Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών

Αλγόριθμοι και Πολυπλοκότητα

Διδάσκοντες: Δημήτρης Φωτάκης, Δώρα Σούλιου

1η Σειρά Γραπτών Ασκήσεων - Ημ/νία Παράδοσης 18/11/2021

Άσκηση 1: Αναδρομικές Σχέσεις (1.5 μον.)

Να υπολογίσετε την τάξη μεγέθους $\Theta(\,)$ των λύσεων των παρακάτω αναδρομικών σχέσεων. Για όλες τις σχέσεις, να θεωρήσετε ότι $T(1) = \Theta(1)$.

1. $T(n) = 4\,T(n/2) + \Theta(n^2 \log n)$
2. $T(n) = 5\,T(n/2) + \Theta(n^2 \log n)$
3. $T(n) = T(n/4) + T(n/2) + \Theta(n)$
4. $T(n) = 2\,T(n/4) + T(n/2) + \Theta(n)$
5. $T(n) = T(n^{1/2}) + \Theta(\log n)$
6. $T(n) = T(n/4) + \Theta(\sqrt{n})$.

Άσκηση 2: Προθεματική Ταξινόμηση (3 μον.)

Έστω μη ταξινομημένος πίνακας $A[1..n]$ αρχικοποιημένος σε μια μετάθεση του συνόλου $\{1, \dots, n\}$. Θέλουμε να ταξινομήσουμε τον πίνακα A χρησιμοποιώντας **μόνο** *προθεματικές περιστροφές*. Μια προθεματική περιστροφή υλοποιείται επιλέγοντας κάποιο $k \le n$ και αντιστρέφοντας τη σειρά των k αριστερότερων στοιχείων του πίνακα, ως εξής:

$$\Big[\underline{A[1], A[2], \dots, A[k-1], A[k]}, A[k+1], \dots, A[n]\Big] \;\rightarrow$$
$$\Big[\underline{A[k], A[k-1], \dots, A[2], A[1]}, A[k+1], \dots, A[n]\Big]$$

Στόχος μας είναι να ταξινομήσουμε τον πίνακα A (σε αύξουσα σειρά) κάνοντας όσο το δυνατόν λιγότερες προθεματικές περιστροφές στην χειρότερη περίπτωση.

(α) Να διατυπώσετε αλγόριθμο που ταξινομεί τον A χρησιμοποιώντας το πολύ $2n$ προθεματικές περιστροφές. Εδώ η ανάλυσή σας οφείλει να είναι κατά το δυνατόν ακριβής και να προσδιορίζει πλήρως και τις σταθερές (π.χ., το $2n - 3$ θεωρείται καλύτερο από το $2n$).

(β) Θεωρούμε τώρα ότι οι επιτρεπτές μας κινήσεις είναι οι *προσημασμένες προθεματικές περιστροφές*. Δηλαδή, κάθε φορά που ένα στοιχείο του τρέχοντος πίνακα συμμετέχει σε μία περιστροφή, αντιστρέφουμε το πρόσημό του:

$$\Big[\underline{A[1], A[2], \dots, A[k-1], A[k]}, A[k+1], \dots, A[n]\Big] \;\rightarrow$$
$$\Big[\underline{-A[k], -A[k-1], \dots, -A[2], -A[1]}, A[k+1], \dots, A[n]\Big]$$

Ο στόχος μας παραμένει να ταξινομήσουμε τα στοιχεία έτσι ώστε να εμφανίζονται τελικά σε αύξουσα σειρά και με τα αρχικά (θετικά) τους πρόσημα, παρόλο που στα ενδιάμεσα βήματα, κάποια πρόσημα μπορεί (προσωρινά) να γίνουν αρνητικά.

Να διατυπώσετε αλγόριθμο που ταξινομεί τον πίνακα A χρησιμοποιώντας το πολύ $3n$ προσημασμένες προθεματικές περιστροφές.

