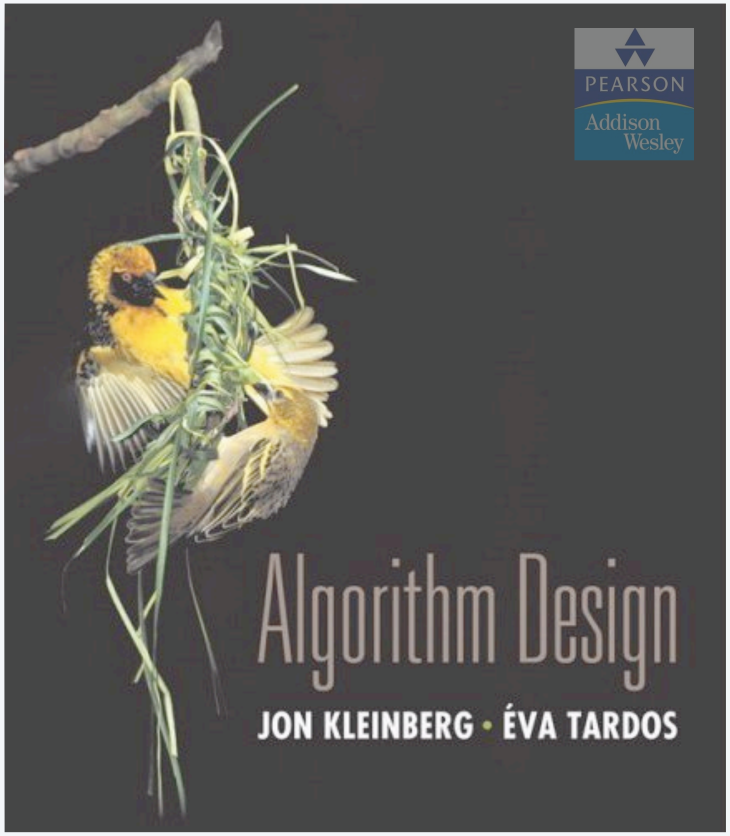


SECTION 6.7

6. DYNAMIC PROGRAMMING II

- *Hirschberg's algorithm*

Sequence alignment in linear space

Theorem. There exist an algorithm to find an optimal alignment in $O(mn)$ time and $O(m+n)$ space.

- Clever combination of divide-and-conquer and dynamic programming.
- Inspired by idea of Savitch from complexity theory.

Programming Techniques — G. Manacher, Editor

A Linear Space Algorithm for Computing Maximal Common Subsequences

D.S. Hirschberg
Princeton University

The problem of finding a longest common subsequence of two strings has been solved in quadratic time and space. An algorithm is presented which will solve this problem in quadratic time and in linear space.

