

**CECS401**  
**Fundamentals of Spoken Language Processing**

**Note-13**  
**Tuesday 10/12/99**

## I. Speech Pattern Comparison using Dynamic Time Alignment

### Determine the time-warping path $\phi$

For best match between two templates  $X$  and  $Y$ , the path  $\phi$  is defined as the one that minimizes  $d_\phi(X, Y)$ , i.e.,

$$\phi = \arg \min_{\phi'} d_{\phi'}(X, Y)$$

The minimum distance is denoted as

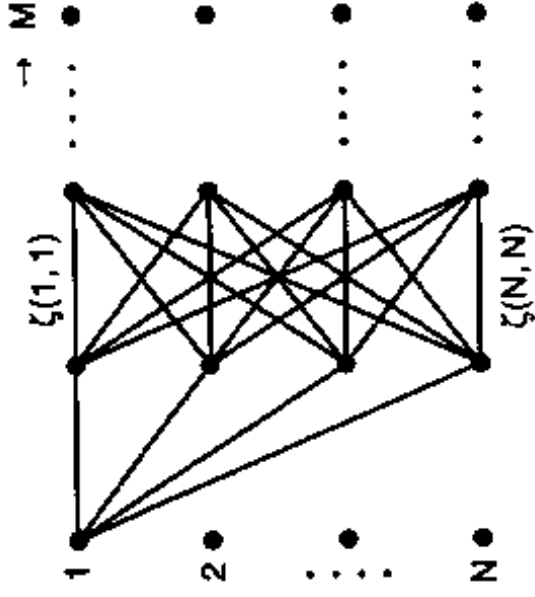
$$d(X, Y) = d_\phi(X, Y)$$

The solution of  $\phi$  requires high-dimensional optimization.

**Dynamic programming** is an efficient mathematical optimization method that turns certain multidimensional optimization problems into multistage decision problems.

Example: the synchronous sequential decision problem

Find the optimal sequence of a fixed number ( $M$ ) moves, starting from a point  $i$  and ending at a point  $j$ , with the minimum cost.



## Definitions

- **Policy:**  
a decision rule to determine which point to visit next
- **Cost of one-step move:**  $\zeta(\cdot, \cdot)$   
e.g.,  $\zeta(l, m)$  is the cost of moving from the point  $l$  to the point  $m$  in one step.
- **Minimum cost of  $s$ -step move:**  $\phi_s(\cdot, \cdot)$   
e.g.,  $\phi_s(i, j)$  is the minimum cost of moving from the point  $i$  to point  $j$  in  $s$  steps.

## The principle of optimality:

an optimal policy has the property that, whatever the initial state and decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.

After  $m - 1$  moves, there are  $N$  costs  $\phi_{m-1}(1, l)$ ,  $1 \leq l \leq N$ .

In the  $m$ th move, the point 2 must be reached from the point  $l^*$  that satisfies

$$l^* = \arg \min_{1 \leq l \leq N} [\phi_{m-1}(1, l) + \zeta(l, 2)]$$

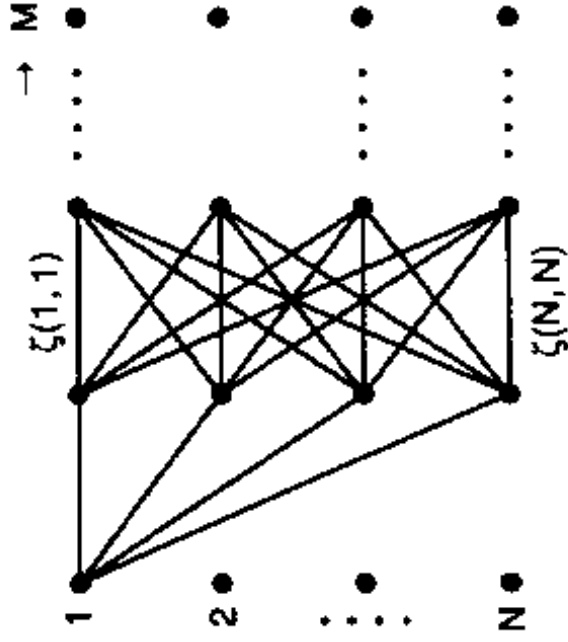
$$\phi_m(1, 2) = \phi_{m-1}(1, l^*) + \zeta(l^*, 2)$$

### Example

Determine the optimal three-step path that starts and ends at the point 1.

