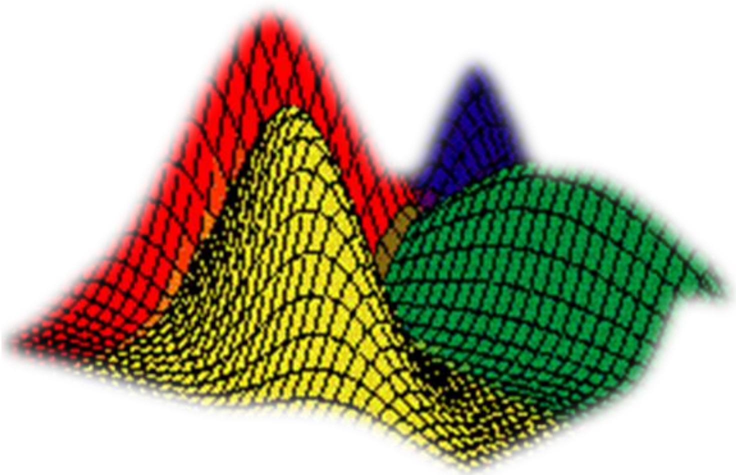
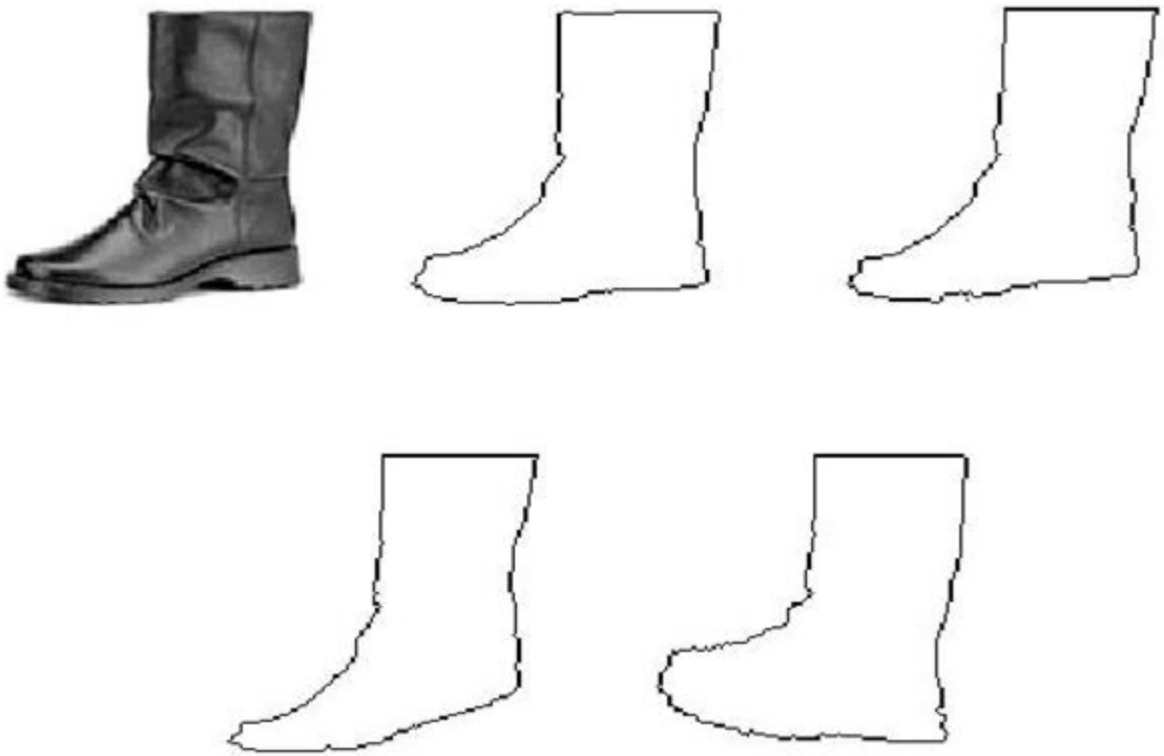


CSE 473
Pattern Recognition

Template Matching



Template Matching

- Typical Applications
  - Speech Recognition
  - Motion Estimation in Video Coding
  - Data Base Image Retrieval
  - Written Word Recognition
  - Bioinformatics

Template Matching

- The Goal:
  - Given a set of reference patterns known as TEMPLATES,
  - find the best match for unknown pattern
  - each class represented by a single typical pattern.

- requires an appropriate “measure” to quantify similarity or matching.

Template Matching

- The cost "measure":
  - deviations between the template and the test pattern.

Template Matching

- The cost "measure":
  - deviations between the template and the test pattern.
  - For example:
    - The word beauty may have been read as beeauty or beuty, etc., due to errors.
    - The same person may speak the same word differently.

Template Matching Methods

- Optimal path searching techniques
- Correlation
- Deformable models

TM using Optimal Path Searching

- Representation: Represent the template by a sequence of measurement vectors or string patterns

Template: $\underline{r}(1), \underline{r}(2), \ldots, \underline{r}(I)$

Test pattern: $\underline{t}(1), \underline{t}(2), \ldots, \underline{t}(J)$

TM using Optimal Path Searching

Template: $\underline{r}(1), \underline{r}(2), \ldots, \underline{r}(I)$

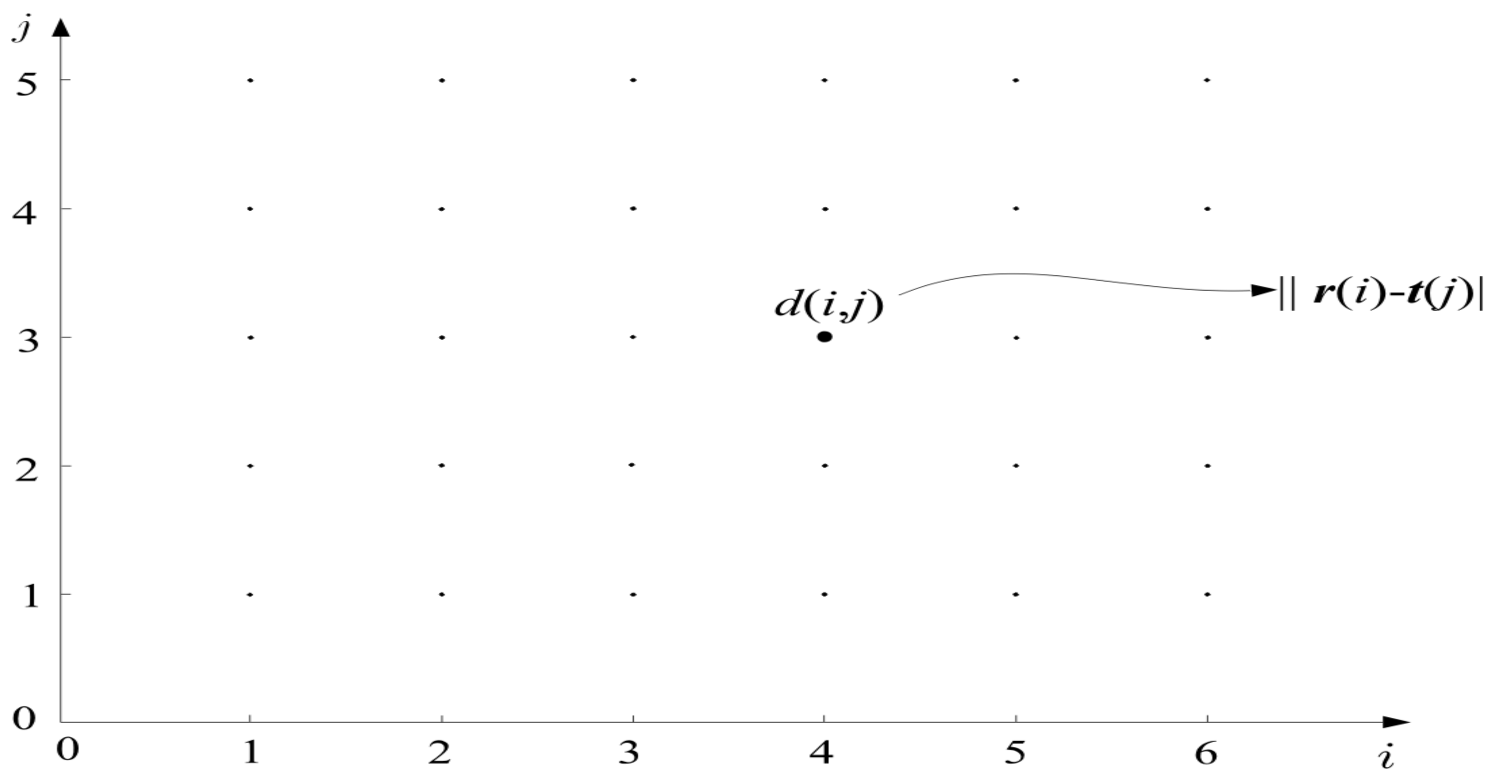
Test pattern: $\underline{t}(1), \underline{t}(2), \ldots, \underline{t}(J)$

- In general $I \neq J$

- We need to find an appropriate distance measure between test and reference patterns.

