (4.1.8)$$

$$E[\|A - CUR\|_2] \le \|A - A_k\|_2 + ((\frac{4}{c})^{1/4} + (\frac{k}{r})^{1/2})\|A\|_F. \quad (4.1.9)$$

Έπειτα αν ορίσουμε $\eta_c = 1 + \sqrt{8\,log(\frac{1}{\delta_c})}$ και $\delta = \delta_r + \delta_c$, τότε με πιθανότητα τουλάχιστον 1-δ έχουμε ότι

$$\|A - CUR\|_F \leq \|A - A_k\|_F + ((\frac{4k\eta_c^2}{c})^{1/4} + (\frac{k}{\delta_r^2 r})^{1/2})\|A\|_F, \quad (4.1.10)$$

$$\|A - CUR\|_2 \leq \|A - A_k\|_2 + ((\frac{4\eta_c^2}{c})^{1/4} + (\frac{k}{\delta_r^2 r})^{1/2})\|A\|_F. \quad (4.1.11)$$

Απόδειξη.

Χρησιμοποιώντας την τριγωνική ανισότητα έχουμε ότι

$$\|A - CUR\|_\xi \leq \left\|A - H_k H_k^T A\right\|_\xi + \left\|H_k H_k^T A - CUR\right\|_\xi \quad (4.1.12)$$

Για ξ=2,F, από τα Λήμματα 4.1.2 και 4.1.2 έχουμε ότι:

$$\|A - CUR\|_\xi \leq \left\|A - H_k H_k^T A\right\|_\xi + \left\|H_k^T A - H_k^T (D_R S_R)^T D_R S_R A\right\|_F \quad (4.1.13)$$

$$= \left\|A - H_k H_k^T A\right\|_\xi + \left\|H_k^T A - \tilde{H}_k^T \tilde{A}\right\|_F. \quad (4.1.14)$$

Η τελευταία σχέση προκύπτει από τις σχέσεις (4.1.4).

Σημειώνουμε ότι η επιλογή των στηλών ικανοποιεί τις απαιτήσεις του αλγορίθμου SVD γραμμικού χρόνου που είδαμε στην προηγούμενη ενότητα [3.11]. Έτσι, από το Θεώρημα 3.1.2, έχουμε:

$$\|A - CUR\|_F \leq \|A - A_k\|_F + (4k)^{1/4}\|AA^T - CC^T\|_F^{1/2} + \left\|H_k^T A - \tilde{H}_k^T \tilde{A}\right\|_F, \quad (4.1.15)$$

$$\|A - CUR\|_2 \leq \|A - A_k\|_2 + \sqrt{2}\|AA^T - CC^T\|_F^{1/2} + \left\|H_k^T A - \tilde{H}_k^T \tilde{A}\right\|_F. \quad (4.1.16)$$

Σημειώνουμε ότι οι πιθανότητες με τις οποίες γίνεται η επιλογή των στηλών είναι της μορφής

$$p_k = \frac{|A^{(k)}||B_{(k)}|}{\sum_{k'=1}^{n}|A^{(k')}||B_{(k')}|}$$

με $B = A^T$. Έτσι είναι βέλτιστες και ισχύει ότι $E[\|AA^T - CC^T\|_F] \leq \frac{1}{\sqrt{c}}\|A\|_F^2$.

Επιπλέον, αν οι πιθανότητες με τις οποίες γίνεται η επιλογή των σειρών δεν είναι βέλτιστες $p_k = \frac{|B_{(k)}|^2}{\|B\|_F^2}$, τότε αφού $\left\|H_k^T\right\|_F = \sqrt{k}$, έχουμε ότι

$$E[\left\|H_k^T A - \tilde{H}_k^T \tilde{A}\right\|_F] \leq \sqrt{\frac{k}{r}}\|A\|_F \quad (4.1.17)$$

Έτσι από τις σχέσεις (4.1.15) και (4.1.16) , χρησιμοποιώντας την ανισότητα Jensen και το Θεώρημα 2.1.1 αποδεικνύονται οι σχέσεις (4.1.8) και (4.1.9).

Για τις σχέσεις (4.1.10) και (4.1.11) αρχικά ορίζουμε το ε_ξ, ξ= c , r ως εξής:

$$\varepsilon_c: \|AA^T - CC^T\|_F \leq \frac{\eta_c}{\sqrt{c}}\|A\|_F^2,$$

$$\varepsilon_r: \left\|AA^T - \tilde{H}_k^T \tilde{A}\right\|_F \leq \frac{1}{\delta_r}\sqrt{\frac{k}{r}}\|A\|_F.$$

Από το Θεώρημα 2.1.1 έχουμε ότι $Pr[\varepsilon_c] \geq 1 - \delta_c$. Χρησιμοποιώντας την ανισότητα Markov στη σχέση $\left\|H_k^T A - \tilde{H}_k^T \tilde{A}\right\|_F$ και τη σχέση (4.1.17) έχουμε ότι

$$Pr[\left\|H_k^T A - \tilde{H}_k^T \tilde{A}\right\|_F \geq \frac{1}{\delta_r}\sqrt{\frac{k}{r}}\|A\|_F] \leq \delta_r$$

και έτσι, $Pr[\varepsilon_r] \geq 1 - \delta_r$.

Το θεώρημα αποδεικνύεται χρησιμοποιώντας τις σχέσεις (4.1.15) και (4.1.16) και δεδομένου ότι $\varepsilon_c \cap \varepsilon_r$. ■

Θεώρημα 4.1.2.