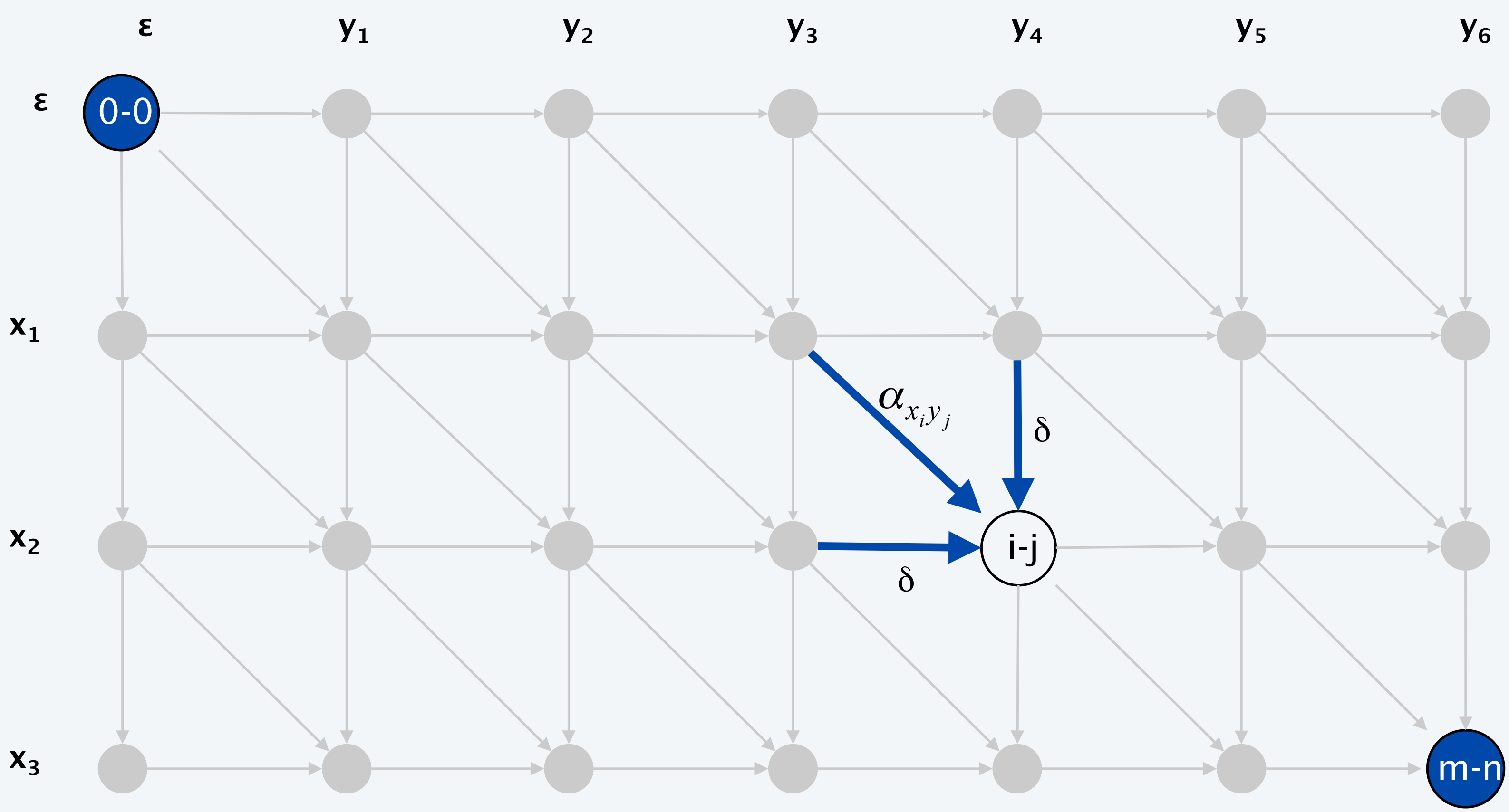
Key Words and Phrases: subsequence, longest common subsequence, string correction, editing

CR Categories: 3.63, 3.73, 3.79, 4.22, 5.25

Hirschberg's algorithm

Edit distance graph.

- Let $f(i, j)$ be shortest path from $(0,0)$ to (i, j).
- Lemma: $f(i, j) = OPT(i, j)$ for all i and j.



Hirschberg's algorithm

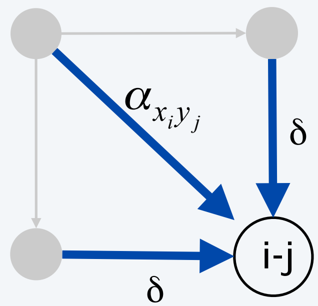
Edit distance graph.

- Let $f(i, j)$ be shortest path from $(0,0)$ to (i, j).
- Lemma: $f(i, j) = OPT(i, j)$ for all i and j.

Pf of Lemma. [by strong induction on $i + j$]

- Base case: $f(0, 0) = OPT(0, 0) = 0$.
- Inductive hypothesis: assume true for all (i', j') with $i' + j' < i + j$.
- Last edge on shortest path to (i, j) is from $(i - 1, j - 1)$, $(i - 1, j)$, or $(i, j - 1)$.
- Thus,

