CHAPTER III

PSEUDOMETRICS FOR TIME SERIES

This is the main chapter of the thesis. It consists of two guidelines and examples of applications of the guidelines.

3.1 Condensations of Distances

First of all, our notion of distance condensation should not be confused with the concept condensing by Devroye et al. (1996, chap. 19), where the data points are eliminated such that the classification is kept unchanged. Having a set of structured data, we usually have a simple distance measure that is easy to compute but gives undesirable classification accuracy when used with the nearest neighbor algorithm. Sometimes we want to allow variations of the two objects in a controllable manner so that they become more similar before we decide how different the two objects are. For example, two signals whose shapes are almost the same but one arrives a second later than the other should be considered almost the same without taking the time shift into account.

With a set of *morphs* allowed to be made to objects before being compared, we can always define another distance function.

Definition 3. *Let Ω be a distance space with the distance d, and $\mathcal{M}$ be a set of functions from Ω to Ω. The distance*

$$\Delta_{d,\mathcal{M}}(x,y) := \inf_{\mu,\nu \in \mathcal{M}} d(\mu x, \nu y) \ ,$$

is called the condensation of d with respect to $\mathcal{M}$.

Note that the value of $\Delta_{d,\mathcal{M}}$ is never greater than d. The idea of condensation is depicted in Figure 3.1; one may think that each point x in the space Ω is mapped to $\{\mu x | \mu \in \mathcal{M}\}$, the set of its all possible morphs by functions in $\mathcal{M}$, and the new distance is the distance between such sets (the distance between sets in this sense is the distance between their closest elements).

For a set of objects, two main components constitute a good distance measure; a base pseudometric and a set of *morph* operations of the objects with desirable properties. We give the definition of such morph operations below.

Definition 4. *Let $\mathcal{M}$ be a set of functions from Ω to Ω. $\mathcal{M}$ is said to be* complete *when,*

i) the identity map is in $\mathcal{M}$,

ii) for each $\mu, \nu \in \mathcal{M}$, the composition $\nu\mu$ is in $\mathcal{M}$,

iii) for each $\mu_1, \nu_1 \in \mathcal{M}$, there are $\mu_2, \nu_2 \in \mathcal{M}$ such that $\mu_2\mu_1 = \nu_2\nu_1$.

We write an application of a function μ in $\mathcal{M}$ to an element x using the prefix notation μx. Compositions are read from right to left i.e. $\mu_2\mu_1 x$ is the result of an application of μ_2 to $\mu_1 x$.

Condition iii) in the definition above is weaker than the requirement that the composition of functions in $\mathcal{M}$ is commutative, i.e. $\nu\mu = \mu\nu$ for every μ, ν in $\mathcal{M}$. It is also weaker than requiring that every function in $\mathcal{M}$ has an inverse. So if a set of functions over Ω is a group, it is always complete in this sense.

Intuitively, with a *complete* set of morphs, two objects morphed from the same object remains similar, in a sense that they are the same up to some further morphing. Time shifts are an example of a complete set of operations.

The following definitions are based on how a whole set of objects change their distance among each other when an operation is applied to the whole set.

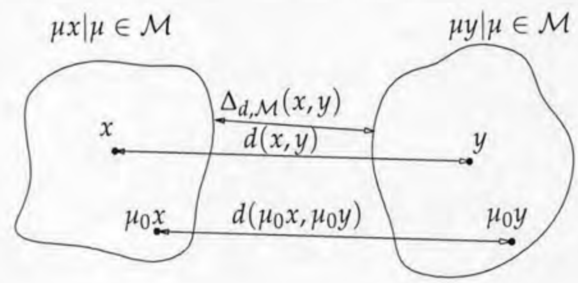

Figure 3.1: The condensation Δ_d measured between point x and y when $\mathcal{M}$ preserves d, μ_0 is a function in $\mathcal{M}$.

Definition 5. *Let Ω be a pseudometric space equipped with a pseudometric d, and $\mathcal{M}$ be a set of functions from Ω to itself. We say that $\mathcal{M}$* preserves *d if and only if,*

$$\forall x, y \in \Omega \quad \forall \mu \in \mathcal{M} \qquad d(\mu x, \mu y) = d(x, y) \ ,$$

$\mathcal{M}$ contracts *d if and only if,*

$$\forall x, y \in \Omega \quad \forall \mu \in \mathcal{M} \qquad d(\mu x, \mu y) \leq d(x, y) \ ,$$

$\mathcal{M}$ expands *d if and only if,*

$$\forall x, y \in \Omega \quad \forall \mu \in \mathcal{M} \qquad d(\mu x, \mu y) \geq d(x, y) \ .$$

We may say that $\mathcal{M}$ is contractive or expansive when the associated distance is implicitly known.

In Figure 3.1 we illustrate that μ_0 preserves the distance d. One may perceive d as the spatial distance on the paper and μ_0 as the translation by a certain amount to the southeast direction, and translating two points at the same time keeps their distance.

It turns out that if we want a subadditive condensation distance wrt. complete morphs, we should focus our interest on *contractive* ones.

Theorem 3. *Let (Ω, d) be a pseudometric space and $\mathcal{M}$ be a complete set of morph operations on Ω. Then the condensation of d wrt. $\mathcal{M}$ is a pseudometric if $\mathcal{M}$ contracts d.*

Proof. For brevity we write Δ to denote the condensation of d wrt. $\mathcal{M}$ throughout the proof.

