Αλγοριθμοι και Πολυπλοκοτητα
Μαθημα 7

Ηλίας Κουτσουπιάς

Τμήμα Πληροφορικής και Τηλεπικοινωνιών
Πανεπιστήμιο Αθηνών
elias@@di.uoa.gr

30/03/2009

Εύρεση k-οστού στοιχείου

Το πρόβλημα της εύρεσης k-οστού στοιχείου Δίνεται ακολουθία ακεραίων $a_0, \ldots, a_{n-1}$ και k. Να βρεθεί η τιμή του k-οστού στοιχείου.

- Αν $k = 0$, να βρεθεί το μικρότερο στοιχείο
- Αν $k = 1$, να βρεθεί το δεύτερο μικρότερο στοιχείο, κοκ,
- Αν $k = len(a) - 1$, να βρεθεί το μέγιστο στοιχείο

Παράδειγμα

- Είσοδος: $a = [7, 1, 2, 6, 9, 4, 6]$, $k = 2$
- Έξοδος: 4

Εύρεση k-οστού στοιχείου


```
def order_statistics(k, a):
    """It returns the k-th smallest element of list a. For k=0, it
    returns the minimum element, and for k=len(a)-1, it returns the
    maximum element."""
    if len(a)==1:
        return a[0]
    else:
        m = partition(a[0], a, 1, len(a))
        a[0], a[m-1] = a[m-1], a[0]
        if k==m-1:
            return a[m-1]
        elif (k<m):
            return order_statistics(k, a[:m])
        else:
            return order_statistics(k-m, a[m:])
```

Χρόνος: $O(n)$?

Ευρεση k-οστου στοιχειου - ορθοτητα

- Η partition(a[0], a, 1, len(a)), αφήνει το στοιχείο x=a[0] στην άκρη και χωρίζει τον υπόλοιπο πίνακα σε 2 μέρη. Τα στοιχεία που είναι μικρότερα ή ίσα του x είναι στις θέσεις a[1:m] και τα μεγαλύτερα από x στις θέσεις a[m:n].
- Μετά την αντιμετάθεση των a[0] και a[m-1], ο πίνακας είναι χωρισμένος σε τρία μέρη:
  - τα στοιχεία στις θέσεις a[0:m-1] είναι μικρότερα ή ίσα με x.
  - το a[m-1] είναι ίσο με x.
  - τα στοιχεία στις θέσεις a[m:n] είναι μεγαλύτερα του x.

  Ανάλογα με τη σχέση του k και το m ο αλγόριθμος ψάχνει αναδρομικά το k στοιχείο στο μεσαίο, αριστερό, ή δεξιό τμήμα.
- Για το δεξιό τμήμα χρειάζεται να βρούμε το στοιχείο $k - m$ αντί για k, επειδή αγνοούμε τα m αριστερότερα στοιχεία.

Πως αναλύουμε αναδρομικούς αλγορίθμους;

- Ορίζουμε τη συνάρτηση χρόνου εκτέλεσης $T(n)$. Λέμε δηλαδή: "Έστω $T(n)$ ο χρόνος εκτέλεσης για είσοδο μήκους n".
- Γράφουμε μια αναδρομική σχέση
- Τη λύνουμε και βρίσκουμε το $T(n)$.

Όπως για παράδειγμα αναλύσαμε την mergesort σε προηγούμενο μάθημα.

