

Chapter 6

Dynamic Programming



Slides by Kevin Wayne. Copyright © 2005 Pearson-Addison Wesley. All rights reserved.

Algorithmic Paradigms

Greed. Build up a solution incrementally, myopically optimizing some local criterion.

Divide-and-conquer. Break up a problem into two sub-problems, solve each sub-problem independently, and combine solution to sub-problems to form solution to original problem.

Dynamic programming. Break up a problem into a series of overlapping sub-problems, and build up solutions to larger and larger sub-problems.

Dynamic Programming Applications

Areas.

- Bioinformatics.
- Control theory.
- Information theory.
- Operations research.
- Computer science: theory, graphics, AI, systems,

Some famous dynamic programming algorithms.

- Viterbi for hidden Markov models.
- Unix diff for comparing two files.
- Smith-Waterman for sequence alignment.
- Bellman-Ford for shortest path routing in networks.
- Cocke-Kasami-Younger for parsing context free grammars.

Fibonacci Sequence



Fibonacci Sequence

- Sequence defined by
 - $a_1 = 1$
 - $a_2 = 1$
 - $a_n = a_{n-1} + a_{n-2}$

- How should you compute the Fibonacci sequence?
- Recursive algorithm:

```
Fib(n)
```

- 1. **If** n = 1 or n = 2, **then**
- 2. return 1
- 3. Else
- 4. a = Fib(n-1)
- 5. b = Fib(n-2)
- 6. **return** a+b

Running Time?

Review Question

• Prove that the solution to the recurrence $T(n)=T(n-1)+T(n-2)+\Theta(1)$ is exponential in n.

Review Question

• Prove that the solution to the recurrence $T(n)=T(n-1)+T(n-2)+\Theta(1)$ is exponential in n.

Easy to show: $\Omega\left(\left(\sqrt{2}\right)^n\right)$ by inspecting the recursion tree Later in the course (if time permits): $\Theta(\phi^n)$ where $\phi\approx 1.618$ is the golden ratio

Computing Fibonacci Sequence Faster

- Observation: Lots of redundancy! The recursive algorithm only solves n-1 different subproblems
- "Memoization": Store the values returned by recursive calls in a sub-table
- Resulting Algorithm:

```
Fib(n)

1. If n =1 or n=2, then

2. return 1

3. Else

4. f[1]=1; f[2]=1

5. For i=3 to n

6. f[i] ← f[i-1]+f[i-2]

7. return f[n]
```

• Running Time?

O(n) if integer operations take constant time.

Computing Fibonacci Sequence Faster

• Observation: Fibonacci recurrence is linear

$$\begin{pmatrix} a_n \\ a_{n-1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_{n-1} \\ a_{n-2} \end{pmatrix} = \dots = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^{n-2} \begin{pmatrix} a_2 \\ a_1 \end{pmatrix}$$

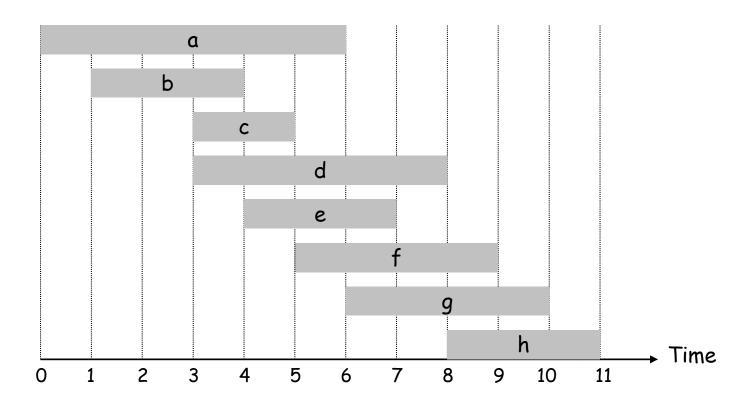
- Can compute Aⁿ using only O(log n) matrix multiplications; each one takes O(1) integer multiplications and additions.
- Total running time?
 - O(log n) integer operations. Exponential improvement!
- Exercise: how big an improvement if we count bit operations?
 - Multiplying k-bit numbers takes $O(k \log k)$ time.
- How many bits needed to write down a_n ?