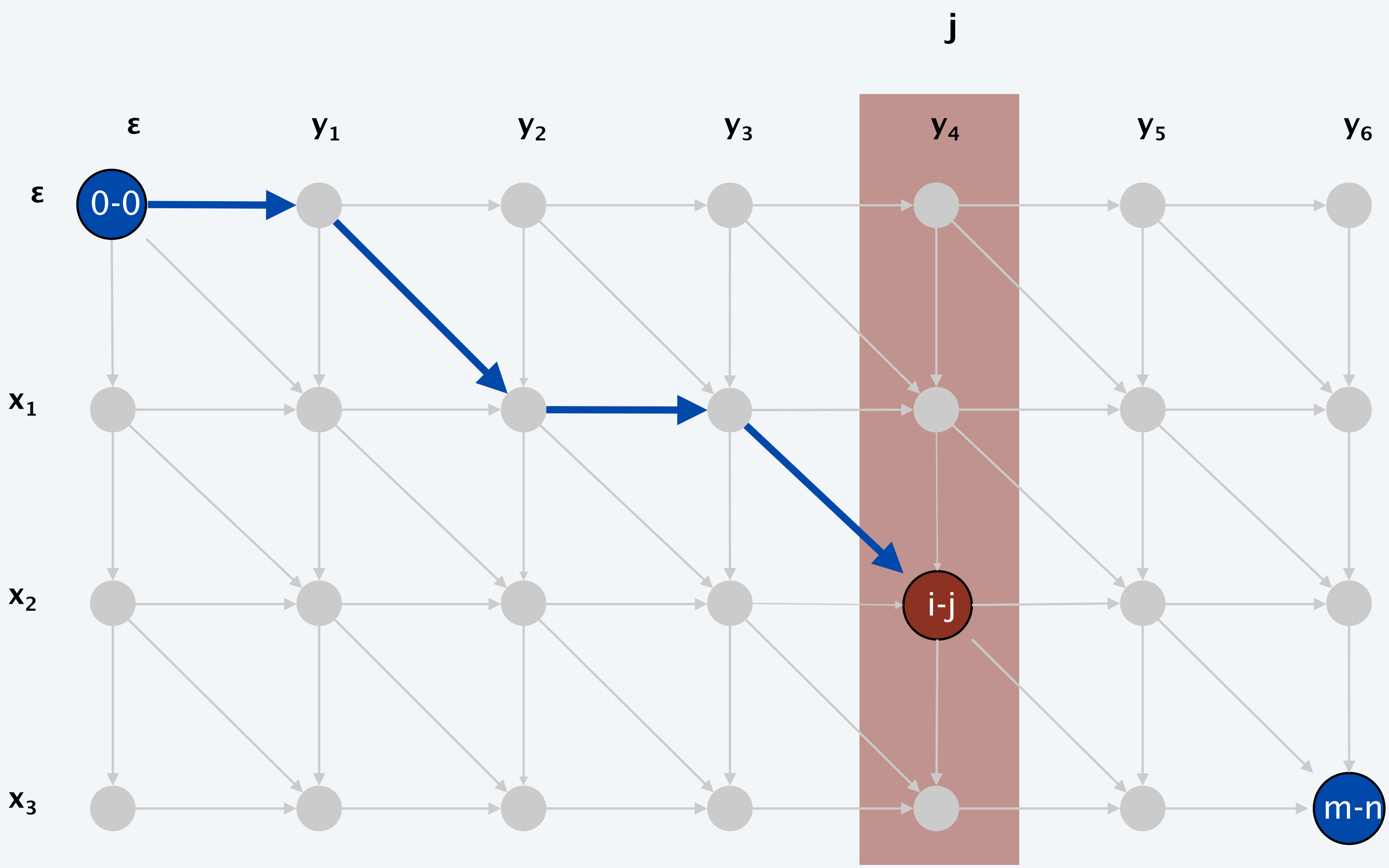
$$\begin{aligned} f(i,j) &= \min\{\alpha_{x_i y_j} + f(i-1, j-1),\ \delta + f(i-1, j),\ \delta + f(i, j-1)\} \\ &= \min\{\alpha_{x_i y_j} + OPT(i-1, j-1),\ \delta + OPT(i-1, j),\ \delta + OPT(i, j-1)\} \\ &= OPT(i,j) \ \blacksquare \end{aligned}$$



Hirschberg's algorithm

Edit distance graph.