### Example

Assuming  $N$  points, consider moving from the point 1 to the point 2 in  $m$  steps. Given that  $m - 1$  moves have been made, determine the next step to reach the point 2.



step-1	step-2		step-3	total cost
(1,1)	(i,1)	$1 \leq i \leq N$	(1,1)	$\zeta(1, i) + \zeta(i, 1) + \zeta(1, 1)$
(1,2)	(i,2)	$1 \leq i \leq N$	(2,1)	$\zeta(1, i) + \zeta(i, 2) + \zeta(2, 1)$
...	...	...	...	...
(1,N)	(i,N)	$1 \leq i \leq N$	(N,1)	$\zeta(1, i) + \zeta(i, N) + \zeta(N, 1)$

**Minimum path cost:**

$$\begin{aligned}\phi_3(1, 1) &= \min_{1 < j \leq N} \left[ \min_{1 \leq i \leq N} [\zeta(1, i) + \zeta(i, j)] + \zeta(j, 1) \right] \\ &= \min_{1 < j < N} [\phi_2(1, j) + \zeta(j, 1)]\end{aligned}$$

**Pointers to optimal path:**

$$j^* = \arg \min_{1 \leq j \leq N} [\phi_2(1, j) + \zeta(j, 1)]$$

$$i^*(j^*) = \arg \min_{1 \leq i \leq N} [\zeta(1, i) + \zeta(i, j^*)]$$

**Optimal path:**

$$\phi = ((1, i^*(j^*)), (i^*(j^*), j^*), (j^*, 1))$$

The pointers to the optimal path can be recorded by the array

$$\xi_m(n) = \arg \min_{1 \leq j < N} [\phi_{m-1}(i, j) + \zeta(j, n)]$$

$$1 \leq m \leq M, 1 \leq n \leq N$$

**The synchronous sequential decision algorithm**  
for solving the  $\phi_M(i, j)$  problem

**Step-1. Initialization**

$$\phi_1(i, n) = \zeta(i, n)$$

$$\xi_1(n) = i$$

$$1 \leq n \leq N$$

**Step-2. Recursion**

$$\phi_{m+1}(i, n) = \min_{1 \leq l \leq N} [\phi_m(i, l) + \zeta(l, n)]$$

$$\xi_{m+1}(n) = \arg \min_{1 \leq l \leq N} [\phi_m(i, l) + \zeta(l, n)]$$

$$1 \leq n \leq N, 1 \leq m \leq M - 2$$

**Step-3. Termination**

$$\phi_M(i, j) = \min_{1 \leq l \leq N} [\phi_{M-1}(i, l) + \zeta(l, j)]$$

$$\xi_M(j) = \arg \min_{1 \leq l \leq N} [\phi_{M-1}(i, l) + \zeta(l, j)]$$

#### Step-4. Backtracking

$$\begin{aligned} i_M &= j \\ i_{M-1} &= \xi_M(i_M) && (i_{M-1}, j) \\ i_{M-2} &= \xi_{M-1}(i_{M-1}) && (i_{M-2}, i_{M-1}) \\ &\vdots && \vdots \\ i_1 &= \xi_2(i_2) && (i_1, i_2) \\ i &= \xi_1(i_1) && (i, i_1) \end{aligned}$$

**Optimal path:**

$$\phi = ((i, i_1), (i_1, i_2), \dots, (i_{M-2}, i_{M-1}), (i_{M-1}, j))$$

Comments on the  $\phi_M(i, j)$  problem:

- In direct optimization, the number of paths to be searched is proportional to  $N^M$ .
- In DP multistage optimization,  $N^2$  paths are searched at each stage, the total number of search paths is proportional to  $M \times N^2$ .
- Each stage in DP summarizes  $N$  optimal paths that end at the  $N$  points of the current stage.
- To construct an  $m$ —stage optimal path, the required past information is the  $(m - 1)$ —stage optimal paths.

