# A Kernel for Hierarchical Parameter Spaces

Frank Hutter and Michael A. Osborne
fh@informatik.uni-freiburg.de and mosb@robots.ox.ac.uk

September 7, 2013

**Abstract**

We define a family of kernels for mixed continuous/discrete hierarchical parameter spaces and show that they are positive definite.

## 1 Introduction

We aim to do inference about some function $g$ with domain (input space) $\mathcal{X}$. $\mathcal{X} = \prod_{i=1}^{D} \mathcal{X}_i$ is a $D$-dimensional input space, where each individual dimension is either bounded real or categorical, that is, $\mathcal{X}_i$ is either $[l_i, u_i] \subset \mathbb{R}$ (with lower and upper bounds $l_i$ and $u_i$, respectively) or $\{v_{i,1}, \ldots, v_{i,m_i}\}$.

Associated with $\mathcal{X}$, there is a DAG structure $\mathcal{D}$, whose vertices are the dimensions $\{1, \ldots, D\}$. $\mathcal{X}$ will be restricted by $\mathcal{D}$: if vertex $i$ has children under $\mathcal{D}$, $\mathcal{X}_i$ must be categorical. $\mathcal{D}$ is also used to specify when each input is *active* (that is, relevant to inference about $g$). In particular, we assume each input dimension is only active under some instantiations of its ancestor dimensions in $\mathcal{D}$. More precisely, we define $D$ functions $\delta_i \colon \mathcal{X} \to \mathcal{B}$, for $i \in \{1, \ldots, D\}$, and where $\mathcal{B} = \{\text{true}, \text{false}\}$. We take

$$\delta_i(\underline{x}) = \delta_i\big(\underline{x}(\text{anc}_i)\big), \tag{1}$$

where $\text{anc}_i$ are the ancestor vertices of $i$ in $\mathcal{D}$, such that $\delta_i(\underline{x})$ is true only for appropriate values of those entries of $\underline{x}$ corresponding to ancestors of $i$ in $\mathcal{D}$. We say $i$ is active for $\underline{x}$ iff $\delta_i(\underline{x})$.

Our aim is to specify a kernel for $\mathcal{X}$, *i.e.*, a positive semi-definite function $k \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. We will first specify an individual kernel for each input dimension, *i.e.*, a positive semi-definite function $k_i \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. $k$ can then be taken as either a sum,

$$k(\underline{x}, \underline{x}') = \sum_{i=1}^{D} k_i(\underline{x}, \underline{x}'), \tag{2}$$

product,

$$k(\underline{x}, \underline{x}') = \prod_{i=1}^{D} k_i(\underline{x}, \underline{x}'), \tag{3}$$

or any other permitted combination, of these individual kernels. Note that each individual kernel $k_i$ will depend on an input vector $\underline{x}$ only through dependence on $x_i$ and $\delta_i(\underline{x})$,

$$k_i(\underline{x}, \underline{x}') = \tilde{k}_i\big(x_i, \delta_i(\underline{x}), x_i', \delta_i(\underline{x}')\big). \tag{4}$$

That is, $x_j$ for $j \neq i$ will influence $k_i(\underline{x}, \underline{x}')$ only if $j \in \mathrm{anc}_i$, and only by affecting whether $i$ is active.

Below we will construct pseudometrics $d_i \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+$: that is, $d_i$ satisfies the requirements of a metric aside from the identity of indiscernibles. As for $k_i$, these pseudometrics will depend on an input vector $\underline{x}$ only through dependence on both $x_i$ and $\delta_i(\underline{x})$. $d_i(\underline{x}, \underline{x}')$ will be designed to provide an intuitive measure of how different $g(\underline{x})$ is from $g(\underline{x}')$. For each $i$, we will then construct a (pseudo-)isometry $f_i$ from $\mathcal{X}$ to a Euclidean space ($\mathbb{R}^2$ for bounded real parameters, and $\mathbb{R}^m$ for categorical-valued parameters with $m$ choices). That is, denoting the Euclidean metric on the appropriate space as $d_E$, $f_i$ will be such that

$$d_i(\underline{x}, \underline{x}') = d_{\mathrm{E}}\big(f_i(\underline{x}), f_i(\underline{x}')\big) \tag{5}$$

for all $\underline{x}, \underline{x}' \in \mathcal{X}$. We can then use our transformed inputs, $f_i(\underline{x})$, within any standard Euclidean kernel $\kappa$. We'll make this explicit in Proposition 2.