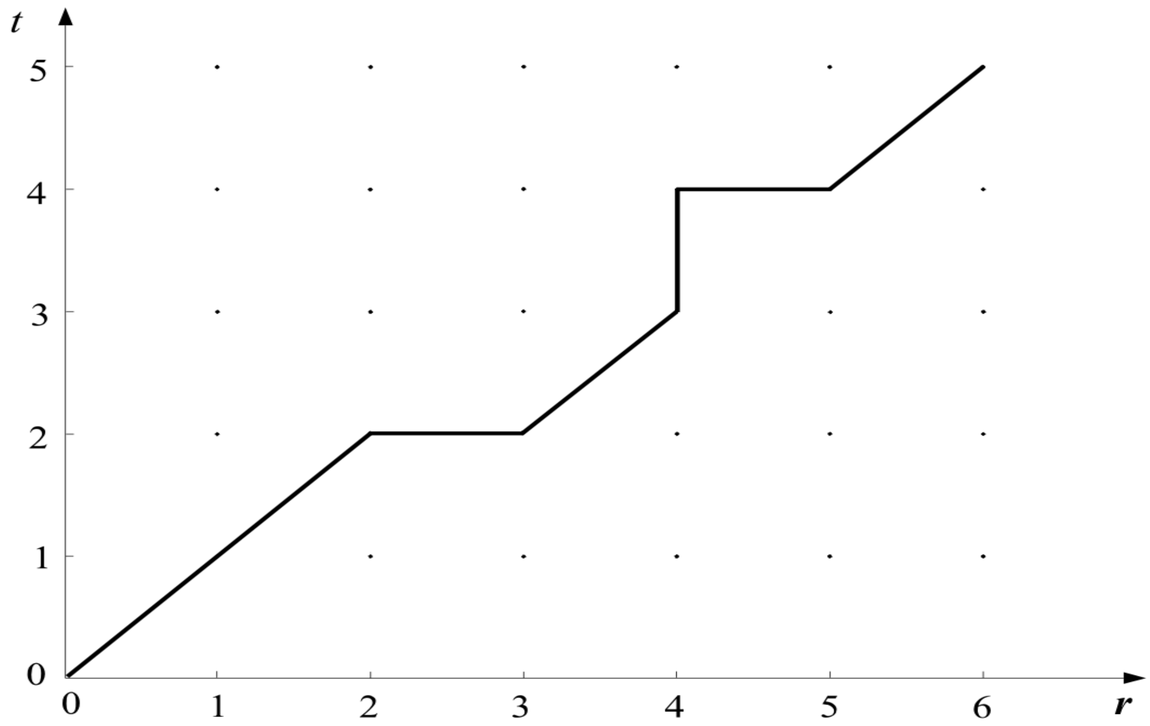
TM using Optimal Path Searching

- Form a grid with I points (template) in horizontal and J points (test) in vertical
- Each point (i,j) of the grid measures the distance between $\underline{r}(i)$ and $\underline{t}(j)$



TM using Optimal Path Searching

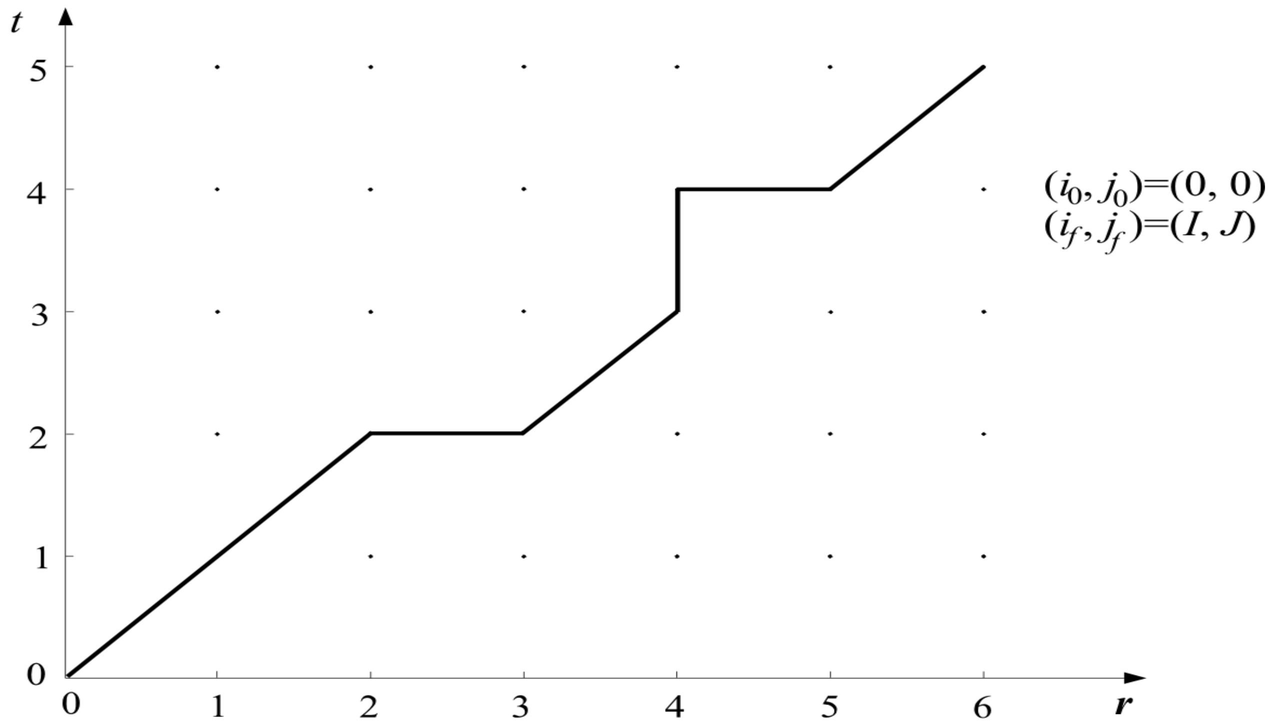
- Path: A path through the grid, from an initial node (i_0, j_0) to a final one (i_f, j_f), is an ordered set of nodes $(i_0, j_0), (i_1, j_1), (i_2, j_2) \ldots (i_k, j_k) \ldots (i_f, j_f)$



TM using Optimal Path Searching

- Path: A path is complete path if:

$$(i_0, j_0) = (0, 0), (i_1, j_1), (i_2, j_2), \ldots, (i_f, j_f) = (I, J)$$

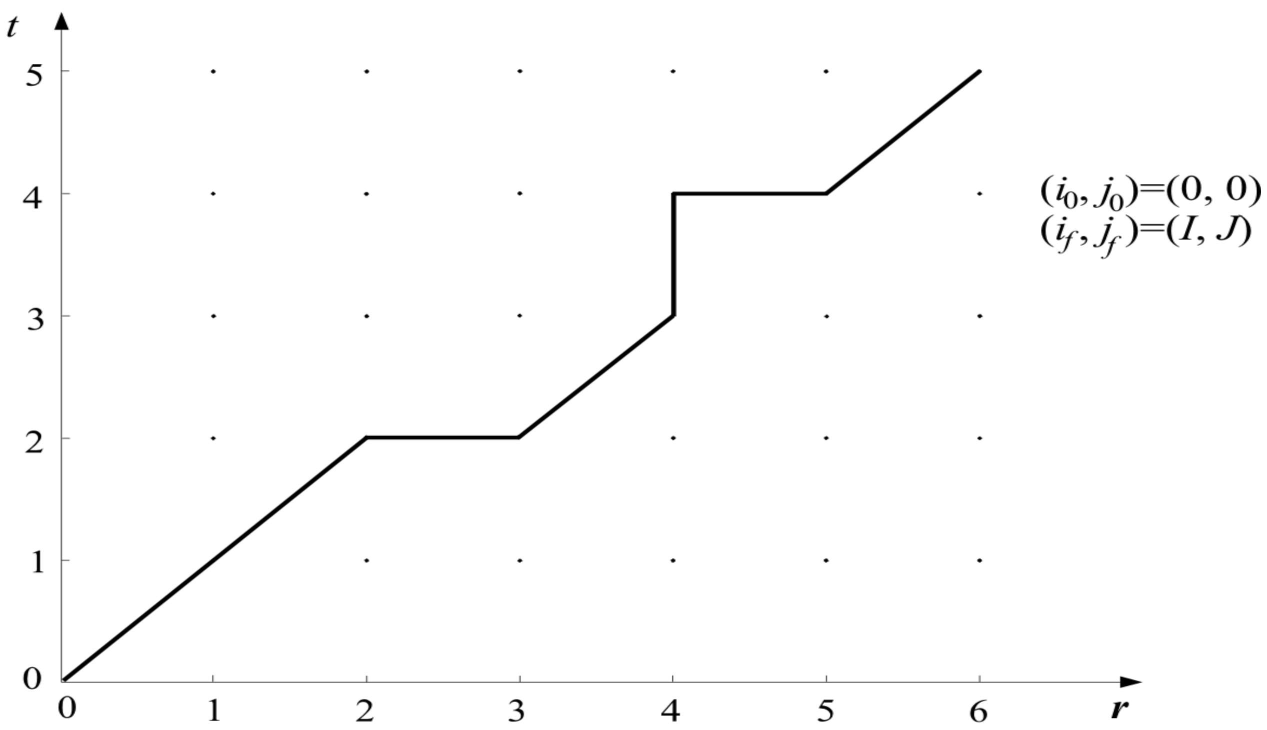


TM using Optimal Path Searching

– Each path is associated with a cost

$$D = \sum_{k=0}^{K-1} d(i_k, j_k)$$

where K is the number of nodes across the path



TM using Optimal Path Searching

- Let the cost up to node (i_k, j_k) be $D(i_k, j_k)$
- By convention
  - $D(0, 0)=0$
  - $d(0,0)=0$