(γ) Θα λέμε ότι ένα ζεύγος στοιχείων $(A_t[i], A_t[i+1])$ σε έναν ενδιάμεσο πίνακα A_t (που προκύπτει από τον $A = A_0$ μετά από t διαδοχικές προσημασμένες προθεματικές περιστροφές) είναι *συμβατό* όταν τα στοιχεία $A_t[i]$ και $A_t[i+1]$ έχουν το ίδιο πρόσημο, η απόλυτη τιμή των στοιχείων $A_t[i]$ και $A_t[i+1]$ διαφέρει κατά 1 (άρα οι φυσικοί αριθμοί $|A_t[i]|$ και $|A_t[i+1]|$ είναι διαδοχικοί) και τα στοιχεία $A_t[i]$ και $A_t[i+1]$ είναι ταξινομημένα στον πίνακα A_t (λαμβάνοντας υπόψη και το πρόσημό τους).

Για παράδειγμα στον πίνακα $[4, -2, -1, 3]$, το ζευγάρι $(-2, -1)$ είναι συμβατό (και κανένα άλλο), γιατί $-2 < -1$ και τα $|-2| = 2$ και $|-1| = 1$ είναι διαδοχικοί φυσικοί αριθμοί.

1. Να δείξετε ότι για κάθε ενδιάμεσο πίνακα A_t, που δεν είναι ο $[-1, -2, \ldots, -n]$, μπορούμε να κατασκευάσουμε ένα νέο συμβατό ζεύγος έπειτα από 2 το πολύ προσημασμένες προθεματικές περιστροφές. Να θεωρήσετε, καταχρηστικά, τη μετακίνηση του μεγαλύτερου στοιχείου του A_t στην τελευταία θέση, με θετικό πρόσημο, ως δημιουργία συμβατού ζεύγους.
2. Να διατυπώσετε αλγόριθμο που ταξινομεί τον A χρησιμοποιώντας το πολύ $2n$ προσημασμένες προθεματικές περιστροφές.

Άσκηση 3: Υπολογισμός Κυρίαρχων Θέσεων (1.5 μον.)

Θεωρούμε πίνακα $A[1 \ldots n]$ με n φυσικούς αριθμούς. Για κάθε $i = 2, \ldots, n$, η θέση που *κυριαρχεί* της θέσης i στον πίνακα A είναι η πλησιέστερη θέση που προηγείται της i και η τιμή της ξεπερνά την τιμή $A[i]$. Τυπικά, θεωρώντας ότι $A[0] = \infty$, η θέση που *κυριαρχεί* της θέσης i στον A είναι η μέγιστη θέση j, $0 \leq j < i$, για την οποία ισχύει $A[j] > A[i]$ (βλ. ότι η θέση που κυριαρχεί της 1 είναι η 0). Να διατυπώσετε αποδοτικό αλγόριθμο που υπολογίζει τη θέση που κυριαρχεί της θέσης i, για κάθε $i = 1, \ldots, n$, στον πίνακα $A[1 \ldots n]$. Να αιτιολογήσετε την ορθότητα και την υπολογιστική πολυπλοκότητα του αλγορίθμου σας.

Άσκηση 4: Φόρτιση Ηλεκτρικών Αυτοκινήτων (1.5 μον.)

Διαχειριζόμαστε ένα σταθμό φόρτισης ηλεκτρικών αυτοκινήτων. Γνωρίζουμε ότι στη διάρκεια των επόμενων T χρονικών μονάδων πρόκειται να μας επισκεφθούν για φόρτιση n ηλεκτρικά αυτοκίνητα, καθώς επίσης και το χρόνο άφιξης a_i κάθε αυτοκινήτου i. Υποθέτουμε ότι από τη στιγμή που ένα αυτοκίνητο συνδέεται σε μία υποδοχή φόρτισης, η φόρτισή του διαρκεί 1 χρονική μονάδα. Στα πρόσφατα διαφημιστικά μας μηνύματα, κυριαρχεί η υπόσχεση ότι η καθυστέρηση ενός αυτοκινήτου στον σταθμό μας δεν ξεπερνά τις d χρονικές μονάδες. Με άλλα λόγια, υποσχόμαστε ότι κάθε αυτοκίνητο i, ολοκληρώνει τη φόρτισή του μέχρι τη χρονική στιγμή $a_i + d$. Δεδομένων του d και των χρόνων άφιξης $1 \leq a_1 \leq a_2 \leq \cdots \leq a_n \leq T$ των n ηλεκτρικών αυτοκινήτων, θέλουμε να υπολογίσουμε το ελάχιστο πλήθος s^* υποδοχών φόρτισης που χρειάζονται για να τηρήσουμε την υπόσχεσή μας. Να διατυπώσετε αποδοτικό αλγόριθμο που υπολογίζει το s^*. Να αιτιολογήσετε την ορθότητα και την υπολογιστική πολυπλοκότητα του αλγορίθμου σας.