Assume that (Ω, d) is a pseudometric space and $\mathcal{M}$ is a complete set of morph operations on contracting d. Obviously, $\Delta(x, x) = 0$ for every x in Ω. The symmetry of Δ follows from the symmetry of d. Let $x, y, z \in \Omega$, $\varepsilon > 0$. By definition of Δ there are some $\mu_1, \mu_2, \nu_1, \nu_2 \in \mathcal{M}$ such that the following hold,

$$d(\mu_1 x, \mu_2 y) < \Delta(x,y) + \frac{\varepsilon}{2} , \tag{3.1}$$

$$d(\nu_1 y, \nu_2 z) < \Delta(y,z) + \frac{\varepsilon}{2} . \tag{3.2}$$

Since $\mathcal{M}$ is complete, there are $\mu_0, \nu_0 \in \mathcal{M}$ making $\mu_0\mu_2 y = \nu_0\nu_1 y$. By definition,

$$\begin{aligned}
\Delta(x,z) &\leq d(\mu_0\mu_1 x, \nu_0\nu_2 z) \\
&\leq d(\mu_0\mu_1 x, \mu_0\mu_2 y) + d(\nu_0\nu_1 y, \nu_0\nu_2 z) \\
&\leq d(\mu_1 x, \mu_2 y) + d(\nu_1 y, \nu_2 z) \\
&< \Delta(x,y) + \Delta(y,z) + \varepsilon .
\end{aligned}$$

The second inequality follows from $\mu_0\mu_2 y = \nu_0\nu_1 y$, the equality is by the assumption that $\mathcal{M}$ contracts d, and the last inequality follows from (3.1) and (3.2).

This is true for arbitrary $\varepsilon > 0$, so $\Delta(x,z) \leq \Delta(x,y) + \Delta(y,z)$. ■

3.1.1 Examples

A simple and trivial example of this kind of pseudometrics is the condensation of the Euclidean distance in a vector space wrt. arbitrary rotations about the origin. The resulting metric is just the difference between the lengths of its two arguments.

As another example, we construct a condensation of the ℓ^∞ metric wrt. the *stretch* operations defined below.

Definition 6. *Let V be the set of all finite sequences. For each $k \in \mathbb{N}, \mathbf{x} \in V$ define $\sigma_k : V \to V$ by,*

$$\sigma_k \mathbf{x} = \begin{cases} [x_1, \ldots, x_{k-1}, x_k, x_k, x_{k+1}, \ldots, x_l] & \text{if } k < l, \\ [x_1, \ldots, x_{k-1}, x_l, x_l] & \text{if } k = l, \\ \mathbf{x} & \text{otherwise} . \end{cases}$$

Let $\mathcal{S}$ be the set containing every finite compositions of morph operations in $\{\sigma_k\}_{k\in\mathbb{N}} \cup \{\mathbb{1}\}$.

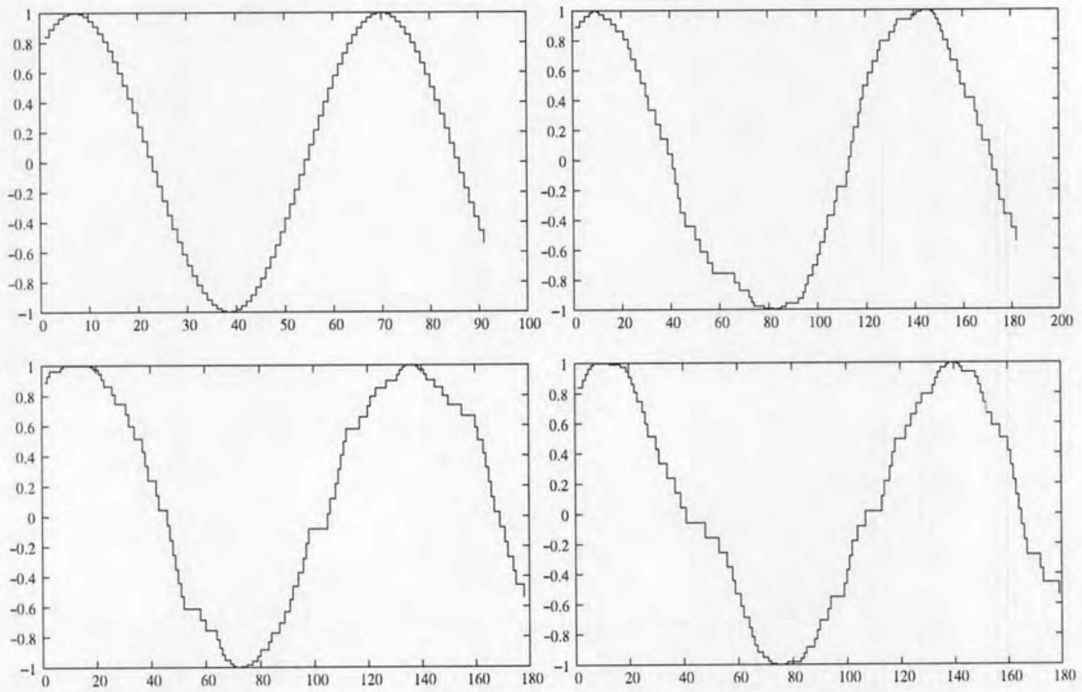

Figure 3.2: A visualization of a time series and its possible stretches. The original time series of length 91 is the top-left. The rest are some of its possible stretches to twice the original length.

One can verify that $\mathcal{S}$ preserves ℓ^∞ metric. Figure 3.2 shows how stretches of a time series may look like.

As a consequence, the following function is a pseudometric on V,

$$\delta_1(\mathbf{x}, \mathbf{y}) = \inf_{\mu,\nu\in\mathcal{S}} \|\mu\mathbf{x} - \nu\mathbf{y}\|_\infty . \tag{3.3}$$

Where $\|\mathbf{x} - \mathbf{y}\|_\infty$ is implicitly defined to be ∞ when $\#\mathbf{x} \neq \#\mathbf{y}$.

Similar to DTW, the distance above can be written in terms of its partial answers.

$$\delta_1(\mathbf{s}, \mathbf{t}) = \max\left(|s_1 - t_1|, \min \begin{cases} \delta_1(\mathbf{s}_\sim, \mathbf{t}_\sim), \\ \delta_1(\mathbf{s}_\sim, \mathbf{t}), \\ \delta_1(\mathbf{s}, \mathbf{t}_\sim) \end{cases} \right), \tag{3.4}$$

```
DELTA1(S[1..n], T[1..m])
W[0][0] ← |S[1] − T[1]|
W[0][1..m], W[1..n][0] ← ∞
for i ← 1 to n
    do for j ← 1 to m
           do μ ← min {W[i − 1][j], W[i − 1][j − 1], W[i][j − 1]}
              W[i][j] ← max {|A[i] − B[j]|, μ}
return W[n][m]
```

Figure 3.3: The algorithm computing δ_1

where $\delta_1([s_1], \mathbf{t}) = \delta_1(\mathbf{t}, [s_1]) = \max\{|s_1 - t_1|, \ldots, |s_1 - t_l|\}$ for the initial cases. $\delta_1(\mathbf{s}, \mathbf{t})$ can be computed in $O(\#\mathbf{s}\#\mathbf{t})$.

