# A Proofs

We provide the proofs for the following reformulated statements from Section 4.2:

1. Given an optimal reference set, the error of CID reduces linearly with the maximal distance of adjacent points in the reference set decreasing.

2. For an optimal reference set obtained with a uniform sampling approach, the CID's error tends to zero with the increasing number of points sampled.

## A.1 Proof Statement 1

The proofs are provided for $q = 1$ from the Definition of CID. The proof for $q > 1$ can be obtained with similar derivations. We start with defining two reference sets $G$ and $F$, obtained by performing sampling within the search domain $D \subseteq \mathbb{R}^n$ resulting in a sparse and dense arrangement/distribution of samples respectively.

Let be $S$ the number of all separated regions of the DOI, then, since the reference sets are optimal, both reference sets witness every separated continuous region $r_s$, $s = 1, \ldots, S$.

From the Definition of CID, the following holds:

$$CID(A, Z) = \frac{1}{|Z|} \sum_{z \in Z} d_z = \frac{1}{|Z|} \sum_{s=1}^{S} \sum_{z \in (Z \cap r_s)} d_z \tag{1}$$

, i.e.,

$$CID(A, Z) = \sum_{s=1}^{S} CID(A, Z \cap r_s) \tag{2}$$

Computing CID for the test set A, using the reference sets $G$ and $F$, we obtain:

$$CID_G = \frac{1}{|G|} \sum_{i=1}^{|G|} d_i^G \tag{3}$$

$$CID_F = \frac{1}{|F|} \sum_{j=1}^{|F|} d_i^F \tag{4}$$

From Equation 2, w.l.o.g., let's assume that $S = 1$, and let us denote $r_1$ as $r$. The difference between the CID values for the corresponding reference sets is then:

1

$$CID_F - CID_G = \frac{1}{|G|} \sum_{i=1}^{|G|} d_i^G - \frac{1}{|F|} \sum_{j=1}^{|F|} d_i^F \qquad (5)$$

We partition the connected region $r$ using a Voronoi diagram built on $G$ into $|G|$ Voronoi cells $c_i$, $i = 1, \ldots, |G|$. Then, we assign to each reference point $f_j \in F$ a Voronoi cell $c_i$ if the reference point lies within the cell, and denote the reference point as $f_k^i$. If $f_j$ lies on a boundary of a cell, we assign it to any of the neighbouring cells. Let be $R^*$ the maximal distance of adjacent points in $G$. Note, that two points in a reference set are called *adjacent* if the respective sampled points are adjacent. Each of the cells lies within a ball $B(g_i; R_i)$, and the radius of each of the balls is bounded by $R^*$, i.e. $R_i \leq R^*$. As $f_k^i$ belongs to the cell $c_i$ with the center $g_i$, the distance between them is bounded: $\|\vec{\xi_{ik}}\| \leq R^*$, where $\vec{\xi_{ik}}$ is the relative position between $f_k^i$, and $g_i$.

By applying the triangle equality of vector addition for the relative positions from the reference points $f_j \equiv f_i^k$ and $g_i$ to the test point yields:

$$\vec{d_j^F} = \vec{d_i^G} + \vec{\xi_{ik}} \qquad (6)$$

Using Equation 5, $CID_F$ from Equation 4 can be rewritten as:

$$CID_F = \frac{1}{|F|} \sum_{i=1}^{|G|} \sum_{k=1}^{K_i} \|\vec{d_i^G} + \vec{\xi_{ik}}\|, \qquad (7)$$

where $K_i$ is the number of points $f_k^i$ belonging to $c_i$.

With Equations 7 and 5, the following holds:

$$CID_F - CID_G = \frac{1}{|F|} \sum_{i=1}^{|G|} \sum_{k=1}^{K_i} \|\vec{d_i^G} + \vec{\xi_{ik}}\| - \frac{1}{|G|} \sum_{i=1}^{|G|} \|\vec{d_i^G}\| \qquad (8)$$

The norm $\|\cdot\|$ satisfies the following inequalities:

- Triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$

- Reverse triangle inequality: $\|x + y\| \geq \|\|x\| - \|y\|\|$

Applying the triangle inequality to Equation 8 yields:

$$CID_F - CID_G \leq \frac{1}{|F|} \sum_{i=1}^{|G|} \sum_{k=1}^{K_i} \|\vec{d_i^G}\| - \frac{1}{|G|} \sum_{i=1}^{|G|} \|\vec{d_i^G}\| + \frac{1}{|F|} \sum_{i=1}^{|G|} \sum_{k=1}^{K_i} \|\vec{\xi_{ik}}\| \underset{*}{=}$$
$$\frac{1}{|F|} \sum_{i=1}^{|G|} (K_i - K^*) \|\vec{d_i^G}\| + \frac{1}{|F|} \sum_{i=1}^{|G|} \sum_{k=1}^{K_i} \|\vec{\xi_{ik}}\| \overset{|F| \to \infty}{\underset{**}{\leq}} C(|G|) + R^*. \qquad (9)$$

Note:

\* : For the equality, $K^*$ is the average number of points $f_k^i$ within a cell $c_i$, for $i = 1, \ldots |G|$, i.e. $K^* = |F|/|G|$.

