14-C2

排序

希尔排序：Shell序列 + 输入敏感性

邓俊辉
deng@tsinghua.edu.cn

让上帝的归上帝，凯撒的归凯撒

Shell's Sequence, 1959: $\mathcal{H}_{shell} = \{ 1, 2, 4, 8, 16, 32, 64, \dots, 2^k, \dots \}$

- 实际上，采用 $\mathcal{H}_{\text{shell}}$，在最坏情况下需要运行 $\Omega(n^2)$ 时间...
- 考查由子序列 A = unsort[0, 2^{N-1}) 和 B = unsort[2^{N-1}, 2^N) 交错而成的序列

11	4	14	3	10	0	15	1	9	6	8	7	13	2	12	5

- 在做2-sorting时，A、B各成一列；故此后必然各自有序

8	0	9	1	10	2	11	3	12	4	13	5	14	6	15	7

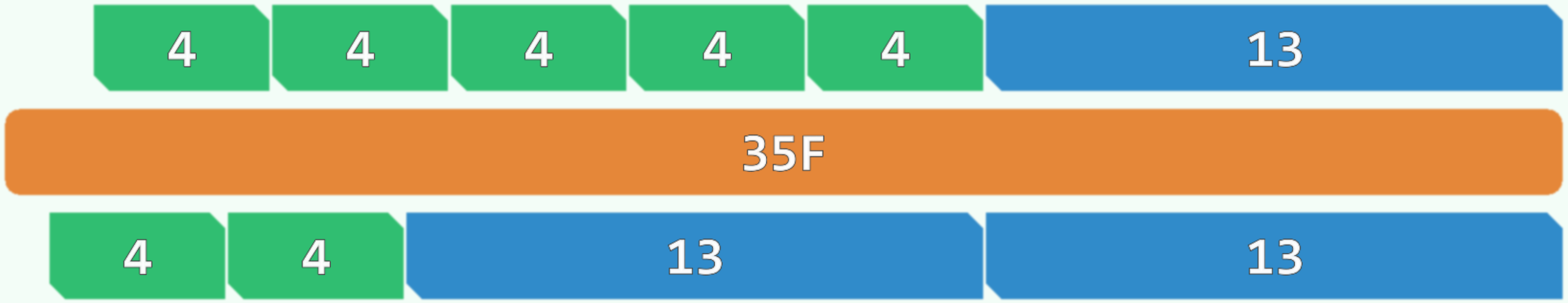
- 然而其中的逆序对依然很多，最后的1-sorting仍需 $1 + 2 + 3 + \cdots + 2^{N-1} = \Omega(n^2/4)$ 时间
- 问题的根源在于，$\mathcal{H}_{\text{shell}}$中各项并不互素，甚至相邻项也非互素

Postage Problem

❖ The postage for a letter is **50F**, and a postcard **35F**

But there are only stamps of **4F** and **13F** available

❖ Possible to stamp the letter and the postcard EXACTLY?



❖ Given a postage P, determine whether $P \in \{ n \cdot 4 + m \cdot 13 \mid n, m \in \mathcal{N} \}$

Linear Combination

- Let $g, h \in \mathcal{N}$

  For any $n, m \in \mathcal{N}$, $n \cdot g + m \cdot h$ is called a **linear combination** of g and h

- Denote $\mathbf{C}(g,h) \;=\; \{\, ng + mh \mid n, m \in \mathcal{N} \,\}$

  $\mathbf{N}(g,h) \;=\; \mathcal{N} \backslash \mathbf{C}(g,h)$ //numbers that are NOT combinations of g and h

  $\mathbf{x}(g,h) \;=\; \max\{\, \mathbf{N}(g,h) \,\}$ //always exists?

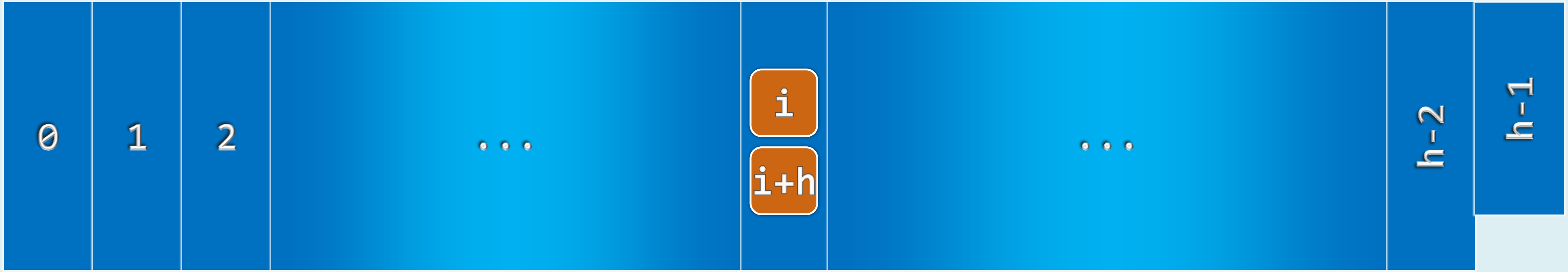
- Theorem: when g and h are **relatively prime**, we have

