ساختمان داده‌ها (۲۲۸۲۲)

مدرس: حسین بومری

[زمستان ۹۹]

سوال ۵: مرتب سازی خطی — نگارنده: آئیریا محمدی

۱

اثبات. برهان خلف فرض کنیم چنین آرایه‌ای را می‌توان در زمان خطی با الگوریتمی به اسم lin-sort مرتب کرد.

الگوریتم زیر را تبدیل هر آرایه دلخواه به شکل مورد نظر ارائه می‌دهیم

$f(B, n)$

```
condition = true // true for greater, false for smaller
for i = 1 : n
    if condition and B[i] < B[i-1]
        swap B[i], B[i-1]
    if not condition and B[i] > B[i-1]
        swap B[i], B[i-1] condition = not condition
```

تحلیل زمانی الگوریتم الگوریتم هر عضو از آرایه A را یک بار می‌بیند و در هر دور $O(1)$ عملیات انجام می‌دهد. پس از $O(n)$ است.

درستی الگوریتم اثبات با استقرا.

برای $n = 1$ درستی آن بدیهی است ($B = A = \{x\}$) .

اگر برای آرایه‌های با اندازه کمتر از n الگوریتم را درست فرض کنیم، برای آرایه به اندازه n خواهیم داشت: ؟؟؟

حال می‌توانیم هر آرایه دلخواه را با الگوریتم زیر مرتب کنیم:

$sort(A, n)$

```
f(A, n) // O(n)
lin-sort(A, n) // O(n)
```

تحلیل زمانی و درستی الگوریتم از مرتبه‌زمانی $O(n)$ است و درستی آن بنا بر درستی الگوریتم‌های استفاده شده و برقراری شرط مطرح شده درست است.

در حالی که می‌دانیم نمی‌توان چنین الگوریتمی با مرتبه زمانی کمتر از $O(n \cdot logn)$ داشت. پس فرض غلط است و الگوریتم lin-sort نمی‌تواند وجود داشته باشد.

دایره مورد نظر را به n بخش با مساحت‌های مساوی تقسیم می‌کنیم. مساحت هر بخش برابر با $\pi(1^2)/n = \pi/n$ خواهد بود. برای دایره مرکزی خواهیم داشت

$$\pi r^2 = \pi/n \to r = \sqrt{1/n} \tag{1}$$

و شعاع دایره ناحیه دوم برابر خواهد بود با

$$\pi R^2 - \pi r^2 = \pi R^2 - \pi/n = \pi/n \to R = \sqrt{2/n} \tag{2}$$

و برای هر ناحیه شماره i داریم $(r = r_{i-1}, R = r_i)$

$$\pi R^2 - \pi r^2 = \pi/n \to R = \sqrt{\frac{1}{n} + r^2} \tag{3}$$

تعریف:

$$r_i = \text{شعاع دایره i ام}$$

ادعا: شعاع دایره محیطی ناحیه i برابر است با $\sqrt{\frac{i}{n}}$

اثبات با استقرا:

برای پایه استقرا، صحت فرض را به ازای $i = 2$ در بالا نشان دادیم، و فرض را برای $i - 1$ دایره اول درست فرض می‌کنیم.

اثبات برای i

$$R = \sqrt{\frac{1}{n} + r^2} = \sqrt{\frac{1}{n} + (\sqrt{\frac{i-1}{n}})^2} = \sqrt{\frac{1}{n} + \frac{i-1}{n}} = \sqrt{\frac{i}{n}} \tag{4}$$

حال مسئله اصلی را حل می‌کنیم.

با توجه به فرض سوال، از آن‌جایی که متوسط تعداد نقاط در هر ناحیه متناسب با مساحت آن ناحیه است اگر دایره را به ناحیه‌های هم مساحت تقسیم کنیم به طور میانگین در هر ناحیه یک نقطه خواهیم داشت.