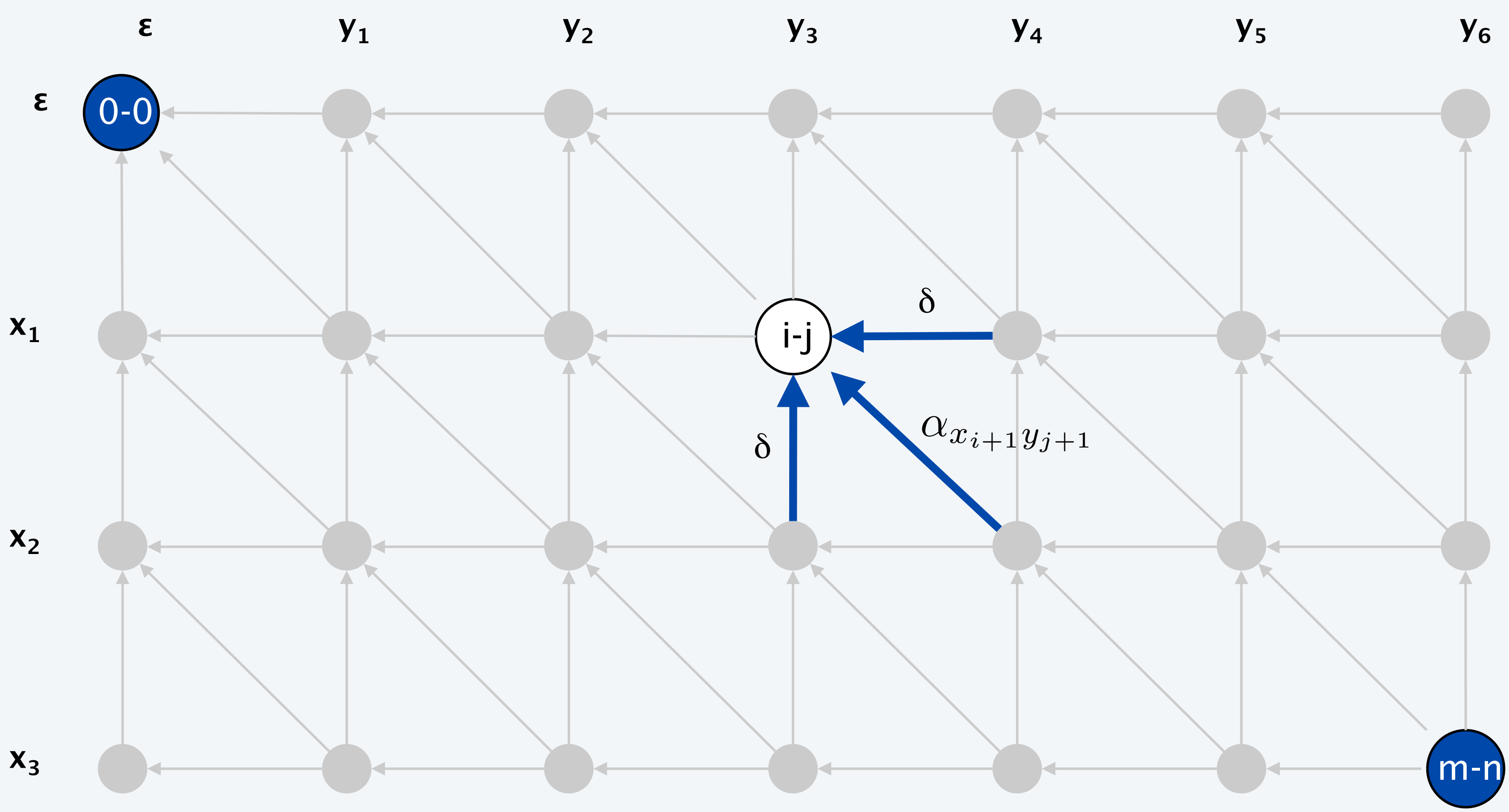
- Let $f(i, j)$ be shortest path from $(0,0)$ to (i, j).
- Lemma: $f(i, j) = OPT(i, j)$ for all i and j.
- Can compute $f(\bullet, j)$ for any j in $O(mn)$ time and $O(m + n)$ space.



Hirschberg's algorithm

Edit distance graph.