**Definition 1.** *A function* $\kappa \colon \mathbb{R}^+ \to \mathbb{R}$ *is* a positive semi-definite covariance function over Euclidean space *if* $K \in \mathbb{R}^{N\times N}$*, defined by*

$$K_{m,n} = \kappa\big(d_{\mathrm{E}}(\underline{y}_m, \underline{y}_n)\big), \quad \textit{for } \underline{y}_m, \underline{y}_n \in \mathbb{R}^P, \quad m, n = 1, \ldots, N,$$

*is positive semi-definite for any* $\underline{y}_1, \ldots, \underline{y}_N \in \mathbb{R}^P$.

A popular example of such a $\kappa$ is the exponentiated quadratic, for which $\kappa(\delta) = \sigma^2 \exp(-\frac{1}{2}\frac{\delta^2}{\lambda^2})$; another popular choice is the rational quadratic, for which $\kappa(\delta) = \sigma^2(1 + \frac{1}{2\alpha}\frac{\delta^2}{\lambda^2})^{-\alpha}$.

**Proposition 2.** *Let* $\kappa$ *be a positive semi-definite covariance function over Euclidean space and let* $d_i$ *satisfy Equation 5. Then,* $k_i \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+$*, defined by*

$$k_i(\underline{x}, \underline{x}') = \kappa\big(d_i(\underline{x}, \underline{x}')\big)$$

*is a positive semi-definite covariance function over input space* $\mathcal{X}$.

*Proof.* We need to show that for any $\underline{x}_1, \ldots, \underline{x}_N \in \mathcal{X}$, $K \in \mathbb{R}^{N\times N}$ defined by

$$K_{m,n} = \kappa\big(d_i(\underline{x}_m, \underline{x}_n)\big), \quad \text{for } \underline{x}_m, \underline{x}_n \in \mathcal{X}, \quad m, n = 1, \ldots, N,$$

is positive semi-definite. Now, by the definition of $d_i$,

$$K_{m,n} = \kappa\Big(d_{\mathrm{E}}\big(f_i(\underline{x}_m), f_i(\underline{x}_n)\big)\Big) = \kappa\big(d_{\mathrm{E}}(\underline{y}_m, \underline{y}_n)\big)$$

where $\underline{y}_m = f_i(\underline{x}_m)$ and $\underline{y}_n = f_i(\underline{x}_n)$ are elements of $\mathbb{R}^P$. Then, by assumption that $\kappa$ is a positive semi-definite covariance function over Euclidean space, $K$ is positive semi-definite. $\square$

We'll now define pseudometrics $d_i$ and associated isometries $f_i$ for both the bounded real and categorical cases.

## 2 Bounded Real Dimensions

Let's first focus on a bounded real input dimension $i$, i.e., $\mathcal{X}_i = [l_i, u_i]$. To emphasize that we're in this real case, we explicitly denote the pseudometric as $d_i^{\mathrm{r}}$ and the (pseudo-)isometry from $(\mathcal{X}, d_i)$ to $\mathbb{R}^2, d_{\mathrm{E}}$ as $f_i^{\mathrm{r}}$. For the definitions, recall that $\delta_i(\underline{x})$ is true iff dimension $i$ is active given the instantiation of $i$'s ancestors in $\underline{x}$.

$$d_i^{\mathrm{r}}(\underline{x}, \underline{x}') = \begin{cases} 0 & \text{if } \delta_i(\underline{x}) = \delta_i(\underline{x}') = \text{false} \\ \omega_i & \text{if } \delta_i(\underline{x}) \neq \delta_i(\underline{x}') \\ \omega_i \sqrt{2} \sqrt{1 - \cos(\pi \rho_i \frac{x_i - x_i'}{u_i - l_i})} & \text{if } \delta_i(\underline{x}) = \delta_i(\underline{x}') = \text{true.} \end{cases}$$

$$f_i^{\mathrm{r}}(\underline{x}) = \begin{cases} [0, 0]^{\intercal} & \text{if } \delta_i(\underline{x}) = \text{ false} \\ \omega_i [\sin \pi \rho_i \frac{x_i}{u_i - l_i}, \cos \pi \rho_i \frac{x_i}{u_i - l_i}]^{\intercal} & \text{otherwise.} \end{cases} .$$