By Equation (3.4), one can be convinced that the distance function above is computed by the algorithm in Figure 3.3.

The following lemma will aid our further discussion. Via the lemma we will show, by a proof sketch motivated by examples, that the quantity in Equation (3.3) is equal to the quantity defined recursively in Equation (3.4). The proof of the lemma will be deferred until we end the proof sketch.

Lemma 1. *Let* $\mathbf{s}$ *and* $\mathbf{t}$ *be two finite sequences. If* $\mathbf{s}'$ *and* $\mathbf{t}'$ *are stretches of* $\mathbf{s}$ *and* $\mathbf{t}$ *respectively, i.e.* $\mathbf{s}' \in \mathcal{S}(\mathbf{s})$ *and* $\mathbf{t}' \in \mathcal{S}(\mathbf{t})$*, such that* $\#\mathbf{s}' = \#\mathbf{t}' > \#\mathbf{s} + \#\mathbf{t} - 1$*. Then there are* $\mathbf{s}''$ *and* $\mathbf{t}''$ *such that* $\mathbf{s}'' \in \mathcal{S}(\mathbf{s})$*,* $\mathbf{t}'' \in \mathcal{S}(\mathbf{t})$*,* $\#\mathbf{s}'' = \#\mathbf{t}'' < \#\mathbf{s}' = \#\mathbf{t}'$ *and* $\|\mathbf{s}'' - \mathbf{t}''\|_p \leq \|\mathbf{s}' - \mathbf{t}'\|_p$*. Where* $p \in [1, \infty]$*.*

The two quantities in Equations (3.3) and (3.4) will be called the LHS and the RHS respectively.

I) Similar to DTW, after solving for the RHS, we have an optimal *warping path*,

$$(i_1, j_1), (i_2, j_2), \ldots, (i_N, j_N) \ . \tag{3.5}$$

The restrictions of the optimal path is the same. The only difference is the associated cost. In this case the cost associated with the warping path is

$$\max_{k \in \{1,\ldots,N\}} \left| s_{i_k} - t_{j_k} \right| \ .$$

II) From the warping path we can construct two *stretched* time series whose lengths are equal and not greater than $\#\mathbf{s} + \#\mathbf{t} - 1$. Furthermore, their distance as computed by ℓ^∞ metric is equal to its associated cost. The two time series constructed are $[s_{i_1}, \ldots, s_{i_N}]$ and $[t_{j_1}, \ldots, t_{j_N}]$. For example, from the warping path in figure 2.1 we construct $[1, 1, 0, 0]$ and $[1, 2, 0, 0]$, which are stretches of $[1, 0, 0]$ and $[1, 2, 0]$ respectively.

III) Conversely, for each pair of *stretched* $\mathbf{s}$ and $\mathbf{t}$ whose lengths are equal and do not exceed $\#\mathbf{s} + \#\mathbf{t} - 1$, we can construct a warping path with the associated cost equal to the ℓ^∞ distance between those two stretched time series. For example, suppose $\mathbf{s}$ and $\mathbf{t}$ are $[1, 2, 0]$ and $[1, 0, 0]$, the stretched time series $\mathbf{s}'$ and $\mathbf{t}'$ are $[1, 2, 0, 0]$ and $[1, 1, 0, 0]$. Noticing that $\mathbf{s}'$ and $\mathbf{t}'$ are $[s_1, s_2, s_3, s_3]$ and $[t_1, t_1, t_2, t_3]$, we can construct the warping path $(1, 1), (2, 1), (3, 2), (3, 3)$ having the desired associated cost.

IV) Therefore the set of all possible costs associated with a warping path of $\mathbf{s}$ and $\mathbf{t}$ is the same as the set of all possible ℓ^∞ distance between $\mathbf{s}' \in \mathcal{S}(\mathbf{s})$ and $\mathbf{t}' \in \mathcal{S}(\mathbf{t})$ of equal lengths not exceeding $\#\mathbf{s} + \#\mathbf{t} - 1$. We conclude that RHS can be viewed as the minimum cost among all distances of such pair of stretched time series.

V) So by Lemma 1, we have LHS $\geq$ RHS. On the other hand, since LHS is an infimum taken over bigger set than that of RHS, LHS $\leq$ RHS. Hence LHS equals RHS.

Now we proof the lemma.

Proof of Lemma 1. We will prove by induction on the lengths of $\mathbf{s}$ and $\mathbf{t}$, the base case when $\#\mathbf{s}, \#\mathbf{t} \leq 2$ can be readily checked.

Assume that the statement holds for every **s** and **t** such that $\#\mathbf{s} \leq M-1$ and $\#\mathbf{t} \leq N$, it remains to show that the statement is true for any **u** and **t** such that $\#\mathbf{u} = M$ and $\#\mathbf{v} = N$.