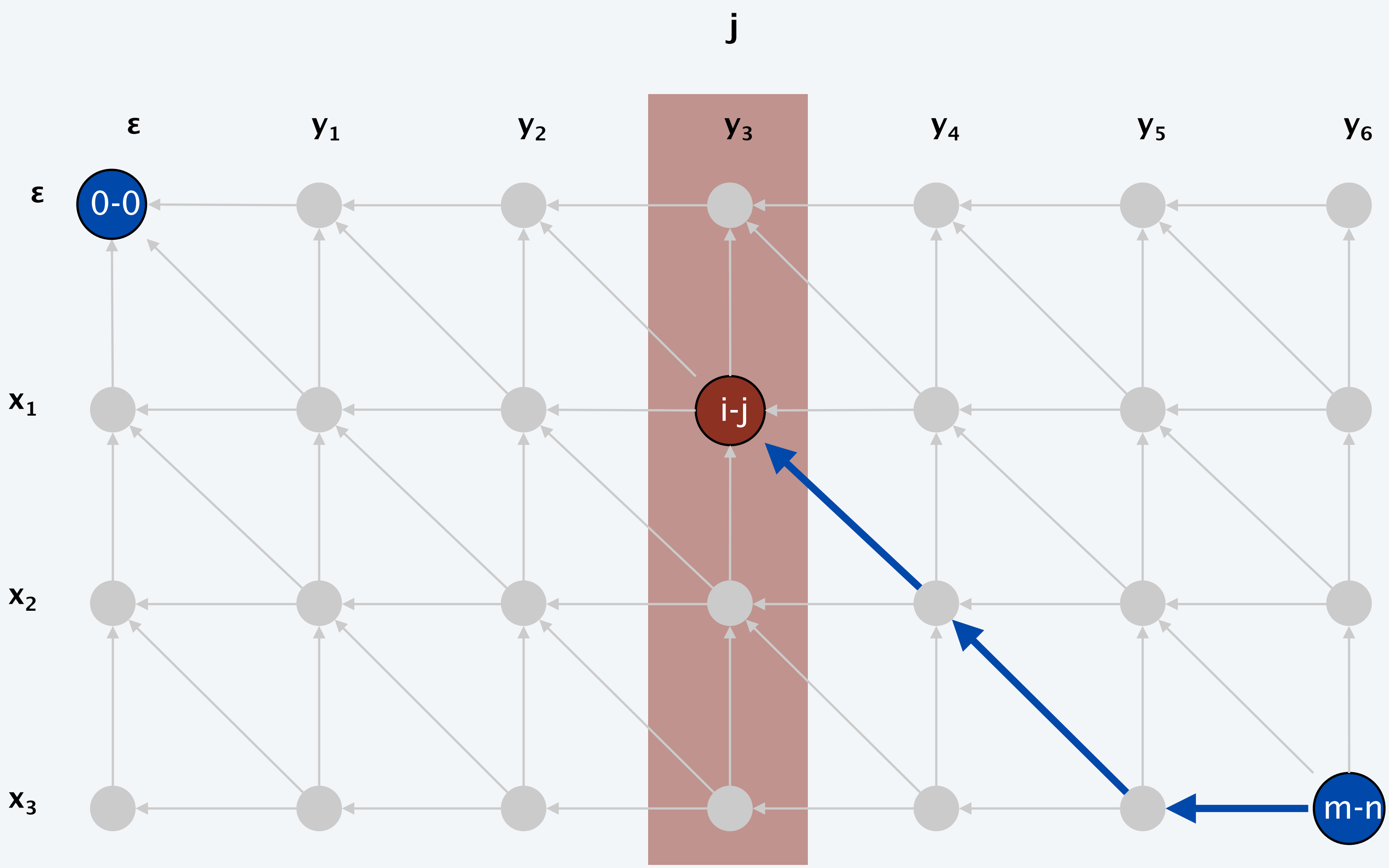
- Let $g(i, j)$ be shortest path from (i, j) to (m, n).
- Can compute by reversing the edge orientations and inverting the roles of $(0, 0)$ and (m, n).