Παραδειγμα

- Ας αναλύσουμε τον αναδρομικό αλγόριθμο εύρεσης k-οστού σημείου με διαμέριση.
- Τα πράγματα περιπλέκονται από το γεγονός πως δεν μπορούμε να ξέρουμε εκ των προτέρων πόσο θα είναι το μήκος του πίνακα που απομένει, δηλαδή σε ποιό σημείο γίνεται η διαμέριση.
  Θα αναλύσουμε το χρόνο του για τα εξής σενάρια:
  1. Τη χειρότερη περίπτωση: Διαμέριση σε 1 και $n-1$ στοιχεία.
  2. Την 'καλύτερη' περίπτωση: Διαμέριση σε $n/2$ και $n/2$ στοιχεία.
  3. Την 'έτσι-κι-έτσι' περίπτωση: Διαμέριση σε $n/3$ και $2n/3$ στοιχεία.
  4. Την τυχαία περίπτωση: Η διαμέριση είναι σε m και $n-m$ στοιχεία, όπου m είναι τυχαία και ομοιόμορφα κατανεμημένο στο $\{1,\ldots,n-1\}$. Δηλαδή: $Pr(m=a)=1/(n-1)$ για κάθε $a=1,\ldots,n-1$.

Παραδειγμα - Χειροτερη περιπτωση

- Έστω $T(n)$ ο χρόνος του αλγορίθμου για πίνακες με n στοιχεία.
- Στη χειρότερη περίπτωση, η αναδρομική κλήση παίρνει χρόνο $T(n-1)$.
- Η διαμέριση παίρνει χρόνο $\Theta(n)$.
- Οι υπόλοιπες εντολές παίρνουν σταθερό χρόνο.
- Τότε,
$$T(n) = T(n-1) + n$$

Γράφουμε n και όχι $\Theta(n)$. $\Theta(n)$ είναι σύνολο και όχι συνάρτηση.
Ποια η λύση αυτής της αναδρομικής εξίσωσης;
Η ακριβής λύση εξαρτάται από την τιμή $T(1)$.
Αλλά ανεξάρτητα από την τιμή της $T(1)$, η λύση είναι $T(n) = \Theta(n^2)$. (Θα μάθουμε αργότερα το γιατί.)

Παραδειγμα - ‘Καλυτερη περιπτωση

- Έστω $T(n)$ ο χρόνος του αλγορίθμου για πίνακες με n στοιχεία.
- Υποθέτουμε πως η διαμέριση γίνεται στη μέση. Η αναδρομική κλήση παίρνει χρόνο $T(n/2)$.
- Η διαμέριση παίρνει πάλι χρόνο $\Theta(n)$.
- Οι υπόλοιπες εντολές παίρνουν πάλι σταθερό χρόνο.
- Τότε,

$$T(n) = T\left(\frac{n}{2}\right) + n$$

Θυμόμαστε: Γράφουμε n και όχι $\Theta(n)$.
Ποια η λύση αυτής της αναδρομικής εξίσωσης;
Η ακριβής λύση εξαρτάται πάλι από την τιμή $T(1)$.
Αλλά ανεξάρτητα από την τιμή της $T(1)$, η λύση είναι $T(n) = \Theta(n)$. Πολύ μικρότερη από την χειρότερη περίπτωση.

Παραδειγμα - Ἑτσι-κι-ετσι περιπτωση

- Έστω $T(n)$ ο χρόνος του αλγορίθμου για πίνακες με n στοιχεία.
- Υποθέτουμε πως η διαμέριση γίνεται σε $\frac{1}{3} - \frac{2}{3}$. Η αναδρομική κλήση παίρνει χρόνο $T(2n/3)$ (υποθέτουμε πως καταλήγουμε στο μεγάλο τμήμα).
- Η διαμέριση παίρνει πάλι χρόνο $\Theta(n)$.
- Οι υπόλοιπες εντολές παίρνουν πάλι σταθερό χρόνο.
- Τότε,

$$T(n) = T\left(\frac{2n}{3}\right) + n$$

Ποια η λύση αυτής της αναδρομικής εξίσωσης;
Η λύση είναι πάλι $T(n) = \Theta(n)$. Η διαφορά με την καλύτερη περίπτωση βρίσκεται στον κρυμμένο συντελεστή του Θ.