Άσκηση 5: Επιλογή (2.5 μον.)

(α) Έστω πολυσύνολο (multiset) S με n θετικούς ακέραιους που όλοι είναι μικρότεροι ή ίσοι δεδομένου ακεραίου M. Έχουμε πρόσβαση (μόνο) στην κατανομή F_S των στοιχείων της συλλογής. Συγκεκριμένα, έχουμε στη διάθεσή μας συνάρτηση $F_S(\ell)$ που για κάθε φυσικό ℓ, επιστρέφει το πλήθος των στοιχείων του S που δεν ξεπερνούν το ℓ, δηλ. $F_S(\ell) = \big|\{x \in S : x \leq \ell\}\big|$. Να διατυπώσετε αποδοτικό αλγόριθμο που δέχεται ως είσοδο φυσικό k, $1 \leq k \leq n$, και υπολογίζει (καλώντας την F_S) το k-οστό μικρότερο στοιχείο του S. Να αιτιολογήσετε την ορθότητα του αλγορίθμου σας και να προσδιορίσετε το πλήθος των απαιτούμενων κλήσεων στην F_S (στη χειρότερη περίπτωση). Προσπαθήστε το πλήθος των κλήσεων στην F_S να μην εξαρτάται από το n (μπορεί όμως να εξαρτάται από το M).

(β) Έστω πίνακας διαφορετικών θετικών ακεραίων $A[1 \dots n]$ και έστω M το μέγιστο στοιχείο του A. Θεωρούμε το πολυσύνολο S που αποτελείται από όλες τις μη αρνητικές διαφορές ζευγών στοιχείων του A. Δηλαδή, έχουμε ότι:

$$S = \{A[i] - A[j] : i \neq j \text{ και } A[i] > A[j]\}$$

Να διατυπώσετε αποδοτικό αλγόριθμο που υπολογίζει το k-οστό μικρότερο στοιχείο του S. Να προσδιορίσετε την υπολογιστική πολυπλοκότητα του αλγορίθμου σας (συναρτήσει των n και M) και να αιτιολογήσετε την ορθότητά του. *Υπόδειξη:* Προσπαθήστε να υλοποιήσετε αποδοτικά την F_S και να χρησιμοποιήσετε τον αλγόριθμο του (α).

Άσκηση 6 – Bonus: Ερωτήματα Ανήκειν

Βασιζόμενοι σε hashing, θέλουμε να αναπτύξουμε μια δομή δεδομένων που διατηρεί μια σύνοψη ενός συνόλου S, αποτελούμενου από n θετικούς φυσικούς αριθμούς, και υλοποιεί αποδοτικά (πρακτικά, σε σταθερό χρόνο) τις παρακάτω δύο λειτουργίες: "πρόσθεσε το x στο S" και "έλεγξε αν $x \in S$". Για αυτό τον σκοπό, διατηρούμε μια σύνοψη του S σε πίνακα A με m θέσεις, μεγέθους ενός δυαδικού ψηφίου η καθεμία, και χρησιμοποιούμε συναρτήσεις hash $h : \mathbb{N}_+ \to [m]$ τέτοιες ώστε $\mathbb{P}\text{rob}[h(x) = j] = 1/m$, για κάθε $x \in \mathbb{N}_+$ και για κάθε $j \in [m]$. Απαιτούμε, τέλος, η χρήση τυχαιότητας να μην οδηγεί σε false negatives (δηλ. κάθε αρνητική απάντηση στο ερώτημα "$x \in S$?" πρέπει να είναι σωστή), ενώ μπορεί να οδηγεί σε false positives (δηλ. μπορεί μια θετική απάντηση στο ερώτημα "$x \in S$?" να είναι λανθασμένη) με μικρή όμως πιθανότητα.