  $\mathbf{x}(g,h) \;=\; (g-1)\cdot(h-1)-1 \;=\; gh - g - h$

  e.g. $\mathbf{x}(3,7) \;=\; 11,\quad \mathbf{x}(4,9) \;=\; 23,\quad \mathbf{x}(\boxed{4},\boxed{13}) \;=\; \boxed{35},\quad \mathbf{x}(5,14) \;=\; 51$

h-sorting & h-ordered

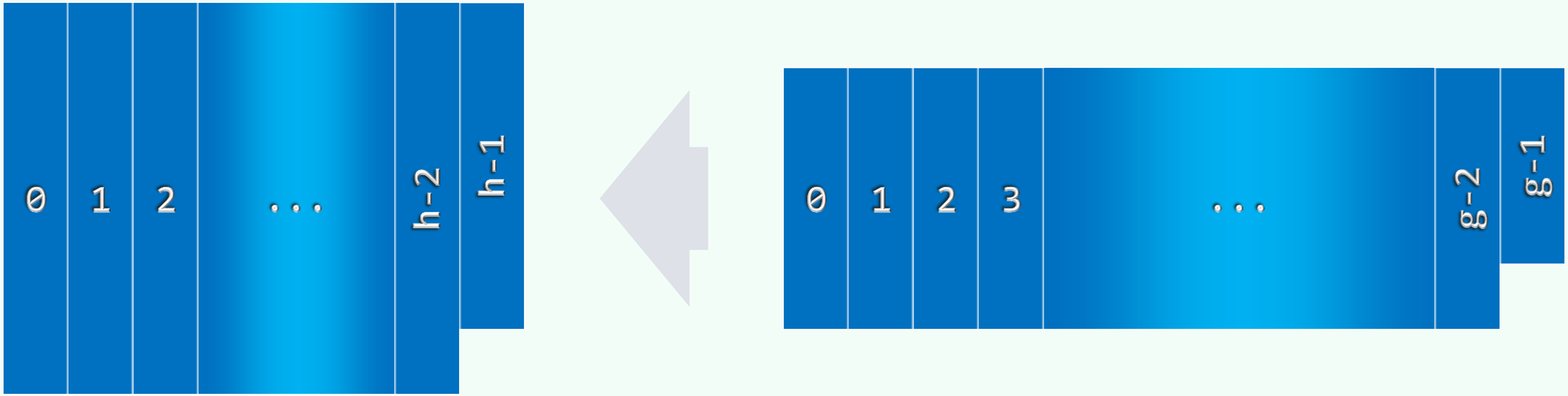
- A sequence S[0,n) is called h-ordered if $S[i] \leq S[i+h], \forall\, 0 \leq i < n-h$
- A 1-ordered sequence is sorted
- h-sorting: an h-ordered sequence is obtained by
  - arranging S into a 2D matrix with h columns and
  - sorting each column respectively