Έστω ένας πίνακας $A \in R^{mxn}$ και έστω C, U και R οι πίνακες που κατασκευάζονται από τον αλγόριθμο LinearTimeCUR επιλέγοντας c στήλες του πίνακα A με πιθανότητες $\{q_j\}_{j=1}^{n}$ και r γραμμές του πίνακα A με πιθανότητες $\{p_i\}_{i=1}^{m}$. Θεωρούμε ότι $p_i = \left|A_{(i)}\right|^2 / \|A\|_F^2$ και $q_j = \left|A^{(j)}\right|^2 / \|A\|_F^2$, και έστω $\varepsilon, \varepsilon' > 0$ με $\varepsilon = \varepsilon'/2$.

Αν $c \geq 4k/\varepsilon^4$ και $r \geq k/\varepsilon^2$, τότε

$$E[\|A - CUR\|_F] \leq \|A - A_k\|_F + \varepsilon' \|A\|_F, \tag{4.1.18}$$

και αν $c \geq 4/\varepsilon^4$ και $r \geq k/\varepsilon^2$ τότε

$$E[\|A - CUR\|_2] \leq \|A - A_k\|_2 + \varepsilon' \|A\|_F. \tag{4.1.19}$$

Στη συνέχεια αν θέσουμε $\eta_c = 1 + \sqrt{8\,log(1/\delta_c)}$, $\delta = \delta_r + \delta_c$ και αν $c \geq 4k\eta_c^2/\varepsilon^4$, $r \geq k/\delta_r^2 \varepsilon^2$ τότε με πιθανότητα τουλάχιστον 1-δ έχουμε ότι

$$\|A - CUR\|_F \leq \|A - A_k\|_F + \varepsilon' \|A\|_F, \tag{4.1.20}$$

και αν $c \geq 4\eta_c^2/\varepsilon^4$και αν $r \geq k/\delta_r^2 \varepsilon^2$ τότε με πιθανότητα τουλάχιστον 1-δ έχουμε ότι

$$\|A - CUR\|_2 \leq \|A - A_k\|_2 + \varepsilon' \|A\|_F. \tag{4.1.21}$$

Τα αποτελέσματα των Θεωρημάτων 4.1.1 και 4.1.2 και για τις δύο νόρμες, Frobenius και Ευκλείδεια, ισχύουν για όλα τα k και έχει ιδιαίτερο ενδιαφέρον όταν ο πίνακας A προσεγγίζεται καλά από έναν low rank πίνακα, να μπορούμε να επιλέξουμε k=O(1) και να πετυχαίνουμε καλή προσέγγιση.

Στη συνέχεια αφού $\|A - A_t\|_2 \leq \|A\|_F/\sqrt{t}$ για όλα τα t=1,2,…,r, τα όρια σε σχέση με την Ευκλείδεια νόρμα έχουν την εξής ενδιαφέρουσα ιδιότητα. Από τη σχέση (4.1.19), βλέπουμε ότι

$$E[\|A - CUR\|_2] \leq (1/\sqrt{k} + \varepsilon')\|A\|_F,$$

και παρόμοια για τη σχέση (4.1.21). Έτσι από τις υποθέσεις του Θεωρήματος 2, εάν επιλέξουμε $k = 1/\varepsilon'^2$ και $\varepsilon'' = 2\varepsilon'$, τότε έχουμε

$$E[\|A - CUR\|_2] \leq \varepsilon'' \|A\|_F, \tag{4.1.22}$$

και με πιθανότητα τουλάχιστον 1-δ,

$$\|A - CUR\|_2 \leq \varepsilon'' \|A\|_F. \tag{4.1.23}$$

4.2. ΑΛΓΟΡΙΘΜΟΣ CUR ΣΤΑΘΕΡΟΥ ΧΡΟΝΟΥ

4.2.1. Ο αλγόριθμος

Ο αλγόριθμος ConstantTimeCUR είναι παρόμοιος με τον LinearTimeCUR επομένως θα τονίσουμε μόνο τα κύρια χαρακτηριστικά του δίνοντας έμφαση στις ομοιότητες και στις διαφορές μεταξύ των αλγορίθμων.

Έχοντας έναν πίνακα $A \in R^{mxn}$ θέλουμε να υπολογίσουμε μια προσέγγιση του. Εύκολα υπολογίζουμε τον πίνακα A' που είναι decomposable ως $A' = C\tilde{U}R \in R^{mxn}$ ο οποίος ικανοποιεί τις απαιτήσεις (i)-(v) που αναφέραμε προηγουμένως. Ο πρόσθετος χώρος στη RAM και ο χρόνος για να υπολογίσουμε τον $\tilde{U}$ είναι O(1).

Ο αλγόριθμος σταθερού χρόνου CUR που παρουσιάζεται στο σχήμα το πετυχαίνει αυτό σχηματίζοντας αρχικά έναν πίνακα $C \in R^{mxc}$ κλιμακοποιώντας κατάλληλα ένα τυχαίο δείγμα c

στηλών του πίνακα *A*. Έπειτα σχηματίζει έναν πίνακα $R \in R^{rxn}$ κλιμακοποιώντας κατάλληλα ένα τυχαία επιλεγμένο δείγμα r γραμμών του πίνακα *A*. Τέλος, υπολογίζει τον πίνακα $\Psi \in R^{rxc}$ από τον *C* διαλέγοντας τις ίδιες τυχαία επιλεγμένες σειρές που χρησιμοποιήθηκαν στην κατασκευή του πίνακα *R*.

Αλγόριθμος CUR σταθερού χρόνου

Είσοδος $A \in R^{mxn}$, r ,c, k $\in Z^+$ τέτοια ώστε $1 \leq r \leq m$, $1 \leq c \leq n$, και $1 \leq k \leq \min(r, c)$
$\{p_i\}_{i=1}^m$ *τέτοιο ώστε* $p_i \geq 0$ and $\sum_{i=1}^m p_i = 1$, και $\{q_j\}_{j=1}^n$ τέτοιο ώστε $q_j \geq 0$ και $\sum_{j=1}^n q_j = 1$.