Let $\mathbf{u}' \in \mathcal{S}(\mathbf{u}), \mathbf{v}' \in \mathcal{S}(\mathbf{v}), \#\mathbf{u}' = \#\mathbf{v}' > \#\mathbf{u} + \#\mathbf{v} - 1$. Then $\mathbf{u}'$ can be written as $\mathbf{u}'_\mathbf{h}\mathbf{u}'_\mathbf{t}$ where every element of $\mathbf{u}'_\mathbf{h}$ equals u_1 and the first element of $\mathbf{u}'_\mathbf{t}$ equals u_2. Since $\mathbf{v}'$ is a stretch of **v**, $\mathbf{v}'\langle\#\mathbf{u}'_\mathbf{h}\rangle$ must be some element in **v**, say v_k. There are two possibilities of $\mathbf{v}'\langle\#\mathbf{u}'_\mathbf{h}+1\rangle$. For the case $\mathbf{v}'\langle\#\mathbf{u}'_\mathbf{h}+1\rangle = v_k$, $\mathbf{u}'$ and $\mathbf{v}'$ are aligned as,

$$\mathbf{u}' = [\underbrace{u_1, \ldots, u_1}_{\mathbf{u}'_\mathbf{h}}, \underbrace{u_2, \ldots, u_M}_{\mathbf{u}'_\mathbf{t}}]$$
$$\mathbf{v}' = [\underbrace{v_1, \ldots, v_k}_{\mathbf{v}'_\mathbf{h}}, \underbrace{v_k, \ldots, v_N}_{\mathbf{v}'_\mathbf{t}}] \ .$$

Define $\mathbf{v}_\mathbf{h}$ as the the sequence $[v_1, v_2, \ldots, v_k]$ and $\mathbf{v}_\mathbf{t}$ as the tail of **v** from the k-th element onwards. One can check that $\mathbf{u}'_\mathbf{t} \in \mathcal{S}(\mathbf{u}_\sim)$ and $\mathbf{v}'_\mathbf{t} \in \mathcal{S}(\mathbf{v}_\mathbf{t})$. Since $\#\mathbf{u}_\sim \leq M-1$ and $\#\mathbf{v}_\mathbf{t} \leq N$, by the assumption, there are $\mathbf{u}''_\mathbf{t}$ and $\mathbf{v}''_\mathbf{t}$ such that

$$\begin{aligned}\#\mathbf{u}''_\mathbf{t} = \#\mathbf{v}''_\mathbf{t} &= \#\mathbf{u}_\sim + \#\mathbf{v}_\mathbf{t} - 1 \\ &= (\#\mathbf{u} - 1) + (\#\mathbf{v} - k + 1) - 1 \\ &= \#\mathbf{u} + \#\mathbf{v} - k - 1 \ ,\end{aligned}$$

and $\|\mathbf{u}''_\mathbf{t} - \mathbf{v}''_\mathbf{t}\|_p \leq \|\mathbf{u}'_\mathbf{t} - \mathbf{v}'_\mathbf{t}\|_p$. Write $\mathbf{u}'' = [u_1]^k\mathbf{u}''_\mathbf{t}$ and $\mathbf{v}'' = \mathbf{v}_\mathbf{h}\mathbf{v}''_\mathbf{t}$, it is easy to see that $\mathbf{u}'' \in \mathcal{S}(u)$ and $\mathbf{v}'' \in \mathcal{S}(v)$. Furthermore, $\mathbf{u}'' = \mathbf{v}'' = \#\mathbf{u} + \#\mathbf{v} - 1$ and

$$\begin{aligned}\|\mathbf{u}'' - \mathbf{v}''\|_p^p &= \|[u_1]^k - \mathbf{v}_\mathbf{h}\|_p^p + \|\mathbf{u}''_\mathbf{t} - \mathbf{v}''_\mathbf{t}\|_p^p \\ &\leq \|[u_1]^{\#\mathbf{u}'_\mathbf{h}} - \mathbf{v}'_\mathbf{h}\|_p^p + \|\mathbf{u}'_\mathbf{t} - \mathbf{v}'_\mathbf{t}\|_p^p \\ &= \|\mathbf{u}' - \mathbf{v}'\|_p^p \ ,\end{aligned}$$

for $p \in [1, \infty)$, and,

$$\begin{aligned} \|\mathbf{u}'' - \mathbf{v}''\|_\infty &= \max\left(\|[u_1]^k - \mathbf{v_h}\|_\infty, \|\mathbf{u}''_\mathbf{t} - \mathbf{v}''_\mathbf{t}\|_\infty\right) \\ &\le \max\left(\|[u_1]^{\#\mathbf{u}'_\mathbf{h}} - \mathbf{v}'_\mathbf{h}\|_\infty, \|\mathbf{u}'_\mathbf{t} - \mathbf{v}'_\mathbf{t}\|_\infty\right) \\ &= \|\mathbf{u}' - \mathbf{v}'\|_\infty \ . \end{aligned}$$

So $\|\mathbf{u}'' - \mathbf{v}''\|_p \le \|\mathbf{u}' - \mathbf{v}'\|_p$.

For the case $\mathbf{v}'\langle \#\mathbf{u}'_\mathbf{h} + 1\rangle = v_{k+1}$, we can proceed through a similar argument and have $\mathbf{u}''' \in \mathcal{S}(u)$ and $\mathbf{v}''' \in \mathcal{S}(v)$ whose lengths are equal to $\#\mathbf{u} + \#\mathbf{v} - 2$ and $\|\mathbf{u}''' - \mathbf{v}'''\|_p \le \|\mathbf{u}' - \mathbf{v}'\|_p$. ■

As a by-product of the previous discussion we have an alternative characterization of the DTW, for $p \in [1, \infty)$,

$$\mathrm{DTW}(\mathbf{x}, \mathbf{y}) = \inf_{\mu, \nu \in \mathcal{S}} \|\mu\mathbf{x} - \nu\mathbf{y}\|_p \ . \tag{3.6}$$

Where $\|\mathbf{x} - \mathbf{y}\|_p$ is implicitly defined to be ∞ when $\#\mathbf{x} \neq \#\mathbf{y}$.

Indeed, DTW is the condensation of ℓ^p metric wrt. the set of stretch operations $\mathcal{S}$, but since $\mathcal{S}$ does not preserve the ℓ^p metric for $p \in [1, \infty)$ Theorem 3 will not guarantee that DTW is subadditive and it is actually not subadditive as we have seen in Chapter 2.

The next example is the condensations of ℓ^p metrics wrt. *gap* insertions.

Definition 7. *Let γ be a real number and W be the set of sequences with constant tail γ, i.e. the sequences of the form $(x_1, \ldots, x_l, \gamma, \gamma, \ldots)$. Define the map $\iota_0 : W \to W$ by,*

$$\iota_0\mathbf{x} = (\gamma, x_1, x_2, \ldots) \ .$$

For $k \in \mathbb{N}$ define $\iota_k : W \to W$ by,

$$\iota_k\mathbf{x} = (x_1, \ldots, x_k, \gamma, x_{k+1}, \ldots) \ .$$

Precisely, $\iota_k\mathbf{x}\langle k+1\rangle = \gamma$, $\iota_k\mathbf{x}\langle i\rangle = \mathbf{x}\langle i\rangle$ for $1 \le i \le k$, and $\iota_k\mathbf{x}\langle j\rangle = \mathbf{x}\langle j-1\rangle$ for

$j \geq k+2$.