6.1 Weighted Interval Scheduling

Weighted Interval Scheduling

Weighted interval scheduling problem.

- $_{\scriptscriptstyle h}$ Job j starts at $s_{\scriptscriptstyle
 m j}$, finishes at ${\sf f}_{\scriptscriptstyle
 m j}$, and has weight or value ${\sf v}_{\scriptscriptstyle
 m j}$.
- Two jobs compatible if they don't overlap.
- Goal: find maximum weight subset of mutually compatible jobs.

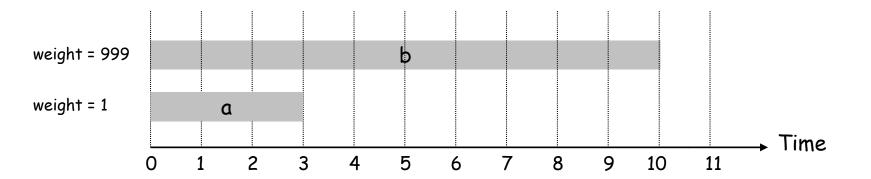


Unweighted Interval Scheduling Review

Recall. Greedy algorithm works if all weights are 1.

- Consider jobs in ascending order of finish time.
- Add job to subset if it is compatible with previously chosen jobs.

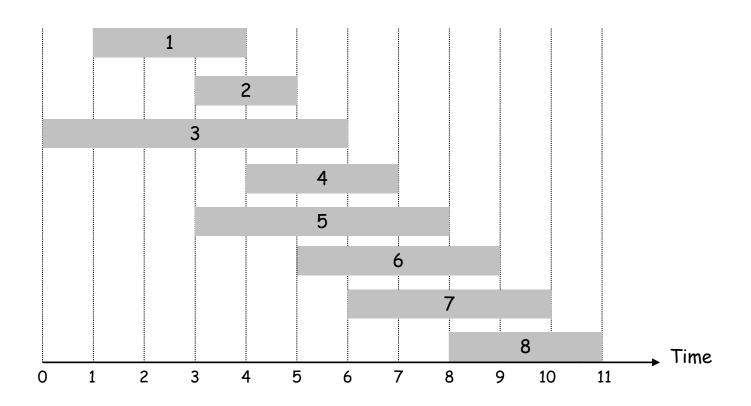
Observation. Greedy algorithm can fail spectacularly if arbitrary weights are allowed.



Weighted Interval Scheduling

Notation. Label jobs by finishing time: $f_1 \le f_2 \le ... \le f_n$. Def. p(j) = largest index i < j such that job i is compatible with j.

Ex:
$$p(8) = 5$$
, $p(7) = 3$, $p(2) = 0$.



Dynamic Programming: Binary Choice

Notation. OPT(j) = value of optimal solution to the problem consisting of job requests 1, 2, ..., j.

- Case 1: OPT selects job j.
 - can't use incompatible jobs { p(j) + 1, p(j) + 2, ..., j 1 }
 - must include optimal solution to problem consisting of remaining compatible jobs 1, 2, ..., p(j)

 optimal substructure
- Case 2: OPT does not select job j.
 - must include optimal solution to problem consisting of remaining compatible jobs 1, 2, ..., j-1

$$OPT(j) = \begin{cases} 0 & \text{if } j = 0\\ \max \{ v_j + OPT(p(j)), OPT(j-1) \} & \text{otherwise} \end{cases}$$