Έξοδος: $\tilde{U} \in R^{cxr}$ και μία περιγραφή του C $\in R^{mxc}$ και R$\in R^{rxn}$

1. Για t=1 μέχρι c
   (a) Διάλεξε $j_t \in \{1, \dots, n\}$ με Pr [j_t= α] $= q_\alpha$ και αποθήκευσε $\{(j_t, q_{j_t}): t = 1, \dots, c\}$
   (b) $C^{(t)} = A^{(j_t)} / \sqrt{cq_{j_t}}$
2. Διάλεξε $\{\pi_i\}_{i=1}^m$ τέτοιο ώστε $\pi_i = |C_i|^2 / \|C\|_F^2$
3. Για t=1 μέχρι w
   (a) Διάλεξε $i_t \in 1, \dots, m$ με Pr $[i_t = \alpha]$= π_α , α=1,..,m
   (b) θέτουμε $W_{(t)} = C_{(i_t)} / \sqrt{w\pi_{i_t}}$
   Υπολόγισε $W^T W$ και το SVD του; $W^T W = \sum_{t=1}^2 \sigma_t^2(W) z^t z^{t^T}$.
4. Αν $\|\cdot\|_F$ bound is desired , set γ=ε/100k
   Else if a $\|\cdot\|_2$ bound is desired , set γ=ε/100.
5. $l = \min\{k, \max\{t: \sigma_t^2(W) \geq \gamma \|W\|_F^2\}\}$.
6. Κράτα τις ιδιάζουσες τιμές $\{\sigma_t(W)\}_{t=1}^l$ και τα αντίστοιχα ιδιάζοντα διανύσματα $\{z^t\}_{t=1}^l$.
7. Για t=1 μέχρι r
   (a) Διάλεξε $i_t \in \{1, \dots, m\}$ με Pr[$i_t = \alpha$]= p_α και αποθήκευσε $\{(i_t, p_{i_t}): t = 1, \dots, r\}$
   (b) $R_{(t)} = A_{(i_t)} / \sqrt{rp_{i_t}}$
   (c) $\Psi_{(t)} = C_{(i_t)} / \sqrt{rp_{i_t}}$
8. Έστω $\tilde{\Phi} = \sum_{t=1}^l \frac{1}{\sigma_t^2(W)} z^t z^{t^T}$ και $\tilde{U} = \tilde{\Phi}\Psi^T$
9. Υπολόγισε $\tilde{U}$, c column labels $\{(j_t, q_{j_t}): t = 1, \dots, c\}$ and r row labels $\{(i_t, p_{i_t}): t = 1, \dots, r\}$

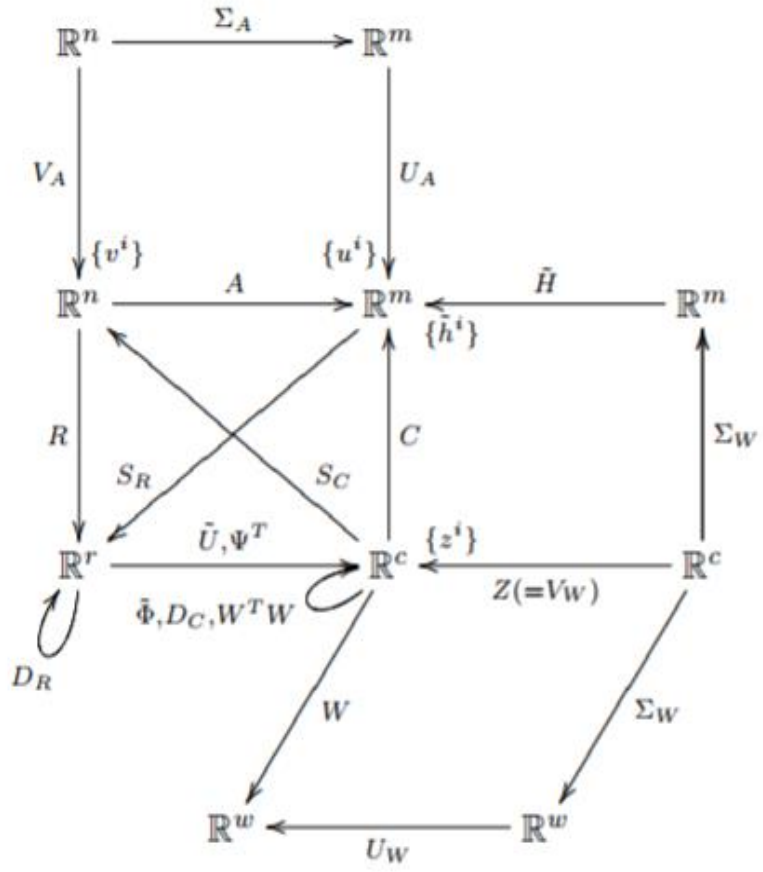
Έχοντας τον πίνακα *C* ο αλγόριθμος επιλέγει τυχαία και κλιμακοποιεί *w* σειρές του πίνακα *C* για να κατασκευάσει τον πίνακα W $\in R^{wxc}$ και έπειτα υπολογίζει τις μεγαλύτερες l ιδιάζουσες τιμές $\sigma_t^2(W)$, t=1,…, l, και τα αντίστοιχα ιδιάζοντα διανύσματα z^t, t= 1,…, l, του πίνακα $W^T W$.

Σημειώνουμε ότι αυτές είναι επίσης προσεγγίσεις ως προς τα τετράγωνα των ιδιαζουσών τιμών και των αντίστοιχων ιδιαζόντων διανυσμάτων του πίνακα *C*. Χρησιμοποιώντας τις ποσότητες αυτές ο πίνακας $\tilde{\Phi} \in R^{cxc}$ μπορεί να οριστεί ως

$$\tilde{\Phi} = \sum_{t=1}^l \frac{1}{\sigma_t^2(W)} z^t z^{t^T}, \qquad (4.2.1)$$

όπου ο πίνακας $\tilde{U} \in R^{rxc}$ κατασκευάζεται ως $\tilde{U} = \tilde{\Phi}\Psi^T$.