Let $\mathcal{I}_\gamma$ be the set of all finite compositions of operations in $\{\iota_k\}_{k\geq 0} \cup \{\mathbb{1}\}$. We sometimes write $\mathcal{I}_\gamma$ as $\mathcal{I}$ when there is no need to specify the value γ.

W can be thought of as either a set of infinite sequences with constant tails as in the definition above or as a set of finite sequences and let the distance between sequences of different lengths be ∞ and use the ℓ^p metric as the distance if the sequences are of the same length.

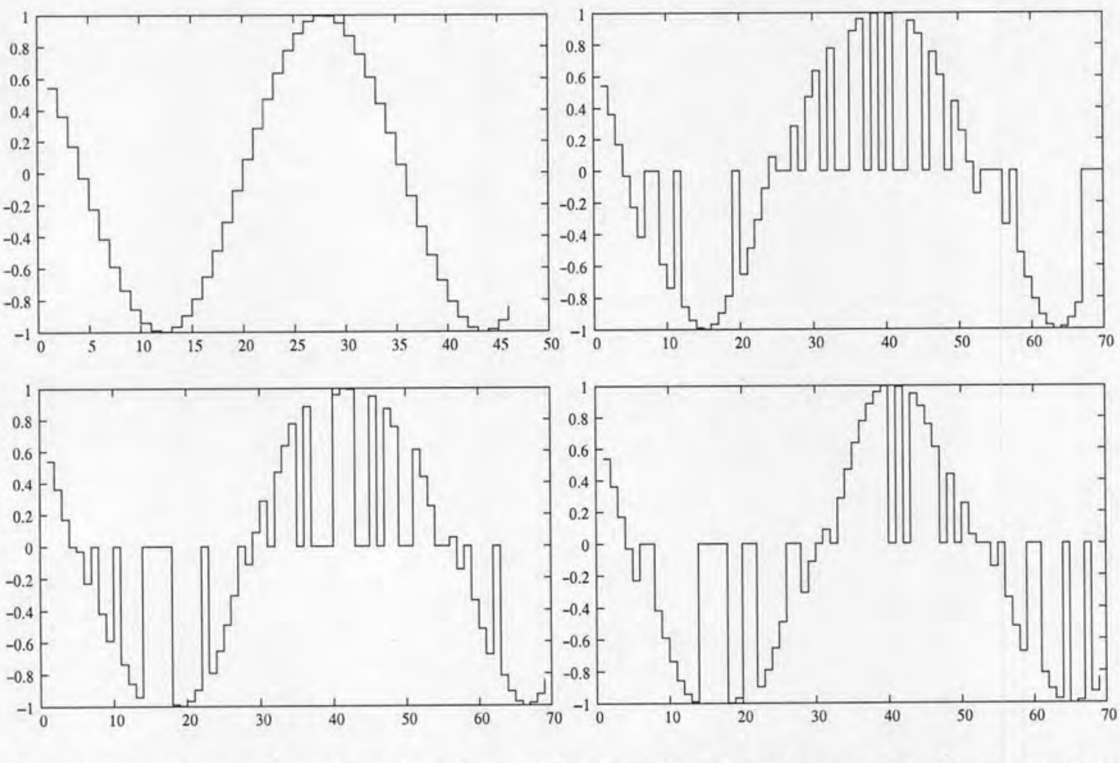

Figure 3.4: A visualization of a time series and its possible results after insertion operations with functions in the class $\mathcal{I}$. The original time series is the top-left. The rest are some of its possible results after gap insertions. The gap value is 0.

An illustration of gap insertions is in Figure 3.4. One can check that $\mathcal{I}$ is complete and it preserves ℓ^p metrics. Hence, the condensation of ℓ^p metrics wrt. $\mathcal{I}$,

$$\delta_2^p(\mathbf{x}, \mathbf{y}) = \inf_{\iota,\kappa \in \mathcal{I}} \|\iota\mathbf{x} - \kappa\mathbf{y}\|_p \ , \tag{3.7}$$

is a pseudometric. It can be computed by dynamic programming using the relation below,

$$\delta_2^p(\mathbf{s},\mathbf{t})^p = \min \begin{cases} |s_1 - t_1|^p + \delta_2^p(\mathbf{s}_\sim, \mathbf{t}_\sim)^p, \\ |\gamma - t_1|^p + \delta_2^p(\mathbf{s}, \mathbf{t}_\sim)^p, \\ |s_1 - \gamma|^p + \delta_2^p(\mathbf{s}_\sim, \mathbf{t})^p \ . \end{cases} \tag{3.8}$$

Where $\delta_2^p([], \mathbf{s})^p = \delta_2^p(\mathbf{s}, [])^p = \sum_{i=1}^{l} |\gamma - s_i|^p$ for the base cases. $\delta_2(\mathbf{s}, \mathbf{t})$ can be computed in $O(\#\mathbf{s}\#\mathbf{t})$.

The fact that the recurrence relation solves the distance defined in Equation (3.7) follows from an argument similar to that used to explain the case of δ_1. However, note that the recurrence relation computes $\inf_{\iota,\kappa\in\mathcal{I}} \|\iota\mathbf{x} - \kappa\mathbf{y}\|_p^p$. Since a^p is strictly increasing when a is nonnegative, it is possible to minimize $\|\iota\mathbf{x} - \kappa\mathbf{y}\|_p^p$ instead of $\|\iota\mathbf{x} - \kappa\mathbf{y}\|_p$. Indeed one can check that

$$\delta_2^p(\mathbf{x}, \mathbf{y}) = \inf_{\iota,\kappa\in\mathcal{I}} \|\iota\mathbf{x} - \kappa\mathbf{y}\|_p = \left\{ \inf_{\iota,\kappa\in\mathcal{I}} \|\iota\mathbf{x} - \kappa\mathbf{y}\|_p^p \right\}^{\frac{1}{p}} \ .$$

Mark that the value of p can be any real number, but δ_2 will be subadditive if $p \in [1, \infty)$ because then the ℓ^p distance will be subadditive and δ_2 will be a condensation of a metric. For the particular case of $p = 1$, this is the distance function known as ERP (Chen and Ng, 2004) that we mentioned in 2.9.