Weighted Interval Scheduling: Brute Force

Brute force algorithm.

```
Input: n, s_1,...,s_n, f_1,...,f_n, v_1,...,v_n

Sort jobs by finish times so that f_1 \leq f_2 \leq ... \leq f_n.

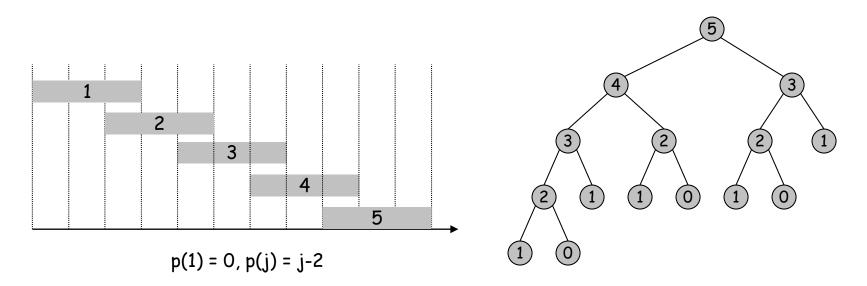
Compute p(1), p(2), ..., p(n)

Compute-Opt(j) {
   if (j = 0)  
       return 0
   else
      return max(v_j + Compute-Opt(p(j)), Compute-Opt(j-1))
}
```

Weighted Interval Scheduling: Brute Force

Observation. Recursive algorithm fails spectacularly because of redundant sub-problems \Rightarrow exponential algorithms.

Ex. Number of recursive calls for family of "layered" instances grows like Fibonacci sequence.



Weighted Interval Scheduling: Memoization

Memoization. Store results of each sub-problem in a cache; lookup as needed.

```
Input: n, s_1, ..., s_n, f_1, ..., f_n, v_1, ..., v_n
Sort jobs by finish times so that f_1 \le f_2 \le ... \le f_n.
Compute p(1), p(2), ..., p(n)
for j = 1 to n
   M[j] = empty \leftarrow global array
M[j] = 0
M-Compute-Opt(j) {
   if (M[j] is empty)
       M[j] = max(w_i + M-Compute-Opt(p(j)), M-Compute-Opt(j-1))
   return M[j]
```

Weighted Interval Scheduling: Running Time

Claim. Memoized version of algorithm takes O(n log n) time.

- Sort by finish time: O(n log n).
- Computing $p(\cdot)$: O(n) after sorting by start time.
- M-Compute-Opt(j): each invocation takes O(1) time and either
 - (i) returns an existing value M[j]
 - (ii) fills in one new entry M[j] and makes two recursive calls
- Progress measure Φ = # nonempty entries of M[].
 - initially Φ = 0, throughout $\Phi \leq$ n.
 - (ii) increases Φ by $1 \Rightarrow$ at most 2n recursive calls.
- Overall running time of M-Compute-Opt(n) is O(n). •

Remark. O(n) if jobs are pre-sorted by start and finish times.

Weighted Interval Scheduling: Finding a Solution

- Q. Dynamic programming algorithms computes optimal value. What if we want the solution itself?
- A. Do some post-processing.

```
Run M-Compute-Opt(n)
Run Find-Solution(n)

Find-Solution(j) {
   if (j = 0)
      output nothing
   else if (v<sub>j</sub> + M[p(j)] > M[j-1])
      print j
      Find-Solution(p(j))
   else
      Find-Solution(j-1)
}
```

of recursive calls \leq n \Rightarrow O(n).

Weighted Interval Scheduling: Bottom-Up

Bottom-up dynamic programming. Unwind recursion.

```
Input: n, s_1,...,s_n, f_1,...,f_n, v_1,...,v_n

Sort jobs by finish times so that f_1 \leq f_2 \leq ... \leq f_n.

Compute p(1), p(2), ..., p(n)

Iterative-Compute-Opt {
    M[0] = 0
    for j = 1 to n
        M[j] = max(v_j + M[p(j)], M[j-1])
}
```

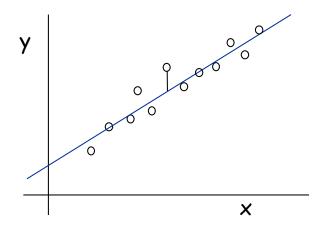
6.3 Segmented Least Squares

Segmented Least Squares

Least squares.