Σημειώνουμε ότι οι πίνακες *C* και *R* δεν υπολογίζονται ρητά. Αντίθετα υπολογίζεται ο πίνακας σταθερού μεγέθους $\tilde{U}$ και μόνο ένας σταθερός αριθμός δυαδικών ψηφίων αποθηκεύονται προκειμένου να καθοριστεί ποιες στήλες και ποιες γραμμές του πίνακα *A* θα επιλεγούν για την κατασκευή του πίνακα *C* και *R*, αντίστοιχα.



Σχήμα 5. Διάγραμμα αλγορίθμου CUR σταθερού χρόνου.

Στον σταθερό πρόσθετο χρόνο οι στήλες του πίνακα $\tilde{H}_l$ είναι απλά προσεγγίσεις των k αριστερών ιδιαζόντων διανυσμάτων του πίνακα C αλλά ακόμα έχουμε ότι $\tilde{H}_l^T \tilde{H}_l \approx I_{l.}$ Για να προσδιοριστεί ποσοτικά αυτό θέτουμε $Z_{a,\beta} \in R^{cx(\beta-\alpha+1)}$ να είναι ο πίνακας $W^T W$ και ο πίνακας $T \in R^{lxl}$ είναι ο διαγώνιος πίνακας με στοιχεία $T_{tt=1/\sigma_t(W).}$ Αν ορίσουμε τον πίνακα $\Delta \in R^{lxl}$ ως εξής:

$$\Delta = T Z_{1,l}^T (C^T C - W^T W) Z_{1,l} T, \tag{4.2.2}$$

τότε από το Λήμμα 4.2.3, έχουμε

$$\left\| \tilde{H}_l \tilde{H}_l^T A - C\tilde{U}\tilde{R} \right\|_F \leq (1 + \|\Delta\|_F^{\frac{1}{2}}) \left\| \tilde{H}_l^T A - \tilde{H}_l^T (D_R S_R)^T D_R S_R A \right\|_F \tag{4.2.3}$$

και το Λήμμα 4.2.2 δείχνει ότι $\tilde{H}_l^T \tilde{H}_l = I_l + \Delta$. Έτσι για γραμμικό χρόνο έχουμε ότι

$$\left\| A - C\tilde{U}R \right\|_\xi \leq \left\| A - \tilde{H}_l \tilde{H}_l^T A \right\|_\xi + \left\| \tilde{H}_l \tilde{H}_l^T A - C\tilde{U}R \right\|_\xi$$

για ξ=2,F και έπειτα, θα βρούμε ένα όριο για κάθε έναν από τους δύο όρους ξεχωριστά.

4.2.2. Ανάλυση της εκτέλεσης και του χρόνου λειτουργίας

Στον αλγόριθμο CUR σταθερού χρόνου οι πιθανότητες δειγματοληψίας $\{p_i\}_{i=1}^m$ και $\{q_j\}_{j=1}^n$ αν επιλεγούν σύμφωνα με το Θεώρημα 3.1.1 , μπορούν να υπολογιστούν σε ένα πέρασμα με χωρική και χρονική πολυπλοκότητα O(c+r) χρησιμοποιώντας τον αλγόριθμο select.

Έχοντας τις στήλες του πίνακα A που θα επιλεγούν δεν κατασκευάζεται στην πραγματικότητα ο πίνακας C αλλά περνάμε στο δεύτερο στάδιο δειγματοληψίας όπου επιλέγονται w σειρές του πίνακα C με πιθανότητες $\{\pi_i\}_{i=1}^m$ προκειμένου να κατασκευαστεί ο πίνακας W. Αυτό απαιτεί χωρική και χρονική πολυπλοκότητα O(w). Έπειτα σε ένα τρίτο πέρασμα κατασκευάζεται ο πίνακας W. Αυτό απαιτεί χωρική και χρονική πολυπλοκότητα O(cw).

Παρομοίως μια περιγραφή του πίνακα R κατασκευάζεται στο ίδιο τρίτο πέρασμα έχοντας χωρική και χρονική πολυπλοκότητα $O(r)$. Έπειτα αφού έχει κατασκευαστεί ο πίνακας W υπολογίζεται ο πίνακας $W^T W$ με $O(c^2 w)$ και υπολογίζεται η SVD του $W^T W$ με χρονική πολυπλοκότητα $O(c^3)$. Ο πίνακας Ψ υπολογίζεται στο ίδιο τρίτο πέρασμα διαλέγοντας τις ίδιες r σειρές του πίνακα C που είχαν επιλεγεί και στην κατασκευή του πίνακα R από τον A. Αυτό απαιτεί χρονική πολυπλοκότητα $O(cr)$. Ο πίνακας $\tilde{\Phi}$ υπολογίζεται με χρονική πολυπλοκότητα $O(c^2 k)$ και έπειτα ο πίνακας $U = \tilde{\Phi}\Psi^T$ με $O(c^2 r)$. Έτσι αν τα c, r και k επιλεγούν σαν σταθερές, συνολικά ο αλγόριθμος CUR σταθερού χρόνου απαιτεί συνολικά χωρική και χρονική πολυπλοκότητα $O(1)$ και ικανοποιούνται οι συνθήκες (i)-(iii).

4.2.3 Ανάλυση του σταδίου δειγματοληψίας

Πριν αποδείξουμε το βασικό θεώρημα της ενότητας αυτής θα αποδείξουμε πρώτα κάποια σημαντικά λήμματα.