Hirschberg's algorithm

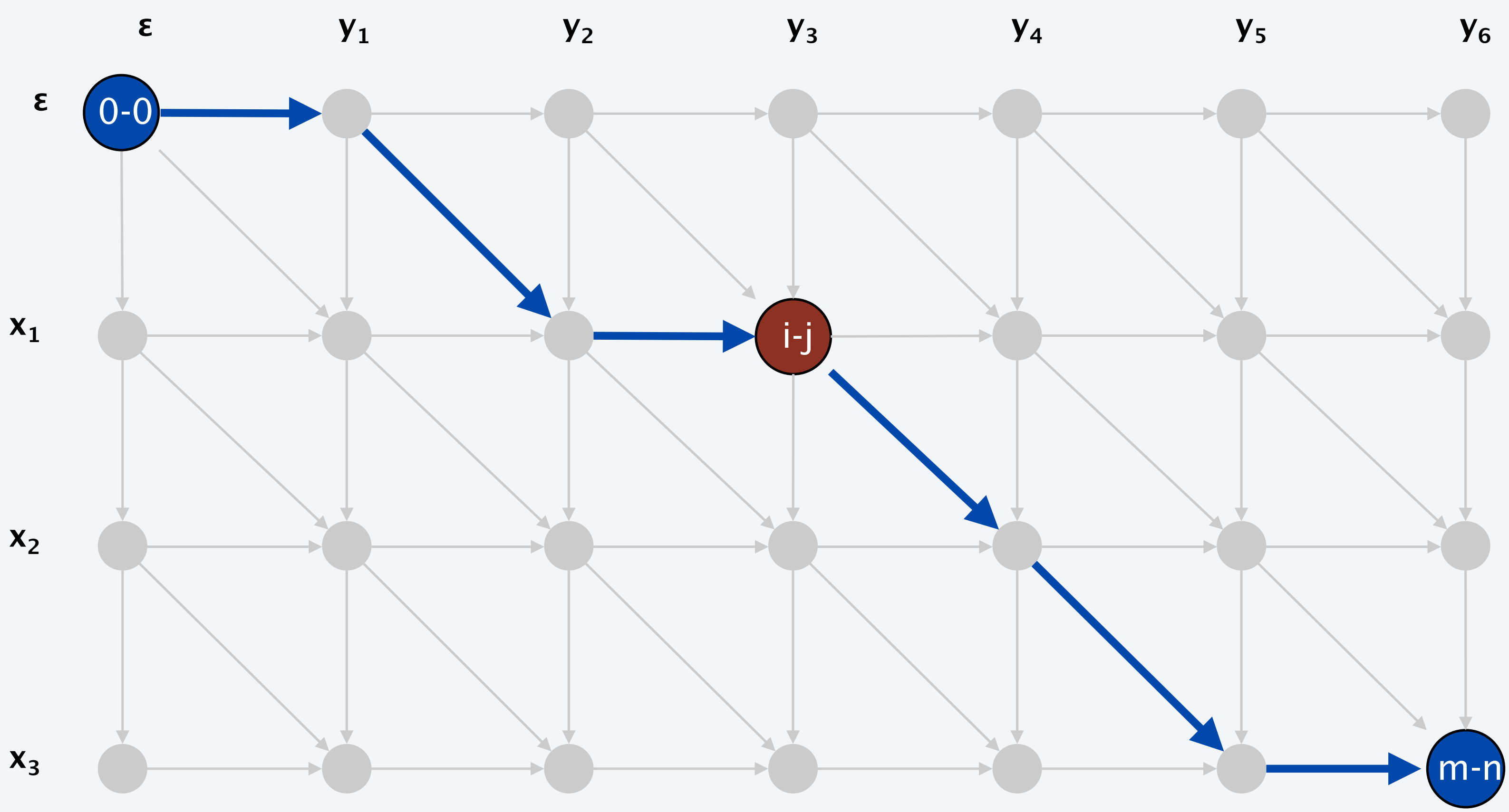
Edit distance graph.

- Let $g(i, j)$ be shortest path from (i, j) to (m, n).
- Can compute $g(\bullet,\ j)$ for any j in $O(mn)$ time and $O(m + n)$ space.