Although our formal arguments do not rely on this, Proposition 5 in the appendix shows that $d_i^{\mathrm{r}}$ is a pseudometric. This pseudometric is defined by two parameters: $\omega_i \in [0, 1]$ and $\rho_i \in [0, 1]$. We firstly define

$$\omega_i = \prod_{j \in \mathrm{anc}_i \cup \{i\}} \gamma_j, \tag{6}$$

where $\gamma_j \in [0, 1]$. This encodes the intuitive notion that differences on lower levels of the hierarchy count less than differences in their ancestors.

Also note that, as desired, if $i$ is inactive for both $\underline{x}$ and $\underline{x}'$, $d_i^{\mathrm{r}}$ specifies that $g(\underline{x})$ and $g(\underline{x}')$ should not differ owing to differences between $x_i$ and $x_i'$. Secondly, if $i$ is active for both $\underline{x}$ and $\underline{x}'$, the difference between $g(\underline{x})$ and $g(\underline{x}')$ due to $x_i$ and $x_i'$ increases monotonically with increasing $|x_i - x_i'|$. Parameter $\rho_i$ controls whether differing in the activity of $i$ contributes more or less to the distance than differing in $x_i$ should $i$ be active. If $\rho = 1/3$, and if $i$ is inactive for exactly one of $\underline{x}$ and $\underline{x}'$, $g(\underline{x})$ and $g(\underline{x}')$ are as different as is possible due to dimension $i$; that is, $g(\underline{x})$ and $g(\underline{x}')$ are exactly as different in that case as if $x_i = l_i$ and $x_i' = u_i$. For $\rho > 1/3$, $i$ being active for both $\underline{x}$ and $\underline{x}'$ means that $g(\underline{x})$ and $g(\underline{x}')$ could potentially be more different than if $i$ was active in only one of them. For $\rho < 1/3$, the converse is true.[1]

We now show that $d_i^{\mathrm{r}}$ and $f_i^{\mathrm{r}}$ can be plugged into a positive semi-definite kernel over Euclidean space to define a valid kernel over space $\mathcal{X}$.

**Proposition 3.** *Let $\kappa$ be a positive semi-definite covariance function over Euclidean space. Then, $k_i \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+$, defined by*

$$k_i(\underline{x}, \underline{x}') = \kappa\big(d_i^{\mathrm{r}}(\underline{x}, \underline{x}')\big)$$

*is a positive semi-definite covariance function over input space $\mathcal{X}$.*

[1] Note that $\underline{x}$ and $\underline{x}'$ must differ in at least one ancestor dimension of $i$ in order for $\delta_i(\underline{x}) \neq \delta_i(\underline{x}')$ to hold, such that in the final kernel combining kernels $k_i$ due to each dimension $i$, differences in the activity of dimension $i$ are penalized both in kernel $k_i$ and in the distance for the kernel of the ancestor dimension causing the difference in $i$'s activity.

*Proof.* Due to Proposition 2, we only need to show that, for any two inputs $\underline{x}, \underline{x}' \in \mathcal{X}$, the isometry condition $d_{\text{E}}\big(f_i^{\text{r}}(\underline{x}), f_i^{\text{r}}(\underline{x}')\big) = d_i^{\text{r}}(\underline{x}, \underline{x}')$ holds.

We use the abbreviation $\alpha = \pi\rho_i \frac{x_i}{u_i - l_i}$ and $\alpha' = \pi\rho_i \frac{x_i'}{u_i - l_i}$ and consider the following three possible cases of dimension $i$ being active or inactive in $\underline{x}$ and $\underline{x}'$.

Case 1: $\delta_i(\underline{x}) = \delta_i(\underline{x}') =$ false. In this case, we trivially have

$$d_{\text{E}}(f_i^{\text{r}}(\underline{x}), f_i^{\text{r}}(\underline{x}')) = d_{\text{E}}([0,0]^{\intercal}, [0,0]^{\intercal}) = 0 = d_i^{\text{r}}(\underline{x}, \underline{x}').$$

Case 2: $\delta_i(\underline{x}) \neq \delta_i(\underline{x}')$. In this case, we have

$$d_{\text{E}}(f_i^{\text{r}}(\underline{x}), f_i^{\text{r}}(\underline{x}')) = d_{\text{E}}([\sin\alpha, \cos\alpha]^{\intercal}, [0,0]^{\intercal}) = \sqrt{\omega_i^2(\sin^2\alpha + \cos^2\alpha)} = \omega_i = d_i^{\text{r}}(\underline{x}, \underline{x}'),$$

and symmetrically for $d_{\text{E}}([0,0]^{\intercal}, [\sin\alpha, \cos\alpha]^{\intercal})$.