Παραδειγμα - Τυχαια περιπτωση

- Έστω $T(n)$ ο χρόνος του αλγορίθμου για πίνακες με n στοιχεία.
- Υποθέτουμε πως η διαμέριση γίνεται σε m και $n-m$ στοιχεία, όπου m τυχαίος ακέραιος ομοιόμορφα κατανεμημένος στο $\{1, 2, \ldots, n-1\}$.
- Η αναδρομική κλήση παίρνει χρόνο $T(m)$.
- Η διαμέριση παίρνει πάλι χρόνο $\Theta(n)$.
- Οι υπόλοιπες εντολές παίρνουν πάλι σταθερό χρόνο.
- Τότε,

$$T(n) = \frac{1}{n-1}T(1) + \frac{1}{n-1}T(2) + \cdots + \frac{1}{n-1}T(n-1) + n$$

Ποια η λύση αυτής της αναδρομικής εξίσωσης;
Η λύση είναι πάλι $T(n) = \Theta(n)$. Η διαφορά με τις προηγούμενες αναλύσεις (την 'καλύτερη' και την 'έτσι-κι-έτσι') βρίσκεται πάλι στον κρυμμένο συντελεστή του Θ.

Αναδρομικες εξισωσεις

- Πολλές φορές στα προγράμματα κάνουμε χρήση της αναδρομής.
- Ο χρόνος εκτέλεσης του αλγορίθμου μπορεί να γραφτεί εύκολα σαν μια αναδρομική εξίσωση.
- Π.χ. στην δυαδική αναζήτηση αν $T(n)$ ο χρόνος υπολογισμού τότε ισχύει $T(n) = T(n/2) + 2$ αφού θέλουμε μια πράξη για την εύρεση του μεσαίου στοιχείου, μια πράξη για την σύγκριση με το στοιχείο της αναζήτησης και την επίλυση ενός αντίστοιχου προβλήματος διάστασης $n/2$.
- Η παραπάνω αναδρομή για να έχει λύση θα πρέπει να γνωρίζουμε το χρόνο εκτέλεσης στην περίπτωση που η αναδρομή τερματίζει. Στο προηγούμενο παράδειγμα γνωρίζουμε πως $T(0) = 1$ αφού όταν δεν έχουμε ένα στοιχείο επιστρέφεται απλά False.

Πως λυνουμε μια αναδρομικη εξισωση;

- **1$^{\text{ος}}$ τρόπος)** Μαντεύουμε την **ακριβή** λύση και την επιβεβαιώνουμε.

Παραδειγμα

Έστω $T(n) = T(n-1) + 2n - 1$ και $T(1) = 1$.
Υπολογίζουμε

$$\begin{aligned} T(1) &= 1 \\ T(2) &= 1 + 3 = 4 \\ T(3) &= 4 + 5 = 9 \\ T(4) &= 9 + 7 = 16 \end{aligned}$$

Μαντεύουμε $T(n) = n^2$.
Επιβεβαιώνουμε $n^2 = (n-1)^2 + 2n - 1$. ✓

Προσοχη με τους συμβολισμους O, Θ, Ω

- Προσοχή : Χρειάζεται να έχουμε ακριβή λύση και όχι λύση με συμβολισμό Θ (ή O ή Ω). Το Θ δεν είναι συνάρτηση και δεν μπορούμε να κάνουμε πράξεις. Αλλιώς μπορεί να καταλήξουμε σε λάθος αποτέλεσμα με αυτό.

Παραδειγμα (Λαθους)

Έστω $T(n) = T(n-1) + 2n - 1$ και $T(1) = 1$. Μαντεύουμε $T(n) = \Theta(n)$. Επιβεβαιώνουμε $T(n) = \Theta(n)$. Φαίνεται σωστό γιατί $\Theta(n-1) + 2n - 1$ είναι $\Theta(n)$. Αλλά είναι **λάθος**: Άλλη η σταθερά c στο Θ του αριστερού μέρους και άλλη η σταθερά στο $\Theta(n-1) + 2n - 1$ του δεξιού μέρους.

