OPTIMAL INTERVAL PARTITIONING AT GIVEN POINTS

Mikhail Y. Kovalyov (Minsk, Belarus)
Sergey Sevastyanov (Novosibirsk, Russia),
Bertrand Lin (Hsinchu, Taiwan),
Feng-Jang Hwang (Sydney, Australia)
1. Formulation.
2. Scheduling interpretation.
4. Existing results.
1. FORMULATION

Problem $K$-segments

Cut into $K$ segments

$N-1$ cut points
Minimize $\text{MaxSize} - \text{MinSize}$
(balancing criterion)
Identical Parallel Machine Scheduling:
Cut given job sequence \((1,\ldots,N)\) into \(K\) segments and assign each segment to a machine.

Minimize \(\text{MaxMachineLoad - MinMachineLoad}\)
3. MOTIVATION

Efficient Retrieval of E-documents (Zobel et al. 1995)

Find E-docs containing (indexed) TEXT

To answer queries effectively, E-docs are ranked. Ranking techniques are biased – some favor large docs, some small – and behave poorly if sizes vary. Experiments shown that partitioning large docs into “almost equal” parts improves retrieval effectiveness.

“War and Peace”:

\[
p_1 \quad p_2 \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad p_N
\]

Ends of paragraphs
Fare zoning for public transport

Partition sequence of consecutive route segments into $K$ zones so that zone lengths are maximally close to each other.
Road construction works

Partition consecutive road intervals between entry/exit points into K segments and assign each segment to a separate construction team such that team workloads are balanced.
4. EXISTING RESULTS

Imre Barany, Victor Goldberg (2013) **Block Partitions of Sequences** [arXiv:1308.2452v3][math.CO]

Scaled $p_j \leq 1$:
- ($\text{MaxSize}-\text{MinSize}) \leq 2$: round "ideal" cut points $k\sum_{j=1}^{N} p_j/K, \ k=1,\ldots, K-1,$ down to the nearest ideal cut points
- ($\text{MaxSize}-\text{MinSize}) \leq 1$: $O(KN^3)$ algorithm

- Optimal solution: ???
5. DYNAMIC PROGRAMMING 1

Richard Bellman (Eye of the Hurricane, autobiography): “I wanted to get across that this was dynamic, this was multi-stage, this was time-varying . . . . Let’s take a word that has an absolutely precise meaning, namely dynamic, in the classical physical sense. It also has a very interesting property that it is impossible to use the word, dynamic, in a pejorative sense . . . . It was something that not even a Congressman could object to. So I used it as an umbrella for my activities.”

Partial solution $S_k=(s_1,\ldots,s_k)$: **state** $(n,k,t)$

$t$ – future (upper bound on) max segment size,

$\Delta(n,k,t)=\min_{\text{MaxSize} \leq t}(t-\text{MinSize})$
Partial solution $S_k = (s_1, \ldots, s_k)$: state $(n, k, t)$

- $t$ – future (upper bound on) maximum segment size,
- $\Delta(n, k, t) = \min_{\text{MaxSize} \leq t} (t - \text{MinSize})$
- $P(j, n) = \sum_{i=j}^{n} p_i$, $B = \{ P(j, n) \mid 1 \leq j \leq n \leq N \}$, $t \in B$

**Init:**

\[
\Delta(n, 1, t) = \begin{cases} 
  t - P(1, n), & \text{if } t \geq P(1, n) \\
  \infty, & \text{otherwise}
\end{cases}
\]

**Recu:**

\[
\Delta(n, k, t) = \min_{0 \leq j \leq n, \ P(j+1, n) \leq t} \{ \max\{t - P(j+1, n), \Delta(j, k-1, t)\} \}
\]

\[
F^* = \min_{t \in B} \{ \Delta(N, K, t) \}
\]
Feasibility problem $K$-segments:

- Order all distinct segment sizes: $L_1 < L_2 < \ldots < L_A$, $A \leq N^2$.
- For interval $[L_a, L_b]$, $1 \leq a \leq b \leq A$, in $O(T)$ verify
- **Property**: All segment sizes $\in [L_a, L_b]$. Satisfied $\Rightarrow$ feasible.