Case 3: $\delta_i(\underline{x}) = \delta_i(\underline{x}') =$ true. We have:

$$\begin{aligned}
d_{\text{E}}(f_i^{\text{r}}(\underline{x}), f_i^{\text{r}}(\underline{x}')) &= d_{\text{E}}(\omega_i[\sin\alpha, \cos\alpha]^{\intercal}, \omega_i[\sin\alpha', \cos\alpha']^{\intercal}) \\
&= \omega_i\sqrt{(\sin\alpha - \sin\alpha')^2 + (\cos\alpha - \cos\alpha')^2} \\
&= \omega_i\sqrt{\sin^2\alpha - 2\sin\alpha\sin\alpha' + \sin^2\alpha' + \cos^2\alpha - 2\cos\alpha\cos\alpha' + \cos^2\alpha'} \\
&= \omega_i\sqrt{(\sin^2\alpha + \cos^2\alpha) + (\sin^2\alpha' + \cos^2\alpha') - 2(\sin\alpha\sin\alpha' + \cos\alpha\cos\alpha')} \\
&= \omega_i\sqrt{1 + 1 - 2\cos(\alpha - \alpha')} \qquad (7) \\
&= \omega_i\sqrt{2}\sqrt{1 - \cos(\pi\rho_i\frac{x_i - x_i'}{u_i - l_i})} = d_i^{\text{r}}(\underline{x}, \underline{x}'),
\end{aligned}$$

where (7) follows from the previous line by using the identity

$$\cos(a - b) = \cos a \cos b + \sin a \sin b.$$

□

# 3 Categorical Dimensions

Now let's define $f_i^{\text{c}}$ and $d_i^{\text{c}}$ for the case that the input $\mathcal{X}_i = \{v_{i,1}, \ldots, v_{i,m_i}\}$ is categorical with $m_i$ possible values. Proceeding as above, we define a pseudometric $d_i^{\text{c}}$ on $\mathcal{X}$ and an isometry from $(\mathcal{X}, d_i^{\text{c}})$ to $(\mathbb{R}^{m_i}, d_{\text{E}}^{m_i})$, and show that we can combine these with a kernel over Euclidean space to construct a valid kernel over space $\mathcal{X}$.

$$d_i^{\text{c}}(\underline{x}, \underline{x}') = \begin{cases} 0 & \text{if } \delta_i(\underline{x}) = \delta_i(\underline{x}') = \text{false} \\ \omega_i & \text{if } \delta_i(\underline{x}) \neq \delta_i(\underline{x}') \\ \omega_i \frac{\sqrt{2}\rho}{1 + (m_i - 1)(1-\rho)^2} \mathbb{I}_{x_i \neq x_i'} & \text{if } \delta_i(\underline{x}) = \delta_i(\underline{x}') = \text{true.} \end{cases}$$

$$f_i^{\mathrm{c}}(\underline{x}) \quad = \quad \begin{cases} \underline{0} \in \mathbb{R}^{m_i} & \text{if } \delta_i(\underline{x}) = \text{ false} \\ \omega_i \, \frac{\underline{e_j} + (1-\rho)\sum_{l \neq j} \underline{e_l}}{\sqrt{1+(m_i-1)(1-\rho)^2}} & \text{if } \delta_i(\underline{x}) = \text{ true and } x_i = v_{i,j}, \end{cases}$$

where $\underline{e_j} \in \mathbb{R}^{m_i}$ is the $j$th unit vector: zero in all dimensions except $j$, where it is $1$. Note that

$$\sqrt{1 + (m_i - 1)(1 - \rho)^2} = \left\| \underline{e_j} + (1 - \rho) \sum_{l \neq j} \underline{e_l} \right\|. \tag{8}$$