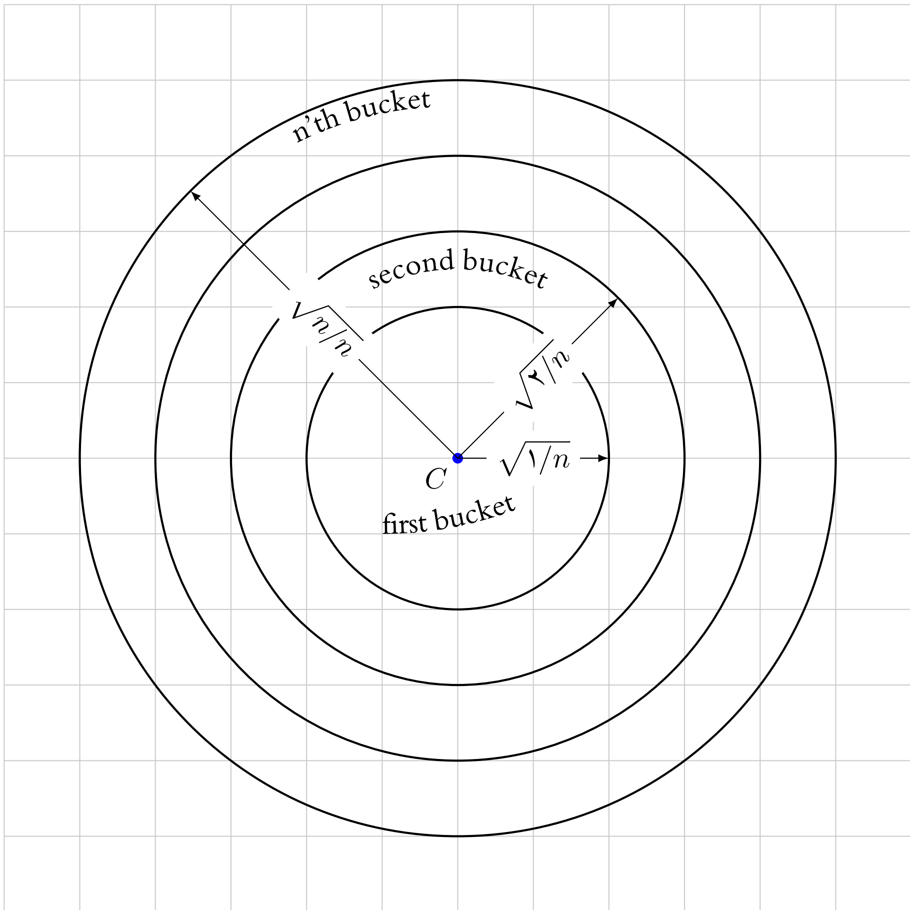
اگر از الگوریتم bucket sort استفاده کنیم:

$$T(n) = \Theta(n) + \sum_{i=1}^{n} (n_i)^2 \tag{5}$$

که n_i تعداد نقاط در ناحیه i است.

و اگر امید ریاضی مرتبه زمانی را حساب کنیم

$$E[T(n)] = E[\Theta(n) + \sum_{i=1}^{n} (n_i)^2] \tag{6}$$



شکل ۱: دایره را به n ناحیه با مساحت مساوی تقسیم کردیم

بعد دو مرحله استفاده از خاصیت خطی‌بودن امیدریاضی

$$E[T(n)] = E[\Theta(n)] + \sum_{i=1}^{n} E[n_i^2] \tag{۷}$$

اگر تعریف کنیم

$$X_{ij} = \begin{cases} 1 & \text{point i being in area j} \\ 0 & \text{otherwise} \end{cases} \tag{۸}$$

خواهیم داشت

$$n_j = \sum_{i=0}^{n} X_{ij} \tag{۹}$$

پس

$$E[n_i^2] = E[(\sum_{j=0}^{n} X_{ji})^2] = E[\sum_{j=0}^{n} X_{ji}^2] + E[\sum_{j=0}^{n}\sum_{k=0}^{n} 2X_{ji}X_{ki}] \tag{۱۰}$$

برای بخش اول از آن‌جایی که احتمال وجود نقطه در هر ناحیط برابر و برابر با $\frac{1}{n}$ است داریم

$$E[X_{ji}^2] = (1 - \frac{1}{n}) * 0^2 + \frac{1}{n} * 1^2 = \frac{1}{n} \tag{۱۱}$$

$$E[\sum_{j=0}^{n} X_{ji}^2] = \sum_{j=0}^{n} E[X_{ji}^2] = n \cdot \frac{1}{n} = 1 \tag{۱۲}$$

و برای بخش دوم با توجه با استقلال X_{ji} و X_{ki} داریم

$$E[\sum_{j=0}^{n}\sum_{k=0}^{n} 2X_{ji}X_{ki}] = \sum_{j=0}^{n}\sum_{k=0}^{n} 2E[X_{ji}]E[X_{ki}] = \binom{n}{2} \cdot 2 \cdot \frac{1}{n} \cdot \frac{1}{n} = \frac{n(n-1)}{n^2} = 1 - \frac{1}{n} \tag{۱۳}$$

و در نتیجه

$$E[n_i^2] = 2 - \frac{1}{n} \tag{۱۴}$$

در نهایت خواهیم داشت

$$E[T(n)] = E[\Theta(n)] + \sum_{i=1}^{n} 2 - \frac{1}{n} = \Theta(n) + 2n - n(1/n) = \Theta(n) + O(n) \tag{۱۵}$$