TM using Optimal Path Searching

- The equation

$$D = \sum_{k=0}^{K-1} d(i_k, j_k)$$

assumes that each node has been associated with some cost

TM using Optimal Path Searching

- The equation

$$D = \sum_{k=0}^{K-1} d(i_k, j_k)$$

assumes that each node has been associated with some cost

- However, each transition (i_{k-1}, j_{k-1}) to (i_k, j_k) may also associate with a cost
- The new equation is:

$$D = \sum_{k} d(i_k, j_k | i_{k-1}, j_{k-1})$$

TM using Optimal Path Searching

$$D = \sum_k d(i_k, j_k | i_{k-1}, j_{k-1})$$

- Search for the path with the optimal cost $D_{opt.}$
- The matching cost between template $\underline{r}$ and test pattern $\underline{t}$ is $D_{opt.}$
- Costly operation
- Needs efficient computation

Bellman's Optimality Principle

- Optimal path:

$$(i_0, j_0) \xrightarrow{opt} (i_f, j_f)$$

Bellman's Optimality Principle

- Optimal path:

$$(i_0, j_0) \xrightarrow{opt} (i_f, j_f)$$

- Let *(i,j)* be an intermediate node, i.e.

$$(i_0, j_0) \to \ldots \to (i, j) \to \ldots \to (i_f, j_f)$$

Bellman's Optimality Principle

- Optimal path:

$$(i_0, j_0) \xrightarrow{opt} (i_f, j_f)$$

- Let (i,j) be an intermediate node, i.e.

$$(i_0, j_0) \rightarrow \ldots \rightarrow (i, j) \rightarrow \ldots \rightarrow (i_f, j_f)$$

Then, write the optimal path through (i, j)

$$(i_0, j_0) \xrightarrow[(i,j)]{opt} (i_f, j_f)$$

Bellman's Optimality Principle

- Bellman's Principle:

$(i_0, j_0) \xrightarrow{opt} (i_f, j_f)$ can be obtained as

$$(i_0, j_0) \xrightarrow{opt} (i, j) \oplus (i, j) \xrightarrow{opt} (i_f, j_f)$$

- meaning: The overall optimal path from (i_0,j_0) to (i_f,j_f) through (i,j) is the concatenation of the optimal paths from (i_0,j_0) to (i,j) and from (i,j) to (i_f,j_f)

Bellman's Optimality Principle

- Bellman's Principle:

$$(i_0, j_0) \xrightarrow{opt} (i_f, j_f) \Leftrightarrow (i_0, j_0) \xrightarrow{opt} (i, j) \oplus (i, j) \xrightarrow{opt} (i_f, j_f)$$

- Let $D_{opt.}(i_{k\text{-}1}, j_{k\text{-}1})$ is the optimal path to reach $(i_{k\text{-}1}, j_{k\text{-}1})$ from (i_0, j_0), then Bellman's principle is stated as:

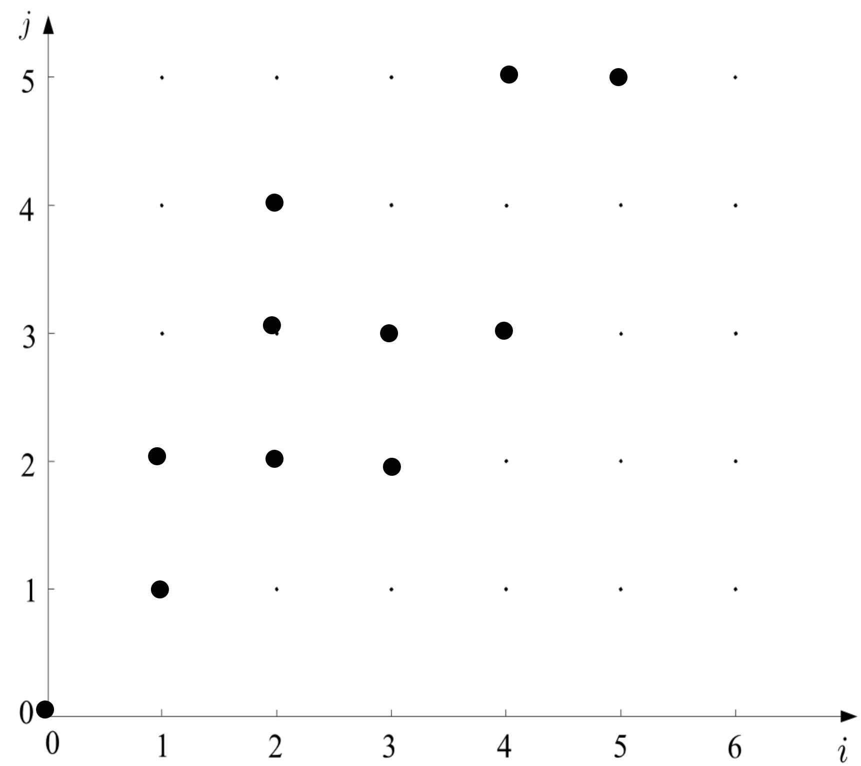
$$D_{opt}(i_k, j_k) = opt\{D_{opt}(i_{k-1}, j_{k-1}) + d(i_k, j_k \mid i_{k-1}, j_{k-1})\}$$

Bellman's Optimality Principle

$$D_{opt}(i_k, j_k) = opt\{D_{opt}(i_{k-1}, j_{k-1}) + d(i_k, j_k \mid i_{k-1}, j_{k-1})\}$$

- We don't need to search the whole space to find the optimal path

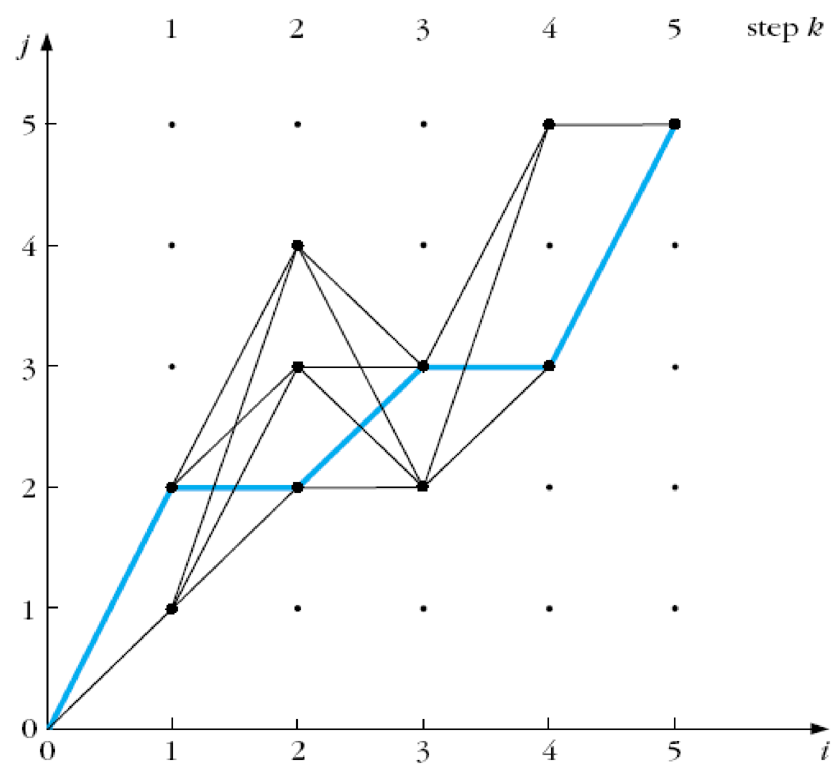
- Global and local constraints may be imposed to reduce the search space



Bellman's Optimality Principle

$$D_{opt}(i_k, j_k) = opt\{D_{opt}(i_{k-1}, j_{k-1}) + d(i_k, j_k \mid i_{k-1}, j_{k-1})\}$$