Again, although our analysis does not require it, we prove in Proposition 6 (see appendix) that $d_i^{\mathrm{c}}$ is a pseudometric. Our pseudometric is again defined by two hyperparameters. Firstly, $\omega_i \in [0, 1]$ is exactly as defined in (6), and similarly allows higher-level inputs to attain greater importance. Similarly, $\rho_i \in [0, 1]$ allows control of to what extent differing in the activity of $i$ affects the distance relative to the influence of differing in $x_i$ should $i$ be active. In particular, for

$$\rho_i^* = \frac{\sqrt{2} - 2 + 2m_i - \sqrt{6 - 4\sqrt{2} + 4(\sqrt{2} - 1)m_i}}{2(m_i - 1)}, \tag{9}$$

$\rho_i < \rho_i^*$ implies that differing in the activity of $i$ is more significant, whereas $\rho_i > \rho_i^*$ implies the converse. The special case $\rho_i = 0$ dictates that differing in $x_i$ has no influence on the distance; $\rho_i = 1$ assigns maximal importance to differing in $x_i$.

**Proposition 4.** *Let $\kappa$ be a positive semi-definite covariance function over Euclidean space. Then, $k_i \colon \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+$, defined by*

$$k_i(\underline{x}, \underline{x}') = \kappa\big(d_i^{\mathrm{c}}(\underline{x}, \underline{x}')\big)$$

*is a positive semi-definite covariance function over input space $\mathcal{X}$.*

*Proof.* We proceed as in the proof of Proposition 3 to show that, for any two inputs $\underline{x}, \underline{x}' \in \mathcal{X}$, the isometry condition $d_{\mathrm{E}}^{m_i}(f_i^{\mathrm{c}}(\underline{x}), f_i^{\mathrm{c}}(\underline{x}')) = d_i^{\mathrm{c}}(\underline{x}, \underline{x}')$ holds.

Case 1: $\delta_i(\underline{x}) = \delta_i(\underline{x}') =$ false. In this case, we trivially have

$$d_{\mathrm{E}}^{m_i}(f_i^{\mathrm{r}}(\underline{x}), f_i^{\mathrm{r}}(\underline{x}')) = d_{\mathrm{E}}^{m_i}(\underline{0}, \underline{0}) = 0 = d_i^{\mathrm{r}}(\underline{x}, \underline{x}').$$

Case 2: $\delta_i(\underline{x}) \neq \delta_i(\underline{x}')$. In this case, we have

$$d_{\mathrm{E}}^{m_i}(f_i^{\mathrm{c}}(\underline{x}), f_i^{\mathrm{c}}(\underline{x}')) = d_{\mathrm{E}}^{m_i}\left(\omega_i \, \frac{\underline{e_j} + (1-\rho)\sum_{l \neq j} \underline{e_l}}{\|\underline{e_j} + (1-\rho)\sum_{l \neq j} \underline{e_l}\|}, \underline{0}\right) = \omega_i = d_i(\underline{x}, \underline{x}'),$$

and symmetrically for $d_{\mathrm{E}}\left(\underline{0}, \omega_i \, \frac{\underline{e_j} + (1-\rho)\sum_{l \neq j} \underline{e_l}}{\|\underline{e_j} + (1-\rho)\sum_{l \neq j} \underline{e_l}\|}\right)$.

Case 3: $\delta_i(\underline{x}) = \delta_i(\underline{x}') =$ true. If $x_i = x_i' = v_{i,j}$, we have

$$d_{\mathrm{E}}^{m_i}(f_i^{\mathrm{c}}(\underline{x}), f_i^{\mathrm{c}}(\underline{x}')) \quad = \quad d_{\mathrm{E}}^{m_i}\big(f_i^{\mathrm{c}}(\underline{x}), f_i^{\mathrm{c}}(\underline{x})\big) = 0 = d_i^{\mathrm{c}}(\underline{x}, \underline{x}').$$