## The Use of DP in DTW for Speech Recognition

DP

number of points:  $N$

number of stages:  $M$

minimum cost:  $\phi_M(1, N)$

optimal path:  $\phi$

DTW

time-axis for template  $Y$ :  $T_y$

time-axis for template  $X$ :  $T_x$

distance between  $X$  and  $Y$ :  $d(X, Y)$

time-warping function:  $\phi = (\phi_x, \phi_y)$

Constraints in DTW:

- endpoint
- monotonicity
- local continuity
- global range
- slope weighting

### Endpoint constraint

Align the beginning and ending points of  $X$  and  $Y$ , respectively, i.e.,

$$\begin{aligned}\phi_x(1) &= 1, \phi_y(1) = 1 \\ \phi_x(T) &= T_x, \phi_y(T) = T_y\end{aligned}$$

### Monotonicity constraint

Only forward time-warping is allowed, i.e.,

$$\begin{aligned}\phi_x(k+1) &\geq \phi_x(k) \\ \phi_y(k+1) &\geq \phi_y(k)\end{aligned}$$

### Local continuity constraint

Drastic warping could cause the loss of phonetic information, and hence warping slope should be limited.

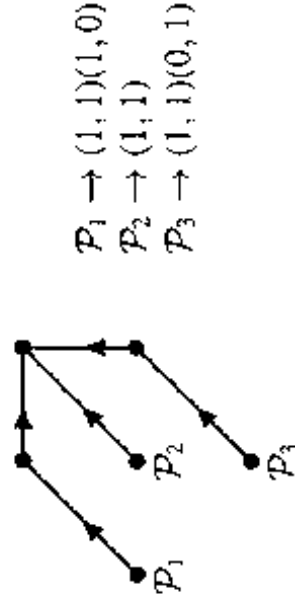
In continuous speech, certain phonetic unit could last for only two frames, a reasonable local constraint is therefore

$$\phi_x(k+1) - \phi_x(k) \leq 2$$

$$\phi_y(k+1) - \phi_y(k) \leq 2$$

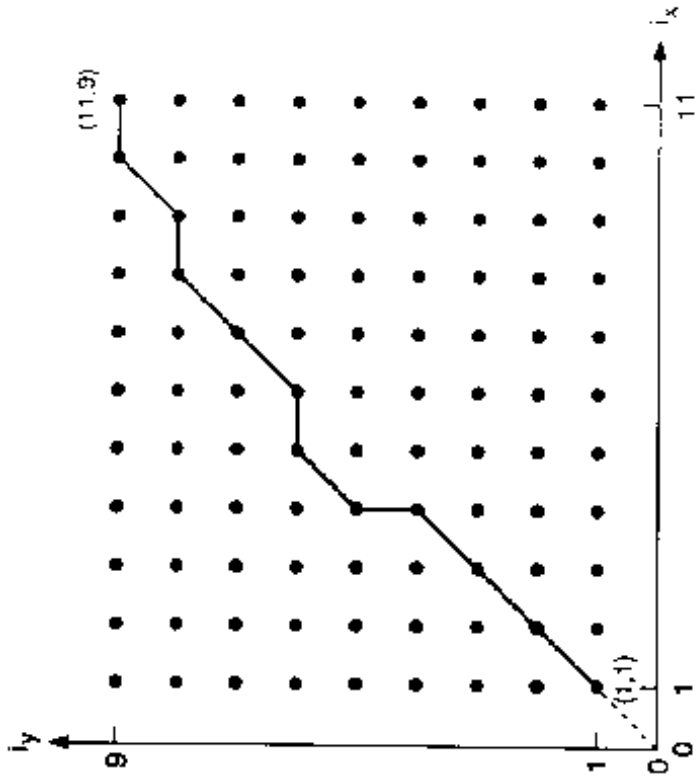
Using local path patterns as continuity constraint

