
2 Creșterea funcțiilor

Ordinul de creștere a timpului de execuție al unui algoritm, definit în capitolul 1 dă o caracterizare simplă pentru eficiența algoritmilor și ne dă în același timp posibilitatea de a compara performanțele relative ale unor algoritmi alternativi. De îndată ce dimensiunea n a datelor de intrare devine suficient de mare, algoritmul de sortare prin interclasare, cu timpul de execuție $\Theta(n \lg n)$ în cazul cel mai defavorabil, este mai performant decât sortarea prin inserție, al cărei timp de execuție în cazul cel mai defavorabil este $\Theta(n^2)$. Deși uneori se poate determina exact timpul de execuție al unui algoritm, așa cum am făcut pentru sortarea prin inserție în capitolul 1, în general, o precizie suplimentară nu merită efortul făcut. Pentru date de intrare suficient de mari, constantele multiplicative și termenii de ordin inferior din expresia exactă a unui timp de execuție sunt dominați de efectele dimensiunii datelor de intrare în sine.

Când examinăm date de intrare de dimensiuni suficient de mari, astfel încât numai ordinul de creștere a timpului de execuție să fie relevant, studiem eficiența *asimptotică* a algoritmilor. Cu alte cuvinte, suntem interesați de modul de creștere a timpului de execuție a unui algoritm o dată cu mărirea dimensiunii datelor de intrare, *în cazul limit* când dimensiunea datelor de intrare crește nemărginit. În mod uzual, un algoritm care este asimptotic mai eficient va fi cea mai bună alegere, cu excepția cazului în care datele de intrare sunt de dimensiune foarte mică.

Acest capitol oferă mai multe metode standard pentru simplificarea analizei asimptotice a algoritmilor. Secțiunea următoare începe prin a defini mai multe tipuri de “notație asimptotică”, dintre care am întâlnit deja Θ -notația. Vor fi prezentate mai multe convenții de notație care vor fi utilizate pe parcursul cărții și în final vom trece în revistă comportamentul funcțiilor care apar frecvent în analiza algoritmilor.

2.1. Notația asimptotică

Notațiile pe care le utilizăm pentru a descrie timpul asimptotic de execuție al unui algoritm sunt definite cu ajutorul unor funcții al căror domeniu de definiție este de fapt mulțimea numerelor naturale $\mathbb{N} = \{0, 1, 2, \dots\}$. Astfel de notații sunt convenabile pentru a descrie funcția timp de execuție în cazul cel mai defavorabil $T(n)$ care în mod uzual este definită doar pentru dimensiuni întregi ale datelor de intrare. Totuși, este necesar uneori să folosim *în mod abuziv* notația asimptotică într-o varietate de moduri. De exemplu, notația se extinde cu ușurință la domeniul numerelor reale sau, alternativ, se poate restrânge la o submulțime a mulțimii numerelor naturale. Este important totuși să înțelegem sensul exact al notației pentru ca atunci când este utilizată în mod abuziv să nu fie utilizată în mod *incorect*. În această secțiune se definesc principalele notații asimptotice și de asemenea, se introduc câteva abuzuri de notație mai des întâlnite.

Θ -notația

În capitolul 1 am arătat că timpul de execuție pentru sortarea prin inserție în cazul cel mai defavorabil este $T(n) = \Theta(n^2)$. Să definim acum exact această notație. Pentru o funcție dată

$g(n)$, notăm cu $\Theta(g(n))$, mulțimea de funcții

$$\Theta(g(n)) = \{ f(n) : \text{există constantele pozitive } c_1, c_2 \text{ și } n_0 \text{ astfel încât} \\ 0 \leq c_1 g(n) \leq f(n) \leq c_2 g(n) \text{ pentru orice } n \geq n_0 \}.$$

O funcție $f(n)$ aparține mulțimii $\Theta(g(n))$ dacă există constantele pozitive c_1 și c_2 astfel încât aceasta să fie plasată între $c_1 g(n)$ și $c_2 g(n)$ pentru n suficient de mare. Deși $\Theta(g(n))$ este o mulțime, vom scrie " $f(n) = \Theta(g(n))$ " pentru a indica faptul că $f(n)$ este un membru al acesteia. Acest abuz al utilizării semnelui de egalitate pentru a desemna apartenența la o mulțime poate să pară confuz la început, dar vom vedea mai târziu că are avantajele sale.

Figura 2.1 (a) dă o imagine intuitivă a funcțiilor $f(n)$ și $g(n)$, când $f(n) = \Theta(g(n))$. Pentru toate valorile lui n aflate la dreapta lui n_0 , valoarea lui $f(n)$ este mai mare sau egală cu $c_1 g(n)$ și mai mică sau egală cu $c_2 g(n)$. Cu alte cuvinte, pentru orice $n \geq n_0$, s-ar putea spune că funcția $f(n)$ este egală cu $g(n)$ până la un factor constant. Vom spune că $g(n)$ este o **margine asimptotic tare** pentru $f(n)$.

Definiția lui $\Theta(g(n))$ cere ca orice membru $f(n) \in \Theta(g(n))$ să fie **asimptotic nenegativ**, adică $f(n)$ să fie nenegativă pentru n destul de mare. O funcție asimptotic pozitivă este strict pozitivă pentru orice n suficient de mare. Prin urmare, funcția $g(n)$ însăși ar trebui să fie nenegativă, altfel mulțimea $\Theta(g(n))$ este vidă. Din acest motiv vom presupune în continuare că orice funcție utilizată în cadrul Θ -notației este asimptotic nenegativă. Această presupunere rămâne valabilă și pentru celelalte notații asimptotice definite în acest capitol.

În capitolul 1 am introdus o noțiune neriguroasă de Θ -notație care ducea la neglijarea termenilor de ordin inferior și a coeficientului constant al termenului de ordin maxim. Să justificăm pe scurt această imagine intuitivă, utilizând de data aceasta definiția, pentru a arăta că $\frac{1}{2}n^2 - 3n = \Theta(n^2)$. Pentru a face aceasta, trebuie determinate constantele pozitive c_1 , c_2 și n_0 astfel încât

$$c_1 n^2 \leq \frac{1}{2}n^2 - 3n \leq c_2 n^2$$

pentru orice $n \geq n_0$. Împărțind cu n^2 , se obține

$$c_1 \leq \frac{1}{2} - \frac{3}{n} \leq c_2.$$

Inegalitatea din dreapta va fi verificată pentru orice $n \geq 1$ dacă alegem $c_2 \geq 1/2$. În același fel, inegalitatea din stânga va fi verificată pentru orice valoare $n \geq 7$ dacă alegem $c_1 \leq 1/14$. Astfel, luând $c_1 = 1/14$, $c_2 = 1/2$ și $n_0 = 7$, se poate verifica faptul că $\frac{1}{2}n^2 - 3n = \Theta(n^2)$. Evident, sunt posibile și alte alegeri pentru aceste constante, însă faptul important aici este doar acela că acestea există. Să mai observăm că aceste constante depind de funcția $\frac{1}{2}n^2 - 3n$; o altă funcție aparținând lui $\Theta(n^2)$ va cere de regulă alte constante.

Putem utiliza definiția formală și pentru a verifica faptul că $6n^3 \neq \Theta(n^2)$. Să presupunem, în scopul obținerii unei contradicții, că există c_2 și n_0 astfel încât $6n^3 \leq c_2 n^2$ pentru orice $n \geq n_0$. De aici, $n \leq c_2/6$, inegalitate care nu poate fi adevărată pentru n oricât de mare deoarece c_2 este o constantă.

Intuitiv, termenii de ordin inferior ai unei funcții asimptotic pozitive pot fi ignorați în determinarea unor margini asimptotic strânse, deoarece aceștia sunt nesemnificativi pentru valori mari ale lui n . O fracțiune infimă a termenului de rang maxim este suficientă pentru a domina termenii de rang inferior. Astfel, dând lui c_1 o valoare ceva mai mică decât coeficientul termenului

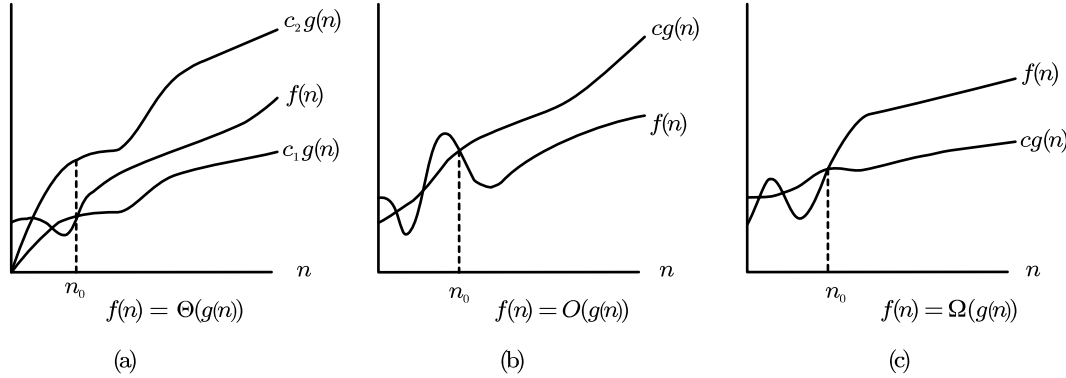


Figura 2.1 Exemple grafice pentru notațiile Θ , O și Ω . În fiecare parte, valoarea indicată a lui n_0 este cea mai mică valoare posibilă; orice valoare mai mare va funcționa, de asemenea. **(a)** Θ -notația mărginește o funcție până la factori constanți. Scriem $f(n) = \Theta(g(n))$ dacă există constantele pozitive n_0 , c_1 și c_2 astfel încât la dreapta lui n_0 , valoarea lui $f(n)$ se află întotdeauna între $c_1 g(n)$ și $c_2 g(n)$ inclusiv. **(b)** O -notația dă o margine superioară pentru o funcție până la un factor constant. Scriem $f(n) = O(g(n))$ dacă există constantele pozitive n_0 și c astfel încât la dreapta lui n_0 , valoarea lui $f(n)$ este întotdeauna mai mică sau egală cu $c g(n)$. **(c)** Ω -notația dă o margine inferioară pentru o funcție până la un factor constant. Scriem $f(n) = \Omega(g(n))$ dacă există constantele pozitive n_0 și c astfel încât la dreapta lui n_0 , valoarea lui $f(n)$ este întotdeauna mai mare sau egală cu $c g(n)$.

de rang maxim și dând lui c_2 o valoare ceva mai mare, s-ar putea ca inegalitățile din definiția Θ -notației să fie satisfăcute. Coeficientul termenului de rang maxim poate fi de asemenea ignorat, deoarece acesta poate schimba c_1 și c_2 doar cu un factor constant.