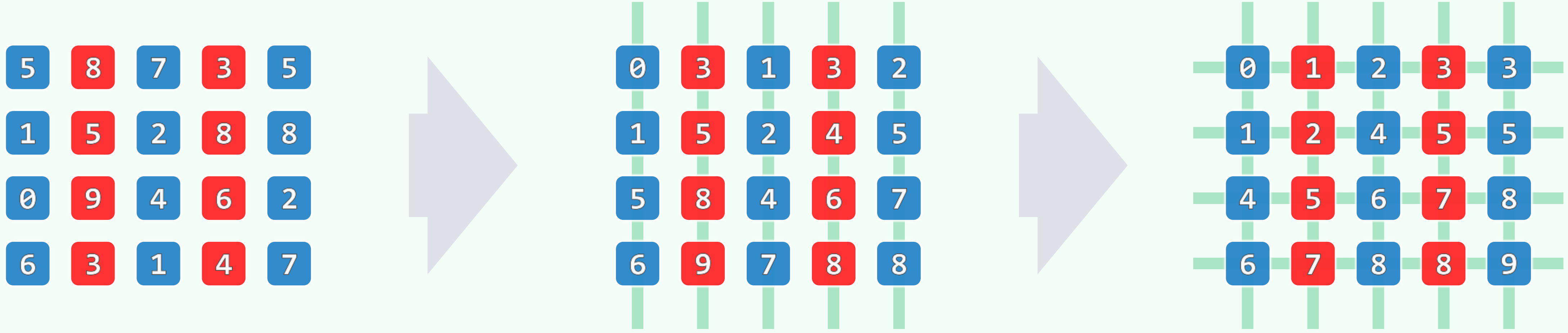
Theorem K

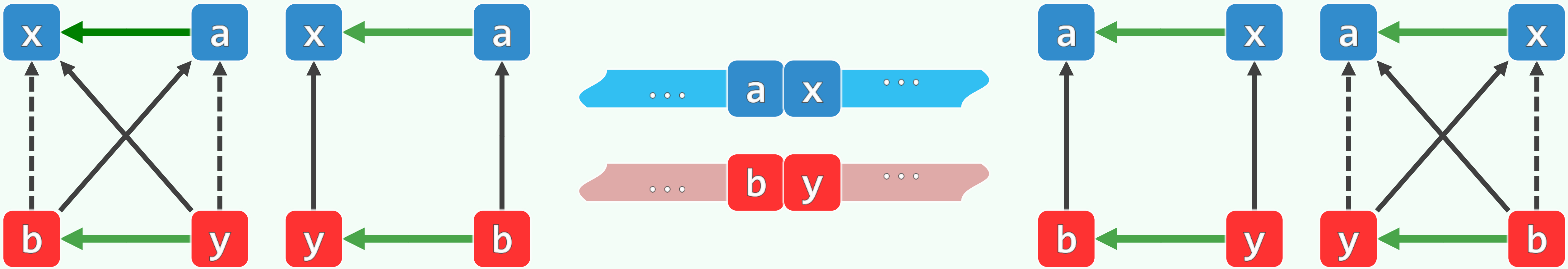
❖ [Knuth, ACP Vol.3 p.90] //习题解析[12-12, 12-13]

A g-ordered sequence REMAINS g-ordered after being h-sorted.