Πως λυνουμε μια αναδρομικη εξισωση;

- **$2^{ος}$ τρόπος)** Ξεδιπλώνουμε την αναδρομή για να βρούμε την γενική μορφή της.

Παραδειγμα

Έστω $T(n) = T(\frac{n}{2}) + n$ και $T(1) = 1$. Την ξεδιπλώνουμε

$$\begin{aligned}
T(n) &= T(\frac{n}{2}) + n \\
&= (T(\frac{n}{4}) + \frac{n}{2}) + n &&= T(\frac{n}{4}) + \frac{3}{2}n \\
&= (T(\frac{n}{8}) + \frac{n}{4}) + \frac{3}{2}n &&= T(\frac{n}{8}) + \frac{7}{4}n \\
&= (T(\frac{n}{16}) + \frac{n}{8}) + \frac{7}{4}n &&= T(\frac{n}{16}) + \frac{15}{8}n
\end{aligned}$$

Παρατηρούμε πως η συνάρτηση έχει τη γενική μορφή

$$T(n) = T(\frac{n}{2^k}) + \frac{2^k - 1}{2^{k-1}}n$$

Παραδειγμα

Παραδειγμα (Συνεχεια)

Αποδεικνύουμε με επαγωγή στο k πως αυτή είναι η σωστή μορφή. Βάση $k = 0$

$$T(n) = T(\frac{n}{2^0}) + \frac{2^0 - 1}{2^{-1}} n \ \checkmark$$

Έστω ότι ισχύει για k. Θα δείξουμε πως ισχύει για $k + 1$:

$$\begin{aligned} T(n) &= T(\frac{n}{2^k}) + \frac{2^k - 1}{2^{k-1}} n \\ &= (T(\frac{n}{2^{k+1}}) + \frac{n}{2^k}) + \frac{2^k - 1}{2^{k-1}} n \\ &= T(\frac{n}{2^{k+1}}) + \frac{1 + 2(2^k - 1)}{2^k} n \\ &= T(\frac{n}{2^{k+1}}) + \frac{2^{k+1} - 1}{2^k} n \checkmark \end{aligned}$$

Παραδειγμα

Παραδειγμα (Συνεχεια)

Έχουμε $T(n) = T(\frac{n}{2^k}) + \frac{2^k-1}{2^{k-1}}n$, για κάθε $k \in \mathbb{N}$ και $T(1) = 1$. Θέτουμε $\frac{n}{2^k} = 1 \Leftrightarrow n = 2^k \Leftrightarrow k = \log n$ και παίρνουμε

$$\begin{aligned} T(n) &= T(1) + \frac{n-1}{\frac{n}{2}}n \\ &= 1 + 2(n-1) \\ &= 2n - 1 \end{aligned}$$

Πως λυνουμε μια αναδρομικη εξισωση;

- **3ος τρόπος)** Μετασχηματίζουμε τη σχέση σε κάποια γνωστή.

Παραδειγμα

Έστω $T(n) = T(\sqrt{n}) + \log n$ και $T(2) = 1$.

Θέτουμε $\begin{array}{ccc} n = 2^k & \text{ή ισοδύναμα} & k = \log n \\ T(n) = f(k) & \text{ή ισοδύναμα} & T(n) = f(\log n) \end{array}$

Ξαναγράφουμε χρησιμοποιώντας f και k (αντί για T και n).

Παρατηρούμε πως $T(\sqrt{n}) = f(\log \sqrt{n}) = f(\frac{1}{2} \log n)$.

Επομένως $f(k) = f(\frac{k}{2}) + k$ και $f(1) = T(2) = 1$. Ξέρουμε τη λύση αυτής της αναδρομικής: $f(k) = 2k - 1$. Και τώρα παίρνουμε τη λύση της αρχικής αναδρομικής:

$T(n) = f(k) = 2k - 1 = 2 \log n - 1$.