Pentru a exemplifica, să considerăm o funcție pătratică oarecare $f(n) = an^2 + bn + c$, în care a , b și c sunt constante iar $a > 0$. Eliminarea termenilor de rang inferior lui 2 și neglijarea constantelor conduce la $f(n) = \Theta(n^2)$. Pentru a arăta același lucru în mod formal, alegem constantele $c_1 = a/4$, $c_2 = 7a/4$ și $n_0 = 2 \cdot \max\{|b|/a, \sqrt{|c|/a}\}$. Cititorul poate arăta că $0 \leq c_1 n^2 \leq an^2 + bn + c \leq c_2 n^2$ pentru orice $n \geq n_0$. În general, pentru orice polinom $p(n) = \sum_{i=0}^d a_i x^i$, unde a_i sunt constante și $a_d > 0$, avem $p(n) = \Theta(n^d)$. (vezi problema 2-1)

Deoarece orice constantă este un polinom de gradul 0, putem exprima orice funcție constantă ca $\Theta(n^0)$ sau $\Theta(1)$. Această ultimă notație este un mic abuz deoarece nu este clar care variabilă tinde la infinit.¹ Vom folosi des notația $\Theta(1)$ înțelegând prin aceasta fie o constantă, fie o funcție constantă.

***O*-notația**

Θ -notația delimitează o funcție asimptotic inferior și superior. Când avem numai *o margine asimptotică superioară*, utilizăm O -notația. Fiind dată o funcție $g(n)$, vom nota cu $O(g(n))$,

¹ Adevărata problemă este aceea că notația noastră obișnuită pentru funcții nu face distincție între funcții și valorile lor. În λ -calcul, parametrii unei funcții sunt clar specificați: funcția n^2 poate fi scrisă ca $\lambda n \cdot n^2$ sau chiar $\lambda r \cdot r^2$. Adoptarea unei notații mai riguroase, pe de altă parte, ar complica manipulările algebrice, așa că preferăm să tolerăm abuzul.

mulțimea

$$O(g(n)) = \{f(n) : \text{există constantele pozitive } c \text{ și } n_0 \text{ astfel încât} \\ 0 \leq f(n) \leq cg(n) \text{ pentru orice } n \geq n_0\}.$$

Vom utiliza O -notația pentru a indica o margine superioară a unei funcții până la un factor constant. Figura 2.1(b) dă imaginea intuitivă aflată în spatele O -notației. Pentru toate valorile lui n aflate la dreapta lui n_0 , valorile corespunzătoare ale lui $f(n)$ sunt egale sau mai mici decât cele ale lui $g(n)$.

Pentru a indica faptul că o funcție $f(n)$ este un membru al lui $O(g(n))$, vom scrie $f(n) = O(g(n))$. Să notăm faptul că $f(n) = \Theta(g(n))$ implică $f(n) = O(g(n))$ deoarece Θ -notația este o noțiune mai puternică decât O -notația. În limbajul teoriei mulțimilor, avem $\Theta(g(n)) \subseteq O(g(n))$. Astfel, demonstrația faptului că orice funcție pătratică $an^2 + bn + c$, unde $a > 0$, în $\Theta(n^2)$, arată de asemenea că orice funcție pătratică este în $O(n^2)$. Ceea ce poate părea și mai surprinzător este că orice funcție *liniar* $an + b$ este în $O(n^2)$, ceea ce se verifică ușor, luând $c = a + |b|$ și $n_0 = 1$.

Unor cititori, care s-au mai întâlnit cu notații de tipul O , ar putea să li se pară ciudat că scriem $n = O(n^2)$. În literatură, O -notația este folosită uneori informal pentru a descrie margini asimptotice strânse, adică ceea ce am definit utilizând Θ -notația. În această carte, totuși, atunci când scriem $f(n) = O(g(n))$, afirmăm, în esență, că un anumit multiplu constant al lui $g(n)$ este o margine asimptotică superioară a lui $f(n)$, fără să spunem nimic despre cât de strânsă este această margine superioară. Distincția dintre marginile asimptotice superioare și marginile asimptotice strânse a devenit standard în literatura consacrată algoritmilor.

Folosind O -notația, se poate adeseori descrie timpul de execuție al unui algoritm inspectând structura globală a algoritmului. De exemplu, structura de ciclu dublu a algoritmului de sortare prin inserție din capitolul 1 asigură imediat o margine superioară de tipul $O(n^2)$ pentru timpul de execuție în cazul cel mai defavorabil: costul ciclului interior este mărginit superior de $O(1)$ (constantă), indicii i, j sunt, ambii, cel mult n iar ciclul interior este executat cel mult o dată pentru fiecare dintre cele n^2 perechi de valori pentru i și j .

Deoarece O -notația descrie o margine superioară, atunci când o folosim pentru delimitarea timpului de execuție în cazul cel mai defavorabil, prin implicație, delimităm, de asemenea, timpul de execuție al algoritmului pentru date de intrare arbitrare. Astfel, delimitarea $O(n^2)$ pentru timpul de execuție în cazul cel mai defavorabil al sortării prin inserție se aplică, în egală măsură, timpului său de execuție pentru orice date de intrare. Delimitarea $\Theta(n^2)$ a timpului de execuție în cazul cel mai defavorabil al sortării prin inserție, totuși, nu implică o delimitare $\Theta(n^2)$ asupra timpului de execuție al sortării prin inserție pentru *orice* date de intrare. De exemplu, am văzut în capitolul 1 că dacă datele de intrare sunt deja sortate, sortarea prin inserție se execută în timpul $\Theta(n)$.

Tehnic, este o inexactitate să se spună că timpul de execuție la sortarea prin inserție este $O(n^2)$ deoarece pentru un n dat timpul real de execuție depinde de forma datelor de intrare de dimensiune n . Prin urmare, timpul de execuție nu este de fapt tocmai o funcție de n . Ceea ce înțelegem atunci când spunem “timpul de execuție este $O(n^2)$ ” este că timpul de execuție în cazul cel mai defavorabil (care este o funcție de n) este $O(n^2)$, sau, echivalent, oricare ar fi datele de intrare de dimensiune n , pentru fiecare valoare a lui n , timpul de execuție pentru acest set de date de intrare este $O(n^2)$.

Ω -notația

La fel cum O -notația furnizează o delimitare asimptotică *superioară* pentru o funcție, Ω -notația furnizează o **delimitare asimptotică inferioară**. Pentru o funcție dată $g(n)$, vom nota cu $\Omega(g(n))$, mulțimea

$$\Omega(g(n)) = \{f(n) : \text{există constantele pozitive } c \text{ și } n_0 \text{ astfel încât} \\ 0 \leq cg(n) \leq f(n) \text{ pentru orice } n \geq n_0\}$$

Imaginea intuitivă în spatele Ω -notației este arătată în figura 2.1(c). Pentru toate valorile n aflate la dreapta lui n_0 , $f(n)$ este mai mare sau egală cu $cg(n)$.

Din definițiile notațiilor asimptotice pe care le-am întâlnit până acum, este ușor de demonstrat următoarea teoremă importantă (vezi exercițiul 2.1-5).

Teorema 2.1 Pentru orice două funcții $f(n)$ și $g(n)$, avem $f(n) = \Theta(g(n))$ dacă și numai dacă $f(n) = O(g(n))$ și $f(n) = \Omega(g(n))$. ■

Ca un exemplu de aplicare a acestei teoreme, faptul că $an^2 + bn + c = \Theta(n^2)$ implică imediat $an^2 + bn + c = O(n^2)$ și $an^2 + bn + c = \Omega(n^2)$. În practică, în loc să utilizăm teorema 2.1 pentru a obține margini asimptotice superioare sau inferioare pornind de la margini asimptotice strânse, așa cum am făcut în exemplul precedent, o utilizăm de obicei pentru a demonstra delimitări asimptotice strânse plecând de la delimitări asimptotice superioare și inferioare. Mărginirea asimptotică tare va fi demonstrată utilizând mărginirea asimptotică, superioară și inferioară.