- Foundational problem in statistic and numerical analysis.
- Given n points in the plane: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.
- Find a line y = ax + b that minimizes the sum of the squared error:

$$SSE = \sum_{i=1}^{n} (y_i - ax_i - b)^2$$



Solution. Calculus \Rightarrow min error is achieved when

$$a = \frac{n \sum_{i} x_{i} y_{i} - (\sum_{i} x_{i}) (\sum_{i} y_{i})}{n \sum_{i} x_{i}^{2} - (\sum_{i} x_{i})^{2}}, \quad b = \frac{\sum_{i} y_{i} - a \sum_{i} x_{i}}{n}$$

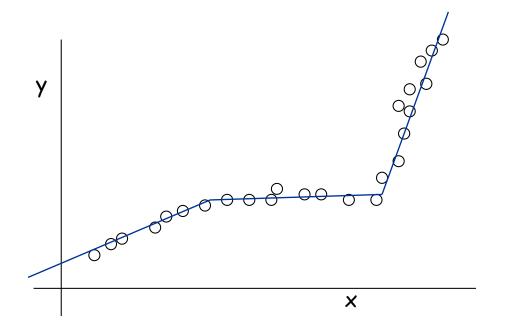
Segmented Least Squares

Segmented least squares.

- Points lie roughly on a sequence of several line segments.
- Given n points in the plane $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ with
- $x_1 < x_2 < ... < x_n$, find a sequence of lines that minimizes f(x).

Q. What's a reasonable choice for f(x) to balance accuracy and parsimony?

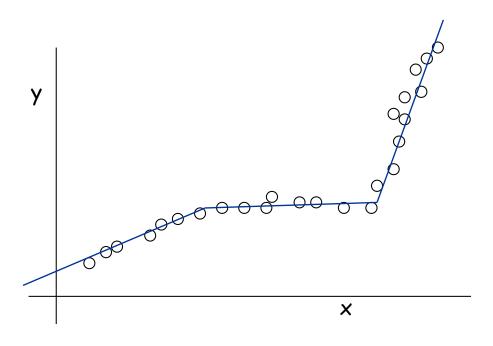
number of lines



Segmented Least Squares

Segmented least squares.

- Points lie roughly on a sequence of several line segments.
- Given n points in the plane $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ with
- $x_1 < x_2 < ... < x_n$, find a sequence of lines that minimizes:
 - the sum of the sums of the squared errors E in each segment
 - the number of lines L
- Tradeoff function: E + c L, for some constant c > 0.



Dynamic Programming: Multiway Choice

Notation.

OPT(j) = minimum cost for points p_1, p_2, \ldots, p_j .

e(i, j) = minimum sum of squares for points $p_i, p_{i+1}, \ldots, p_j$.

To compute OPT(j):

- Last segment uses points p_i , p_{i+1} , ..., p_i for some i.
- Cost = e(i, j) + c + OPT(i-1).

$$OPT(j) = \begin{cases} 0 & \text{if } j = 0\\ \min_{1 \le i \le j} \left\{ e(i,j) + c + OPT(i-1) \right\} & \text{otherwise} \end{cases}$$

Segmented Least Squares: Algorithm

```
INPUT: p_1, ..., p_N \in C
Segmented-Least-Squares() {
   M[0] = 0
   for j = 1 to n
       for i = 1 to j
           compute the least square error e<sub>i</sub> for
           the segment p<sub>i</sub>,..., p<sub>i</sub>
   for j = 1 to n
       M[j] = \min_{1 \le i \le j} (e_{ij} + c + M[i-1])
   return M[n]
```

Running time. $O(n^3)$. \checkmark can be improved to $O(n^2)$ by pre-computing various statistics

Bottleneck = computing e(i, j) for $O(n^2)$ pairs, O(n) per pair using previous formula.