(α) Σχεδιάστε μια τέτοια δομή δεδομένων με χρήση μιας μόνο hash συνάρτησης και υπολογίστε την πιθανότητα μιας false positive απάντησης. Ποια είναι η πιθανότητα για false positive απάντηση αν $m = 8n$;

(β) Σχεδιάστε μια τέτοια δομή δεδομένων με χρήση $k \geq 1$ ανεξάρτητων hash συναρτήσεων και υπολογίστε την πιθανότητα μιας false positive απάντησης. Ποια είναι η βέλτιστη τιμή του k και ποια η αντίστοιχη πιθανότητα για false positive απάντηση αν $m = 8n$;

Άσκηση 7 – Bonus: Χωρισμός σε Ομάδες Σχεδόν Ομοίων

Θεωρούμε σύνολο U στο οποίο έχουμε ορίσει μια έννοια μετρικής απόστασης[1] $d : U \times U \to \mathbb{R}_{\geq 0}$. Βασιζόμενοι σε hashing, θέλουμε να αναπτύξουμε μια μέθοδο χωρισμού των στοιχείων του U σε ομάδες, ώστε στοιχεία που είναι κοντινά με βάση τη d να καταλήγουν στην ίδια ομάδα με σχετικά μεγάλη πιθανότητα, και στοιχεία που είναι μακρινά με βάση τη d να καταλήγουν στην ίδια ομάδα με σχετικά μικρή πιθανότητα (αυτή η τεχνική είναι γνωστή ως Locality-Sensitive Hashing – LSH). Τυπικά, λέμε ότι μια οικογένεια συναρτήσεων $\mathcal{H}$ είναι (d_1, d_2, p_1, p_2)-ευαίσθητη για μια έννοια απόστασης d στο σύνολο U αν για κάθε $h \in \mathcal{H}$ και για κάθε $x, y \in U$, (i) αν $d(x, y) \leq d_1$, τότε $\text{Prob}[h(x) = h(y)] \geq p_1$, ενώ (ii) αν $d(x, y) \geq d_2$, τότε $\text{Prob}[h(x) = h(y)] \leq p_2$.

(α) Να βρείτε μια $(d_1, d_2, 1 - d_1/n, 1 - d_2/n)$-ευαίσθητη οικογένεια συναρτήσεων για την Hamming απόσταση δυαδικών συμβολοσειρών μήκους n, για κάθε $n \geq 1$ και για κάθε $0 \leq d_1 < d_2 \leq n$.

(β) Να βρείτε μια $(d/2, 2d, 1/2, 1/3)$-ευαίσθητη οικογένεια συναρτήσεων για την Ευκλείδεια απόσταση σημείων στο $\mathbb{R} \times \mathbb{R}$, για κάθε $d > 0$.

(γ) Να βρείτε μια $(d_1, d_2, 1 - d_1, 1 - d_2)$-ευαίσθητη οικογένεια συναρτήσεων για την Jaccard απόσταση συνόλων, για κάθε $0 \leq d_1 < d_2 \leq 1$. Η Jaccard απόσταση δύο συνόλων S και T ορίζεται ως $J(S, T) = 1 - |S \cap T|/|S \cup T|$.

[1] Μια απόσταση d είναι *μετρική* αν ικανοποιεί τις παρακάτω ιδιότητες: (i) για κάθε $x \in U$, $d(x, x) = 0$, (ii) για κάθε $x, y \in U$, $d(x, y) = d(y, x)$ (συμμετρία), και (iii) για κάθε $x, y, z \in U$, $d(x, z) \leq d(x, y) + d(y, z)$ (τριγωνική ανισότητα).