Λήμμα 4.2.1.

$$C\tilde{U}R = \tilde{H}_l \tilde{H}_l^T (D_R S_R)^T D_R S_R A.$$

Απόδειξη.

Αφού $\tilde{h}^t = Cz^t/\sigma_t(W)$ για t=1,...,l, έχουμε ότι $C = \sum_{t=1}^{c} \sigma_t(W)\, \tilde{h}^t z^{t^T}$.

Επίσης, έχουμε ότι

$$C\tilde{U}R = C\left(\sum_{t=1}^{l} \frac{1}{\sigma_t^2(W)} z^t z^{t^T}\right) C^T (D_R S_R)^T R$$

$$= \left(\sum_{t=1}^{l} \tilde{h}^t \tilde{h}^{t^T}\right)(D_R S_R)^T R.$$

Το λήμμα αποδείχθηκε αφού $\sum_{t=1}^{l} \tilde{h}^t \tilde{h}^{t^T} = \tilde{H}_l \tilde{H}_l^T$ και R=$D_R S_R A$. ■

Λήμμα 4.2.2.

$$\tilde{H}_l^T \tilde{H}_l = I_l + \Delta.$$

Επιπλέον, για ξ=2,

$$\|\Delta\|_\xi \leq \frac{1}{\gamma \|W\|_F^2} \|C^T C - W^T W\|_\xi.$$

Απόδειξη.

Υπενθυμίζουμε ότι $\tilde{H}_l = CZ_{1,l}T$ και ότι $T^T Z_{1,l}^T W^T W Z_{1,l} T = I_l$ οπότε ,

$$\left\| \tilde{H}_l^T \tilde{H}_l - I_l \right\|_\xi = \left\| T^T Z_{1,l}^T C^T C Z_{1,l} T - T^T Z_{1,l}^T W^T W Z_{1,l} T \right\|_\xi \quad (4.2.4)$$

$$= \left\| T^T Z_{1,l}^T (C^T C - W^T W) Z_{1,l} T \right\|_\xi. \quad (4.2.5)$$

Χρησιμοποιώντας τις ιδιότητες της Ευκλείδειας νόρμας και ειδικότερα τις σχέσεις

$$\|AB\|_\xi \leq \|A\|_2 \|B\|_\xi, \quad (4.2.6)$$

$$\|AB\|_\xi \leq \|A\|_\xi \|B\|_2, \quad (4.2.7)$$

για ξ=2,F, έχουμε ότι

$$\left\| \tilde{H}_l^T \tilde{H}_l - I_l \right\|_\xi \leq \left\| T^T Z_{1,l}^T \right\|_2 \|C^T C - W^T W\|_\xi \left\| Z_{1,l} T \right\|_2 \quad (4.2.8)$$

$$\leq \|T\|_2^2 \|C^T C - W^T W\|_\xi \quad (4.2.9)$$

$$\leq \max_{t=1,..,l} \left(\frac{1}{\sigma_t^2(W)}\right) \|C^T C - W^T W\|_\xi, \quad (4.2.10)$$

Αφού $\left\|Z_{1,l}\right\|_2 = 1$. Το λήμμα αποδεικνύεται αφού $\sigma_t^2(W) \geq \gamma\|W\|_F^2$ για t= 1,..,l. ■

Στο επόμενο λήμμα, θα αποδείξουμε τη σχέση (4.2.3).

Λήμμα 4.2.3.

$$\left\|\widetilde{H}_l\widetilde{H}_l^T A - C\widetilde{U}R\right\|_F \leq (1 + \|\Delta\|_F^{\frac{1}{2}})\left\|\widetilde{H}_l^T A - \widetilde{H}_l^T (D_R S_R)^T D_R S_R A\right\|_F.$$

Απόδειξη.

Από το Λήμμα 4.2.1 έχουμε ότι $C\widetilde{U}R = \widetilde{H}_l\widetilde{H}_l^T (D_R S_R)^T D_R S_R A$. Θεωρούμε τον πίνακα $\Omega \in R^{lxn}$ ο οποίος ορίζεται ως

$$\Omega = \widetilde{H}_l^T A - \widetilde{H}_l^T (D_R S_R)^T D_R S_R A.$$

Επίσης, αφού $Tr(XX^T) = \|X\|_F^2$ για τον πίνακα X, έχουμε ότι

$$\begin{aligned}
\left\|\widetilde{H}_l\Omega\right\|_F^2 &= Tr(\Omega^T\widetilde{H}_l^T\widetilde{H}_l\Omega) \\
&= Tr(\Omega^T(I_l + \Delta)\Omega) && (4.2.11) \\
&= \|\Omega\|_F^2 + Tr(\Omega^T\Delta\Omega) \\
&\leq \|\Omega\|_F^2 + \|\Delta\|_2\|\Omega\|_F^2, && (4.2.12)
\end{aligned}$$

όπου η σχέση (4.2.11) προκύπτει από το λήμμα (4.2.2) και η σχέση (4.2.12) ισχύει αφού

$$|Tr(\Omega^T\Delta\Omega) \leq \|\Delta\|_2 Tr(\Omega^T\Omega)|.$$

Άρα

$$\left\|\widetilde{H}_l\widetilde{H}_l^T A - C\widetilde{U}R\right\|_F \leq (1 + \|\Delta\|_F^{\frac{1}{2}})\left\|\widetilde{H}_l^T A - \widetilde{H}_l^T (D_R S_R)^T D_R S_R A\right\|_F$$

και το λήμμα αποδείχθηκε. ■

Τέλος, στο λήμμα που ακολουθεί, θα δείξουμε ότι $\|W\|_F = \|C\|_F = \|A\|_F$ όταν χρησιμοποιούνται οι βέλτιστες πιθανότητες.

Λήμμα 4.2.4.