Example

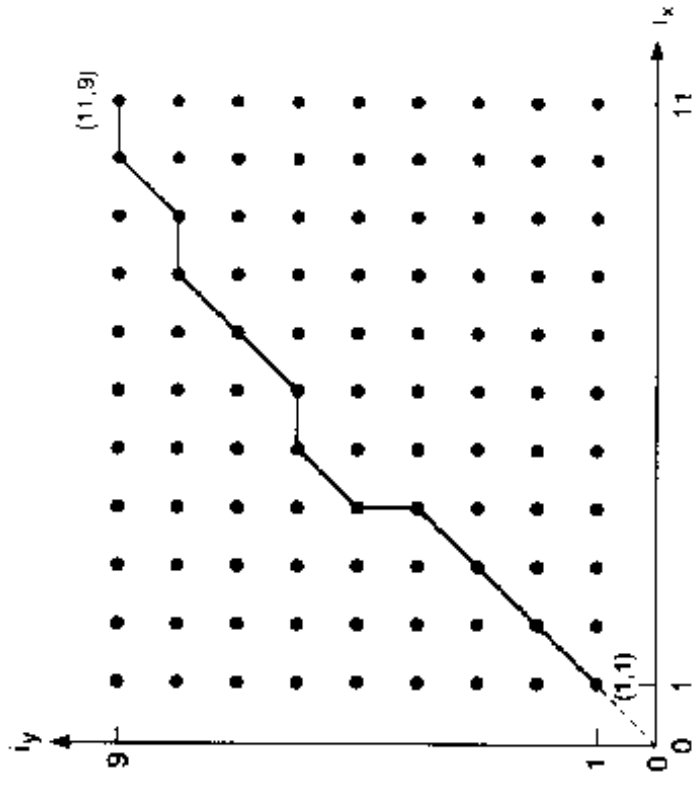


A global warping path is constructed by the three types of local warping paths  $P_1, P_2, P_3$ .

Example: label a global warping path by  $(P_1, P_2, P_3)$ .



Example: label a global warping path by  $(P_1, P_2, P_3)$ .



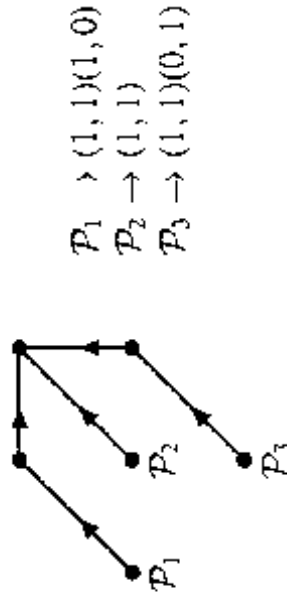
In continuous speech, certain phonetic unit could last for only two frames, a reasonable local constraint is therefore

$$\phi_x(k+1) - \phi_x(k) \leq 2$$

$$\phi_y(k+1) - \phi_y(k) \leq 2$$

Using local path patterns as continuity constraint

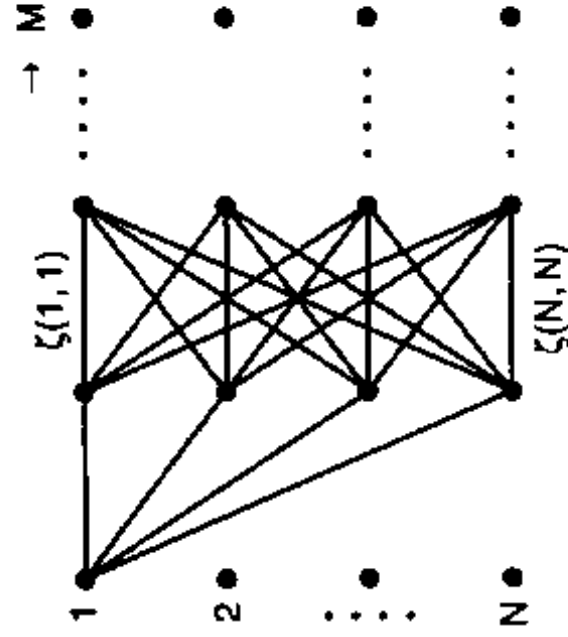
Example



A global warping path is constructed by the three types of local warping paths  $\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3$ .

### Example

Assuming  $N$  points, consider moving from the point 1 to the point 2 in  $m$  steps. Given that  $m - 1$  moves have been made, determine the next step to reach the point 2.



Example: the synchronous sequential decision problem  
Find the optimal sequence of a fixed number ( $M$ ) moves, starting from a point  $i$  and ending at a point  $j$ , with the minimum cost.