Deoarece Ω -notația descrie o margine inferioară, atunci când o utilizăm pentru a evalua timpul de execuție al unui algoritm în cazul cel mai favorabil, în mod implicit, evaluăm și timpul de execuție al algoritmului pentru date de intrare arbitrare. De exemplu, timpul de execuție în cazul cel mai favorabil al sortării prin inserție este $\Omega(n)$, de unde rezultă că timpul de execuție al sortării prin inserție este $\Omega(n)$.

Timpul de execuție al sortării prin inserție se află prin urmare între $O(n)$ și $\Omega(n^2)$ deoarece este situat întotdeauna între o funcție liniară și o funcție pătratică de n . Mai mult, aceste aproximări sunt asimptotic cele mai strânse: de exemplu, timpul de execuție al sortării prin inserție nu este $\Omega(n^2)$ deoarece acesta este $\Theta(n)$ în cazul datelor de intrare deja sortate. Nu este contradictoriu să spunem totuși că acest timp de execuție este $\Omega(n^2)$ în cazul cel mai defavorabil deoarece există cel puțin un tip de date de intrare care impun algoritmului un timp de execuție $\Omega(n^2)$. Atunci când spunem că *timpul de execuție* al unui algoritm este $\Omega(g(n))$, înțelegem că *nu contează ce set particular de date de intrare de dimensiune n este ales pentru fiecare valoare a lui n , timpul de execuție pe acel set de date de intrare este cel puțin o constantă înmulțită cu $g(n)$, pentru n suficient de mare.*

Notația asimptotică în ecuații

Am văzut deja cum poate fi utilizată o notație asimptotică în formulele matematice. De exemplu, la introducerea O -notației, am scris “ $n = O(n^2)$ ”. Am putea scrie, de asemenea $2n^2 + 3n + 1 = 2n^2 + \Theta(n)$. Cum interpretăm astfel de formule?

Atunci când o notație asimptotică se află singură în membrul drept al unei ecuații, ca în $n = O(n^2)$, deja am definit semnul de egalitate ca reprezentând apartenența din teoria mulțimilor: $n \in O(n^2)$. În general, totuși, când notația asimptotică apare într-o relație, o vom interpreta

ca reprezentând o funcție anonimă pe care nu ne obosim să o numim. De exemplu, egalitatea $2n^2 + 3n + 1 = 2n^2 + \Theta(n)$ înseamnă de fapt $2n^2 + 3n + 1 = 2n^2 + f(n)$, unde $f(n)$ este o funcție oarecare din mulțimea $\Theta(n)$. În acest caz, $f(n) = 3n + 1$, care este într-adevăr în $\Theta(n)$.

Folosirea notațiilor asimptotice în aceste situații poate ajuta la eliminarea unor detalii inutile dintr-o ecuație. De exemplu, în capitoul 1, am exprimat timpul de execuție în cazul cel mai defavorabil la sortarea prin interclasare prin relația de recurență

$$T(n) = 2T(n/2) + \Theta(n)$$

Dacă suntem interesați doar de comportarea asimptotică a lui $T(n)$, nu are nici un rost să specificăm exact toți termenii de ordin inferior: aceștia sunt toți presupuși a fi incluși în funcția anonimă notată prin termenul $\Theta(n)$.

Numărul funcțiilor "anonime" într-o expresie este înțeles ca fiind egal cu numărul notațiilor asimptotice care apar în această expresie. De exemplu, în expresia

$$\sum_{i=1}^n O(i)$$

există o singură funcție anonimă (o funcție de i). Această expresie *nu* este echivalentă cu $O(1) + O(2) + \dots + O(n)$, care oricum nu are o interpretare clară.

În anumite cazuri apar notații asimptotice în membrul stâng al unei ecuații, cum ar fi

$$2n^2 + \Theta(n) = \Theta(n^2).$$

Vom interpreta astfel de ecuații utilizând următoarea regulă: *Oricum ar fi alese funcțiile anonime din partea stângă a semnelor egal, există o modalitate de alegere a funcțiilor anonime din partea dreaptă a semnelor egal pentru a realiza egalitatea descrisă de ecuație.* Astfel, sensul exemplului nostru este acela că pentru orice funcție $f(n) \in \Theta(n)$, există o anumită funcție $g(n) \in \Theta(n^2)$ astfel încât $2n^2 + f(n) = g(n)$ pentru orice n . Cu alte cuvinte, membrul drept al ecuației oferă un nivel de detaliu mai scăzut decât membrul stâng.

Un număr de astfel de relații pot fi înălțuite ca în

$$2n^2 + 3n + 1 = 2n^2 + \Theta(n) = \Theta(n^2)$$

Putem interpreta fiecare ecuație separat după regula de mai sus. Prima ecuație spune că există o funcție $f(n) \in \Theta(n)$ astfel încât $2n^2 + 3n + 1 = 2n^2 + f(n)$ pentru orice n . A doua ecuație spune că pentru orice funcție $g(n) \in \Theta(n)$ (așa cum este $f(n)$), există o anumită funcție $h(n) \in \Theta(n^2)$ astfel încât $2n^2 + g(n) = h(n)$ pentru orice n . Să observăm că această interpretare implică $2n^2 + 3n + 1 = \Theta(n^2)$, fapt care coincide cu informația intuitivă pe care o transmite lanțul celor două ecuații de mai sus.

***o*-notația**

Delimitarea asimptotică superioară dată de *O*-notația definită anterior poate sau nu să fie o delimitare asimptotic strânsă. Astfel, relația $2n^2 = O(n^2)$, este o delimitare asimptotic strânsă pe când $2n = O(n^2)$ nu este. Vom utiliza în cele ce urmează *o*-notația pentru a desemna o delimitare superioară care nu este asimptotic strânsă. Formal, vom defini $o(g(n))$ ("o mic de g de n ") ca fiind mulțimea

$$o(g(n)) = \{f(n) \mid \text{pentru orice constantă pozitivă } c > 0 \text{ există o constantă } n_0 > 0 \text{ astfel încât } 0 \leq f(n) < cg(n) \text{ pentru orice } n \geq n_0\}.$$

De exemplu, $2n = o(n^2)$, dar $2n^2 \neq o(n^2)$.

Definițiile pentru O -notație și o -notație sunt similare. Principala diferență este aceea că în $f(n) = O(g(n))$, delimitarea $0 \leq f(n) \leq cg(n)$ are loc pentru o anumită constantă $c > 0$, dar, în $f(n) = o(g(n))$, delimitarea $0 \leq f(n) < cg(n)$ are loc pentru toate constantele $c > 0$. Intuitiv, în cazul o -notației, funcția $f(n)$ devine neglijabilă relativ la $g(n)$ atunci când n tinde la infinit; cu alte cuvinte

$$\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 0. \quad (2.1)$$

Unii autori folosesc această limită ca o definiție pentru o -notație: definiția din această carte impune, de asemenea, funcțiilor anonime să fie asimptotic nenegative.

ω -notația

Prin analogie, ω -notația este față de Ω -notație ceea ce este o -notația față de O -notație. Utilizăm ω -notația pentru a desemna o delimitare asimptotică inferioară care nu este asimptotic strânsă. O cale de a o defini este

$$f(n) \in \omega(g(n)) \text{ dacă și numai dacă } g(n) \in o(f(n)).$$

Formal, totuși, definim $\omega(g(n))$ ("omega mic de g de n ") ca fiind mulțimea

$$\omega(g(n)) = \{f(n) : \text{pentru orice constantă pozitivă } c > 0 \text{ există o constantă } n_0 > 0 \text{ astfel încât } 0 \leq cg(n) < f(n) \text{ pentru orice } n \geq n_0\}.$$

De exemplu, $n^2/2 = \omega(n)$, dar $n^2/2 \neq \omega(n)$. Relația $f(n) = \omega(g(n))$ implică

$$\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = \infty,$$

dacă limita există. Cu alte cuvinte, $f(n)$ devine oricât de mare relativ la $g(n)$ atunci când n tinde la infinit.

Compararea funcțiilor

Multe dintre proprietățile relațiilor dintre numerele reale se aplică și la compararea asimptotică a funcțiilor. Pentru cele ce urmează vom presupune că $f(n)$ și $g(n)$ sunt asimptotic pozitive.

Tranzitivitatea:

$$\begin{array}{llll} f(n) = \Theta(g(n)) & \text{și} & g(n) = \Theta(h(n)) & \text{implică} & f(n) = \Theta(h(n)), \\ f(n) = O(g(n)) & \text{și} & g(n) = O(h(n)) & \text{implică} & f(n) = O(h(n)), \\ f(n) = \Omega(g(n)) & \text{și} & g(n) = \Omega(h(n)) & \text{implică} & f(n) = \Omega(h(n)), \\ f(n) = o(g(n)) & \text{și} & g(n) = o(h(n)) & \text{implică} & f(n) = o(h(n)), \\ f(n) = \omega(g(n)) & \text{și} & g(n) = \omega(h(n)) & \text{implică} & f(n) = \omega(h(n)). \end{array}$$

Reflexivitatea:

$$\begin{array}{ll} f(n) &= \Theta(f(n)), \\ f(n) &= O(f(n)), \\ f(n) &= \Omega(f(n)). \end{array}$$

Simetria:

$$f(n) = \Theta(g(n)) \quad \text{dacă și numai dacă} \quad g(n) = \Theta(f(n)).$$

Antisimetria:

$$\begin{aligned} f(n) = O(g(n)) & \quad \text{dacă și numai dacă} \quad g(n) = \Omega(f(n)), \\ f(n) = o(g(n)) & \quad \text{dacă și numai dacă} \quad g(n) = \omega(f(n)). \end{aligned}$$

Deoarece aceste proprietăți sunt valide pentru notații asimptotice, se poate trasa o analogie între compararea asimptotică a două funcții f și g și compararea asimptotică a două numere reale a și b :

$$\begin{aligned} f(n) = O(g(n)) & \approx a \leq b, \\ f(n) = \Omega(g(n)) & \approx a \geq b, \\ f(n) = \Theta(g(n)) & \approx a = b, \\ f(n) = o(g(n)) & \approx a < b, \\ f(n) = \omega(g(n)) & \approx a > b. \end{aligned}$$

O proprietate a numerelor reale, totuși, nu se transpune la notații asimptotice:

Trihotomia: Pentru orice două numere reale a și b exact una dintre următoarele relații este adevărată: $a < b$, $a = b$ sau $a > b$.