6.4 Knapsack Problem

Knapsack Problem

Knapsack problem.

- Given n objects and a "knapsack."
- Item i weighs $w_i > 0$ kilograms and has value $v_i > 0$.
- Knapsack has capacity of W kilograms.
- Goal: fill knapsack so as to maximize total value.

Ex: { 3, 4 } has value 40.

W = 11

Item	Value	Weight
1	1	1
2	6	2
3	18	5
4	22	6
5	28	7

Greedy: repeatedly add item with maximum ratio v_i / w_i .

Ex: $\{5, 2, 1\}$ achieves only value = $35 \Rightarrow greedy$ not optimal.

Dynamic Programming: False Start

Def. OPT(i) = max profit subset of items 1, ..., i.

- Case 1: OPT does not select item i.
 - OPT selects best of { 1, 2, ..., i-1 }
- Case 2: OPT selects item i.
 - accepting item i does not immediately imply that we will have to reject other items
 - without knowing what other items were selected before i, we don't even know if we have enough room for i

Conclusion. Need more sub-problems!

Dynamic Programming: Adding a New Variable

Def. OPT(i, w) = max profit subset of items 1, ..., i with weight limit w.

- Case 1: OPT does not select item i.
 - OPT selects best of { 1, 2, ..., i-1 } using weight limit w
- Case 2: OPT selects item i.
 - new weight limit = w wi
 - OPT selects best of { 1, 2, ..., i-1 } using this new weight limit

$$OPT(i, w) = \begin{cases} 0 & \text{if } i = 0 \\ OPT(i-1, w) & \text{if } w_i > w \\ \max\{OPT(i-1, w), v_i + OPT(i-1, w-w_i)\} & \text{otherwise} \end{cases}$$

Knapsack Problem: Bottom-Up

Knapsack. Fill up an n-by-W array.

```
Input: n, w_1, ..., w_N, v_1, ..., v_N
for w = 0 to W
   M[0, w] = 0
for i = 1 to n
   for w = 1 to W
      if (w_i > w)
          M[i, w] = M[i-1, w]
      else
          M[i, w] = \max \{M[i-1, w], v_i + M[i-1, w-w_i]\}
return M[n, W]
```

Knapsack Algorithm

W + 1

		0	1	2	3	4	5	6	7	8	9	10	11
n + 1	ф	0	0	0	0	0	0	0	0	0	0	0	0
	{ 1 }	0	1	1	1	1	1	1	1	1	1	1	1
	{ 1, 2 }	0 🔨	1	6	7	7	1 7	7	7	7	7	7	7
	{ 1, 2, 3 }	0	1	6	7	7	18	19	24	25	25	25	2 5
	{1,2,3,4}	0	1	6	7	7	18	22	24	28	29	29	40
	{1,2,3,4,5}	0	1	6	7	7	18	22	28	29	34	34	40

OPT: { 4, 3 }

value = 22 + 18 = 40

W = 11

Item	Value	Weight				
1	1	1				
2	6	2				
3	18	5				
4	22	6				
5	28	7				

Knapsack Problem: Running Time

Running time. $\Theta(n W)$.

- Not polynomial in input size!
- "Pseudo-polynomial."
- Decision version of Knapsack is NP-complete. [Chapter 8]

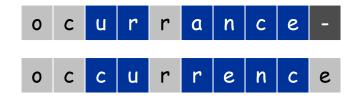
Knapsack approximation algorithm. There exists a polynomial algorithm that produces a feasible solution that has value within 0.01% of optimum. [Section 11.8]

6.6 Sequence Alignment

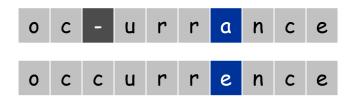
String Similarity

How similar are two strings?