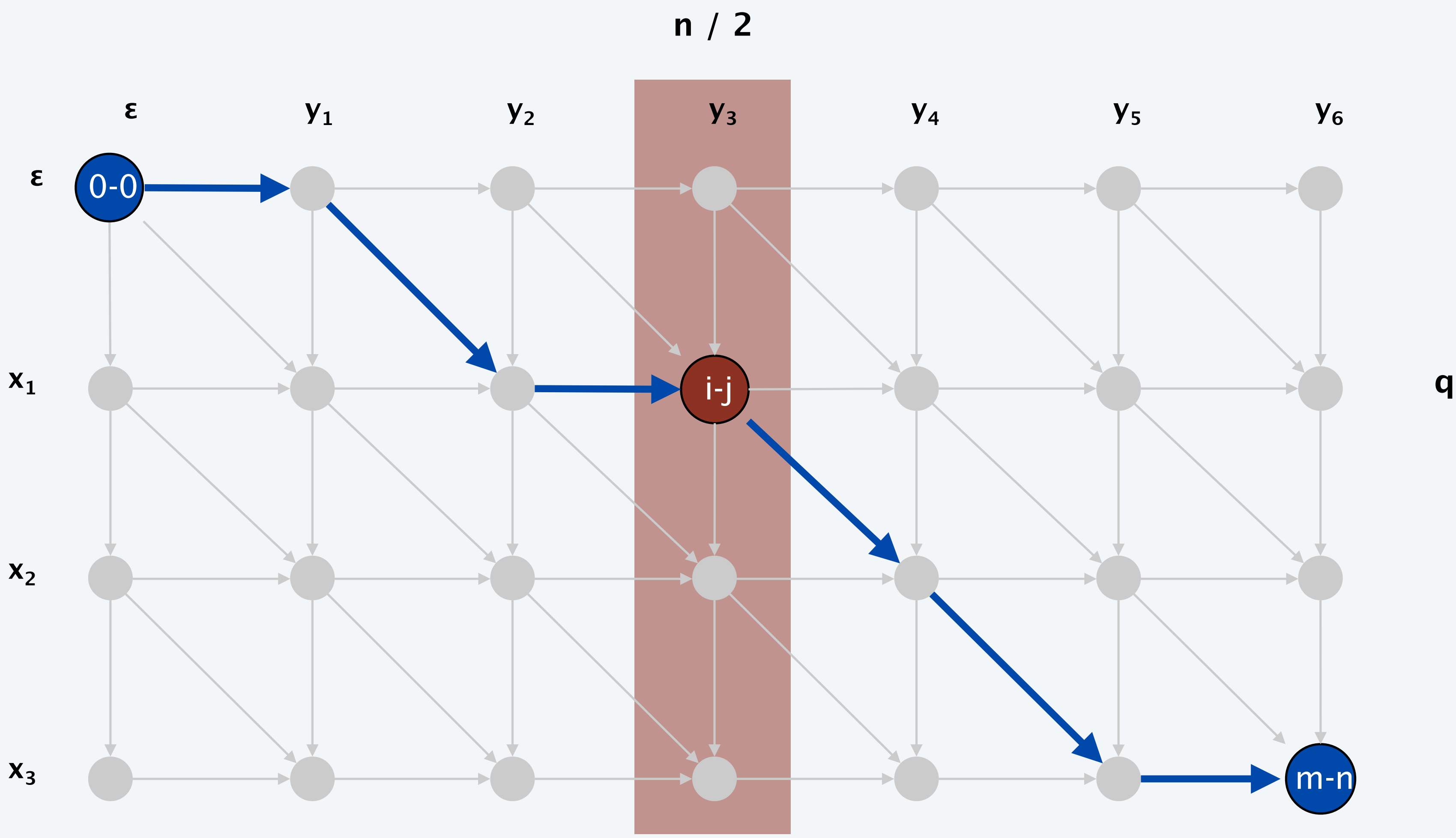
Hirschberg's algorithm

Observation 1. The cost of the shortest path that uses (i, j) is $f(i, j) + g(i, j)$.