Πως λυνουμε μια αναδρομικη εξισωση;

- **4ος τρόπος)** Θυμόμαστε τη λύση κάποιων βασικών αναδρομικών σχέσεων :
  1. Η εξίσωση $T(n) = T(n-1) + n^c$ έχει λύση
  $$T(n) = \Theta(n^{c+1})$$
  2. Η εξίσωση $T(n) = aT(\frac{n}{b}) + n^c$ έχει την εξής λύση.
     - Συγκρίνουμε τους αριθμούς $\frac{\log a}{\log b}$ και c.
     - Αν είναι ίσοι τότε
     $$T(n) = \Theta(n^c \log n)$$
     - Αν διαφέρουν κρατάμε το μεγαλύτερο $c' = \max\{\frac{\log a}{\log b}, c\}$ και τότε
     $$T(n) = \Theta(n^{c'})$$

Παραδειγματα

$$T(n) = T(n-1) + 1 \quad \rightarrow \quad T(n) = \Theta(n)$$

$$T(n) = T(n-1) + n \quad \rightarrow \quad T(n) = \Theta(n^2)$$

$$T(n) = 2T(\frac{n}{2}) + n \quad \rightarrow \quad T(n) = \Theta(n \log n)$$

γιατί $a = 2, b = 2, c = 1$ και $\frac{\log a}{\log b} = 1 = c$

$$T(n) = 4T(\frac{n}{2}) + 1 \quad \rightarrow \quad T(n) = \Theta(n^2)$$

γιατί $a = 4, b = 2, c = 0$ και $\frac{\log a}{\log b} = 2 > c$

$$T(n) = T(\frac{n}{2}) + n \quad \rightarrow \quad T(n) = \Theta(n)$$

γιατί $a = 1, b = 2, c = 1$ και $\frac{\log a}{\log b} = 0 < c$

Ειδικες περιπτωσεις

- Τι κάνουμε με αναδρομικές εξισώσεις της μορφής $T(n) = T(n-1) + n^c \log n$ και $T(n) = aT(\frac{n}{b}) + n^c \log n$;
- Λύνουμε την αναδρομική αγνοώντας το $\log n$ και πολλαπλασιάζουμε τη λύση με $\log n$.
- Η μοναδική εξαίρεση είναι όταν στη δεύτερη αναδρομική έχουμε $c < \frac{\log a}{\log b}$. Σε αυτή την περίπτωση **δεν** πολλαπλασιάζουμε τη λύση με $\log n$, δηλαδή είναι σαν να μην υπάρχει.

Ειδικες περιπτωσεις

Παραδειγμα

Έστω

$$T(n) = 4T(\frac{n}{2}) + n^2 \log n$$

Τότε έχουμε $c \geq \frac{\log a}{\log b}$ και η τελική λύση είναι

$$T(n) = \Theta(n^2(\log n)(\log n)) \quad \text{δηλαδή} \quad T(n) = \Theta(n^2 \log^2 n)$$

Παραδειγμα

Έστω

$$T(n) = 4T(\frac{n}{2}) + n \log n$$

Τότε έχουμε $c < \frac{\log a}{\log b}$ και η τελική λύση είναι

$$T(n) = \Theta(n^2)$$

Βασικες αναδρομικες σχεσεις

- Δεν θα αποδείξουμε την ορθότητα των βασικών αναδρομικών εξισώσεων.
- Η λύση της αναδρομικής σχέσης $T(n) = T(n-1) + n^c$ προκύπτει αν την ξεδιπλώσουμε :

$$\begin{aligned} T(n) &= T(n-1) + n^c \\ &= T(n-2) + (n-1)^c + n^c \\ &= T(n-3) + (n-2)^c + (n-1)^c + n^c \\ &= \vdots \\ &= 1^c + 2^c + \cdots + n^c = \Theta(n^{c+1}) \end{aligned}$$

- Η λύση της αναδρομικής σχέσης $T(n) = aT(n/b) + n^c$ είναι κάπως πιο περίπλοκη.