Deși orice două numere reale pot fi comparate, nu toate funcțiile sunt asimptotic comparabile. Cu alte cuvinte, pentru două funcții $f(n)$ și $g(n)$, se poate întâmpla să nu aibă loc nici $f(n) = O(g(n))$, nici $f(n) = \Omega(g(n))$. De exemplu, funcțiile n și $n^{1+\sin n}$ nu pot fi comparate utilizând notații asimptotice, deoarece valoarea exponentului în $n^{1+\sin n}$ oscilează între 0 și 2, luând toate valorile intermediare.

Exerciții

2.1-1 Fie $f(n)$ și $g(n)$ funcții asimptotic nenegative. Utilizând definiția de bază a Θ -notației, demonstrați că $\max(f(n), g(n)) = \Theta(f(n) + g(n))$.

2.1-2 Demonstrați că pentru orice constante reale a și b , unde $b > 0$, avem

$$(n + a)^b = \Theta(n^b). \quad (2.2)$$

2.1-3 Explicați de ce afirmația “Timpul de execuție al algoritmului A este cel puțin $O(n^2)$ ” este lipsită de conținut.

2.1-4 Este adevărat că $2^{n+1} = O(2^n)$? Este adevărat că $2^{2n} = O(2^n)$?

2.1-5 Demonstrați teorema 2.1.

2.1-6 Demonstrați că timpul de execuție al unui algoritm este $\Theta(g(n))$ dacă și numai dacă timpul său de execuție în cazul cel mai defavorabil este $O(g(n))$ și timpul său de execuție în cazul cel mai favorabil este $\Omega(g(n))$.

2.1-7 Demonstrați că $o(g(n)) \cap \omega(g(n))$ este mulțimea vidă.

2.1-8 Putem extinde notația noastră la cazul a doi parametri n și m care tind la infinit în mod independent, cu viteze diferite. Pentru o funcție dată $g(n, m)$ notăm cu $O(g(n, m))$ mulțimea de funcții

$$O(g(n, m)) = \{f(n, m) : \text{există constantele pozitive } c, n_0 \text{ și } m_0 \text{ astfel încât} \\ 0 \leq f(n, m) \leq cg(n, m) \text{ pentru orice } n \geq n_0 \text{ și } m \geq m_0\}.$$

Dați definițiile corespunzătoare pentru $\Omega(g(n, m))$ și $\Theta(g(n, m))$.

2.2. Notății standard și funcții comune

Această secțiune trece în revistă câteva funcții și notații matematice standard și explorează relațiile dintre ele. Ea ilustrează, de asemenea, utilizarea notațiilor asimptotice.

Monotonie

O funcție $f(n)$ este **monoton crescătoare** dacă $m \leq n$ implică $f(m) \leq f(n)$. Analog, ea este **monoton descrescătoare** dacă $m \leq n$ implică $f(m) \geq f(n)$. O funcție $f(n)$ este **strict crescătoare** dacă $m < n$ implică $f(m) < f(n)$ și **strict descrescătoare** dacă $m < n$ implică $f(m) > f(n)$.

Părți întregi inferioare și superioare

Pentru orice număr real x , notăm cel mai mare întreg mai mic sau egal cu x prin $\lfloor x \rfloor$ (se citește “partea întreagă inferioară a lui x ”) și cel mai mic întreg mai mare sau egal cu x prin $\lceil x \rceil$ (se citește “partea întreagă superioară a lui x ”). Pentru orice x real,

$$x - 1 < \lfloor x \rfloor \leq x \leq \lceil x \rceil < x + 1.$$

Pentru orice întreg n ,

$$\lfloor n/2 \rfloor + \lfloor n/2 \rfloor = n,$$

iar pentru orice întreg n și orice întregi $a \neq 0$ și $b > 0$,

$$\lceil \lceil n/a \rceil / b \rceil = \lceil n/ab \rceil \tag{2.3}$$

și

$$\lfloor \lfloor n/a \rfloor / b \rfloor = \lfloor n/ab \rfloor. \tag{2.4}$$

Funcțiile parte întreagă inferioară și superioară sunt monoton crescătoare.

Polinoame

Fiind dat un întreg pozitiv d , **un polinom în n de gradul d** este o funcție $p(n)$ de forma

$$p(n) = \sum_{i=0}^d a_i n^i,$$

unde constantele a_0, a_1, \dots, a_d sunt **coeficienții** polinomului, iar $a_d \neq 0$. Un polinom este *asimptotic pozitiv* dacă și numai dacă $a_d > 0$. Pentru un polinom asimptotic pozitiv $p(n)$ de grad d avem $p(n) = \Theta(n^d)$. Pentru orice constantă reală $a \geq 0$, funcția n^a este monoton crescătoare, iar pentru orice constantă reală $a \leq 0$ funcția n^a este monoton descrescătoare. Spunem că o funcție $f(n)$ este **polinomial mărginită** dacă $f(n) = n^{O(1)}$, ceea ce este echivalent cu a spune că $f(n) = O(n^k)$ pentru o anumită constantă k (vezi exercițiul 2.2-2).

Exponențiale

Pentru orice $a \neq 0$, m și n reale avem următoarele identități:

$$\begin{aligned} a^0 &= 1, \\ a^1 &= a, \\ a^{-1} &= 1/a, \\ (a^m)^n &= a^{mn}, \\ (a^m)^n &= (a^n)^m, \\ a^m a^n &= a^{m+n}. \end{aligned}$$

Pentru orice n și $a \geq 1$, funcția a^n este monoton crescătoare în n . Când ne va conveni, vom presupune că $0^0 = 1$.

Ratele de creștere ale polinoamelor și exponențialelor pot fi legate prin următorul fapt. Pentru orice constante a și b astfel încât $a > 1$,

$$\lim_{n \rightarrow \infty} \frac{n^b}{a^n} = 0, \quad (2.5)$$

din care putem trage concluzia că

$$n^b = o(a^n).$$

Astfel, orice funcție exponențială având baza strict mai mare decât 1 crește mai repede decât orice funcție polinomială.

Utilizând e pentru a nota $2.71828\dots$, baza funcției logaritm natural, avem, pentru orice x real,

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots = \sum_{i=0}^{\infty} \frac{x^i}{i!}, \quad (2.6)$$

unde “!” desemnează funcția factorial definită mai târziu în această secțiune. Pentru orice x real avem inegalitatea

$$e^x \geq 1 + x, \quad (2.7)$$

unde egalitatea are loc numai pentru $x = 0$. Când $|x| \leq 1$, avem aproximația

$$1 + x \leq e^x \leq 1 + x + x^2. \quad (2.8)$$

Când $x \rightarrow 0$, aproximația lui e^x prin $1 + x$ este destul de bună:

$$e^x = 1 + x + \Theta(x^2).$$

(În această ecuație, notația asimptotică este utilizată pentru a descrie comportamentul la limită pentru $x \rightarrow 0$ mai degrabă decât pentru $x \rightarrow \infty$.) Avem, pentru orice x ,

$$\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n = e^x.$$

Logaritmi

Vom utiliza următoarele notații:

$$\begin{aligned} \lg n &= \log_2 n && (\text{logaritm binar}), \\ \ln n &= \log_e n && (\text{logaritm natural}), \\ \lg^k n &= (\lg n)^k && (\text{exponențierea}), \\ \lg \lg n &= \lg(\lg n) && (\text{compunerea}). \end{aligned}$$

O convenție de notație importantă pe care o vom adopta este aceea că *funcțiile logaritmice se vor aplica numai următorului termen dintr-o formulă*, astfel că $\lg n + k$ va însemna $(\lg n) + k$ și nu $\lg(n + k)$. Pentru $n > 0$ și $b > 1$, funcția $\log_b n$ este strict crescătoare.

Pentru orice numere reale $a > 0, b > 0, c > 0$ și n ,

$$\begin{aligned} a &= b^{\log_b a}, \\ \log_c(ab) &= \log_c a + \log_c b, \\ \log_b a^n &= n \log_b a, \\ \log_b a &= \frac{\log_c a}{\log_c b}, \\ \log_b(1/a) &= -\log_b a, \\ \log_b a &= \frac{1}{\log_a b}, \\ a^{\log_b n} &= n^{\log_b a}. \end{aligned} \quad (2.9)$$

Deoarece schimbarea bazei unui logaritm de la o constantă la o altă constantă schimbă valoarea logaritmului doar printr-un factor constant, vom utiliza adesea notația “ $\lg n$ ” când factorii constanți nu vor fi importanți, așa cum este cazul cu O -notația. Informaticienilor li se pare că 2 este baza cea mai naturală pentru logaritmi, deoarece atât de mulți algoritmi și structuri de date presupun descompunerea unei probleme în două părți.