Έστω ο πίνακας $A \in R^{mxn}$. Ο αλγόριθμος CUR σταθερού χρόνου διαλέγει c στήλες του πίνακα A με πιθανότητες $\{q_j\}_{j=1}^n$ (και έπειτα επιλέγει w γραμμές του πίνακα C με πιθανότητες $\{\{\pi_i\}_{i=1}^m$ για να κατασκευάσει τον πίνακα W) ,στη συνέχεια επιλέγει r γραμμές του πίνακα A με πιθανότητες $\{p_i\}_{i=1}^m$.

Αν $p_i = \left|A_{(i)}\right|^2/\|A\|_F^2$ και $q_j = \left|A^{(j)}\right|^2/\|A\|_F^2$ τότε $\|W\|_F = \|C\|_F = \|A\|_F$.

Απόδειξη.

Αν $p_i = \left|A_{(i)}\right|^2/\|A\|_F^2$, τότε έχουμε ότι $\|C\|_F^2 = \sum_{t=1}^c \left|C^{(t)}\right|^2 = \sum_{t=1}^c \frac{\left|A^{(i_t)}\right|^2}{cp_{i_t}} = \|A\|_F^2$.

Αντίστοιχα, αν $q_j = \left|A^{(j)}\right|^2/\|A\|_F^2$, τότε $\|W\|_F^2 = \sum_{t=1}^w \left|W_{(t)}\right|^2 = \sum_{t=1}^w \frac{\left|C_{(i_t)}\right|^2}{wq_{i_t}} = \|C\|_F^2$. ■

Στη συνέχεια παρουσιάζουμε το βασικό θεώρημα αυτής της ενότητας. Σημειώνουμε ότι στο θεώρημα αυτό θα περιοριστούμε στις πιθανότητες δειγματοληψίας που είναι βέλτιστες , όπως τις ορίσαμε παραπάνω, επιλέγοντας επαρκώς πολλές στήλες και σειρές έτσι ώστε να διασφαλίσουμε ότι το σφάλμα θα είναι μικρότερο από $\varepsilon\|A\|_F$.

Θεώρημα 4.2.1.

Έστω ο πίνακας $A \in R^{mxn}$. Ο αλγόριθμος CUR σταθερού χρόνου διαλέγει c στήλες του πίνακα A με πιθανότητες $\{q_j\}_{j=1}^{n}$ (και έπειτα επιλέγει w γραμμές του πίνακα C με πιθανότητες $\{\{\pi_i\}_{i=1}^{m}$ για να κατασκευάσει τον πίνακα W) ,στη συνέχεια επιλέγει r γραμμές του πίνακα A με πιθανότητες $\{p_i\}_{i=1}^{m}$.

Αν $p_i = \left|A_{(i)}\right|^2 / \|A\|_F^2$ και $q_j = \left|A^{(j)}\right|^2 / \|A\|_F^2$, ορίζουμε $\eta = 1 + \sqrt{8log(\frac{3}{\delta})}$ και $\varepsilon > 0$.

Αν η νόρμα Frobenius είναι η επιθυμητή και συνεπώς ο αλγόριθμος τρέχει με $\gamma=\varepsilon/100k$, τότε αν θέσουμε $c = \Omega\left(\frac{k^2\eta^2}{\varepsilon^8}\right)$,$w = \Omega\left(\frac{k^2\eta^2}{\varepsilon^8}\right)$ και $r = \Omega(\frac{k}{\delta^2\varepsilon^2})$ τότε με πιθανότητα τουλάχιστον 1-δ έχουμε ότι

$$\left\|A - C\tilde{U}R\right\|_F \leq \|A - A_k\|_F + \varepsilon\|A\|_F. \qquad (4.2.13)$$

Αν η Ευκλείδεια νόρμα είναι η επιθυμητή και συνεπώς ο αλγόριθμος τρέχει με $\gamma=\varepsilon/100k$, τότε αν θέσουμε $c = \Omega\left(\frac{\eta^2}{\varepsilon^8}\right)$, $w = \Omega\left(\frac{\eta^2}{\varepsilon^8}\right)$ και $r = \Omega(\frac{k}{\delta^2\varepsilon^2})$, τότε με πιθανότητα τουλάχιστον 1-δ έχουμε ότι

$$\left\|A - C\tilde{U}R\right\|_2 \leq \|A - A_k\|_2 + \varepsilon\|A\|_F. \qquad (4.2.14)$$

Απόδειξη

Ορίζουμε τις σχέσεις:

$$\varepsilon_c: \|AA^T - CC^T\|_F \leq \frac{\eta}{\sqrt{c}}\|A\|_F^2, \qquad (4.2.15)$$

$$\varepsilon_w: \|C^TC - W^TW\|_F \leq \frac{\eta}{\sqrt{w}}\|A\|_F^2, \qquad (4.2.16)$$

$$\varepsilon_r: \left\|\tilde{H}_l^T A - \tilde{H}_l^T (D_R S_R)^T D_R S_R A\right\|_F \leq \frac{3}{\delta\sqrt{r}}\left\|\tilde{H}_l\right\|_F \|A\|_F. \qquad (4.2.17)$$

Χρησιμοποιώντας τα αποτελέσματα του Θεωρήματος 2.1.1 ισχύουν οι σχέσεις (4.2.15) και (4.2.16) με πιθανότητα μεγαλύτερη από $1 - \delta/3$. Θα αποδείξουμε ότι και η σχέση (4.2.17) ισχύει με πιθανότητα μεγαλύτερη από $1 - \delta/3$. Για να το αποδείξουμε αυτό θα πρέπει να δείξουμε ότι

$$E[\left\|\tilde{H}_l^T A - \tilde{H}_l^T (D_R S_R)^T D_R S_R A\right\|_F] \leq \frac{1}{\sqrt{r}}\left\|\tilde{H}_l^T\right\|_F \|A\|_F \qquad (4.2.18)$$

δεδομένου ότι $\Pr[\varepsilon_r] \geq 1 - \delta/3$.