\*\* : For the inequality, presuming we have a uniform reference set $F$ of an infinite size, the first sum represents the deviation from the average of the volume of each Voronoi cell. As shown in [2], the deviation reduces with the increasing number of points. Furthermore, the reduction is even higher for higher dimensions [1]. Note, that when the reference set is obtained by using an equidistant sampling technique (e.g., grid sampling), the variance of the volume of a Voronoi cell is equal to zero, as the volumes are equal for each Voronoi cell, which means that the first sum is equal to zero. In other words, the first sum represents the uniform properties of the reference sets, and the more uniform the reference set, the smaller the variance is. We presume that the set of sampled points, and therefore the reference set $G$ are uniform enough to neglect the first sum. The second sum is bounded, as the size of each cell is bounded by the respective radius of a ball.

Applying the reverse triangle inequality on Equation 8, the following holds:

$$CID_F - CID_G \geq \frac{1}{|F|} \sum_{i=1}^{|G|} \sum_{k=1}^{K_i} |d_i^G - \|\vec{\xi_{ik}}\|| - \frac{1}{|G|} \sum_{i=1}^{|G|} d_i^G \underset{*}{=}$$

$$\frac{1}{|F|} \sum_{i=1}^{|G|} \sum_{k=1}^{K_i^+} (d_i^G - \|\vec{\xi_{ik}}\|) + \frac{1}{|F|} \sum_{i=1}^{|G|} \sum_{k=1}^{K_i^-} (\|\vec{\xi_{ik}}\| - d_i^G) - \frac{1}{|F|} \sum_{i=1}^{|G|} K^* d_i^G = \quad (10)$$

$$\frac{1}{|F|} \sum_{i=1}^{|G|} (K_i^+ - K^*) d_i^G - \frac{1}{|F|} \sum_{i=1}^{|G|} \sum_{k}^{K_i^+} \|\vec{\xi_{ik}}\| + \frac{1}{|F|} \sum_{i=1}^{|G|} \sum_{k}^{K_i^-} (\|\vec{\xi_{ik}}\| - d_i^G)$$

Note:

\* : For the equality, we separate the sum of the moduli over the reference points within the cell into two sums: where the expression under the modulus is positive ($K_i^+$ terms), and negative ($K_i^-$ terms).

Now, we consider each of the sums from Equation 10 separately. For the second sum holds:

$$\frac{1}{|F|} \sum_{i=1}^{|G|} \sum_{k}^{K_i^+} \|\vec{\xi_{ik}}\| \leq \frac{1}{|F|} \sum_{i=1}^{|G|} \sum_{k}^{K_i} \|\vec{\xi_{ik}}\| \leq R^*, \quad (11)$$

as the size of each cell is bounded by the respective radius of a ball.

To assess the rest of the sums, let us study the dependence of $K_i^-$ and $K_i^+$ on $R^*$. For a fixed test set, the following holds:

$$\exists G_0 : \forall G : |G| > |G_0| \implies K_i^- \leq R^* \cdot K^*. \quad (12)$$

3

As $K_i^+ + K_i^- = K_i$, and for a uniform equidistant arrangement of samples, $K_i = K^*$, the following holds:

$$K_i^+ = K^* - K_i^- \geq K^* - R^* \cdot K^* = (1 - R^*) \cdot K^* \tag{13}$$

If the arrangement of samples is not structured, the additional terms involving the variances of the cell volume can be neglected similarly to Equation 9. Then, from Equation 12, for the third sum holds:

$$\frac{1}{|F|} \sum_{i=1}^{|G|} \sum_{k}^{K_i^-} (\|\vec{\boldsymbol{\xi}_{ik}}\| - d_i^G) \leq C' \cdot \frac{1}{K^*} \cdot R^* \cdot K^* = C' \cdot R^*; \ C' > 0. \tag{14}$$

Similarly, for the first sum in the Equation 13 holds:

$$\left| \frac{1}{|F|} \sum_{i=1}^{|G|} (K_i^+ - K^*) \cdot d_i^G \right| \leq C'' \cdot \frac{1}{K^*} \cdot R^* \cdot K^* = C'' \cdot R^*; \ C'' > 0. \tag{15}$$

Then for the Equation 10 holds:

$$CID_F - CID_G \geq -(C' + C'' + 1) \cdot R^* \geq -C''' \cdot R^*; \ C''' > 0. \tag{16}$$

Finally, from Equation 16 and Equation 9 follows:

$$|CID_F - CID_G| \leq C^* \cdot R^*, \ C^* = \max(C''', C). \tag{17}$$

Equation 17 implies that the error reduces linearly with the maximal distance between adjacent points in the reference set decreasing, which proofs the Statement 1.

## A.2  Proof Statement 2

Because the sampling strategy is uniform, the maximum distance between adjacent points in the reference set $R^*$ decreases with the increasing number of sampled points. Given, that the reference set is optimal, statement 2 follows from Equation 17.

# References

[1]    Luc Devroye et al. "On the measure of Voronoi cells". In: *arXiv e-prints*, arXiv:1512.04267 (Dec. 2015), arXiv:1512.04267.

[2]    E. N. Gilbert. "Random Subdivisions of Space into Crystals". In: *The Annals of Mathematical Statistics* (1962), pp. 958 –972.