Ordering Problem

Given an array holding lengths of files, $L[1 \ldots n]$, create an output array $M[1 \ldots n]$ specifying the order in which the files should be stored to minimize access time.

- cost to read the k^{th} file if $M[i] = i$

$$\text{cost}(k) = \sum_{i=1}^{k} L[i]$$

- assume that all files are accessed with the same probability

$$E[\text{cost}] = \sum_{k=1}^{n} \frac{\text{cost}(k)}{n} = \sum_{k=1}^{n} \sum_{i=1}^{k} \frac{L[i]}{n}$$

- suppose we permute the order of the files
  - let $\pi(i)$ be the index of the file stored at position i on the tape

$$E[\mathrm{cost}(\pi)] = \sum_{k=1}^{n} \frac{\mathrm{cost}(k)}{n} = \sum_{k=1}^{n} \sum_{i=1}^{k} \frac{L[\pi(i)]}{n}$$

How should we choose our permutation?

smallest to largest

Theorem

$E[\text{cost}]$ is minimized when

$\forall i \quad L[\pi(i)] \leq L[\pi(i+1)]$

Proof

- assume for contradiction that π minimizes expected cost but $L[\pi(i)] > L[\pi(i+1)]$ for some i
- let $\pi(i) = a$ and $b = \pi(i+1)$

- If we swap a and b in π
  - the cost of accessing file a increases by $L[b]$
  - the cost of accessing file b decreases by $L[a]$
- the overall expected cost changes by

$$\frac{L[b] - L[a]}{n}$$

- since $L[b] < L[a]$ this is a decrease in the expected cost.
- contradicting the assumption that π is minimal.

Unequal File Access Frequencies

Suppose the files are not equally likely to be accessed. For simplicity assume we are given an array $F[1 \ldots n]$ of access frequencies for each file. File i will be accessed exactly $F[i]$ times over the lifetime of the tape.

- Does our greedy strategy still work?

$L = [1, 2]$

$F = [0, 1]$

greedy strategy $M = [1, 2]$

total cost: 3

optimal $M = [2, 1]$

total cost = 2

Goal: minimize the lifetime cost to access files.

$$\text{Lifetime-cost}(\pi) = \sum_{k=1}^{n} \left(F[\pi(k)] \, \text{cost}(k) \right)$$

$$= \sum_{k=1}^{n} \left(F[\pi(k)] \sum_{i=1}^{k} L[\pi(i)] \right)$$

$$= \sum_{k=1}^{n} \sum_{i=1}^{k} \left(F[\pi(k)] \, L[\pi(i)] \right)$$

What should our greedy strategy be?

- highest frequency first

$L = [100, 1]$

$F = [10, 9]$

greedy $M = [1, 2]$

cost $= 1909$

optimal $M = [2, 1]$

cost $= 1019$

Better Greedy Strategy

$$\frac{\text{length}}{\text{frequency}}$$

Theorem

lifetime cost is ~~minimed~~ minimized when

$$\forall i \quad \frac{L[\pi(i)]}{F[\pi(i)]} \le \frac{L[\pi(i+1)]}{F[\pi(i+1)]}$$

Proof

- assume for contradiction that π that minimizes lifetime cost but

$$\frac{L[\pi(i)]}{F[\pi(i)]} > \frac{L[\pi(i+1)]}{F[\pi(i+1)]}$$

- let $a = \pi(i)$ and $b = \pi(i+1)$
- If we swap a and b
  - cost of accessing file a increases by $L[b]$
  - cost of accessing file b decreases by $L[a]$
  - the overall lifetime cost change by

$$F[a]L[b] - F[b]L[a]$$

$$-\frac{L[b]}{F[b]} < \frac{L[a]}{F[a]} \Rightarrow F[a]L[b] - F[b]L[a]$$

is negative

contradicting the assumption that π minimizes lifetime cost.