Ισχυριζόμαστε ότι το ε_r ισχύει με πιθανότητα μεγαλύτερη από 1-δ/3. Για να το αποδείξουμε αυτό πρέπει να δείξουμε ότι

$$\mathrm{E}[\left\|\tilde{H}_l^T A - \tilde{H}_l^T (D_R S_R)^T D_R S_R A\right\|_F \leq \frac{3}{\delta\sqrt{r}}\left\|\tilde{H}_l\right\|_F \|A\|_F. \qquad (4.2.19)$$

Από τον ισχυρισμό ότι $\Pr[\varepsilon_r] \geq 1 - \delta/3$. Η σχέση (4.2.19) ισχύει από το Θεώρημα 2.1.1 αφού οι πιθανότητες που χρησιμοποιούνται για να επιλεχθούν οι στήλες του πίνακα $\tilde{H}_l$ και οι αντίστοιχες γραμμές του πίνακα A είναι της μορφής $p_k = \frac{|B_k|^2}{\|B\|_F^2}$. Επίσης από τις υποθέσεις του θεωρήματος έχουμε ότι

$$\Pr[\varepsilon_\xi] \geq 1 - \delta/3 \quad \text{για } \xi=c,w,r.$$

Έπειτα από το Λήμμα 4.2.2 και από την ανισότητα Cauchy-Schwartz προκύπτει ότι

$$\left\|\tilde{H}_l\right\|_F^2 = \sum_{t=1}^{l}\left|\tilde{h}^{t^T}\tilde{h}^t\right| = \sum_{t=1}^{l} 1 + \Delta_{tt} \leq k + \sqrt{k}\|\Delta\|_F. \qquad (4.2.20)$$

Αφού $\sqrt{1+x} \leq 1+\sqrt{x}$ για x≥ 0 , από τις σχέσεις (4.2.18) και (4.2.20) και από την υπόθεση για το ε_r συνεπάγεται ότι

$$\left\|\widetilde{H}_l^T A - \widetilde{H}_l^T (D_R S_R)^T D_R S_R A\right\|_F \leq \frac{3}{\delta\sqrt{r}}(\sqrt{k} + k^{\frac{1}{4}}\|\Delta\|_F^{\frac{1}{2}})\|A\|_F. \quad (4.2.21)$$

Συνδυάζοντας την τελευταία σχέση με το Λήμμα 4.2.3 έχουμε ότι

$$\left\|\widetilde{H}_l\widetilde{H}_l^T A - C\widetilde{U}R\right\|_F \leq \frac{3}{\delta\sqrt{r}}(1+\|\Delta\|_F^{\frac{1}{2}})\,(\sqrt{k} + k^{\frac{1}{4}}\|\Delta\|_F^{\frac{1}{2}})\|A\|_F$$

$$\leq (\frac{9k}{\delta^2 r})^{1/2}(1+\|\Delta\|_F^{\frac{1}{2}})^{\,2}\|A\|_F \quad (4.2.22)$$

$$\leq \left(\frac{9k}{\delta^2 r}\right)^{\frac{1}{2}}(1+3\|\Delta\|_F^{\frac{1}{2}})\,\|A\|_F \quad (4.2.23)$$

$$\leq \left(\frac{9k}{\delta^2 r}\right)^{\frac{1}{2}}\left(1+\frac{3\|C^TC-W^TW\|_F^{\frac{1}{2}}}{\sqrt{\gamma}\|W\|_F}\right)\|A\|_F. \quad (4.2.24)$$

Ας εξετάσουμε αρχικά τη σχέση (4.2.13) για το όριο της νόρμας Frobenius. Υπενθυμίζουμε ότι $\gamma=\varepsilon/100k$. Θα χρησιμοποιήσουμε την τριγωνική ανισότητα για να δείξουμε ότι

$$\left\|A - C\widetilde{U}R\right\|_F \leq \left\|A - \widetilde{H}_l\widetilde{H}_l^T A\right\|_F + \left\|\widetilde{H}_l\widetilde{H}_l^T A - C\widetilde{U}R\right\|_F \quad (4.2.25)$$

και φράσουμε κάθε όρο ξεχωριστά. Αρχικά χρησιμοποιούμε τις πιθανότητες $\{q_j\}_{j=1}^n$ και τις τιμές των $c,w= \Omega\,(k^2\eta^2/\varepsilon^8)$ και από τη συνθήκη $\varepsilon_c \cap \varepsilon_W$ έχουμε ότι

$$\left\|A - \widetilde{H}_l\widetilde{H}_l^T A\right\|_F \leq \|A - A_k\|_F + \frac{\varepsilon}{2}\|A\|_F. \quad (4.2.26)$$

Στη συνέχεια, χρησιμοποιούμε επίσης τις πιθανότητες $\{p_i\}_{i=1}^m$ και τις τιμές των $r = \Omega(k/\delta^2\varepsilon^2)$ και $w= \Omega\,(k^2\eta^2/\varepsilon^8)$ και από τη σχέση (4.2.24), το Λήμμα 4.2.4 και τη συνθήκη $\varepsilon_r \cap \varepsilon_W$

$$\left\|\widetilde{H}_l\widetilde{H}_l^T A - C\widetilde{U}R\right\|_F \leq \frac{\varepsilon}{2}\|A\|_F. \quad (4.2.27)$$

Επίσης, από τη συνθήκη $\varepsilon_c \cap \varepsilon_W \cap \varepsilon_r$ η οποία έχει πιθανότητα τουλάχιστον 1-δ και συνδυάζοντας τις σχέσεις (4.2.26) και (4.2.27), παρατηρούμε ότι αποδεικνύεται η σχέση (4.2.13).