Există o dezvoltare în serie simplă pentru $\ln(1+x)$ când $|x| < 1$:

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \frac{x^5}{5} - \dots$$

Avem, de asemenea, următoarele inegalități pentru $x > -1$:

$$\frac{x}{1+x} \leq \ln(1+x) \leq x, \quad (2.10)$$

unde egalitatea are loc numai pentru $x = 0$.

Spunem că o funcție $f(n)$ este **polilogaritmă mărginită** dacă $f(n) = \lg^{O(1)} n$. Putem lega creșterile polinoamelor și polilogaritmilor înlocuind pe n cu $\lg n$ și pe a cu 2^a în ecuația (2.5), obținând

$$\lim_{n \rightarrow \infty} \frac{\lg^b n}{(2^a)^{\lg n}} = \lim_{n \rightarrow \infty} \frac{\lg^b n}{n^a} = 0.$$

Din această limită, putem conchide că

$$\lg^b n = o(n^a)$$

pentru orice constantă $a > 0$. Astfel, orice funcție polinomială pozitivă crește mai repede decât orice funcție polilogaritmă.

Factoriali

Notăția $n!$ (se citește “ n factorial”) este definită pentru numere întregi $n \geq 0$ prin

$$n! = \begin{cases} 1 & \text{dacă } n = 0, \\ n \cdot (n-1)! & \text{dacă } n > 0. \end{cases}$$

Astfel, $n! = 1 \cdot 2 \cdot 3 \cdots n$.

O margine superioară slabă a funcției factorial este $n! \leq n^n$, deoarece fiecare dintre cei n factori ai produsului factorial este cel mult n . **Aproximația lui Stirling**,

$$n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \left(1 + \Theta\left(\frac{1}{n}\right)\right), \quad (2.11)$$

unde e este baza logaritmilor naturali, ne dă o margine superioară mai strânsă, dar și o margine inferioară. Utilizând aproximația lui Stirling, putem demonstra că

$$\begin{aligned} n! &= o(n^n), \\ n! &= \omega(2^n), \\ \lg(n!) &= \Theta(n \lg n). \end{aligned}$$

Următoarele delimitări sunt, de asemenea, valabile pentru orice n :

$$\sqrt{2\pi n} \left(\frac{n}{e}\right)^n \leq n! \leq \sqrt{2\pi n} \left(\frac{n}{e}\right)^{ne^{1/12n}} \quad (2.12)$$

Funcția logaritm iterată

Utilizăm notația $\lg^* n$ (se citește “logaritm stea de n ”) pentru a nota logaritmul iterat, care este definit după cum urmează. Fie funcția $\lg^{(i)} n$ definită recursiv pentru întregii nenegativi i prin

$$\lg^{(i)} n = \begin{cases} n & \text{dacă } i = 0, \\ \lg(\lg^{(i-1)} n) & \text{dacă } i > 0 \text{ și } \lg^{(i-1)} n > 0, \\ \text{nedefinită} & \text{dacă } i > 0 \text{ și } \lg^{(i-1)} n \geq 0 \text{ sau } \lg^{(i-1)} n \text{ este nedefinită.} \end{cases}$$

Calculați diferența între funcția $\lg^{(i)} n$ (funcția logaritm aplicată succesiv de i ori, începând cu argumentul n) și $\lg^i n$ (logaritmul lui n ridicat la puterea i). Funcția logaritm iterată este definită prin

$$\lg^* n = \min \left\{ i \geq 0 : \lg^{(i)} n \leq 1 \right\}.$$

Funcția logaritm iterată este o funcție *foarte* lent crescătoare:

$$\begin{aligned} \lg^* 2 &= 1, \\ \lg^* 4 &= 2, \\ \lg^* 16 &= 3, \\ \lg^* 65536 &= 4, \\ \lg^*(2^{65536}) &= 5. \end{aligned}$$

Deoarece numărul atomilor din universul observabil este estimat la aproximativ 10^{80} , care este mult mai mic decât 2^{65536} , rareori întâlnim o valoare a lui n astfel încât $\lg^* n > 5$.

Numere Fibonacci

Numerele Fibonacci sunt definite prin următoarea recurență:

$$F_0 = 0, F_1 = 1, F_i = F_{i-1} + F_{i-2} \quad \text{pentru } i \geq 2. \quad (2.13)$$

Astfel, fiecare număr Fibonacci este suma celor două numere Fibonacci anterioare, rezultând secvența

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, \dots$$

Numerele Fibonacci sunt legate de **raportul de aur** ϕ și de conjugatul său $\hat{\phi}$, care sunt date de următoarele formule:

$$\phi = \frac{1 + \sqrt{5}}{2} = 1.61803\dots, \hat{\phi} = \frac{1 - \sqrt{5}}{2} = -.61803\dots \quad (2.14)$$

Mai precis, avem

$$F_i = \frac{\phi^i - \hat{\phi}^i}{\sqrt{5}}, \quad (2.15)$$

ceea ce se poate demonstra prin inducție (exercițiul 2.2-7). Deoarece $|\hat{\phi}| < 1$, avem $|\hat{\phi}^i|/\sqrt{5} < 1/\sqrt{5} < 1/2$, astfel că cel de-al i -lea număr Fibonacci este egal cu $\phi^i/\sqrt{5}$ rotunjit la cel mai apropiat întreg. Astfel, numerele Fibonacci cresc exponențial.

Exerciții

2.2-1 Arătați că dacă $f(n)$ și $g(n)$ sunt funcții monoton crescătoare, atunci la fel sunt și funcțiile $f(n) + g(n)$ și $f(g(n))$, iar dacă $f(n)$ și $g(n)$ sunt în plus nenegative, atunci și $f(n) \cdot g(n)$ este monoton crescătoare.

2.2-2 Utilizați definiția O -notației pentru a arăta că $T(n) = n^{O(1)}$ dacă și numai dacă există o constantă $k > 0$ astfel încât $T(n) = O(n^k)$.

2.2-3 Demonstrați ecuația (2.9).

2.2-4 Demonstrați că $\lg(n!) = \Theta(n \lg n)$ și că $n! = o(n^n)$.

2.2-5 ★ Este funcția $\lceil \lg n \rceil!$ polinomial mărginită? Este funcția $\lceil \lg \lg n \rceil!$ polinomial mărginită?

2.2-6 ★ Care este asimptotic mai mare: $\lg(\lg^* n)$ sau $\lg^*(\lg n)$?

2.2-7 Demonstrați prin inducție că cel de-al i -lea număr Fibonacci verifică egalitatea $F_i = (\phi^i - \hat{\phi}^i)/\sqrt{5}$, unde ϕ este raportul de aur iar $\hat{\phi}$ este conjugatul său.

2.2-8 Demonstrați că pentru $i \geq 0$ cel de-al $(i + 2)$ -lea număr Fibonacci verifică $F_{i+2} \geq \phi^i$.

Probleme

2-1 Comportamentul asimptotic al polinoamelor

Fie

$$p(n) = \sum_{i=0}^d a_i p^i,$$

unde $a_d > 0$ un polinom de gradul d în n și fie k o constantă. Utilizați proprietățile notațiilor asimptotice pentru a demonstra următoarele proprietăți.

a. Dacă $k \geq d$, atunci $p(n) = O(n^k)$.

b. Dacă $k \leq d$, atunci $p(n) = \Omega(n^k)$.

c. Dacă $k = d$, atunci $p(n) = \Theta(n^k)$.

d. Dacă $k > d$, atunci $p(n) = o(n^k)$.

e. Dacă $k < d$, atunci $p(n) = \omega(n^k)$.

2-2 Creșteri asimptotice relative

Indicați, pentru fiecare pereche de expresii (A, B) din tabelul următor, dacă A este O , o , Ω , ω sau Θ de B . Presupuneți că $k \geq 1$, $\epsilon > 0$ și $c > 1$ sunt constante. Răspunsul trebuie să fie sub forma unui tabel, cu “da” sau “nu” scrise în fiecare rubrică.

	A	B	O	o	Ω	ω	Θ
a.	$\lg^k n$	n^ϵ					
b.	n^k	c^n					
c.	\sqrt{n}	$n^{\sin n}$					
d.	2^n	$2^{n/2}$					
e.	$n^{\lg m}$	$m^{\lg n}$					
f.	$\lg(n!)$	$\lg(n^n)$					

2-3 Ordonarea după ratele de creștere asimptotică

- a. Ordonati următoarele funcții după ordinea de creștere; mai precis, găsiți un aranjament de funcții g_1, g_2, \dots, g_{30} care să verifice $g_1 = \Omega(g_2)$, $g_2 = \Omega(g_3)$, \dots , $g_{29} = \Omega(g_{30})$. Partiționați lista în clase de echivalență astfel încât $f(n)$ și $g(n)$ să fie în aceeași clasă de echivalență dacă și numai dacă $f(n) = \Theta(g(n))$.

$\lg(\lg^* n)$	$2^{\lg^* n}$	$(\sqrt{2})^{\lg n}$	n^2	$n!$	$(\lg n)!$
$\left(\frac{3}{2}\right)^n$	n^3	$\lg^2 n$	$\lg(n!)$	2^{2^n}	$n^{1/\lg n}$
$\ln \ln n$	$\lg^* n$	$n \cdot 2^n$	$n^{\lg \lg n}$	$\ln n$	1
$2^{\lg n}$	$(\lg n)^{\lg n}$	e^n	$4^{\lg n}$	$(n+1)!$	$\sqrt{\lg n}$
$\lg^*(\lg n)$	$2^{\sqrt{2 \lg n}}$	n	2^n	$n \lg n$	$2^{2^{n+1}}$

- b. Dați un exemplu de o singură funcție nenegativă $f(n)$ astfel încât pentru toate funcțiile $g_i(n)$ de la punctul (a), $f(n)$ să nu fie nici $O(g_i(n))$ nici $\Omega(g_i(n))$.