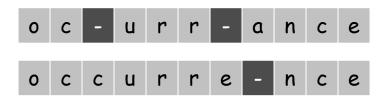
- ocurrance
- occurrence



6 mismatches, 1 gap



1 mismatch, 1 gap



0 mismatches, 3 gaps

Edit Distance

Applications.

- Basis for Unix diff.
- Speech recognition.
- Computational biology.

Edit distance. [Levenshtein 1966, Needleman-Wunsch 1970]

- Gap penalty δ ; mismatch penalty α_{pq} .
- Cost = sum of gap and mismatch penalties.

$$\alpha_{TC} + \alpha_{GT} + \alpha_{AG} + 2\alpha_{CA}$$
 $2\delta + \alpha_{CA}$

Sequence Alignment

Goal: Given two strings $X = x_1 x_2 ... x_m$ and $Y = y_1 y_2 ... y_n$ find alignment of minimum cost.

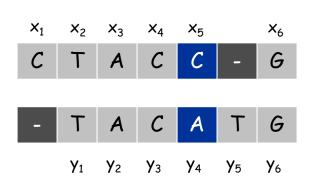
Def. An alignment M is a set of ordered pairs x_i - y_j such that each item occurs in at most one pair and no crossings.

Def. The pair x_i-y_j and $x_{i'}-y_{j'}$ cross if i < i', but j > j'.

$$cost(M) = \underbrace{\sum_{(x_i, y_j) \in M} \alpha_{x_i y_j}}_{\text{mismatch}} + \underbrace{\sum_{i: x_i \text{ unmatched}} \delta + \sum_{j: y_j \text{ unmatched}} \delta}_{\text{gap}}$$

Ex: CTACCG VS. TACATG.

Sol:
$$M = x_2 - y_1, x_3 - y_2, x_4 - y_3, x_5 - y_4, x_6 - y_6.$$



Sequence Alignment: Problem Structure

Def. OPT(i, j) = min cost of aligning strings $x_1 x_2 ... x_i$ and $y_1 y_2 ... y_j$.

- Case 1: OPT matches x_i-y_j .
 - pay cost of x_i - y_j + min cost of aligning two strings $x_1 x_2 \dots x_{i-1}$ and $y_1 y_2 \dots y_{j-1}$
- Case 2a: OPT leaves x_i unmatched.
 - pay gap for x_i and min cost of aligning $x_1 x_2 \dots x_{i-1}$ and $y_1 y_2 \dots y_j$
- Case 2b: OPT leaves y unmatched.
 - pay gap for y_j and min cost of aligning $x_1\,x_2\,\ldots\,x_i$ and $y_1\,y_2\,\ldots\,y_{j-1}$

$$OPT(i, j) = \begin{cases} j\delta & \text{if } i = 0 \\ \alpha_{x_i y_j} + OPT(i-1, j-1) & \text{otherwise} \\ \delta + OPT(i, j-1) & \text{otherwise} \\ \delta + OPT(i, j-1) & \text{if } j = 0 \end{cases}$$

Sequence Alignment: Algorithm

```
Sequence-Alignment(m, n, x_1x_2...x_m, y_1y_2...y_n, \delta, \alpha) {
   for i = 0 to m
      M[0, i] = i\delta
   for j = 0 to n
      M[j, 0] = j\delta
   for i = 1 to m
       for j = 1 to n
          M[i, j] = min(\alpha[x_i, y_i] + M[i-1, j-1],
                            \delta + M[i-1, j],
                            \delta + M[i, j-1]
   return M[m, n]
```

Analysis. $\Theta(mn)$ time and space.

English words or sentences: $m, n \le 10$.

Computational biology: m = n = 100,000.10 billions ops OK, but 10GB array?