Hirschberg's algorithm

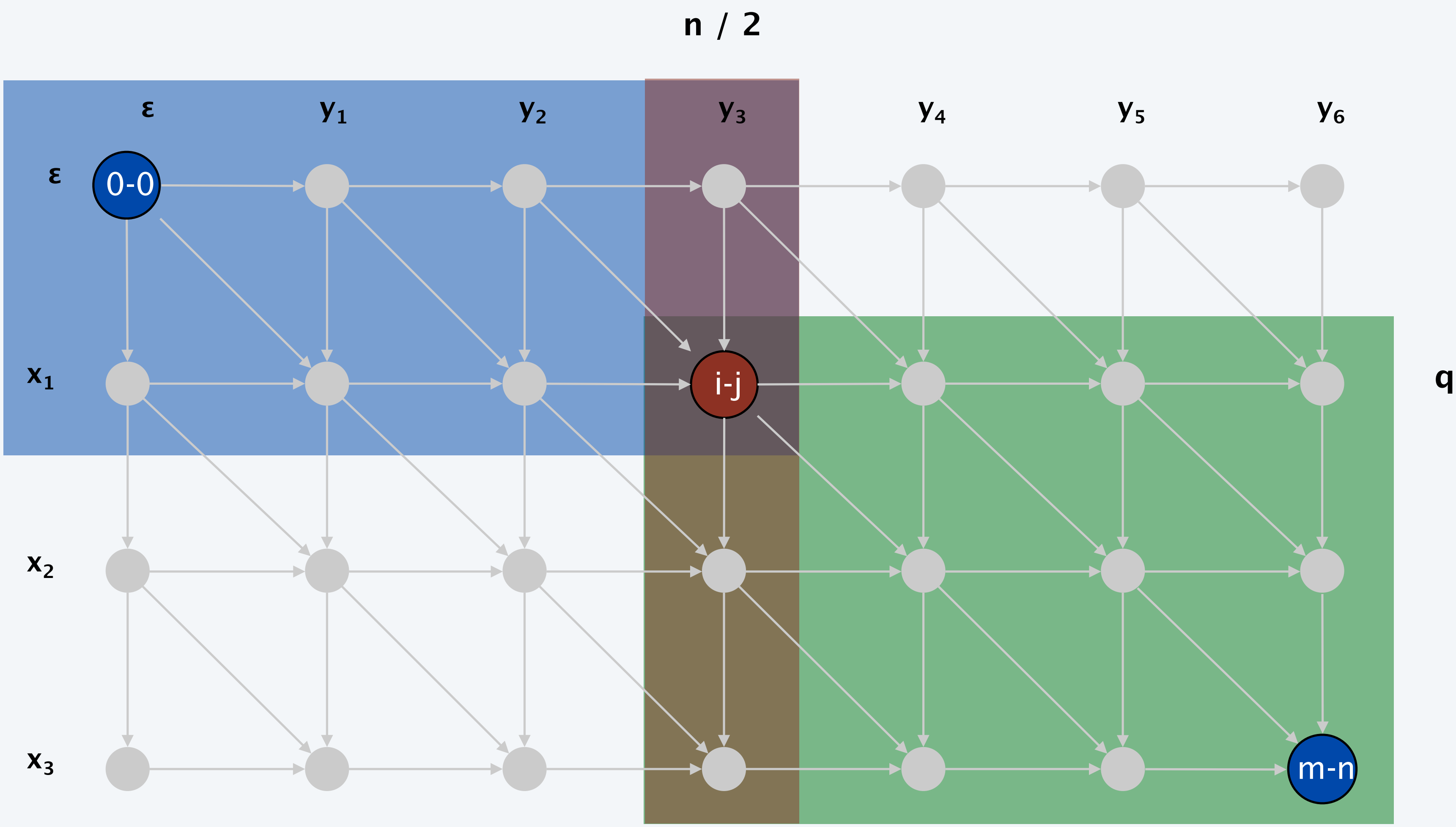
Observation 2. let q be an index that minimizes $f(q, n/2) + g(q, n/2)$.
Then, there exists a shortest path from $(0, 0)$ to (m, n) uses $(q, n/2)$.



Hirschberg's algorithm

Divide. Find index q that minimizes $f(q, n/2) + g(q, n/2)$; align x_q and $y_{n/2}$.

Conquer. Recursively compute optimal alignment in each piece.



Hirschberg's algorithm: running time analysis warmup

Theorem. Let $T(m, n)$ = max running time of Hirschberg's algorithm on strings of length at most m and n. Then, $T(m, n) = O(m\,n \log n)$.

Pf. $T(m, n) \leq 2\,T(m, n/2) + O(m\,n)$

$$\Rightarrow \quad T(m, n) = O(m\,n \log n).$$

Remark. Analysis is not tight because two subproblems are of size $(q, n/2)$ and $(m - q, n/2)$. In next slide, we save $\log n$ factor.

Hirschberg's algorithm: running time analysis

Theorem. Let $T(m, n)$ = max running time of Hirschberg's algorithm on strings of length at most m and n. Then, $T(m, n) = O(m n)$.

Pf. [by induction on n]

- $O(m n)$ time to compute $f(\bullet,\ n/2)$ and $g(\bullet,\ n/2)$ and find index q.
- $T(q, n/2) + T(m - q, n/2)$ time for two recursive calls.
- Choose constant c so that:
$$\begin{aligned} T(m, 2) &\le c\,m \\ T(2, n) &\le c\,n \\ T(m, n) &\le c\,m\,n + T(q, n/2) + T(m-q, n/2) \end{aligned}$$
- Claim. $T(m, n) \le 2\,c\,m\,n$.
- Base cases: $m = 2$ or $n = 2$.
- Inductive hypothesis: $T(m, n) \le 2\,c\,m\,n$ for all (m', n') with $m' + n' < m + n$.
$$\begin{aligned} T(m, n) &\le T(q, n/2) + T(m-q, n/2) + c\,m\,n \\ &\le 2\,c\,q\,n/2 \;+\; 2\,c\,(m-q)\,n/2 \;+\; c\,m\,n \\ &= c\,q\,n \;+\; c\,m\,n \;-\; c\,q\,n \;+\; c\,m\,n \\ &= 2\,c\,m\,n \;\blacksquare \end{aligned}$$