2-4 Proprietăți ale notațiilor asimptotice

Fie $f(n)$ și $g(n)$ funcții asimptotic pozitive. Demonstrați următoarele conjecturi sau demonstrați că ele nu sunt adevărate.

- $f(n) = O(g(n))$ implică $g(n) = O(f(n))$.
- $f(n) + g(n) = \Theta(\min(f(n), g(n)))$.
- $f(n) = O(g(n))$ implică $\lg(f(n)) = O(\lg(g(n)))$, unde $\lg(g(n)) > 0$ și $f(n) \geq 1$ pentru orice n suficient de mare.
- $f(n) = O(g(n))$ implică $2^{f(n)} = O(2^{g(n)})$.
- $f(n) = O((f(n))^2)$.
- $f(n) = O(g(n))$ implică $g(n) = \Omega(f(n))$.
- $f(n) = \Theta(f(n/2))$.
- $f(n) + o(f(n)) = \Theta(f(n))$.

2-5 Variații ale lui O și Ω

Unii autori definesc Ω într-un mod ușor diferit de al nostru; să utilizăm $\overset{\infty}{\Omega}$ (se citește “omega infinit”) pentru această definiție alternativă. Spunem că $f(n) = \overset{\infty}{\Omega}(g(n))$ dacă există o constantă pozitivă c astfel încât $f(n) \geq cg(n) \geq 0$ pentru infinit de mulți n .

- Arătați că pentru oricare două funcții $f(n)$ și $g(n)$ care sunt asimptotic nenegative, fie $f(n) = O(g(n))$, fie $f(n) = \overset{\infty}{\Omega}(g(n))$ fie amândouă, în timp ce această afirmație nu este adevărată dacă utilizăm Ω în loc de $\overset{\infty}{\Omega}$.
- Descrieți avantajele și dezavantajele potențiale ale utilizării lui $\overset{\infty}{\Omega}$ în locul lui Ω pentru a descrie timpii de execuție ai programelor.

Unii autori îl definesc și pe O într-un mod ușor diferit; să utilizăm O' pentru definiția alternativă. Spunem că $f(n) = O'(g(n))$ dacă și numai dacă $|f(n)| = O(g(n))$.

- c.** Ce se întâmplă cu fiecare dintre direcțiile lui “dacă și numai dacă” din teorema 2.1 cu această nouă definiție?

Unii autori definesc \tilde{O} (se citește “o moale”) ca însemnând O cu factorii logaritmici ignorați:

$$\tilde{O}(g(n)) = \{f(n) \quad : \quad \text{există constantele pozitive, } c, k \text{ și } n_0 \text{ astfel încât} \\ 0 \leq f(n) \leq cg(n) \lg^k(n) \text{ pentru orice } n \geq n_0\}.$$

- d.** Definiți $\tilde{\Omega}$ și $\tilde{\Theta}$ într-un mod analog. Demonstrați proprietatea analogă corespunzătoare teoremei 2.1.

2-6 Funcții iterate

Operatorul de iterație “*” utilizat în funcția \lg^* poate fi aplicat funcțiilor monoton crescătoare pe mulțimea numerelor reale. Pentru o funcție f care satisface $f(n) < n$ definim funcția $f^{(i)}$ recursiv pentru numerele întregi i nenegative prin

$$f^{(i)}(n) = \begin{cases} f(f^{(i-1)}(n)) & \text{dacă } i > 0, \\ n & \text{dacă } i = 0. \end{cases}$$

Pentru o constantă dată $c \in \mathbf{R}$, definim funcția iterată f_c^* prin

$$f_c^*(n) = \min\{i \geq 0 : f^{(i)}(n) \leq c\},$$

care nu trebuie să fie bine-definită în toate cazurile. Cu alte cuvinte, cantitatea $f_c^*(n)$ este numărul de aplicații iterate ale funcției f necesare pentru a reduce argumentul la c sau la mai puțin.

Pentru fiecare dintre următoarele funcții $f(n)$ și constantele c , dați o evaluare cât se poate de strânsă a lui $f_c^*(n)$.

	$f(n)$	c	$f_c^*(n)$
a.	$\lg n$	1	
b.	$n - 1$	0	
c.	$n/2$	1	
d.	$n/2$	2	
e.	\sqrt{n}	2	
f.	\sqrt{n}	1	
g.	$n^{1/3}$	2	
h.	$n/\lg n$	2	

Note bibliografice

Knuth [121] investighează originea O -notației și afirmă că ea apare pentru prima dată într-un text de teoria numerelor al lui P. Bachmann din 1982. o -notația a fost inventată de către E. Landau în 1909 pentru discuția distribuției numerelor prime. Notatiile Ω și Θ au fost introduse de către Knuth [124] pentru a corecta practica din literatură, populară dar derutantă din punct de vedere tehnic, de a utiliza O -notația atât pentru margini superioare cât și pentru margini

inferioare. Multă lume continuă să utilizeze O -notația acolo unde Θ -notația este mai precisă din punct de vedere tehnic. O altă discuție pe tema istoriei și dezvoltării notațiilor asimptotice poate fi găsită în Knuth [121, 124] și Brassard și Bratley [33].

Nu toți autorii definesc notațiile asimptotice în același mod, deși diferitele definiții concordă în cele mai comune situații. Unele dintre definițiile alternative se aplică și la funcții care nu sunt asimptotic nenegative, dacă valorile lor absolute sunt mărginite în mod corespunzător.

Alte proprietăți ale funcțiilor matematice elementare pot fi găsite în orice carte de referință de matematică, cum ar fi Abramowitz și Stegun [1] sau Beyer [27], sau într-o carte de analiză, cum ar fi Apostol [12] sau Thomas și Finney [192]. Knuth [121] conține o mulțime de materiale despre matematica discretă, utilizată în informatică.

3 Sume

Când un algoritm conține o structură de control iterativă cum ar fi un ciclu **cât timp** sau **pentru**, timpul său de execuție poate fi exprimat ca suma timpilor necesari la fiecare execuție a corpului ciclului. De exemplu, am văzut în secțiunea 1.2 că a j -a iterație a sortării prin inserție a necesitat un timp proporțional cu j în cazul cel mai defavorabil. Adunând timpii necesari pentru fiecare iterație, am obținut o sumă (sau serie)

$$\sum_{j=2}^n j.$$

Evaluarea sumei a dat o margine $\Theta(n^2)$ pentru timpul de execuție al algoritmului în cazul cel mai defavorabil. Acest exemplu indică importanța generală a înțelegerii manipulării și delimitării sumelor. (După cum vom vedea în capitolul 4, sumele apar și când utilizăm anumite metode pentru rezolvarea recurențelor.)

Secțiunea 3.1 prezintă mai multe formule de bază în care intervin sume. Secțiunea 3.2 oferă tehnici utile pentru delimitarea sumelor. Formulele din secțiunea 3.1 sunt date fără demonstrație, deși demonstrațiile pentru unele dintre ele sunt prezentate în secțiunea 3.2, pentru a ilustra metodele acelei secțiuni. Majoritatea celorlalte demonstrații pot fi găsite în orice manual de analiză.

3.1. Formule de însumare și proprietăți

Fiind dat un șir de numere a_1, a_2, \dots , suma finită $a_1 + a_2 + \dots + a_n$ poate fi scrisă

$$\sum_{k=1}^n a_k.$$

Dacă $n = 0$, valoarea sumei este definită ca fiind 0. Dacă n nu este un număr întreg, mai presupunem că limita superioară este $\lfloor n \rfloor$. Analog, dacă suma începe cu $k = x$, unde x nu este un întreg, presupunem că valoarea inițială pentru însumare este $\lfloor x \rfloor$. (În general, vom scrie în mod explicit părțile întregi inferioare sau superioare). Valoarea unei serii finite este întotdeauna bine definită, iar termenii săi pot fi adunați în orice ordine.

Fiind dat un șir de numere a_1, a_2, \dots , suma infinită $a_1 + a_2 + \dots$ poate fi scrisă

$$\sum_{k=1}^{\infty} a_k,$$

care este interpretată ca

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n a_k.$$

Dacă limita nu există, seria **diverge**; altfel, ea **converge**. Termenii unei serii convergente nu pot fi întotdeauna adunați în orice ordine. Putem rearanja, totuși, termenii unei **serii absolut convergente**, care este o serie $\sum_{k=1}^{\infty} a_k$ pentru care seria $\sum_{k=1}^{\infty} |a_k|$ converge de asemenea.

Liniaritate

Pentru orice număr real c și orice șiruri finite a_1, a_2, \dots, a_n și b_1, b_2, \dots, b_n

$$\sum_{k=1}^n (ca_k + b_k) = c \sum_{k=1}^n a_k + \sum_{k=1}^n b_k.$$

Proprietatea de liniaritate este verificată, de asemenea, și de seriile infinite convergente.

Proprietatea de liniaritate poate fi exploatată pentru a manipula sumele care încorporează notații asimptotice. De exemplu,

$$\sum_{k=1}^n \Theta(f(k)) = \Theta\left(\sum_{k=1}^n f(k)\right).$$

În această ecuație, Θ -notația din membrul stâng se aplică variabilei k , dar în membrul drept ea se aplică lui n . Astfel de manipulări pot fi aplicate și seriilor infinite convergente.