- We don't need to search the whole space to find the optimal path
- Global and local constraints may be imposed to reduce the search space



Application of TM in Text Matching: The Edit Distance

- The Edit distance
  - It is used for matching written words. Applications:
    - Automatic Editing
    - Text Retrieval

Application of TM in Text Matching: The Edit Distance

- The Edit distance
  - It is used for matching written words. Applications:
    - Automatic Editing
    - Text Retrieval
  - The measure to be adopted for matching, must take into account:
    - Wrongly identified symbols
      e.g. “befuty” instead of “beauty”
    - Insertion errors, e.g. “bearuty”
    - Deletion errors, e.g. “beuty”

The Edit Distance

- Edit distance: **Minimal** total number of changes, *C*, insertions I and deletions R, required to change pattern A into pattern B,

$$D(A,B)=\min_j[C(j)+I(j)+R(j)]$$

where j runs over All possible variations of symbols, in order to convert $A \longrightarrow B$

The Edit Distance

- Edit distance: **Minimal** total number of changes, C, insertions I and deletions R, required to change pattern A into pattern B,

$$D(A,B) = \min_j [C(j) + I(j) + R(j)]$$

where j runs over All possible variations of symbols, in order to convert $A \longrightarrow B$

- *Example*: many ways to change **beuty** to **beauty**

The Edit Distance

- The optimal path search algorithm can be used, provided we know
  - Initial conditions
  - Search space
  - Allowable transitions
  - Distance measure

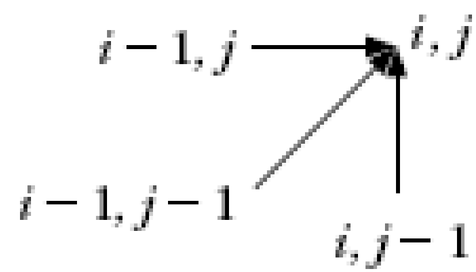
The Edit Distance

- Cost $D(0,0) = 0$,
- Complete path is searched
- Allowable predecessors and costs:

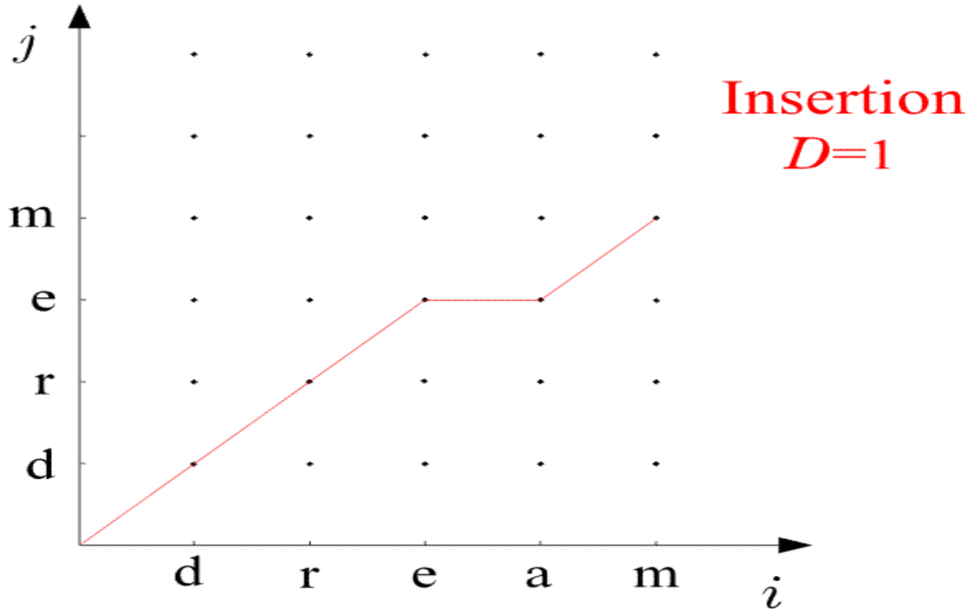
  - $(i-1, j-1) \rightarrow (i, j)$

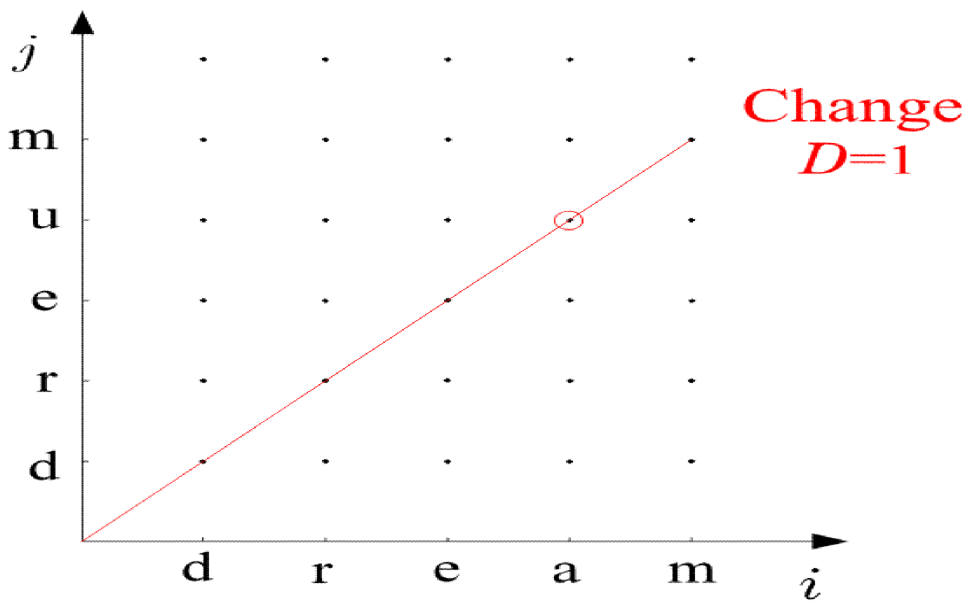
    $$d(i, j \mid i-1, j-1) = \begin{cases} 0, & \text{if } t(i) = r(j) \\ 1, & t(i) \neq r(j) \end{cases}$$

  - Horizontal $d(i, j \mid i-1, j) = 1$
  - Vertical $d(i, j \mid i, j-1) = 1$

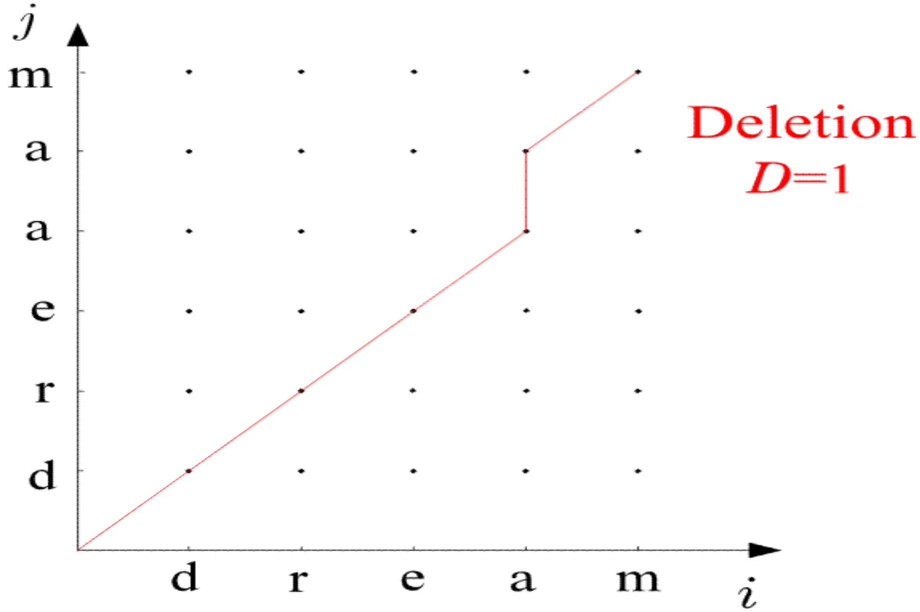


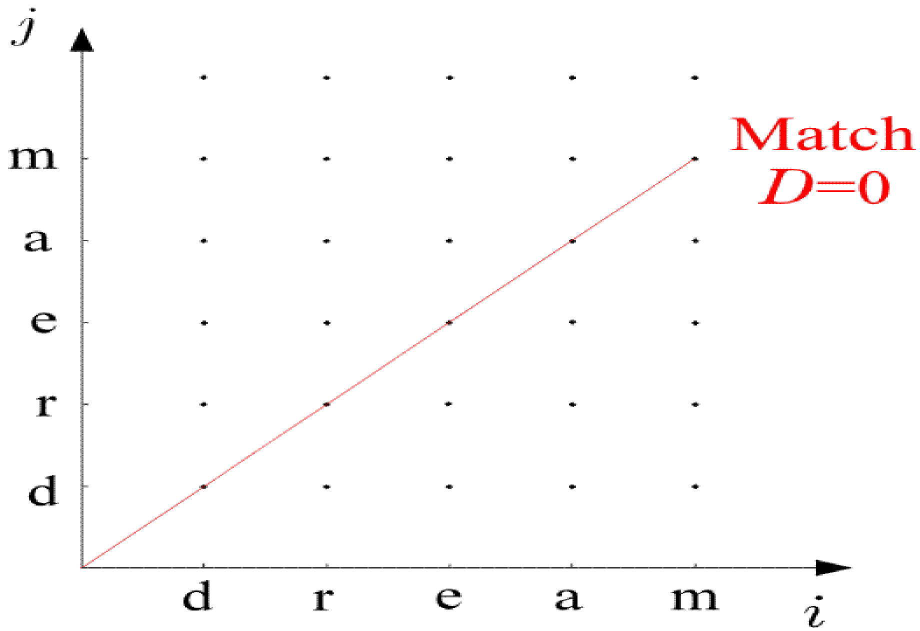
- Examples:





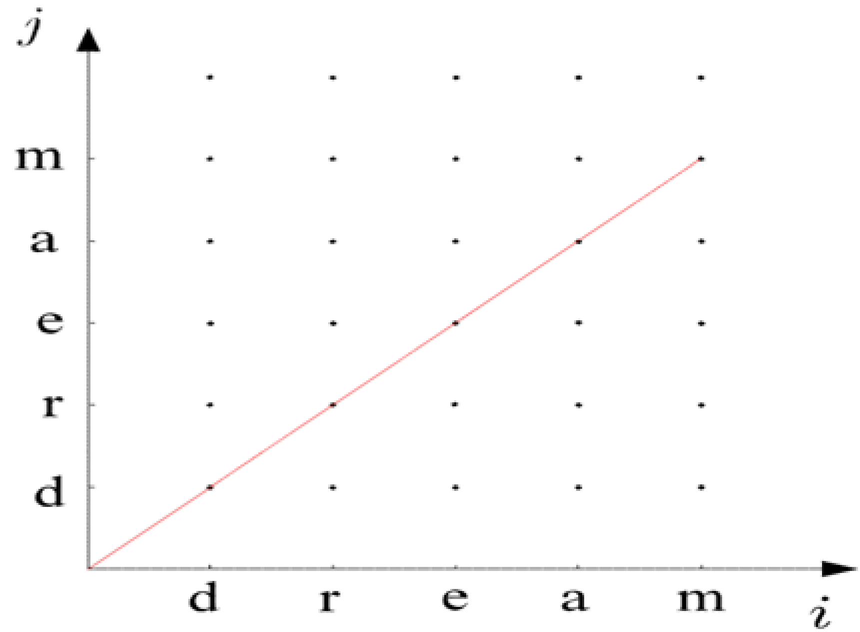
- Examples:





The Edit Distance

- The Algorithm
  - $D(0,0)=0$
  - *For* i=1, *to* I
    - $D(i,0)=D(i\text{-}1,0)+1$
  - *END {FOR}*
  - *For* $j=1$ *to* J
    - $D(0,j)=D(0,j\text{-}1)+1$
  - *END{FOR}*
  - *For* $i=1$ *to* I
    - *For* $j=1$, *to* J
      - $C_1=D(i\text{-}1,j\text{-}1)+d(i,j \mid i\text{-}1,j\text{-}1)$
      - $C_2=D(i\text{-}1,j)+1$
      - $C_3=D(i,j\text{-}1)+1$
      - $D(i,j)=min\ (C_1,C_2,C_3)$
    - *END {FOR}*
  - *END {FOR}*
  - $D(A,B)=D(I,J)$



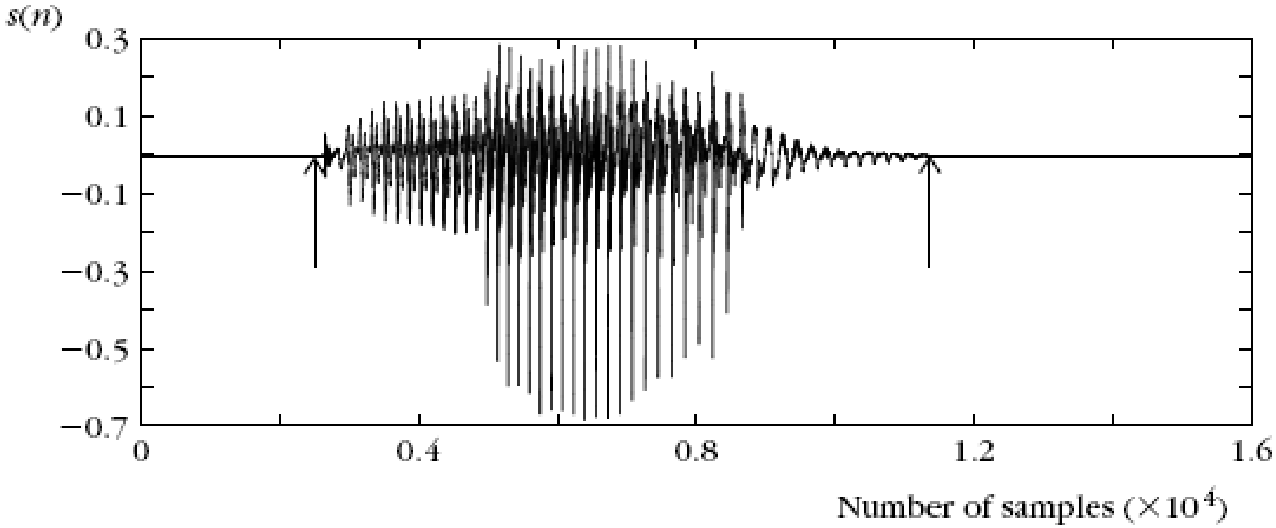
Application of TM in Speech Recognition

- A number of variations
  - Speaker Independent Speech Recognition
  - Speaker Dependent Speech Recognition
  - Continuous Speech Recognition
  - Isolated word recognition (IWR)