If $x_i = v_{i,j} \neq v_{i,j'} = x'_i$, we have

$$\begin{aligned}
d_{\mathrm{E}}(f_i^{\mathrm{c}}(\underline{x}), f_i^{\mathrm{c}}(\underline{x}')) &= d_{\mathrm{E}}^{m_i}\left(\omega_i \, \frac{\underline{e_j} + (1-\rho)\sum_{l\neq j} \underline{e_l}}{\sqrt{1+(m_i-1)(1-\rho)^2}}, \, \omega_i \, \frac{\underline{e'_j} + (1-\rho)\sum_{l\neq j'} \underline{e_l}}{\sqrt{1+(m_i-1)(1-\rho)^2}}\right) \\
&= \omega_i \frac{\sqrt{\big(1-(1-\rho)\big)^2 + \big(1-(1-\rho)\big)^2}}{1+(m_i-1)(1-\rho)^2} \\
&= \omega_i \frac{\sqrt{2}\rho}{1+(m_i-1)(1-\rho)^2} \\
&= d_i^{\mathrm{c}}(\underline{x}, \underline{x}'). \qquad (10)
\end{aligned}$$

□

# A Proof of pseudometric properties

**Proposition 5.** *$d_i^{\mathrm{r}}$ is a pseudometric on $\mathcal{X}$.*

*Proof.* The non-negativity and symmetry of $d_i^{\mathrm{r}}$ are trivially proven. To prove the triangle inequality, consider $\underline{x}, \underline{x}', \underline{x}'' \in \mathcal{X}$.

Case 1: $\delta_i(\underline{x}) = \delta_i(\underline{x}') =$ false, such that $d_i^{\mathrm{r}}(\underline{x}, \underline{x}') = 0$. Here, from non-negativity, clearly $d_i^{\mathrm{r}}(\underline{x}, \underline{x}') = 0 \leq d_i^{\mathrm{r}}(\underline{x}, \underline{x}'') + d_i^{\mathrm{r}}(\underline{x}', \underline{x}'')$.

Case 2: $\delta_i(\underline{x}) \neq \delta_i(\underline{x}')$, such that such that $d_i^{\mathrm{r}}(\underline{x}, \underline{x}') = \omega_i$. Without loss of generality, assume $\delta_i(\underline{x}) =$ true, $\delta_i(\underline{x}') =$ false and $\delta_i(\underline{x}'') =$ true.

$$d_i^{\mathrm{r}}(\underline{x}, \underline{x}'') + d_i^{\mathrm{r}}(\underline{x}', \underline{x}'') = d_i^{\mathrm{r}}(\underline{x}, \underline{x}'') + \omega_i \qquad (11)$$

Hence $d_i^{\mathrm{r}}(\underline{x}, \underline{x}'') + d_i^{\mathrm{r}}(\underline{x}', \underline{x}'') \geq \omega_i = d_i^{\mathrm{r}}(\underline{x}, \underline{x}')$ by non-negativity.

Case 3: $\delta_i(\underline{x}) = \delta_i(\underline{x}') =$ true, such that $d_i^{\mathrm{r}}(\underline{x}, \underline{x}') = \omega_i\sqrt{2}\sqrt{1-\cos(\pi\rho_i \frac{x_i - x'_i}{u_i - l_i})}$. If $\delta_i(\underline{x}'') =$ false,

$$d_i^{\mathrm{r}}(\underline{x}, \underline{x}'') + d_i^{\mathrm{r}}(\underline{x}', \underline{x}'') = 2\omega_i \geq \omega_i\sqrt{2}\sqrt{1-\cos(\pi\rho_i \frac{x_i - x'_i}{u_i - l_i})} = d_i^{\mathrm{r}}(\underline{x}, \underline{x}'). \qquad (12)$$

If $\delta_i(\underline{x}'') =$ true, consider the 'worst' possible case in which, without loss of generality, $x_i = l_i$ and $x'_i = u_i$, such that $d_i^{\mathrm{r}}(\underline{x}, \underline{x}') = 2\omega_i^2$. We define the abbreviation $\beta'' =$