Consider the case when $p = 2$ and $\gamma = 0$, in computing the distance δ_2^2 between $\mathbf{x}$ and $\mathbf{y}$. The quantity $\|\iota\mathbf{x} - \kappa\mathbf{y}\|_2$ is minimized over all possible insertions ι and κ in $\mathcal{I}_0$. Recalling that the square function is strictly increasing when the domain is positive, it is possible to minimize $\|\iota\mathbf{x} - \kappa\mathbf{y}\|_2^2$ instead, and by the polarization identity,

$$\|\iota\mathbf{x} - \kappa\mathbf{y}\|_2^2 = \|\iota\mathbf{x}\|_2^2 + \|\kappa\mathbf{y}\|_2^2 - 2\iota\mathbf{x} \cdot \kappa\mathbf{y} \ .$$

Since $\gamma = 0$ gap insertions keep the norm of time series unchanged, i.e. $\iota\mathbf{x}$ for every ι in $\mathcal{I}_0$ and $\mathbf{x}$ in W. Therefore we have

$$\|\iota\mathbf{x} - \kappa\mathbf{y}\|_2^2 = \|\mathbf{x}\|_2^2 + \|\mathbf{y}\|_2^2 - 2\iota\mathbf{x} \cdot \kappa\mathbf{y} \ .$$

So minimizing the quantity above over all ι and κ in $\mathcal{I}_0$ is the same as maximizing the dot product $\iota\mathbf{x} \cdot \kappa\mathbf{y}$ in the right hand side of the above equation over all ι and κ in $\mathcal{I}_0$.

To summarize, for the case when $p = 2$ and $\gamma = 0$ the distance δ_2^2 between $\mathbf{x}$ and $\mathbf{y}$ can be viewed as the ℓ^2 distance between the *inserted* time series $\iota\mathbf{x}$ and $\kappa\mathbf{y}$ derived from $\mathbf{x}$ and $\mathbf{y}$ such that their similarity as measured by their dot product $\iota\mathbf{x} \cdot \kappa\mathbf{y}$ is maximized.

Inspired by the above discussion, we propose another subadditive condensation base on the idea of maximizing similarity. We first introduce the distance function

$$\angle(\mathbf{x}, \mathbf{y}) = \arccos\left(\frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2}\right),$$

whose geometric interpretation is the measure of the angle between two vectors. One can check that $\angle$ is a pseudometric.

Now we condense the distance function wrt. the *insertions* $\mathcal{I}_0$ defined above with the gap value $\gamma = 0$, giving the condensation

$$\begin{aligned}\delta_3(\mathbf{x}, \mathbf{y}) &= \inf_{\iota,\kappa \in \mathcal{I}_0} \angle(\iota\mathbf{x}, \kappa\mathbf{y}) \\ &= \inf_{\iota,\kappa \in \mathcal{I}_0} \arccos\left(\frac{\iota\mathbf{x} \cdot \kappa\mathbf{y}}{\|\iota\mathbf{x}\|_2 \|\kappa\mathbf{y}\|_2}\right) \\ &= \inf_{\iota,\kappa \in \mathcal{I}_0} \arccos\left(\frac{\iota\mathbf{x} \cdot \kappa\mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2}\right).\end{aligned}$$

The third equality follows from the fact that insertions of zero gaps preserves the norms.

When the gap value γ is zero, it can be checked that the insertions of gaps preserve the distance $\angle$. Hence it follows that δ_3 is a pseudometric.

For every real number $\alpha > 0$ and every pair of time series $\mathbf{s}$ and $\mathbf{s}'$ such that $\mathbf{s}' = \alpha\mathbf{s} = [\alpha s_1, \ldots, \alpha s_l]$ we have $\delta_3(\mathbf{s}, \mathbf{s}') = 0$. To justify this intuitively, if we have a time series $\mathbf{s}$ and another one with the same *shape* but of different scale, then they are not different when measured with the distance δ_3.

Next we briefly discuss a way to compute δ_3. First note that the arccos function is strictly decreasing, this fact can be used to show that

$$\delta_3(\mathbf{x},\mathbf{y}) = \cos\left(\sup_{\iota,\kappa\in\mathcal{I}_0} \frac{\iota\mathbf{x}\cdot\kappa\mathbf{y}}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2}\right) .$$

So one can maximize the quantity $\iota\mathbf{x}\cdot\kappa\mathbf{y}$ over all possible ι and κ in $\mathcal{I}_0$ instead. Writing the quantity $\sup_{\iota,\kappa\in\mathcal{I}_0}\iota\mathbf{x}\cdot\kappa\mathbf{y}$ as $\delta_{iii}(\mathbf{x},\mathbf{y})$, we can then compute the distance δ_3 by $\delta_3(\mathbf{x},\mathbf{y}) = \cos\left(\delta_{iii}(\mathbf{x},\mathbf{y})/\|\mathbf{x}\|_2\|\mathbf{y}\|_2\right)$.

δ_{iii} can be computed using the recurrence relation

$$\delta_{iii}(\mathbf{s},\mathbf{t}) = \max\begin{cases} s_1t_1 + \delta_{iii}(\mathbf{s}_\sim,\mathbf{t}_\sim), \\ \delta_{iii}(\mathbf{s},\mathbf{t}_\sim), \\ \delta_{iii}(\mathbf{s}_\sim,\mathbf{t}) \ . \end{cases}$$

Where $\delta_{iii}([\,],\mathbf{s}) = \delta_{iii}(\mathbf{s},[\,]) = 0$ for the base cases.