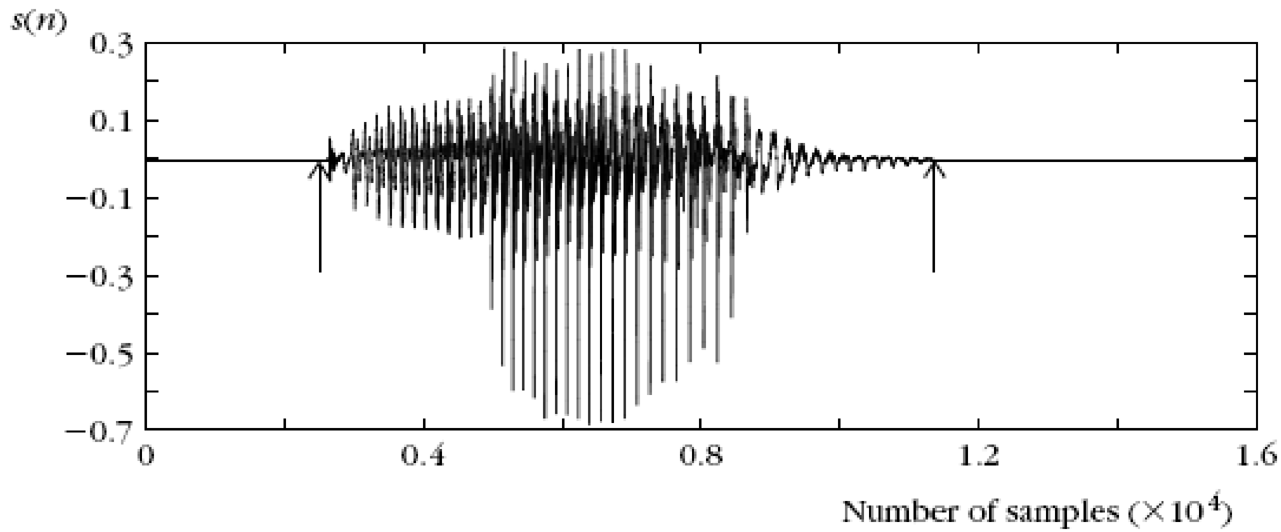
Application of TM in IWR

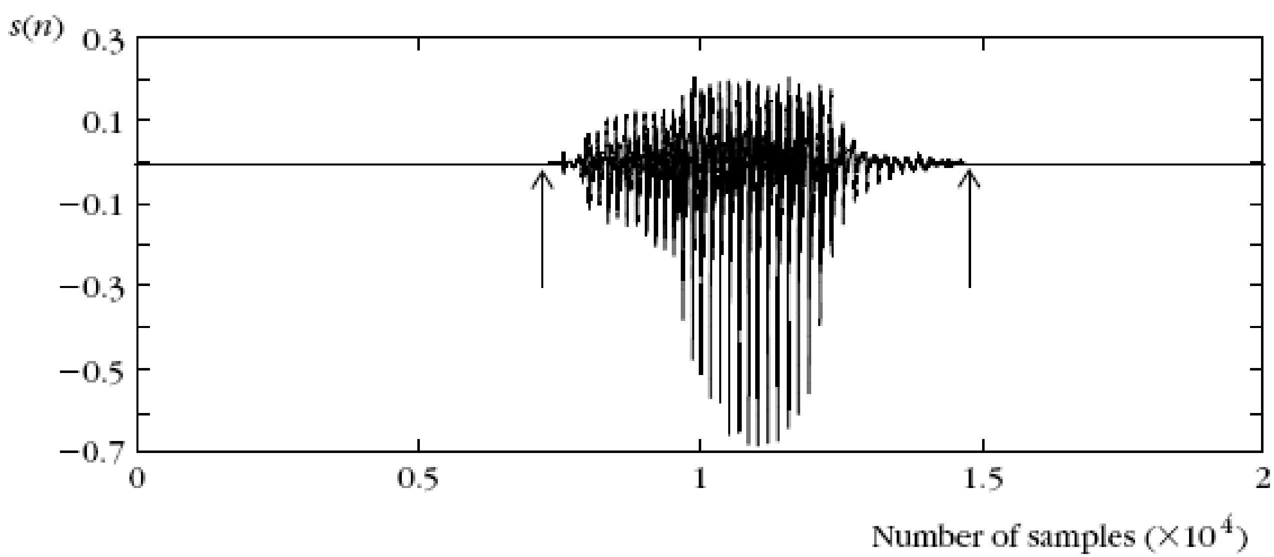
- The goal:
  - Given a number of known spoken words in a data base (reference patterns)
  - find the best match of an unknown spoken word (test pattern).
- Procedure:
  - compare the test word against reference words

Application of TM in IWR



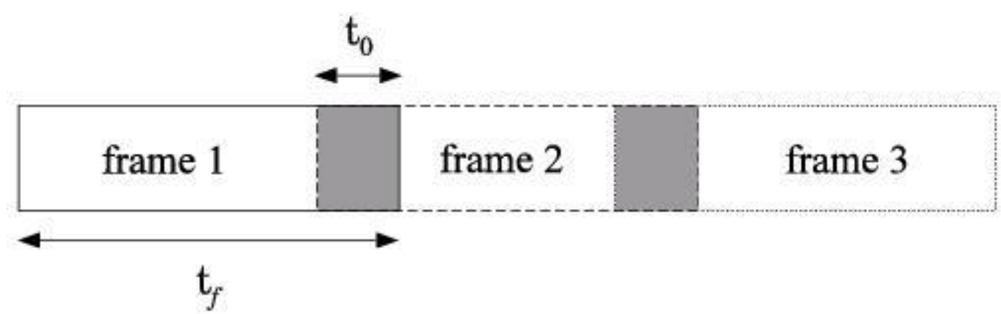
Application of TM in IWR





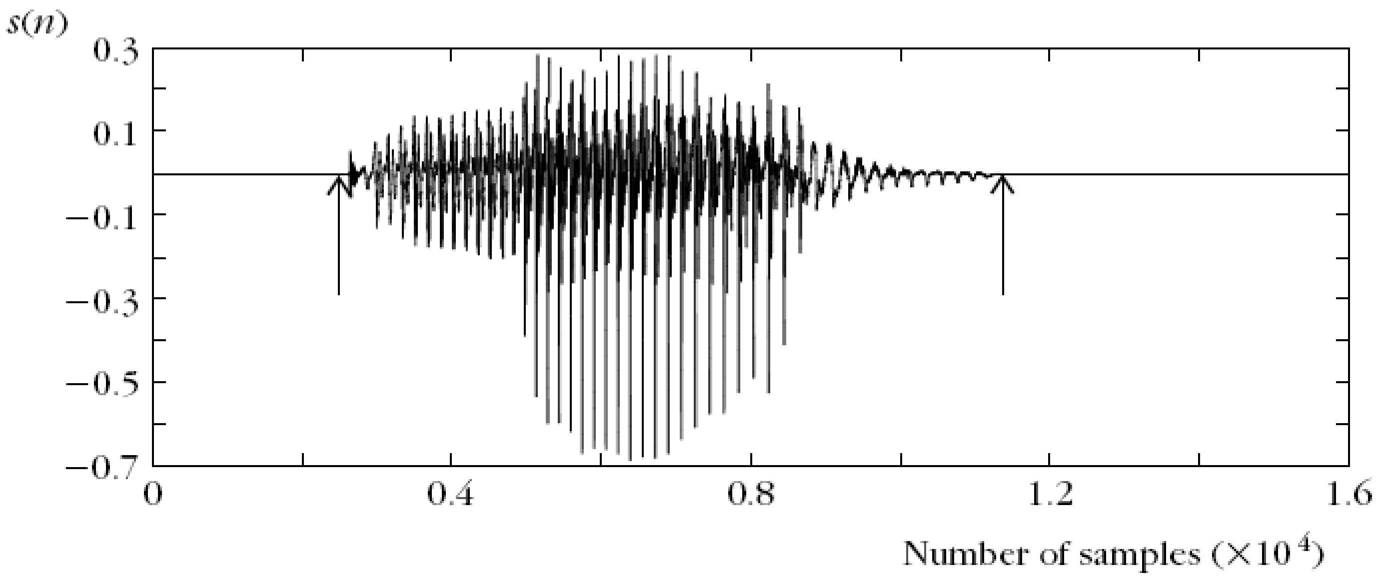
Application of TM in IWR

- The procedure:
  - Express the test and each of the reference patterns as sequences of feature vectors , $\underline{r}(i)$, $\underline{t}(j)$.
  - To this end, divide each of the speech segments in a number of successive frames.



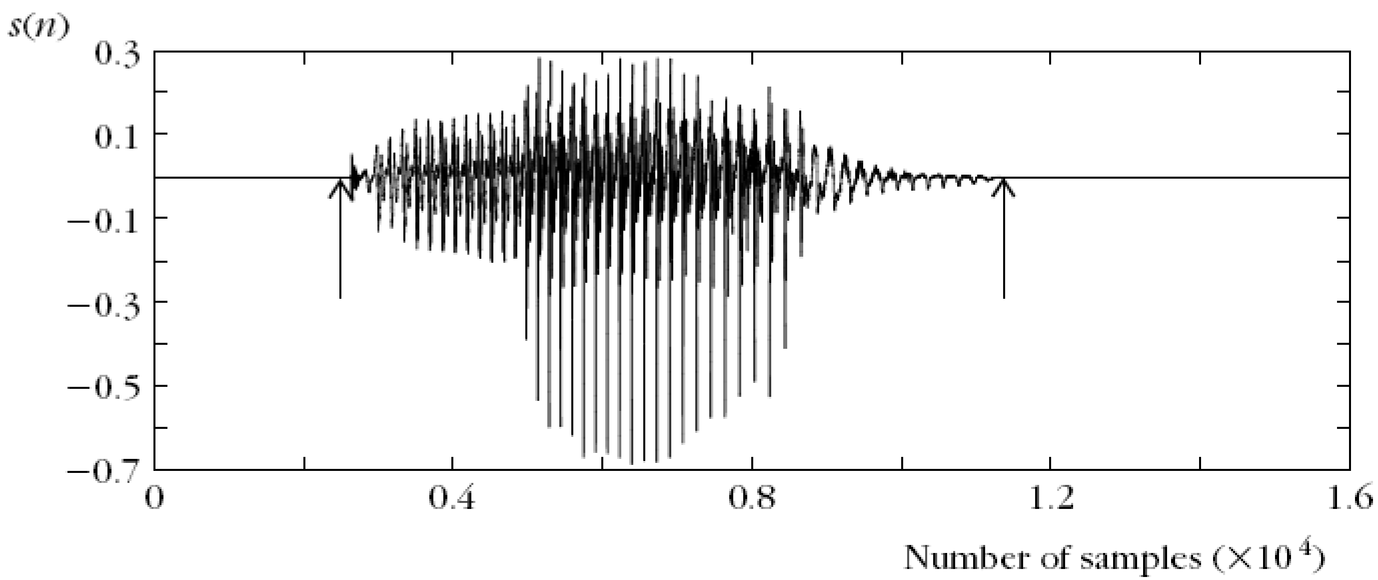
Application of TM in IWR

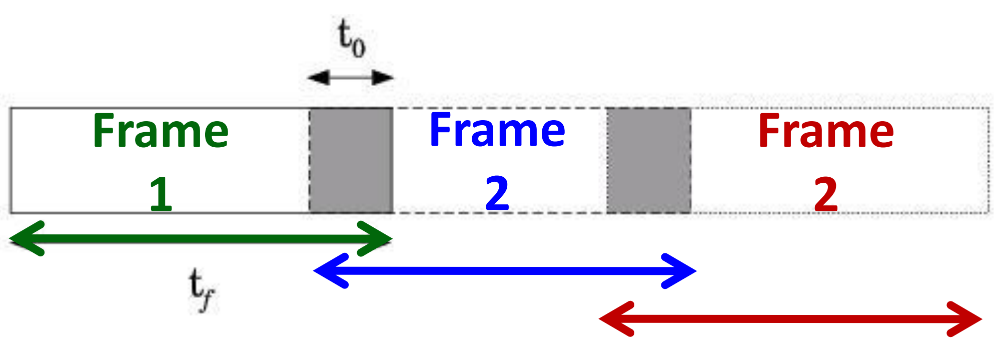
- The procedure:
  - Sample a speech segment from a microphone:



Application of TM in IWR

- The procedure:





$t_f = 512$

$t_0 = 100$

- each frame is represented by a vector of 512 samples

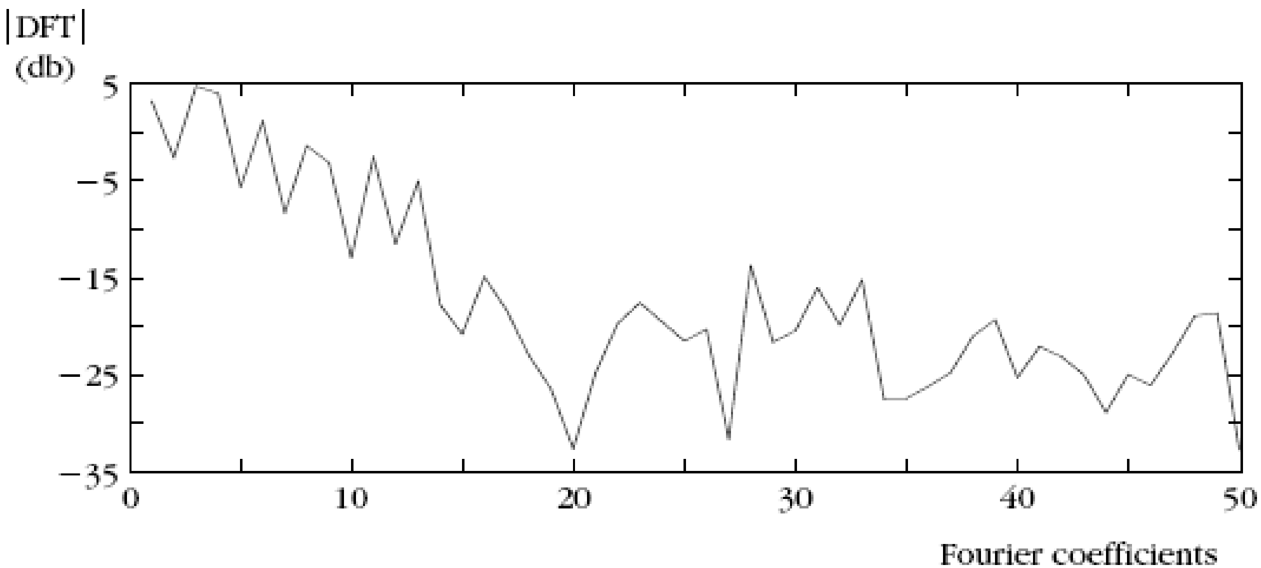
$$\underline{r}(i) = \begin{bmatrix} x_i(0) \\ x_i(1) \\ \ldots \\ \ldots \\ x_i(512) \end{bmatrix}, \; i = 1, \ldots, I \qquad \underline{t}(j) = \begin{bmatrix} x_j(0) \\ x_j(1) \\ \ldots \\ \ldots \\ x_j(512) \end{bmatrix}, \; j = 1, \ldots, J$$

- convert them to DFT

$$DFT(\underline{r}(i)) = DFT(\begin{bmatrix} x_i(0) \\ x_i(1) \\ \dots \\ \dots \\ x_i(512) \end{bmatrix}) = \begin{bmatrix} X_i(0) \\ X_i(1) \\ \dots \\ \dots \\ X_i(512) \end{bmatrix}$$

$$DFT(\underline{t}(j)) = DFT(\begin{bmatrix} x_i(0) \\ x_i(1) \\ \dots \\ \dots \\ x_i(512) \end{bmatrix}) = \begin{bmatrix} X_i(0) \\ X_i(1) \\ \dots \\ \dots \\ X_i(512) \end{bmatrix}$$

- convert them to DFT



- For each frame compute a feature vector. For example, the DFT coefficients and use, say, ℓ of those:

$$\underline{r}(i) = \begin{bmatrix} X_i(0) \\ X_i(1) \\ \dots \\ \dots \\ X_i(\ell - 1) \end{bmatrix}, \quad i = 1, \dots, I \qquad \underline{t}(j) = \begin{bmatrix} X_j(0) \\ X_j(1) \\ \dots \\ \dots \\ X_j(\ell - 1) \end{bmatrix}, \quad j = 1, \dots, J$$

- For each frame compute a feature vector. For example, the DFT coefficients and use, say, ℓ of those:

$$\underline{r}(i) = \begin{bmatrix} X_i(0) \\ X_i(1) \\ \dots \\ \dots \\ X_i(\ell-1) \end{bmatrix}, \; i = 1, \dots, I \qquad \underline{t}(j) = \begin{bmatrix} X_j(0) \\ X_j(1) \\ \dots \\ \dots \\ X_j(\ell-1) \end{bmatrix}, \; j = 1, \dots, J$$

- Choose a cost function associated with each node across a path, e.g., the Euclidean distance

$$\left\| \underline{r}(i_k) - \underline{t}(j_k) \right\| = d(i_k, j_k)$$

- For each frame compute a feature vector. For example, the DFT coefficients and use, say, ℓ of those:

$$\underline{r}(i)=\begin{bmatrix} X_i(0) \\ X_i(1) \\ \dots \\ \dots \\ X_i(\ell-1) \end{bmatrix},\ i=1,\ \dots,\ I \qquad \underline{t}(j)=\begin{bmatrix} X_j(0) \\ X_j(1) \\ \dots \\ \dots \\ X_j(\ell-1) \end{bmatrix},\ j=1,\ \dots,\ J$$

- Choose a cost function associated with each node across a path, e.g., the Euclidean distance

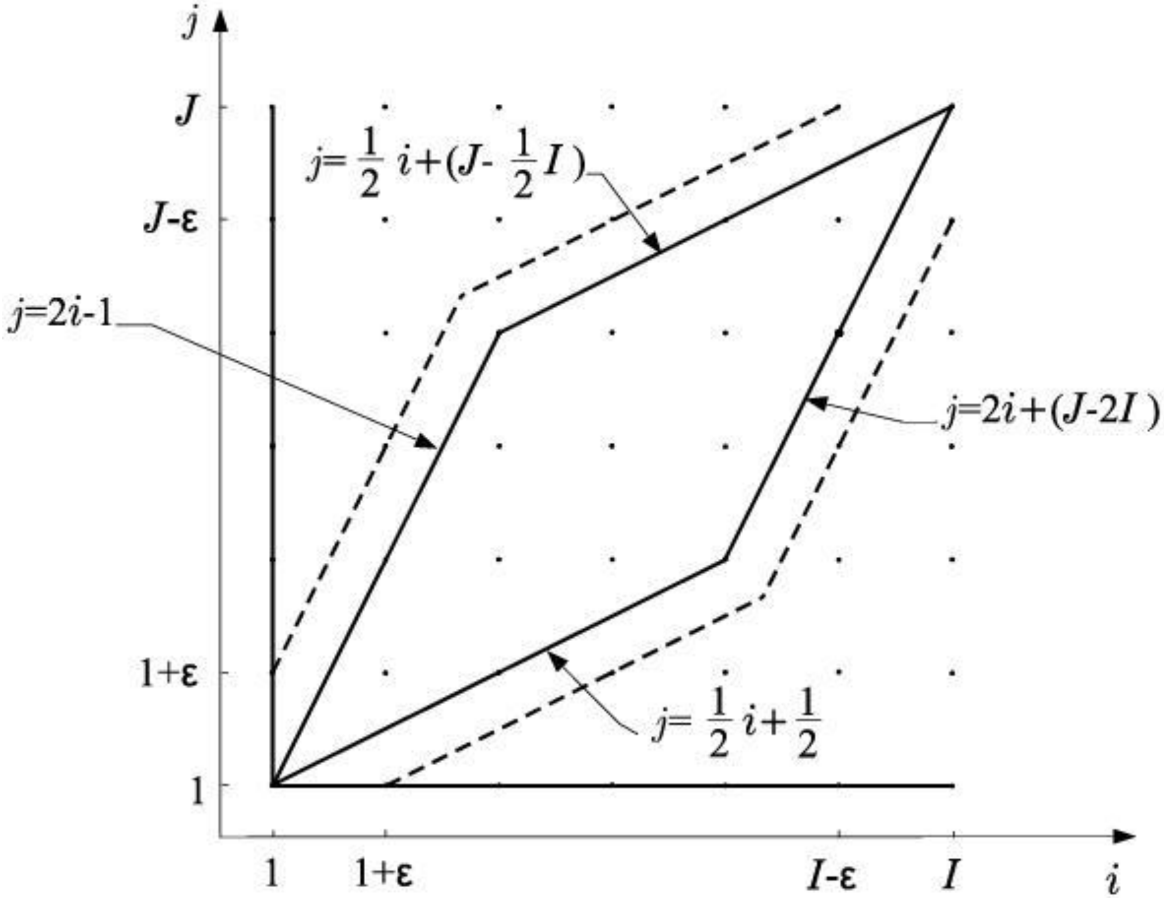
$$\left\|\underline{r}(i_k)-\underline{t}(j_k)\right\|=d(i_k,\ j_k)$$

- find the optimal path in the grid
- Match the test pattern to the reference pattern associated with the optimal path

- Prior to performing the math one has to choose:
  - end point constraints
  - global constraints
  - local constraints
  - distance

– Prior to performing the math one has to choose:

- The global constraints: Defining the region of space within which the search for the optimal path will be performed.