$\frac{x_i''-l_i}{u_i-l_i}$, giving

$$\begin{aligned}
\left(d_i^{\mathrm{r}}(\underline{x},\underline{x}'')+d_i^{\mathrm{r}}(\underline{x}',\underline{x}'')\right)^2 &= 2\omega_i^2\left(\sqrt{1-\cos(\pi\rho_i\beta'')}+\sqrt{1-\cos\bigl(\pi\rho_i(1-\beta'')\bigr)}\right)^2 \\
&= 2\omega_i^2\Bigg(2-\cos(\pi\rho_i\beta'')-\cos\bigl(\pi\rho_i(1-\beta'')\bigr) \\
&\qquad +2\sqrt{\Bigl(1-\cos(\pi\rho_i\beta'')\Bigr)\Bigl(1-\cos\bigl(\pi\rho_i(1-\beta'')\bigr)\Bigr)}\Bigg) \\
&= 2\omega_i^2\left(2+2\sqrt{1+\cos(\pi\rho_i\beta'')\cos\bigl(\pi\rho_i(1-\beta'')\bigr)}\right) \\
&= 4\omega_i^2\bigl(1+|\sin\pi\rho_i\beta''|\bigr) \\
&\geq 4\omega_i^2 = d_i^{\mathrm{r}}(\underline{x},\underline{x}')^2.
\end{aligned} \tag{13}$$

Hence, from non-negativity, we have $d_i^{\mathrm{r}}(\underline{x},\underline{x}'')+d_i^{\mathrm{r}}(\underline{x}',\underline{x}'')\geq d_i^{\mathrm{r}}(\underline{x},\underline{x}')$. □

**Proposition 6.** *$d_i^{\mathrm{c}}$ is a pseudometric on $\mathcal{X}$.*

*Proof.* The non-negativity and symmetry of $d_i^{\mathrm{c}}$ are trivially proven. To prove the triangle inequality, consider $\underline{x},\underline{x}',\underline{x}''\in\mathcal{X}$.

Case 1: $\delta_i(\underline{x})=\delta_i(\underline{x}')=$ false, such that $d_i^{\mathrm{c}}(\underline{x},\underline{x}')=0$. Here, from non-negativity, clearly $d_i^{\mathrm{c}}(\underline{x},\underline{x}')=0\leq d_i^{\mathrm{c}}(\underline{x},\underline{x}'')+d_i^{\mathrm{c}}(\underline{x}',\underline{x}'')$.

Case 2: $\delta_i(\underline{x})\neq\delta_i(\underline{x}')$, such that such that $d_i^{\mathrm{c}}(\underline{x},\underline{x}')=\omega_i$. Without loss of generality, assume $\delta_i(\underline{x})=$ true, $\delta_i(\underline{x}')=$ false and $\delta_i(\underline{x}'')=$ true.

$$d_i^{\mathrm{c}}(\underline{x},\underline{x}'')+d_i^{\mathrm{c}}(\underline{x}',\underline{x}'')=d_i^{\mathrm{c}}(\underline{x},\underline{x}'')+\omega_i \tag{14}$$

Hence $d_i^{\mathrm{c}}(\underline{x},\underline{x}'')+d_i^{\mathrm{c}}(\underline{x}',\underline{x}'')\geq\omega_i=d_i^{\mathrm{c}}(\underline{x},\underline{x}')$ by non-negativity.

Case 3: $\delta_i(\underline{x})=\delta_i(\underline{x}')=$ true, such that $d_i^{\mathrm{c}}(\underline{x},\underline{x}')=\omega_i\frac{\sqrt{2}\rho}{1+(m_i-1)(1-\rho)^2}\mathbb{I}_{x_i\neq x_i'}$. If $\delta_i(\underline{x}'')=$ false,

$$d_i^{\mathrm{c}}(\underline{x},\underline{x}'')+d_i^{\mathrm{c}}(\underline{x}',\underline{x}'')=2\omega_i\geq\omega_i\frac{\sqrt{2}\rho}{1+(m_i-1)(1-\rho)^2}\mathbb{I}_{x_i\neq x_i'}=d_i^{\mathrm{c}}(\underline{x},\underline{x}'). \tag{15}$$

If $\delta_i(\underline{x}'')=$ true,

$$\begin{aligned}
d_i^{\mathrm{c}}(\underline{x},\underline{x}'')+d_i^{\mathrm{c}}(\underline{x}',\underline{x}'') &= \omega_i\frac{\sqrt{2}\rho}{1+(m_i-1)(1-\rho)^2}(\mathbb{I}_{x_i\neq x_i''}+\mathbb{I}_{x_i'\neq x_i''}) \\
&\geq \omega_i\frac{\sqrt{2}\rho}{1+(m_i-1)(1-\rho)^2}\mathbb{I}_{x_i\neq x_i'}=d_i^{\mathrm{c}}(\underline{x},\underline{x}').
\end{aligned} \tag{16}$$

□