Serii aritmetice

Suma

$$\sum_{k=1}^n k = 1 + 2 + \dots + n,$$

care a apărut când am analizat sortarea prin inserție, este o *serie aritmetică* și are valoarea

$$\sum_{k=1}^n k = \frac{1}{2}n(n+1) \tag{3.1}$$

$$= \Theta(n^2) \tag{3.2}$$

Serii geometrice

Pentru orice $x \neq 1$ real, suma

$$\sum_{k=0}^n x^k = 1 + x + x^2 + \dots + x^n$$

este o *serie geometrică* sau *exponențială* și are valoarea

$$\sum_{k=0}^n x^k = \frac{x^{n+1} - 1}{x - 1}. \tag{3.3}$$

Când suma este infinită și $|x| < 1$ avem seria geometrică infinită descrescătoare

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1 - x}. \tag{3.4}$$

Serii armonice

Pentru întregi pozitivi n , al n -lea **număr armonic** este

$$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots + \frac{1}{n} = \sum_{k=1}^n \frac{1}{k} = \ln n + O(1). \quad (3.5)$$

Integrarea și diferențierea seriilor

Se pot obține formule suplimentare prin integrarea sau diferențierea formulelor de mai sus. De exemplu, diferențiind ambii membri ai seriei geometrice infinite (3.4) și înmulțind cu x , obținem

$$\sum_{k=0}^{\infty} kx^k = \frac{x}{(1-x)^2} \quad (3.6)$$

Serii telescopante

Pentru orice șir a_0, a_1, \dots, a_n

$$\sum_{k=1}^n (a_k - a_{k-1}) = a_n - a_0 \quad (3.7)$$

deoarece fiecare dintre termenii a_1, a_2, \dots, a_{n-1} este adunat exact o dată și scăzut exact o dată. Spunem că suma **telescopează**. Analog,

$$\sum_{k=0}^{n-1} (a_k - a_{k+1}) = a_0 - a_n.$$

Ca un exemplu de sumă telescopantă, să considerăm seria

$$\sum_{k=1}^{n-1} \frac{1}{k(k+1)}.$$

Deoarece putem rescrie fiecare termen ca

$$\frac{1}{k(k+1)} = \frac{1}{k} - \frac{1}{k+1},$$

obținem

$$\sum_{k=1}^{n-1} \frac{1}{k(k+1)} = \sum_{k=1}^{n-1} \left(\frac{1}{k} - \frac{1}{k+1} \right) = 1 - \frac{1}{n}.$$

Produse

Produsul finit $a_1 a_2 \cdots a_n$ poate fi scris

$$\prod_{k=1}^n a_k.$$

Dacă $n = 0$, valoarea produsului este definită ca fiind 1. Putem converti o formulă cu un produs într-o formulă cu o sumă, utilizând identitatea

$$\lg \left(\prod_{k=1}^n a_k \right) = \sum_{k=1}^n \lg a_k.$$

Exerciții

3.1-1 Găsiți o formulă simplă pentru $\sum_{k=1}^n (2k - 1)$.

3.1-2 ★ Arătați că $\sum_{k=1}^n 1/(2k - 1) = \ln(\sqrt{n}) + O(1)$ manipulând seria armonică.

3.1-3 ★ Arătați că $\sum_{k=0}^{\infty} (k - 1)/2^k = 0$.

3.1-4 ★ Evaluați suma $\sum_{k=1}^{\infty} (2k + 1)x^{2k}$.

3.1-5 Utilizați proprietatea de liniaritate a însumării pentru a demonstra că $\sum_{k=1}^n O(f_k(n)) = O(\sum_{k=1}^n f_k(n))$.

3.1-6 Demonstrați că $\sum_{k=1}^{\infty} \Omega(f(k)) = \Omega(\sum_{k=1}^{\infty} f(k))$.

3.1-7 Evaluați produsul $\prod_{k=1}^n 2 \cdot 4^k$.

3.1-8 ★ Evaluați produsul $\prod_{k=2}^n (1 - 1/k^2)$.

3.2. Delimitarea sumelor

Există multe tehnici disponibile pentru delimitarea sumelor care descriu timpii de execuție ai algoritmilor. Iată câteva dintre metodele cel mai frecvent utilizate.

Inducția matematică

Cea mai simplă cale de a evalua o serie este utilizarea inducției matematice. De exemplu, să demonstrăm că seria aritmetică $\sum_{k=1}^n k$ are valoarea $\frac{1}{2}n(n + 1)$. Putem verifica ușor pentru $n = 1$, așa că facem ipoteza de inducție că formula este adevărată pentru orice n și demonstrăm că are loc pentru $n + 1$. Avem

$$\sum_{k=1}^{n+1} k = \sum_{k=1}^n k + (n + 1) = \frac{1}{2}n(n + 1) + (n + 1) = \frac{1}{2}(n + 1)(n + 2).$$

Nu trebuie să ghicim valoarea exactă a unei sume pentru a utiliza inducția matematică. Inducția poate fi utilizată și pentru a demonstra o delimitare. De exemplu, să demonstrăm că seria geometrică $\sum_{k=0}^n 3^k$ este $O(3^n)$. Mai precis, să demonstrăm că $\sum_{k=0}^n 3^k \leq c \cdot 3^n$ pentru o anumită constantă c . Pentru condiția inițială $n = 0$ avem $\sum_{k=0}^0 3^k = 1 \leq c \cdot 1$, cât timp $c \geq 1$. Presupunând că delimitarea are loc pentru n , să demonstrăm că are loc pentru $n + 1$. Avem

$$\sum_{k=0}^{n+1} 3^k = \sum_{k=0}^n 3^k + 3^{n+1} \leq c \cdot 3^n + 3^{n+1} = \left(\frac{1}{3} + \frac{1}{c}\right) c \cdot 3^{n+1} \leq c \cdot 3^{n+1}$$

cât timp $(1/3 + 1/c) \leq 1$ sau, echivalent, $c \geq 3/2$. Astfel, $\sum_{k=0}^n 3^k = O(3^n)$, ceea ce am dorit să arătăm.

Trebuie să fim prudenți când utilizăm notații asimptotice pentru a demonstra delimitări prin inducție. Considerăm următoarea demonstrație vicioasă a faptului că $\sum_{k=1}^n k = O(n)$. Desigur, $\sum_{k=1}^1 k = O(1)$. Acceptând delimitarea pentru n , o demonstrăm acum pentru $n + 1$:

$$\begin{aligned} \sum_{k=1}^{n+1} k &= \sum_{k=1}^n k + (n+1) \\ &= O(n) + (n+1) \quad \leftarrow \text{gre it!} \\ &= O(n+1). \end{aligned}$$

Greșeala în argumentație este aceea că O ascunde o “constantă” care crește în funcție de n și deci nu este constantă. Nu am arătat că aceeași constantă e valabilă pentru *to* n .

Delimitarea termenilor

Uneori, o limită superioară bună a unei serii se poate obține delimitând fiecare termen al seriei, și este, de multe ori, suficient să utilizăm cel mai mare termen pentru a-i delimita pe ceilalți. De exemplu, o limită superioară rapidă a seriei aritmetice (3.1) este

$$\sum_{k=1}^n k \leq \sum_{k=1}^n n = n^2.$$

În general, pentru o serie $\sum_{k=1}^n a_k$, dacă punem $a_{max} = \max_{1 \leq k \leq n} a_k$, atunci

$$\sum_{k=1}^n a_k \leq n a_{max}.$$

Tehnica delimitării fiecărui termen dintr-o serie prin cel mai mare termen este o metodă slabă când seria poate fi, în fapt, delimitată printr-o serie geometrică. Fiind dată seria $\sum_{k=0}^n a_k$, să presupunem că $a_{k+1}/a_k \leq r$ pentru orice $k \geq 0$, unde $r < 1$ este o constantă. Suma poate fi delimitată printr-o serie geometrică descrescătoare infinită, deoarece $a_k \leq a_0 r^k$, și astfel,

$$\sum_{k=0}^n a_k \leq \sum_{k=0}^{\infty} a_0 r^k = a_0 \sum_{k=0}^{\infty} r^k = a_0 \frac{1}{1-r}.$$

Putem aplica această metodă pentru a delimita suma $\sum_{k=1}^{\infty} (k/3^k)$. Primul termen este $1/3$, iar raportul termenilor consecutivi este

$$\frac{(k+1)/3^{k+1}}{k/3^k} = \frac{1}{3} \cdot \frac{k+1}{k} \leq \frac{2}{3},$$

pentru toți $k \geq 1$. Astfel, fiecare termen este mărginit superior de $(1/3)(2/3)^k$, astfel încât

$$\sum_{k=1}^{\infty} \frac{k}{3^k} \leq \sum_{k=1}^{\infty} \frac{1}{3} \left(\frac{2}{3}\right)^k = \frac{1}{3} \cdot \frac{1}{1-2/3} = 1.$$

O greșeală banală în aplicarea acestei metode este să arătăm că raportul termenilor consecutivi este mai mic decât 1 și apoi să presupunem că suma este mărginită de o serie geometrică. Un exemplu este seria armonică infinită, care diverge, deoarece

$$\sum_{k=1}^{\infty} \frac{1}{k} = \lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{k} = \lim_{n \rightarrow \infty} \Theta(\lg n) = \infty.$$

Raportul termenilor $k+1$ și k în această serie este $k/(k+1) < 1$, dar seria nu e mărginită de o serie geometrică descrescătoare. Pentru a delimita o serie printr-o serie geometrică, trebuie să arătăm că raportul este mărginit de o constantă subunitară; adică trebuie să existe un $r < 1$, care este *constant*, astfel încât raportul tuturor perechilor de termeni consecutivi nu depășește niciodată r . În seria armonică nu există nici un astfel de r , deoarece raportul devine arbitrar de apropiat de 1.