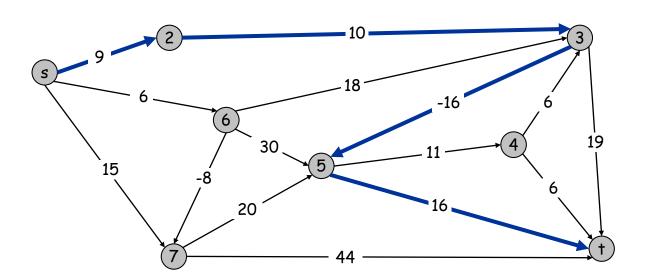
6.8 Shortest Paths

Shortest Paths

Shortest path problem. Given a directed graph G = (V, E), with edge weights c_{vw} , find shortest path from node s to node t.

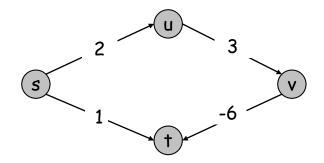
allow negative weights

Ex. Nodes represent agents in a financial setting and c_{vw} is cost of transaction in which we buy from agent v and sell immediately to w.

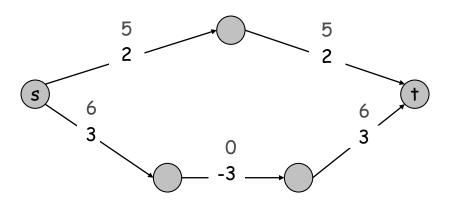


Shortest Paths: Failed Attempts

Dijkstra. Can fail if negative edge costs.

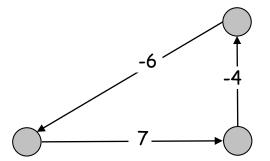


Re-weighting. Adding a constant to every edge weight can fail.

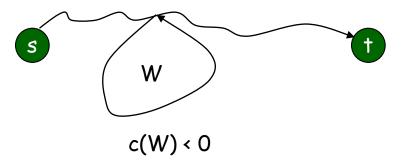


Shortest Paths: Negative Cost Cycles

Negative cost cycle.



Observation. If some path from s to t contains a negative cost cycle, there does not exist a shortest s-t path; otherwise, there exists one that is simple.



Shortest Paths: Dynamic Programming

Def. OPT(i, v) = length of shortest v-t path P using at most i edges.

- Case 1: P uses at most i-1 edges.
 - OPT(i, v) = OPT(i-1, v)
- Case 2: P uses exactly i edges.
 - if (v, w) is first edge, then OPT uses (v, w), and then selects best w-t path using at most i-1 edges

$$OPT(i, v) = \begin{cases} \infty & \text{if } i = 0 \\ \min \left\{ OPT(i-1, v), \min_{(v, w) \in E} \left\{ OPT(i-1, w) + c_{vw} \right\} \right\} & \text{otherwise} \end{cases}$$

Remark. By previous observation, if no negative cycles, then OPT(n-1, v) = length of shortest v-t path.

Shortest Paths: Implementation

```
Shortest-Path(G, t) {
    foreach node v ∈ V
        M[0, v] ← ∞
    M[0, t] ← 0

for i = 1 to n-1
    foreach node v ∈ V
        M[i, v] ← M[i-1, v]
    foreach edge (v, w) ∈ E
        M[i, v] ← min { M[i, v], M[i-1, w] + c<sub>vw</sub> }
}
```

Analysis. $\Theta(mn)$ time, $\Theta(n^2)$ space.

Finding the shortest paths. Maintain a "successor" for each table entry.

Shortest Paths: Practical Improvements

Practical improvements.

- Maintain only one array M[v] = shortest v-t path that we have found so far.
- No need to check edges of the form (v, w) unless M[w] changed in previous iteration.

Theorem. Throughout the algorithm, M[v] is length of some v-t path, and after i rounds of updates, the value M[v] is no larger than the length of shortest v-t path using \leq i edges.

Overall impact.