Θα αποδείξουμε τώρα τη σχέση (4.2.14) που δίνει το όριο της Ευκλείδειας νόρμας.

Υπενθυμίζουμε ότι $\gamma=\varepsilon/100$ και

$$\left\|A - C\widetilde{U}R\right\|_2 \leq \left\|A - \widetilde{H}_l\widetilde{H}_l^T A\right\|_2 + \left\|\widetilde{H}_l\widetilde{H}_l^T A - C\widetilde{U}R\right\|_2. \quad (4.2.28)$$

Αρχικά χρησιμοποιούμε τις πιθανότητες $\{q_j\}_{j=1}^n$ και τις τιμές των c,w= $\Omega\,(\eta^2/\varepsilon^8)$ και από τη συνθήκη $\varepsilon_c \cap \varepsilon_W$ έχουμε ότι

$$\left\|A - \widetilde{H}_l\widetilde{H}_l^T A\right\|_2 \leq \|A - A_k\|_2 + \frac{\varepsilon}{2}\|A\|_F. \quad (4.2.29)$$

Στη συνέχεια χρησιμοποιούμε επίσης τις πιθανότητες $\{p_i\}_{i=1}^m$ και τις τιμές των $r = \Omega(k/\delta^2\varepsilon^2)$ και w= $\Omega\,(\eta^2/\varepsilon^8)$ και από τη σχέση (4.2.24) , το Λήμμα 4.2.4 και τη συνθήκη $\varepsilon_r \cap \varepsilon_W$,

$$\left\|\widetilde{H}_l\widetilde{H}_l^T A - C\widetilde{U}R\right\|_2 \leq \frac{\varepsilon}{2}\|A\|_F. \quad (4.2.30)$$

Επίσης, από τη συνθήκη $\varepsilon_c \cap \varepsilon_W \cap \varepsilon_r$ η οποία έχει πιθανότητα τουλάχιστον 1-δ και συνδυάζοντας τις σχέσεις (4.2.29) και (4.2.30) παρατηρούμε ότι αποδεικνύεται η σχέση (4.2.14). ■

Βιβλιογραφία

[1] D. Achlioptas and F. McSherry, Fast computation of low rank matrix approximations, J. ACM, to appear.

[2] D. Achlioptas and F. McSherry, Fast computation of low rank matrix approximations, in Proceedings of the 33rd Annual ACM Symposium on Theory of Computing, 2001, pp. 611– 618.

[3] R. Bhatia, Matrix Analysis, Springer-Verlag, New York, 1997.

[4] P. Drineas and R. Kannan, Pass efficient algorithms for approximating large matrices, in Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms, 2003, pp. 223–232.
[5] P. Drineas, R. Kannan, and M. W. Mahoney, Fast Monte Carlo algorithms for matrices II: Computing a low-rank approximation to a matrix, SIAM J. Comput., 36 (2006), pp. 158–183.

[6] P. Drineas, R. Kannan, and M. W. Mahoney, Fast Monte Carlo algorithms for matrices III: Computing a compressed approximate matrix decomposition, SIAM J. Comput., 36 (2006), pp. 184–206.

[7] P. Drineas, R. Kannan, and M. W. Mahoney, Fast Monte Carlo Algorithms for Matrices I: Approximating Matrix Multiplication, Tech. Report YALEU/DCS/TR-1269, Department of Computer Science, Yale University, New Haven, CT, 2004.

[8] J. Feigenbaum, S. Kannan, M. Strauss, and M. Viswanathan, An approximate L1difference algorithm for massive data sets, in Proceedings of the 40th Annual IEEE Symposium on the Foundations of Computer Science, 1999, pp. 501–511.

[9] G. H. Golub and C. F. Van Loan, Matrix Computations, Johns Hopkins University Press, Baltimore, MD, 1989.

[10] M. R. Henzinger, P. Raghavan, and S. Rajagopalan, Computing on Data Streams, Tech. Report 1998-011, Digital Systems Research Center, Palo Alto, CA, 1998.

[11] R. A. Horn and C. R. Johnson, Matrix Analysis, Cambridge University Press, New York, 1985.

[12] J. I. Munro and M. S. Paterson, Selection and sorting with limited storage, in Proceedings of the 19th Annual IEEE Symposium on Foundations of Computer Science, 1978, pp. 253– 258.

[13] G. W. Stewart and J. G. Sun, Matrix Perturbation Theory, Academic Press, New York, 1990

[14] J. S. Vitter, Random sampling with a reservoir, ACM Trans. Math. Softw., 11 (1985), pp. 37–57.

[15] O. Alter, P. O. Brown, and D. Botstein, Singular value decomposition for genome-wide expression data processing and modeling, Proc. Natl. Acad. Sci. USA, 97 (2000), pp. 10101– 10106.

[16] M. W. Berry, S. T. Dumais, and G. W. O'Brian, Using linear algebra for intelligent information retrieval, SIAM Rev., 37 (1995), pp. 573–595.

[17] S. T. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, Indexing by latent semantic analysis, J. Amer. Soc. Inform. Sci., 41 (1990), pp. 391–407

[18] C. H. Papadimitriou, P. Raghavan, H. Tamaki, and S. Vempala, Latent semantic indexing: A probabilistic analysis, in Proceedings of the 17th ACM Symposium on Principles of Database Systems, 1998, pp. 159–168.

[19] S. Raychaudhuri, J. M. Stuart, and R. B. Altman, Principal components analysis to summarize microarray experiments: Application to sporulation time series, in Proceedings of the Pacific Symposi

[21] R. A. Horn and C. R. Johnson, Matrix Analysis, Cambridge University Press, New York, 1985.

[22] D. Coppersmith and S. Winograd, Matrix multiplication via arithmetic progressions, J. Symbolic Comput., 9 (1990), pp. 251–280.um on Biocomputing 2000, 2000, pp. 455–466.