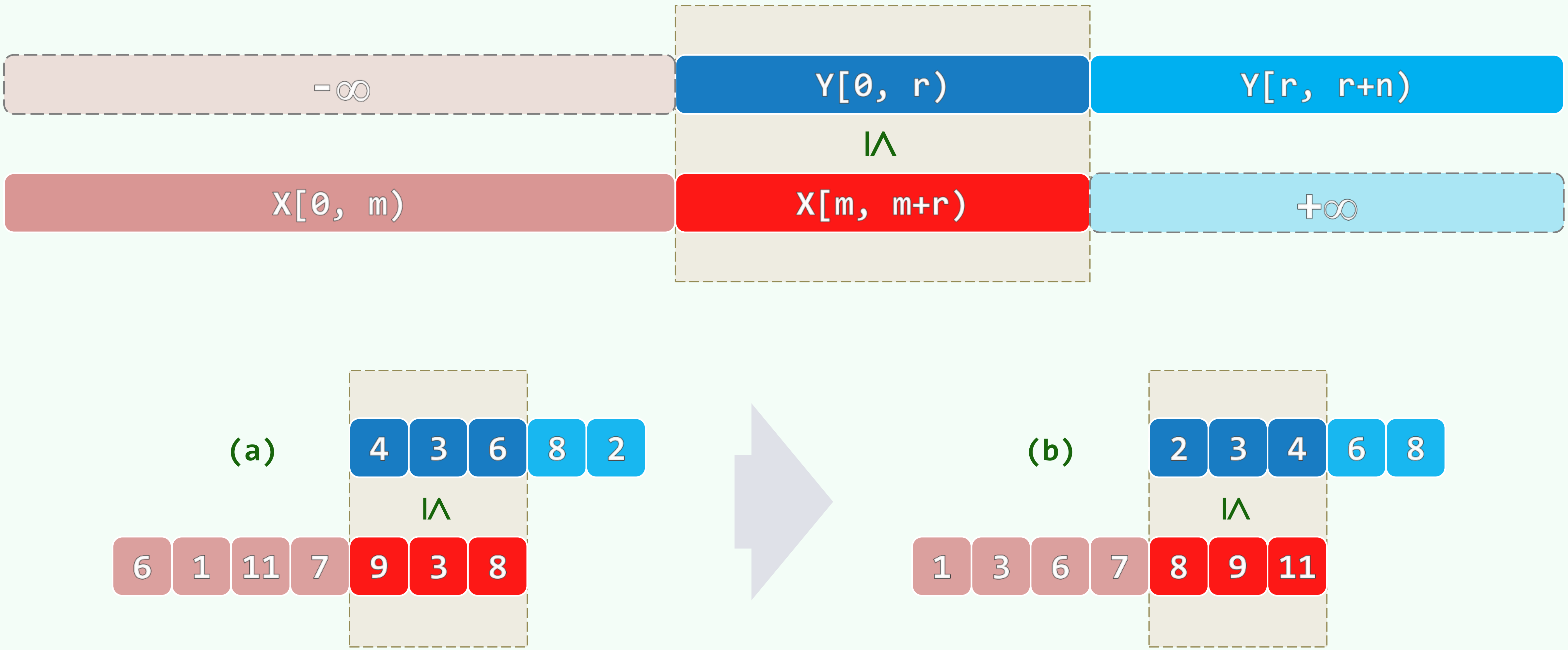


Order Preservation



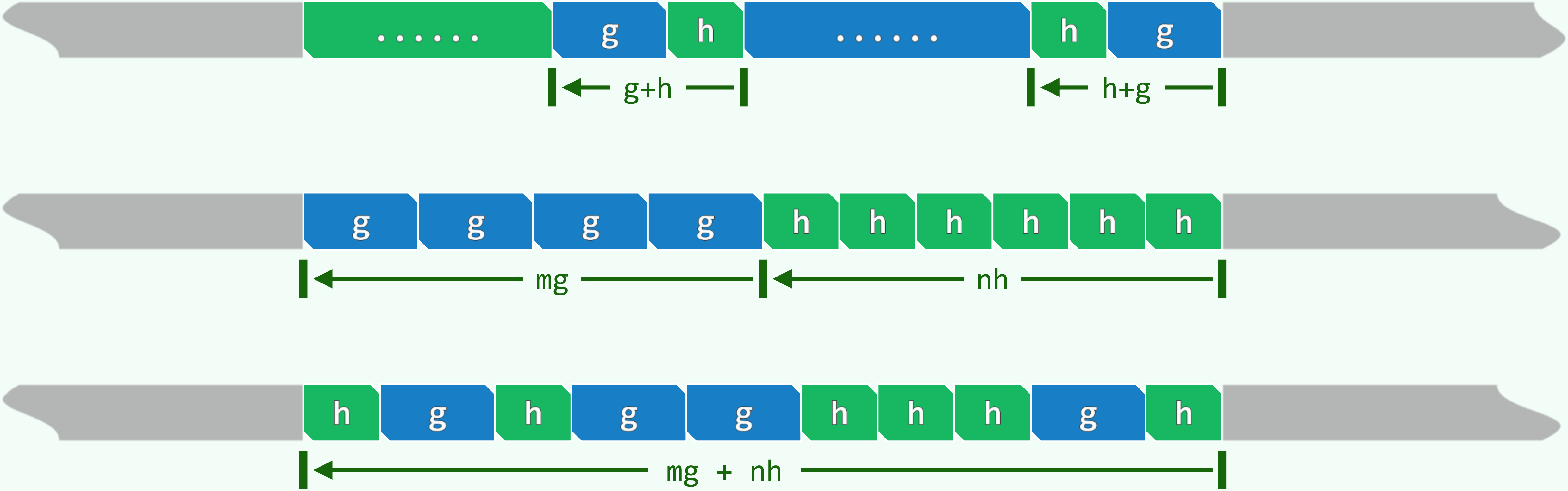


Lemma L



Linear Combination

A sequence that is both g-ordered and h-ordered is called (g,h)-ordered, which must be both (g+h)-ordered and (mg+nh)-ordered for any m, n ∈ N



Inversion

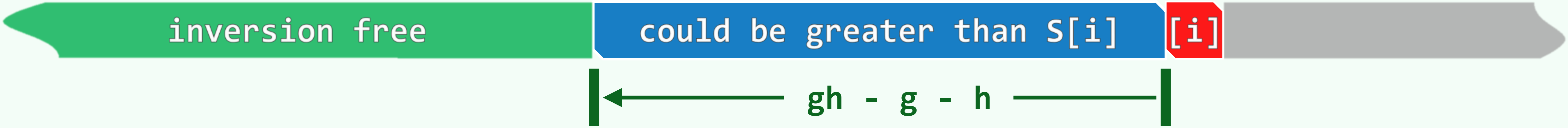
- Let S[0,n) be a (g,h)-ordered sequence, where g and h are **relatively prime**

  Then for all elements S[j] and S[i], we have

$$i - j > \mathbf{x}(g,h) \qquad \text{only if} \qquad S[j] \leq S[i]$$

- This implies that to the **LEFT** of each element,

  only the previous $\mathrm{x}(g,h)$ elements could be **GREATER**



There would be no more than $n \cdot \mathrm{x}(g,h)$ **INVERSIONs** altogether