Partiționarea sumelor

Delimitări pentru o sumare dificilă se pot obține exprimând seria ca o sumă de două sau mai multe serii, partiționând domeniul indicelui și apoi delimitând fiecare dintre seriile rezultate. De exemplu, să presupunem că încercăm să găsim o limită inferioară pentru seria aritmetică $\sum_{k=1}^n k$, despre care s-a arătat deja că are limita superioară n^2 . Am putea încerca să minorăm fiecare termen al sumei prin cel mai mic termen, dar deoarece acest termen este 1, găsim o limită inferioară de n pentru sumă – foarte departe de limita noastră superioară, n^2 .

Putem obține o limită inferioară mai bună, partiționând întâi suma. Să presupunem, pentru comoditate, că n este par. Avem

$$\sum_{k=1}^n k = \sum_{k=1}^{n/2} k + \sum_{k=n/2+1}^n k \geq \sum_{k=1}^{n/2} 0 + \sum_{k=n/2+1}^n (n/2) \geq (n/2)^2 = \Omega(n^2)$$

care este o limită asimptotic strânsă, deoarece $\sum_{k=1}^n k = O(n^2)$.

Pentru o sumă care apare din analiza unui algoritm, putem partiționa adesea suma și ignora un număr constant de termeni inițiali. În general, această tehnică se aplică atunci când fiecare termen a_k din suma $\sum_{k=0}^n a_k$ este independent de n . Atunci pentru orice $k_0 > 0$ constant putem scrie

$$\sum_{k=0}^n a_k = \sum_{k=0}^{k_0-1} a_k + \sum_{k=k_0}^n a_k = \Theta(1) + \sum_{k=k_0}^n a_k,$$

deoarece termenii inițiali ai sumei sunt toți constanți și numărul lor este constant. Apoi, putem utiliza alte metode pentru a delimita $\sum_{k=k_0}^n a_k$. De exemplu, pentru a găsi o limită superioară asimptotică pentru

$$\sum_{k=0}^{\infty} \frac{k^2}{2^k},$$

observăm că raportul termenilor consecutivi este

$$\frac{(k+1)^2/2^{k+1}}{k^2/2^k} = \frac{(k+1)^2}{2k^2} \leq \frac{8}{9},$$

dacă $k \geq 3$. Astfel, suma poate fi partiționată în

$$\sum_{k=0}^{\infty} \frac{k^2}{2^k} = \sum_{k=0}^2 \frac{k^2}{2^k} + \sum_{k=3}^{\infty} \frac{k^2}{2^k} \leq O(1) + \frac{9}{8} \sum_{k=0}^{\infty} \left(\frac{8}{9}\right)^k = O(1),$$

deoarece a doua sumă este o serie geometrică descrescătoare.

Tehnica partiționării sumelor poate fi utilizată pentru a găsi delimitări asimptotice în situații mult mai dificile. De exemplu, putem obține o delimitare $O(\lg n)$ a seriei armonice (3.5)

$$H_n = \sum_{k=1}^n \frac{1}{k}.$$

Ideea este să partiționăm domeniul de la 1 la n în $\lfloor \lg n \rfloor$ părți și să delimităm superior contribuția fiecărei părți prin 1. Astfel,

$$\sum_{k=1}^n \frac{1}{k} \leq \sum_{i=0}^{\lfloor \lg n \rfloor} \sum_{j=0}^{2^i-1} \frac{1}{2^i+j} \leq \sum_{i=0}^{\lfloor \lg n \rfloor} \sum_{j=0}^{2^i-1} \frac{1}{2^i} \leq \sum_{i=0}^{\lfloor \lg n \rfloor} 1 \leq \lg n + 1. \quad (3.8)$$

Aproximarea prin integrale

Când o sumă poate fi exprimată ca $\sum_{k=m}^n f(k)$, unde $f(k)$ este o funcție monoton crescătoare, o putem aproxima prin integrale:

$$\int_{m-1}^n f(x)dx \leq \sum_{k=m}^n f(k) \leq \int_m^{n+1} f(x)dx. \quad (3.9)$$

Justificarea pentru această aproximare este arătată în figura 3.1. Suma este reprezentată prin aria dreptunghiurilor din figură, iar integrala este regiunea hașurată de sub curbă. Când $f(k)$ este o funcție monoton descrescătoare, putem utiliza o metodă similară pentru a obține marginile

$$\int_m^{n+1} f(x)dx \leq \sum_{k=m}^n f(k) \leq \int_{m-1}^n f(x)dx \quad (3.10)$$

Aproximația integrală (3.10) dă o estimare strânsă pentru al n -lea număr armonic. Pentru o limită inferioară, obținem

$$\sum_{k=1}^n \frac{1}{k} \geq \int_1^{n+1} \frac{dx}{x} = \ln(n+1). \quad (3.11)$$

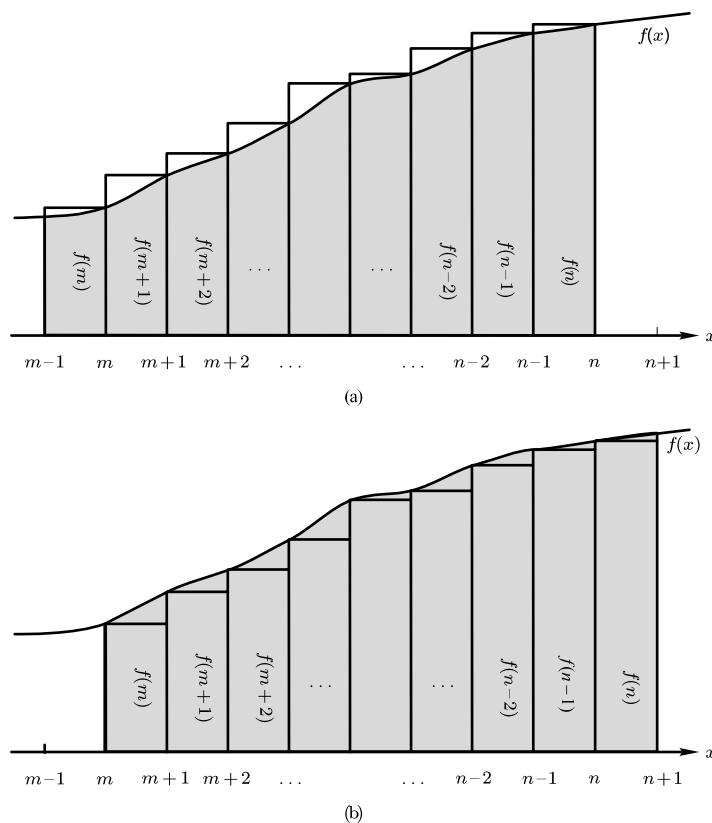


Figura 3.1 Aproximarea lui $\sum_{k=m}^n f(k)$ prin integrale. Aria fiecărui dreptunghi este indicată în interiorul dreptunghiului, iar aria tuturor dreptunghiurilor reprezintă valoarea sumei. Integrala este reprezentată prin aria hașurată de sub curbă. Comparând ariile în **(a)**, obținem $\int_{m-1}^n f(x)dx \leq \sum_{k=m}^n f(k)$, și apoi, deplasând dreptunghiurile cu o unitate la dreapta, obținem $\sum_{k=m}^n f(k) \leq \int_m^{n+1} f(x)dx$ din **(b)**.

Pentru limita superioară, deducem inegalitatea

$$\sum_{k=2}^n \frac{1}{k} \leq \int_1^n \frac{dx}{x} = \ln n,$$

care dă limita

$$\sum_{k=1}^n \frac{1}{k} \leq \ln n + 1 \quad (3.12)$$

Exerciții

3.2-1 Arătați că $\sum_{k=1}^n 1/k^2$ este mărginită superior de o constantă.

3.2-2 Găsiți o margine superioară a sumei

$$\sum_{k=0}^{\lfloor \lg n \rfloor} \lceil n/2^k \rceil.$$

3.2-3 Arătați că al n -lea număr armonic este $\Omega(\lg n)$, partiționând suma.

3.2-4 Aproximați $\sum_{k=1}^n k^3$ printr-o integrală.

3.2-5 De ce nu am utilizat aproximația integrală (3.10) direct pentru $\sum_{k=1}^n 1/k$ pentru a obține o margine superioară pentru al n -lea număr armonic?

Probleme

3-1 Delimitarea sumelor

Datați delimitări asimptotice strânse pentru următoarele sume. Presupuneți că $r \geq 0$ și $s \geq 0$ sunt constante.

a. $\sum_{k=1}^n k^r.$

b. $\sum_{k=1}^n \lg^s k.$

c. $\sum_{k=1}^n k^r \lg^s k.$

Note bibliografice

Knuth [121] este o referință excelentă pentru materialul prezentat în acest capitol. Proprietăți de bază ale seriilor pot fi găsite în orice carte bună de analiză, cum ar fi Apostol [12] sau Thomas și Finney [192].