Solving problem $K$-segments

- Find all *minimal* feasible intervals $[L_a, L_b]$, $1 \leq a \leq b \leq A$. ($# \leq A$)
- Optimal solution corresponds to a *minimum* feasible interval.

$O(AT)$

Problem AllMin:

- Input: numbers 1,\ldots,A.
- Intrl $[L_a, L_b]$, $1 \leq a \leq b \leq A$, can satisfy Property, verified in $O(T)$.
- Property satisfied for Intrl $\Rightarrow$ satisfied for any surrounding Intrl.
- Find all minimal feasible intervals $[L_a, L_b]$, $1 \leq a \leq b \leq A$. 

Solving problem AllMin

Matrix $Y = (y_{ab})$:

$$y_{ab} = 1 \text{ iff } [L_a, L_b] \text{ feasible}$$

Solution = red entries

Sufficient to consider “staircase” with $\leq 2A-1$ entries $\Rightarrow$ $O(AT)$

$A \leq N^2$ for K-segments
7. DP2 FOR FEASIBILITY PROBLEM

Problem Feasible(a,b):
All segment sizes belong to $[L_a,L_b]$?

Equivalent definition:
$S=(s_1,\ldots,s_K)$, $s_k=(i_{k-1}+1,\ldots,i_k)$, $k=1,\ldots,K$, $i_0=0$, is feasible iff

$$i_k \in [A(i_{k-1}),B(i_{k-1})], \ k=1,\ldots,K,$$ where

- $A(i_{k-1})=\min_{n \in [i_{(k-1)}+1,N]} \{ n \mid P(i_{k-1}+1,n) \geq L_a \}$
- $B(i_{k-1})=\max_{n \in [i_{(k-1)}+1,N]} \{ n \mid P(i_{k-1}+1,n) \leq L_b \}$

$[A(i_{k-1}),B(i_{k-1})], \ k=1,\ldots,K,$ belong to wider set of
$[A(j),B(j)], \ j=0,1,\ldots,N-1,$ (some can be $\emptyset$)
calculated in $O(N)$ and enumerated in DP2.
[A(i_{k-1}), B(i_{k-1})], k=1,\ldots,K, belong to wider set of [A(j), B(j)], j=0,1,\ldots,N-1, (some can be $\emptyset$) calculated in $O(N)$ and enumerated in **DP2**

Non-empty [A(j), B(j)] satisfy

$A(j_1) \leq A(j_2) \leq \ldots \leq A(j_h)$ and $B(j_1) \leq B(j_2) \leq \ldots \leq B(j_h)$

$O(N)$ to find $A(j_1), A(j_2), \ldots, A(j_h)$ and $B(j_1), B(j_2), \ldots, B(j_h)$
Partial solution

\[ S = (1, \ldots, n) = (s_1, \ldots, s_k), \ s_r = (i_{r-1} + 1, \ldots, i_r), \ r = 1, \ldots, k: \]

**feasible state** \((k, n)\) if

\[ i_r \in [A(i_{r-1}), B(i_{r-1})], \ r = 1, \ldots, k \]

**\(X_k\) – set of feasible states** \((k, n), \ n = 1, \ldots, N\)

\[ P(j, n) = \sum_{i=j}^{n} p_i \]

**Init:** \(X_1 = \{(1, n) \mid L_a \leq P(1, n) \leq L_b, \ n = 1, \ldots, N\}\)

**Recu:** \(X_k = \{(k, n) \mid n \in [A(j), B(j)], (k-1, j) \in X_{k-1}\} \)

If \(X_K = \{(K, N)\} \Rightarrow [L_a, L_b] \) feasible
≤ N² problems Feasible\([a,b]\), each in \(O(KN)\) => Problem K-segments can be solved in \(O(KN^3)\)

Minimize (MaxSize-MinSize)

Thank you