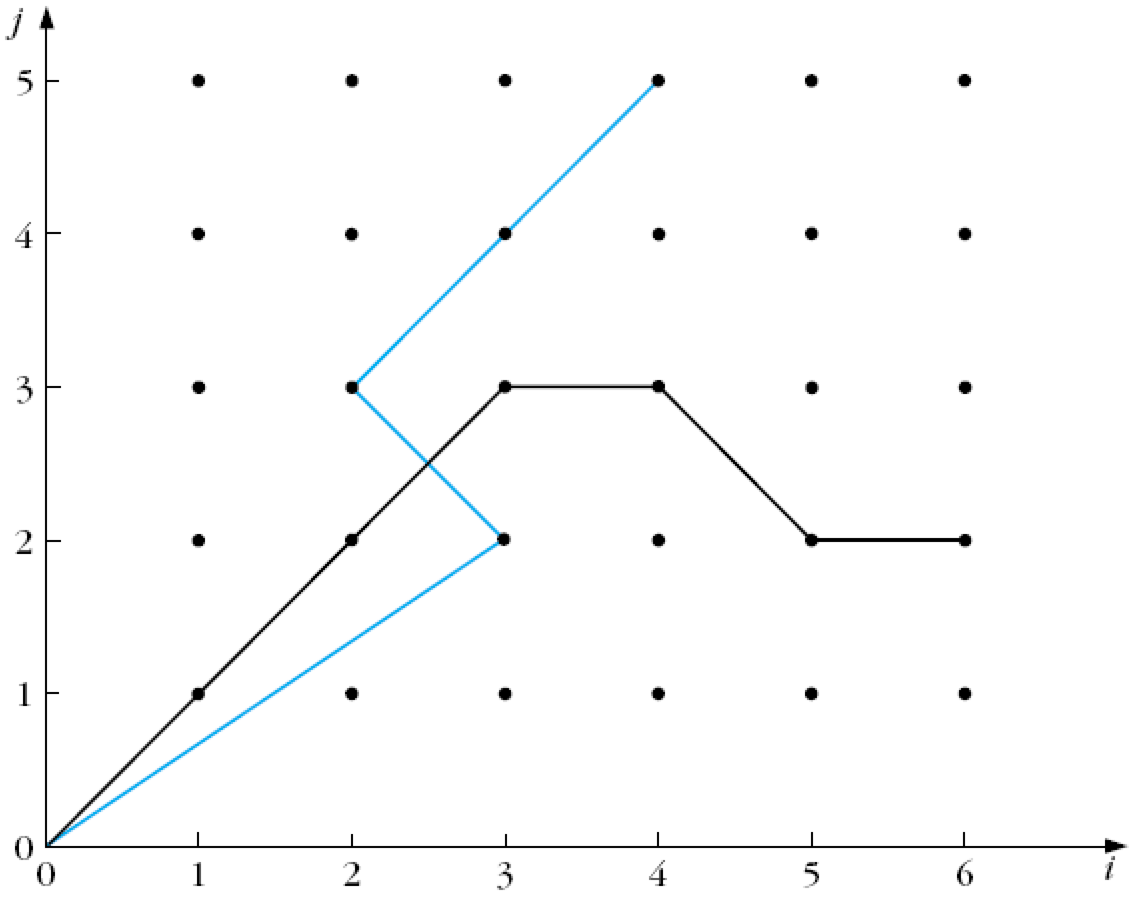
- The local constraints: monotonic path

$$i_{k-1} \leq i_k \quad \text{and} \quad j_{k-1} \leq j_k$$

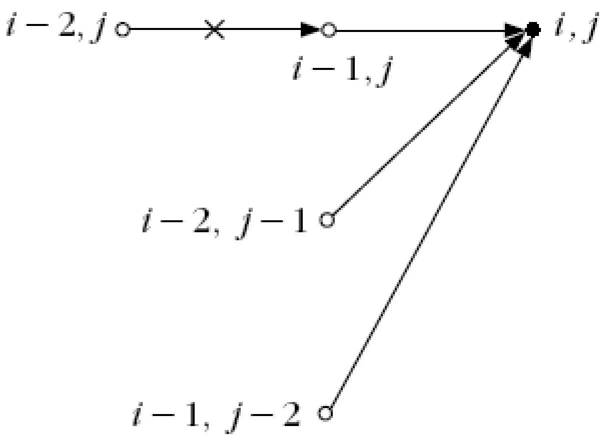
- The local constraints: monotonic path

$$i_{k-1} \leq i_k \quad \text{and} \quad j_{k-1} \leq j_k$$

- Non-monotonic path

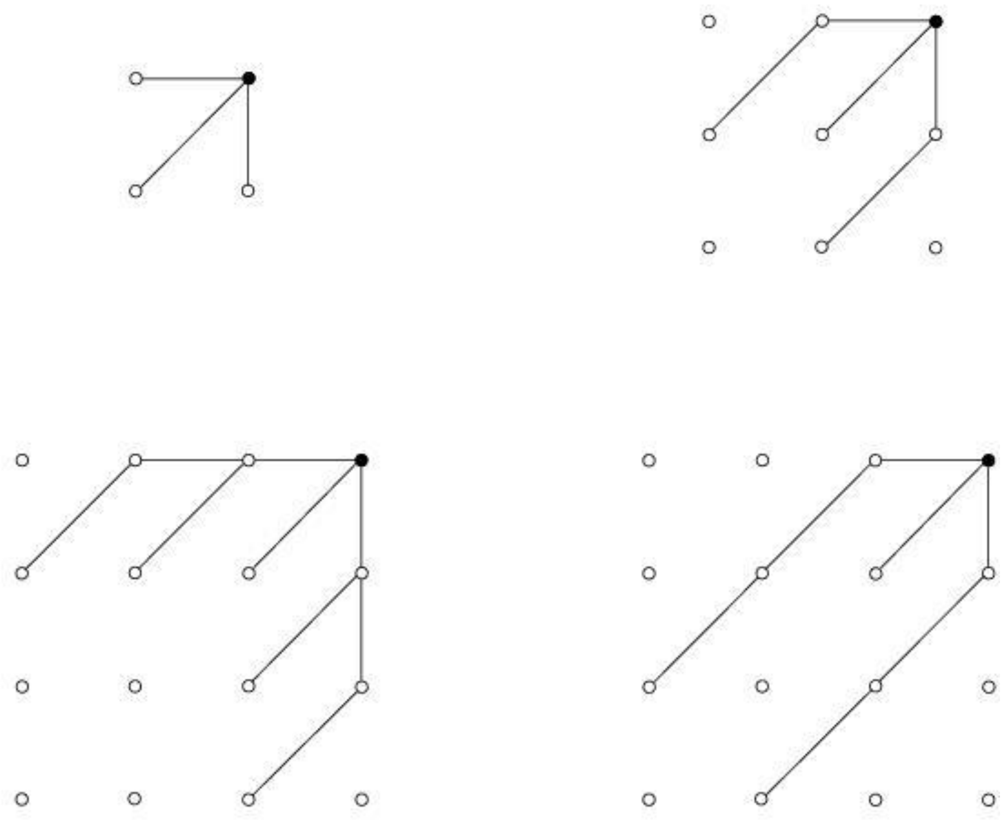


- The local constraints: Defining the type of transitions allowed between the nodes of the grid.



Itakura local constraints

- The local constraints: Defining the type of transitions allowed between the nodes of the grid.



Sakoe and Chiba local constraints

- cost function:
  - Euclidean distance
  - only node distance

$$d(i_k, j_k \mid i_{k-1}, j_{k-1}) = d(i_k, j_k)$$

$$= \left\| \underline{r}(i_k) - \underline{t}(j_k) \right\|$$