Again, $\delta_{iii}(\mathbf{s},\mathbf{t})$ can be computed in $O(\#\mathbf{s}\#\mathbf{t})$ time. Since δ_3 can be computed from δ_{iii} in constant time, δ_3 can also be computed in $O(\#\mathbf{s}\#\mathbf{t})$ time.

3.1.2 Interpolation of Time Series

If a condensation is well defined in the form

$$\Delta_{d,\mathcal{M}}(x,y) = \min_{\mu,\nu\in\mathcal{M}}\|\mu x - \nu y\|,$$

or, to put it differently, if the pair of morph operations yielding minimal distance always exists and the base distance is a norm metric, then we can do interpolation of objects in a certain way. δ_1 and δ_2 are examples of this type of condensation.

Proposition 1. *Let* $\mathbf{a},\mathbf{b}$ *be vectors in a vector space* V *with a norm* $\|\cdot\|$*. Let* $\mathcal{M}$ *be a complete set of morph operations on* V *preserving the* $\|\cdot\|$*-metric. If* $\Delta(\mathbf{x},\mathbf{y}) = \min_{\mu,\nu\in\mathcal{M}}\|\mu x - \nu y\|$ *is well defined, then for any* $\theta\in[0,1]$ *there is* $\mathbf{c}\in V$ *satisfying* $\Delta(\mathbf{a},\mathbf{c}) = \theta\,\Delta(\mathbf{a},\mathbf{b})$ *and* $\Delta(\mathbf{b},\mathbf{c}) = (1-\theta)\,\Delta(\mathbf{a},\mathbf{b})$.

Proof. By assumption there is $\mu_0, \nu_0 \in \mathcal{M}$ yielding $\|\mu_0 \mathbf{a} - \nu_0 \mathbf{b}\| = \Delta(\mathbf{a}, \mathbf{b})$. Write $\mathbf{a}' = \mu_0 \mathbf{a}$ and $\mathbf{b}' = \nu_0 \mathbf{b}$. Let $\mathbf{c} = (1-\theta)\mathbf{a}' + \theta \mathbf{b}'$, then

$$\Delta(\mathbf{a}, \mathbf{c}) \leq \|\mathbf{a}' - \mathbf{c}\| = \theta \|\mathbf{a}' - \mathbf{b}'\| = \theta\, \Delta(\mathbf{a}, \mathbf{b}) \ ,$$
$$\Delta(\mathbf{c}, \mathbf{b}) \leq \|\mathbf{c} - \mathbf{b}'\| = (1-\theta) \|\mathbf{b}' - \mathbf{a}'\| = (1-\theta)\, \Delta(\mathbf{a}, \mathbf{b}) \ .$$

Since Δ is subadditive (by Theorem 3), and by the two inequalities above,

$$\begin{aligned} \Delta(\mathbf{a}, \mathbf{b}) &\leq \Delta(\mathbf{a}, \mathbf{c}) + \Delta(\mathbf{c}, \mathbf{b}) \\ &\leq \theta\, \Delta(\mathbf{a}, \mathbf{b}) + (1-\theta)\, \Delta(\mathbf{a}, \mathbf{b}) = \Delta(\mathbf{a}, \mathbf{b}) \ . \end{aligned}$$

This implies that $\Delta(\mathbf{a}, \mathbf{c}) \geq \theta\, \Delta(\mathbf{a}, \mathbf{b})$ and $\Delta(\mathbf{b}, \mathbf{c}) \geq (1-\theta)\, \Delta(\mathbf{a}, \mathbf{b})$. Together with the first two inequalities, we conclude that $\Delta(\mathbf{a}, \mathbf{c}) = \theta\, \Delta(\mathbf{a}, \mathbf{b})$ and $\Delta(\mathbf{b}, \mathbf{c}) = (1-\theta)\, \Delta(\mathbf{a}, \mathbf{b})$. ■

The proposition says that a way to interpolate between two time series **s** and **t** wrt. $\Delta_{d,\mathcal{M}}$ is by doing linear interpolation between the closest pair of time series among all possible pairs such that one of the pair can be morphed from **s** and the other can be morphed from **t**. Figure 3.5 shows an example of interpolations between two time series wrt. the distance δ_2.

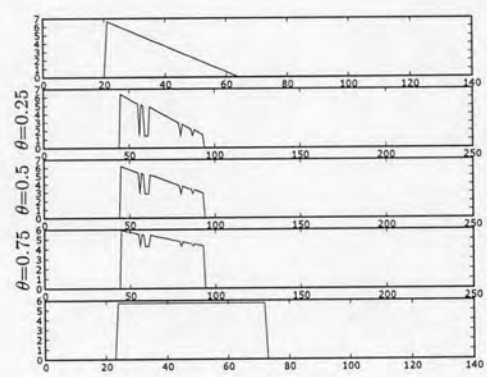

Figure 3.5: The interpolation between two time series wrt. δ_2^1. The value γ is set to zero. When θ is closer to 1 the interpolated time series is closer to the bottom time series when measured with δ_2^1

3.2 Shortcut Distance

Lemma 2. *Given a set Ω and a nonnegative function $d : \Omega \times \Omega \to [0, \infty]$ such that $d(x, x) = 0$ for every x in Ω. The function $\Xi_d : \Omega \times \Omega \to [0, \infty]$ defined by,*

$$\Xi_d(x, y) = \inf \left\{ \sum_{i=1}^{N} d(x_{i-1}, x_i) \middle| x_0, \ldots, x_n \in \Omega, N \in \mathbb{N}, x_0 = x, x_n = y \right\}$$

is subadditive. Ξ_d is called the shortcut *of d.*

Proof. Let p, q, r be any points in Ω. For a fixed $\varepsilon > 0$, by definition there are $\{p = p_0, \ldots, p_M = q = q_0, \ldots, r = q_N\} \subseteq \Omega$ such that,

$$\Xi_d(p, q) + \Xi_d(q, r) + \varepsilon \geq \sum_{i=1}^{M} d(p_{i-1}, p_i) + \sum_{j=1}^{N} d(q_{j-1}, q_j) \ .$$