- Memory: O(m + n).
- Running time: O(mn) worst case, but substantially faster in practice.

Bellman-Ford: Efficient Implementation

```
Push-Based-Shortest-Path(G, s, t) {
   foreach node v \in V {
      M[v] \leftarrow \infty
      successor[v] \leftarrow \phi
   M[t] = 0
   for i = 1 to n-1 {
       foreach node w \in V {
       if (M[w] has been updated in previous iteration) {
          foreach node v such that (v, w) \in E 
              if (M[v] > M[w] + C_{vw}) {
                 M[v] \leftarrow M[w] + c_{vw}
                 successor[v] \leftarrow w
       If no M[w] value changed in iteration i, stop.
```

6.10 Negative Cycles in a Graph

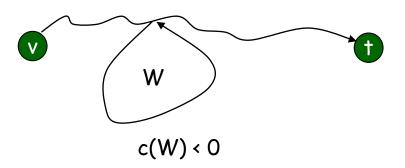
Detecting Negative Cycles

Lemma. If OPT(n,v) = OPT(n-1,v) for all v, then no negative cycles. Pf. Bellman-Ford algorithm.

Lemma. If OPT(n,v) < OPT(n-1,v) for some node v, then (any) shortest path from v to t contains a cycle W. Moreover W has negative cost.

Pf. (by contradiction)

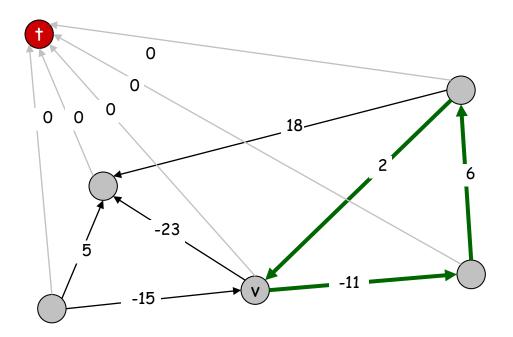
- Since OPT(n,v) < OPT(n-1,v), we know P has exactly n edges.
- By pigeonhole principle, P must contain a directed cycle W.
- Deleting W yields a v-t path with < n edges \Rightarrow W has negative cost.



Detecting Negative Cycles

Theorem. Can detect negative cost cycle in O(mn) time.

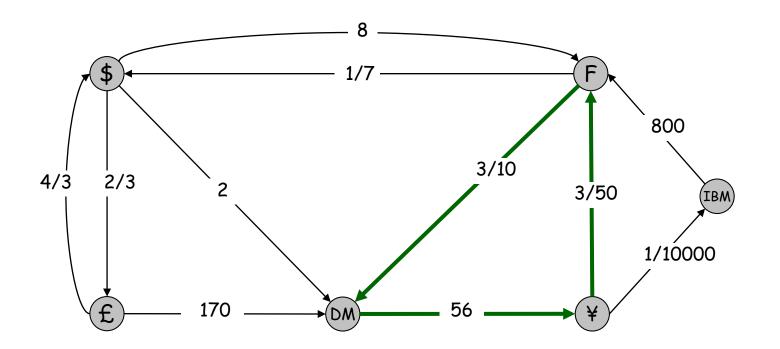
- Add new node t and connect all nodes to t with 0-cost edge.
- Check if OPT(n, v) = OPT(n-1, v) for all nodes v.
 - if yes, then no negative cycles
 - if no, then extract cycle from shortest path from v to t



Detecting Negative Cycles: Application

Currency conversion. Given n currencies and exchange rates between pairs of currencies, is there an arbitrage opportunity?

Remark. Fastest algorithm very valuable!



Detecting Negative Cycles: Summary

Bellman-Ford. O(mn) time, O(m + n) space.

- Run Bellman-Ford for n iterations (instead of n-1).
- Upon termination, Bellman-Ford successor variables trace a negative cycle if one exists.
- See p. 288 for improved version and early termination rule.