Rename the points $q_1, \ldots, q_N$ to $p_{M+1}, \ldots, p_{M+N}$ respectively. Noting that $d(p_M, q_0) = 0$, we can write the above inequality as,

$$\Xi_d(p, q) + \Xi_d(q, r) + \varepsilon \geq \sum_{i=1}^{M+N} d(p_{i-1}, p_i) \geq \Xi_d(p, r) \ .$$

The second inequality holds by definition. Since this is true for arbitrary $\varepsilon > 0$ we conclude that Ξ_d is subadditive. ∎

One can check that if d is symmetric, then Ξ_d is also symmetric. By the lemma that we can always construct a pseudometric from a symmetric distance function by defining its shortcut. Even if the distance function d we have is not symmetric, we can create a symmetric distance function first and use the new symmetric distance. For example, the functions $(x, y) \mapsto \frac{1}{2}d(x, y) + \frac{1}{2}d(y, x)$ and $(x, y) \mapsto \max\{d(x, y), d(y, x)\}$ are always symmetric, and they will be subadditive if d is subadditive. In fact, however, symmetry is not a vital property because one can do nearest neighbor anyway by explicitly specifying the direction when comparing distances.

Note that a subadditive distance function is always a shortcut of some distance function since it is always a shortcut of itself. Together with the lemma this suggests a vague intuition that every metric measures the minimum cost of finite gradual changes from one object to another. In a sense, the shortcut is the length of the *shortest path* between objects, and for subadditive distances, the shortest path is always the *direct* one.

The question of what is the shortcut of DTW is an open question; once the shortcut of DTW is known, the way to efficiently compute it and its classification performance are the next to be enquired.

3.2.1 Examples

By inspecting the expression in Equation (3.8) of δ_2 (ERP) with the assumption that γ in the equation is zero and the base metric is the ℓ^1 metric, it can be shown that $\delta_2^1(\mathbf{x}, \mathbf{y})$ is the minimum cost of the sequence of the following transformations leading $\mathbf{x}$ to $\mathbf{y}$,

- delete x_i from $\mathbf{x}$, resulting a shorter time series, costs $|x_i|$,
- insert a number r into $\mathbf{x}$, resulting a longer time series, costs $|r|$,
- change the value of x_i to v, costs $|x_i - v|$.

The right hand side of Equation (3.8) chooses the minimum among three choices. We can think of the first choice as the cost we have to pay in order to transform $\mathbf{s}$ to $\mathbf{t}$ by changing the value s_1 of the first element of $\mathbf{s}$ to match t_1 and do the best we can for the rest. The second choice is the lowest cost we need to pay if we insert t_1 at the head of $\mathbf{s}$ first. The third choice is for the case of deleting the first element of $\mathbf{s}$ first.

Observe that the minimum cost of transformations from $\mathbf{x}$ to $\mathbf{y}$ is equal to the cost of transformations from $\mathbf{y}$ to $\mathbf{x}$. This is because there is a sequence of transformations from $\mathbf{y}$ to $\mathbf{x}$ with cost c for each sequence that transform $\mathbf{x}$ to $\mathbf{y}$ whose cost is also c. Suppose we have a sequence of transformations making $\mathbf{x}$ become $\mathbf{y}$, we can reverse the order of that sequence and substitute each deletion with an appropriate insertion and vice versa as well as reversing the changes of values. The new sequence of transformations will have equal cost and change **y** to **x**.

For $\pi > 0$, the minimum cost of the sequence of the following morphs leading **x** to **y** is another pseudometric

- delete x_i from **x**, costs $|x_i| + \pi$
- insert a number r into **x**, costs $|r| + \pi$
- change the value of x_i to v, costs $|x_i - v|$

The distance above is more general than the ERP distance because it reduces to ERP when π is zero. The value π can be thought of as the penalty needed to be paid for each insertion, and deletion.

The minimum cost can be found by dynamic programming using the relation below.

$$\delta_4(\mathbf{s},\mathbf{t}) = \min \begin{cases} |s_1 - t_1| + \delta_4(\mathbf{s}_\sim, \mathbf{t}_\sim), \\ \pi + |t_1| + \delta_4(\mathbf{s}, \mathbf{t}_\sim), \\ \pi + |s_1| + \delta_4(\mathbf{s}_\sim, \mathbf{t}) \ . \end{cases}$$

Where $\delta_4([], \mathbf{s}) = \delta_4(\mathbf{s}, []) = l\pi + \sum_{i=1}^{l} |s_i|$ for the initial cases. $\delta_4(\mathbf{s},\mathbf{t})$ can be computed in $O(\#\mathbf{s}\#\mathbf{t})$ time.

In fact we can have a more general form

$$\delta_4^p(\mathbf{s},\mathbf{t})^p = \min \begin{cases} |s_1 - t_1|^p + \delta_4(\mathbf{s}_\sim, \mathbf{t}_\sim)^p, \\ \pi + |t_1|^p + \delta_4(\mathbf{s}, \mathbf{t}_\sim)^p, \\ \pi + |s_1|^p + \delta_4(\mathbf{s}_\sim, \mathbf{t})^p \ . \end{cases}$$

Where $\delta_4^p([], \mathbf{s}) = \delta_4^p(\mathbf{s}, []) = \left(l\pi + \sum_{i=1}^{l} |s_i|^p\right)^{1/p}$ for the initial cases.

We can show that $\delta_4^p(\mathbf{x}, \mathbf{y})$ computes the p-th root of the minimum cost of the sequence of the following transformations leading $\mathbf{x}$ to $\mathbf{y}$,

- delete x_i from $\mathbf{x}$, costs $|x_i|^p + \pi$,
- insert a number r into $\mathbf{x}$, costs $|r|^p + \pi$,
- change the value of x_i to v, costs $|x_i - v|^p$.

Using the fact that $(a+b)^{1/p} \leq a^{1/p} + b^{1/p}$ for $a, b \geq 0$ and $p \geq 1$, one can check that the distance δ_4^p is subadditive for every $p \geq 1$.

As a final remark, we note again that the argument that the set of all time series with all rational values is a countable dense subset will suffice to establish that all of the pseudometric spaces of time series in this chapter are separable. Consequently, Theorem 1 about asymptotic properties applies to k-NN in these pseudometrics.